Inclusion des nouvelles technologies en vue de l'évolution de la disponibilité et de la qualité du service d'un système d'information

Cédric Bombail

CONSERVATOIRE NATIONAL DES ARTS ET METIERS

CENTRE REGIONAL ASSOCIÉ DE TOULOUSE

MÉMOIRE

présenté en vue d'obtenir

le DIPLÔME D'INGÉNIEUR CNAM

SPÉCIALITÉ : INFORMATIQUE

par

Cédric Bombail

Inclusion des nouvelles technologies en vue de l'évolution de la disponibilité et de la qualité du service d'un système d'information.

Soutenu le 28 juin 2011

JURY

Président : Madame le Professeur Anne WEI, CNAM

Membres : Monsieur Hadj BATATIA, Maître de conférences, CNAM
Monsieur Bertrand RAIFF, Maître de conférences, CNAM
Madame Sylvie ROQUES, Directrice de recherche, CNRS
Monsieur Philippe SABY, Ingénieur de recherche, MESR

Remerciements

Je tiens à remercier Sylvie ROQUES, directrice du laboratoire, Boris DINTRANS, président de la commission informatique et Philippe SABY, responsable informatique pour m'avoir accordé leur confiance en me proposant de conduire un projet structurant pour l'informatique du laboratoire d'Astrophysique de Toulouse et Tarbes.

Je les remercie pour m'avoir apporté leur soutien et autorisé à présenter ce travail dans le cadre de ma formation d'ingénieur au CNAM de Toulouse, et plus particulièrement Philippe pour sa disponibilité et ses conseils avisés dans le cadre du déroulement de l'action et de la rédaction du présent mémoire d'ingénieur.

Je remercie également Bertrand RAIFF, mon tuteur CNAM, pour avoir accepté la charge d'encadrer mon mémoire d'ingénieur.

Je tiens aussi à remercier Nicolas PEREZ et Michel L'HOSTIS, mes collègues du service informatique, pour leur disponibilité, la qualité de leur travail et leur soutien sur les tâches quotidiennes du service ; sans leur implication, la refonte du système d'information du laboratoire d'Astrophysique n'aurait probablement pas été possible.

Je remercie enfin et surtout Corinne, ma compagne, pour son soutien et sa patience sans faille durant ces années d'études au CNAM, notamment pendant la période de rédaction de ce document durant laquelle je n'étais pas très disponible.

Liste des abréviations

CICT : Centre Interuniversitaire de Calcul Toulousain
CLUSIF : Club de la Sécurité Informatique Français
CNRS : Centre National de la Recherche Scientifique
DHCP : Dynamic Host Configuration Protocol *
DNS : Domain Name System *
GLPI : Gestion Libre de Parc Informatique (Solution open source de gestion de parc)
GPL : Général Public licence *
IHM : Interface Homme Machine *
LATT : Laboratoire d'Astrophysique de Toulouse et Tarbes
MESR : Ministère de l'Enseignement Supérieur et de la Recherche
MOA : Maîtrise d'ouvrage *
MOE : Maîtrise d'œuvre *
NAS : Network Attached Storage *
NFS : Network File System *
NTP : Network Time Protocol *
OMP : Observatoire Midi-Pyrénées
OSU : Observatoire des Sciences de l'Univers
PCA : Plan de Continuité d'Activité
PL-SQL : Procedural Language / Structured Query Language *
PRA : Plan de Reprise d'Activité
PSI : Politique de Sécurité Informatique
RADIUS: Remote Authentication Dial-In User Service *
RENATER : Réseau National de Télécommunications pour la Technologie l'Enseignement et la Recherche
ROI : Return on Investissement *
SAN: Storage Area Network *
SGBD : Système de Gestion de Bases de Données *
SPOF : Single Point of Failure *
UMR : Unité Mixte de Recherche
UMS : Unité Mixte de Services
UPS : Université Paul Sabatier, Toulouse
VLAN : Virtual Local Area Network *
VM : Virtual Machine *

* *ces termes sont détaillés dans le glossaire technique*.

Glossaire des termes techniques

DHCP : Dynamic Host Configuration Protocol, protocole de configuration dynamique de l'Hôte. Ce protocole permet à un ordinateur qui se connecte sur un réseau d'obtenir dynamiquement sa configuration.

DNS : Domain Name System, système de noms de domaine. Service permettant d'établir une correspondance entre une adresse IP et un nom de domaine.

Fail over : Il s'agit de la capacité à basculer automatiquement à un système redondant ou en veille en cas de défaillance d'un service actif (serveur, système ou réseau).

Firewall : Pare-feu. C'est un logiciel ou un matériel permettant de faire respecter la politique de sécurité du réseau en permettant de protéger un ordinateur ou un réseau d'ordinateurs des intrusions provenant d'un réseau tiers.

GPL : Général Public licence (ou "GNU General Public License"). C'est une licence libre destinée notamment aux œuvres logicielles. Elle a pour but de garantir la liberté de partager et changer toutes les versions d'un programme afin d'assurer qu'il restera libre pour tous les utilisateurs.

Hyperviseur : Logiciel de virtualisation hôte. L'hyperviseur est la plate-forme de virtualisation permettant à plusieurs systèmes d'exploitation de travailler sur une même machine physique de façon parallèle et indépendante.

IHM : Interface Homme Machine. C'est une interface qui permet la communication entre un humain et une machine de façon ergonomique, efficace, et facilitée.

Load balancing : équilibrage de charge. L'équilibrage de charge est une technique pour distribuer un processus entre plusieurs ressources afin d'en optimiser le rendement.

MOA : Maîtrise d'ouvrage. C'est l'entité porteuse du besoin, définissant les objectifs, le calendrier et le budget consacré à un projet. La MOA maîtrise l'idée de base du projet, et représente à ce titre les utilisateurs finaux à qui l'ouvrage est destiné. Le maître d'ouvrage est responsable de l'expression fonctionnelle des besoins.

MOE : Maîtrise d'œuvre. C'est l'entité retenue par la MOA pour réaliser l'ouvrage, dans les conditions de délais, de qualité et de coût fixées par elle. La maîtrise d'œuvre est donc responsable des choix techniques inhérents à la réalisation de l'ouvrage conformément aux exigences de la maîtrise d'ouvrage.

NAS : Network Attached Storage, stockage de données centralisé. Un NAS est un serveur de fichiers autonome relié à un réseau, disposant de son propre système d'exploitation et d'un logiciel de configuration paramétré. Sa principale fonction est le stockage de données en un volume centralisé pour des clients réseau hétérogènes.

NFS : Network File System, système de fichiers en réseau. Ce système de fichiers permet aux hôtes distants de se connecter à des systèmes de fichiers disponibles sur un réseau et de les utiliser comme des systèmes de fichiers locaux.

NTP : Network Time Protocol, peut se traduire par protocole d'heure réseau. Il s'agit d'un protocole permettant de synchroniser, via un réseau informatique, l'horloge locale d'ordinateurs sur la référence d'heure universelle, via une hiérarchie de serveurs synchronisés.

PL-SQL : Procedural Language - Structured Query Language. Il s'agit d'un langage procédural créé par Oracle et utilisé dans le cadre de bases de données relationnelles. Il permet de combiner des requêtes SQL et des instructions procédurales, dans le but de créer des traitements complexes destinés à être stockés sur le serveur de base de données, comme par exemple des procédures stockées ou des déclencheurs.

RADIUS: Remote Authentication Dial-In User Service. Il s'agit d'un protocole client-serveur permettant de centraliser des données d'authentification. Il repose principalement sur un serveur relié à une base d'authentification et un client RADIUS faisant office d'intermédiaire entre l'utilisateur final et le serveur.

ROI : Return on Investissement, retour sur investissement (RSI). Il s'agit de la rentabilité du capital investi, du rendement ou du taux de rendement ou taux de profit, désigné par un ratio financier qui mesure le montant d'argent gagné ou perdu par rapport à la somme initialement investie dans un investissement.

SAN: Storage Area Network, réseau de stockage. Un SAN est un réseau spécialisé permettant de mutualiser des ressources de stockage. Il se différencie des autres systèmes de stockage par un accès bas niveau aux disques. Les baies de stockage n'apparaissent pas comme des volumes partagés sur le réseau mais sont directement accessibles en mode bloc par le système de fichiers des serveurs.

SGBD : Système de Gestion de Bases de Données. Un SGBD un ensemble de services permettant de gérer une base de données en permettant l'accès aux informations à de multiples utilisateurs et une manipulation des données présentes dans la base de données à l'aide d'un LMD (langage de manipulation des données) comme le langage SQL (Structured Query Language).

SPOF : Single Point of Failure, point individuel de défaillance. C'est un point d'un système informatique dont le reste du système dépend et dont la panne entraîne un dysfonctionnement du système.

VLAN : Virtual Local Area Network, réseau local virtuel. C'est un réseau informatique logique indépendant permettant de segmenter et de sécuriser un réseau physique. Plusieurs VLAN peuvent coexister sur un même commutateur réseau.

VM : Virtual Machine, machine virtuelle. Environnement d'exécution émulé, hébergé sur un hyperviseur, qui fourni les mêmes services qu'une machine physique.

Wiki : Un wiki est un site Internet collaboratif dont les pages sont modifiables par les visiteurs autorisés. Au sein du laboratoire d'Astrophysique, il servait à partager la documentation technique du service de façon collaborative entre les intervenants.

Sommaire

I. Introduction

Un système d'information (SI) est un ensemble organisé de ressources (matériels, logiciels, personnel, données et procédures) qui permet de regrouper, de classifier, de traiter et de diffuser de l'information sur un environnement donné [1].
Au sens informatique, il s'agit donc de l'ensemble des éléments participant à la gestion, au traitement, au transport et à la diffusion de l'information d'une organisation.

Le service Informatique, dans lequel le stage correspondant à l'action décrite dans ce mémoire d'ingénieur a été réalisé, est le service informatique du Laboratoire d'Astrophysique de Toulouse-Tarbes, à l'Observatoire Midi-Pyrénées. Ce service regroupait des compétences systèmes et réseaux nécessaires au bon fonctionnement du système d'information du laboratoire. Le système d'information en production en 2009, bien que fonctionnel, n'était pas un système que l'on pouvait qualifier de hautement disponible. Le service informatique avait relevé quelques points d'efforts sur lesquels il devait agir pour améliorer son fonctionnement et sa pérennité.

Ce mémoire traite de l'évolution d'un système d'information par l'inclusion de technologies nouvelles, ayant pour objectif d'augmenter la qualité du service rendu et de concevoir et mettre en production un système à haute disponibilité.

Ce document est présenté en cinq chapitres :

- Le premier chapitre est une présentation rapide de l'environnement dans lequel j'ai réalisé mon stage d'ingénieur et de la problématique existante au moment de ce stage ;
- Le deuxième chapitre présente les concepts qui vont être mis en œuvre pour atteindre l'objectif et la méthodologie que j'ai choisie d'appliquer pour ces intégrations ;
- Le troisième chapitre traite de la conception et du développement d'un prototype du nouveau système d'information permettant de tester et de valider l'articulation des nouveaux concepts d'architecture du système d'information ;
- Le quatrième chapitre détaille les évolutions technologiques et organisationnelles qui ont été apportées à ce prototype pour l'intégrer pleinement en exploitation. Ces évolutions devront permettre de rendre le système hautement disponible ;
- Le cinquième et dernier chapitre conclue ce mémoire en faisant le bilan de l'intégration de la nouvelle organisation et introduit l'évolution en cours à une échelle supérieure.

1. Contexte

Mon stage s'est déroulé de septembre 2009 à décembre 2010 au sein du laboratoire d'Astrophysique de Toulouse -Tarbes, un des laboratoires de recherche de l'Observatoire Midi-Pyrénées (OMP).
L'OMP est un Observatoire des Sciences de l'Univers (OSU) et une composante de l'Université Paul Sabatier (UPS) sous tutelle CNRS, UPS, IRD, et CNES. Il œuvre dans tous les domaines des Sciences de l'Univers, de la planète Terre et de l'Environnement.

Lunette Jean Rösch, Pic du Midi (65)

Sa mission principale est la réalisation d'observations continues et systématiques de la Terre et de l'Univers. Pour ce faire, l'OMP est spécialisé dans le développement et l'utilisation d'instrumentations spatiales, de modélisations numériques et de bases de données spatiales et environnementales.
L'OMP est géographiquement réparti sur les sites de Toulouse, Tarbes, Lannemezan, ainsi que du célèbre Pic du Midi.

Les laboratoires de l'OMP viennent de terminer une phase de restructuration. Au moment de mon stage, l'Observatoire Midi-Pyrénées était composé de huit laboratoires de recherche : le Laboratoire d'Astrophysique de Toulouse - Tarbes (LATT), le Centre d'Etude Spatiale des Rayonnements (CESR), le Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS), le Laboratoire d'Aérologie (LA), le laboratoire de Dynamique Terrestre et Planétaire (DTP), le Laboratoire des Mécanismes et Transferts en Géologie (LMTG), le laboratoire d'Ecologie fonctionnelle et environnement (ECOLAB) et le Centre d'Etudes Spatiales de la Biosphère (CESBIO).

Au 1er janvier 2011, deux fusions de laboratoires ont réduit le nombre d'unités de recherche à six. D'un coté, le LATT, le CESR et une partie du DTP ont fusionné pour faire donner naissance à l'Institut de Recherche en Astronomie et Planétologie (IRAP), le plus important laboratoire d'astrophysique et planétologie de France. L'autre partie du DTP et le LMTG ont fusionné pour devenir le laboratoire de Géosciences Environnement Toulouse (GET).

Jusqu'en 2011, le laboratoire d'Astrophysique se composait, comme les autres laboratoires, de personnel de recherche (chercheurs, doctorants,…) et de personnels techniques (ingénieurs, techniciens et administratifs) pour gérer et administrer ses ressources propres, tant du point de vue de la gestion que de l'informatique.

L'informatique de chaque laboratoire était autonome et s'appuyait sur des structures fédératives pour les ressources mutualisées, les accès au réseau métropolitain (OMP et Université) et au réseau national de la recherche (Renater).

Le Laboratoire d'Astrophysique de Toulouse - Tarbes gérait donc ses propres ressources informatiques, tant au niveau de l'infrastructure technique que des outils scientifiques (serveurs de calculs pour effectuer des modélisations, bases de données sol de satellites spatiaux, etc.). Le réseau et les systèmes informatiques (câblage, équipements actifs, serveurs) étaient administrés par le service informatique du laboratoire de façon autonome.

2. Environnement

2.1 Le service informatique du LATT

Le service Informatique du Laboratoire d'Astrophysique se composait de quatre personnes : Philippe Saby, Ingénieur de Recherche, qui dirigeait le service et a encadré la rédaction de ce mémoire ; Nicolas Perez, Assistant Ingénieur, administrateur systèmes et réseaux ; Michel L'Hostis, administrateur systèmes et réseaux, dont on m'avait confié l'encadrement pour mener à bien mon projet, et moi-même, Cédric Bombail, Ingénieur d'Etudes, administrateur systèmes et réseaux, et auteur de ce mémoire CNAM.

Ce service assurait l'administration système et réseau d'un parc hétérogène et multi-sites (Toulouse, Tarbes et Pic du midi) de 280 machines et d'une trentaine de serveurs. Dans cet environnement, organisé principalement sous LINUX, la mission se structurait en quatre classes de services. Les services orientés utilisateurs, les services orientés "machines", la gestion des infrastructures et l'appui scientifique.

- Le service aux utilisateurs proposait un compte informatique unique, sécurisé et les outils adaptés: messagerie, web, ftp, etc.

- Le service aux "machines" incluait la connexion au réseau, les diagnostics, les dépannages, l'installation et la maintenance des applicatifs scientifiques métiers.

- La gestion des infrastructures systèmes et réseaux proposait un socle technique stable et pérenne bâti sur des distributions Linux, des réseaux SAN et NAS ainsi qu'un système de sauvegarde. Cet ensemble reposait sur une infrastructure réseau protégée par des gardes barrières de périmètres (firewall).

- Pour terminer, l'appui scientifique était une mission essentielle bâtie autour de conseils et d'expertises informatiques, de mise à disposition d'espaces de stockage et de sauvegarde dédiés et d'implication dans certains projets.

Cette structure fut le cadre de mon stage d'ingénieur de septembre 2009 à décembre 2010.

2.2. L'environnement technique du LATT

Un système d'Information représente l'ensemble des éléments participant à la gestion, au traitement, au transport et à la diffusion de l'information au sein d'une organisation. D'un point de vue informatique, le système d'information regroupe l'ensemble des ressources permettant de maintenir, utiliser et faire fonctionner les services offerts.

Pour un service informatique, l'ingénierie des connaissances va donc s'articuler autour des deux composantes :

- La gestion de contenu : il s'agit de la gestion d'informations brutes et de leur transformation en données structurées qui seront la source des services offerts;
- La gestion des accès : c'est la gestion des flux et des protocoles d'échange dans les réseaux de communications internes ou partagés avec les partenaires. Dans le cas d'un service informatique, on parlera de gestion du réseau informatique.

Pour synthétiser, on trouve deux branches principales : les données brutes et l'accès à ces données. Cela constitue, pour le métier d'informaticien, les deux axes principaux à prendre en compte dans la création ou la réorganisation d'un système d'information.

En septembre 2009, l'environnement technique du LATT se composait d'un système d'information et d'une infrastructure systèmes et réseaux détaillée dans la Figure 1.

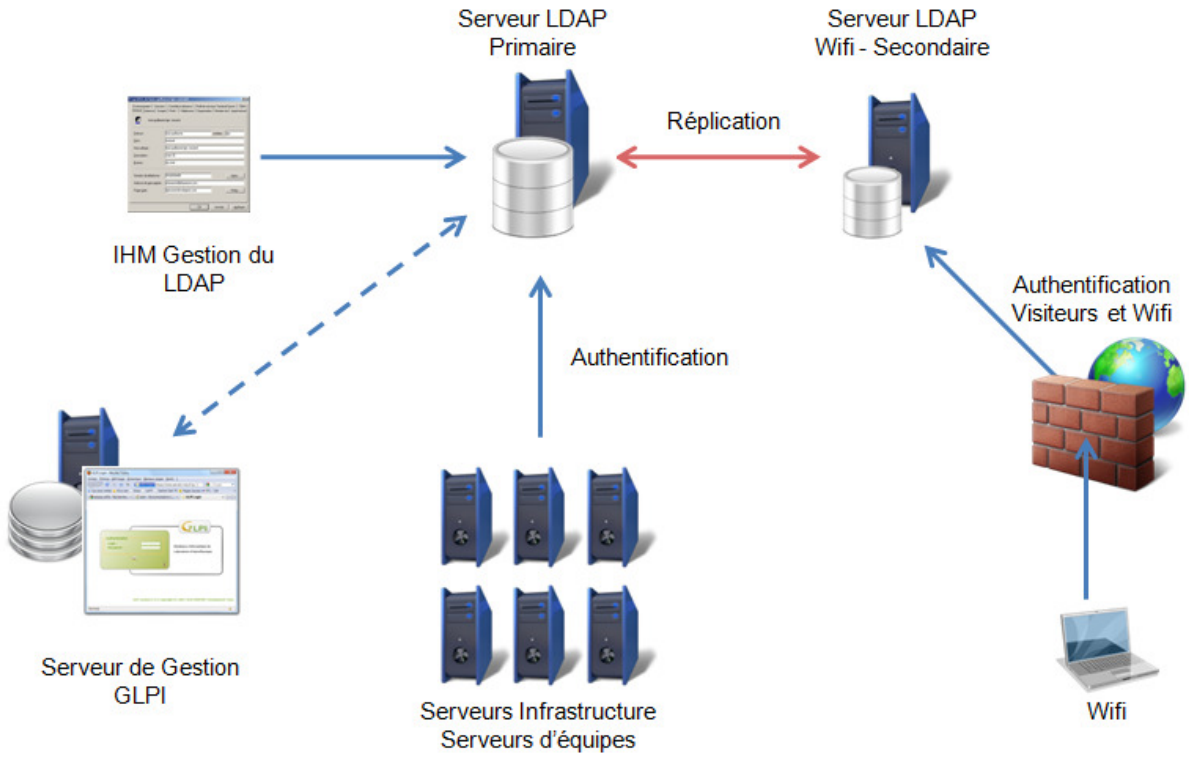


Figure 1 : Architecture du système d 'information du LATT - septembre 2009

La gestion des données de contenu regroupait les données utilisateur et les données du parc informatique.

La gestion des données utilisateur (identification et authentification) était réalisée par l'intermédiaire d'un annuaire LDAP (Lightweight Directory Access Protocol) [RFC 3377].
LDAP est un protocole permettant l'interrogation et la modification des services d'annuaire, reposant sur TCP/IP. S'appuyant sur un schéma de données spécifique, il était la source des données pour les comptes utilisateurs du LATT. L'ensemble des données d'authentification (nom d'utilisateur et mot de passe) et d'identification (téléphone, mél, groupes,…) étaient regroupées dans cet annuaire LDAP.

La gestion des données du parc informatique était réalisée par l'intermédiaire du logiciel « Gestion Libre de Parc Informatique (GLPI) ». Ce logiciel permettait de gérer l'ensemble des matériels informatiques du laboratoire : stations de travail, serveurs, imprimantes, commutateurs réseau,…
Il s'agissait d'un progiciel de gestion de parc incluant des fonctionnalités de type « helpdesk » (centre d'assistance). Les données concernant les stations de travail y étaient donc recensées, mais apparaissaient aussi dans d'autres sources d'information comme dans les fichiers de configuration des services DNS et DHCP.

Du point de l'infrastructure systèmes et réseaux, la grande majorité des serveurs d'infrastructure et d'exploitation scientifique était des machines physiques. A cette époque, le laboratoire souhaitait étudier la piste de la virtualisation afin d'analyser ses avantages et ses inconvénients sur la pérennité du système d'information.
Le réseau informatique, support de la gestion de l'accès aux données, avait été entièrement restructuré en 2009 avec la mise à jour de toute l'infrastructure réseau : câblage (catégorie 7), prises réseau (connecteurs catégorie 6) et matériels actifs (commutateurs et châssis gigabits Ethernet).
Ces matériels actifs offraient des capacités de routage, de contrôle et de gestion dynamique des machines connectées.

2.3. Points d'effort recensés en 2009

Bien que cet environnement technique ai été fonctionnel pendant plusieurs années, diverses pistes d'évolution avaient été recensées par le service informatique. Il s'agissait des points d'effort sur lesquels nous devions intervenir à court ou moyen terme pour améliorer le système :

- La gestion du cycle de vie informatique de l'utilisateur ;
- La cohérence des données dans le système ;
- Le maintien de l'exactitude des données utilisateur ;
- La gestion et la sécurisation des accès au réseau ;
- Le renforcement de la disponibilité du système d'information.

La gestion du cycle de vie informatique de l'utilisateur était traitée de façon automatique et satisfaisante pour l'étape de suppression. Par contre, la création et les mises à jour reposaient uniquement sur les interventions manuelles des membres du service informatique.
Nous détaillons ci-après le processus d'ouverture d'un compte utilisateur (Figure 2).

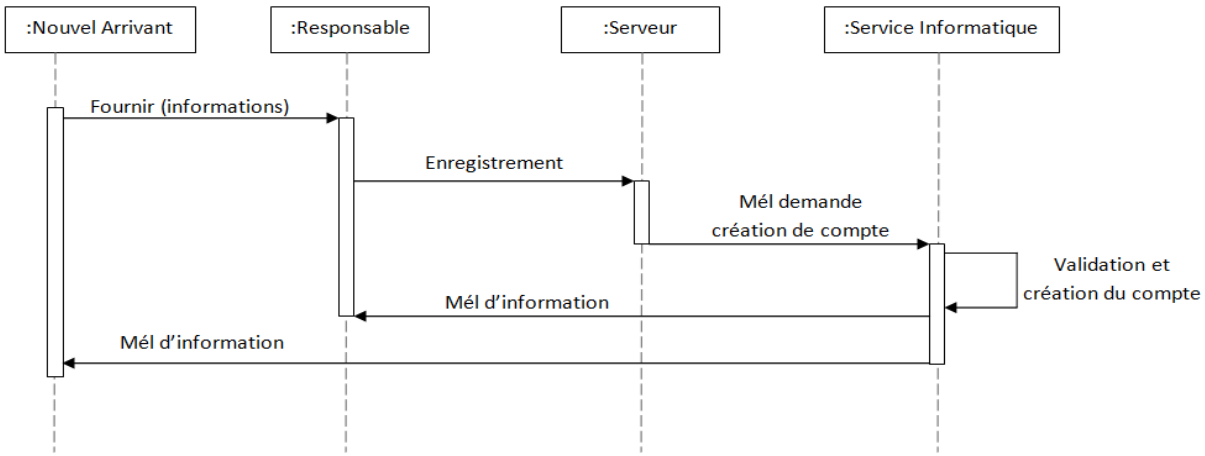


Figure 2 : Ouverture du compte utilisateur au LATT - septembre 2009

Lors de l'arrivée d'un nouvel utilisateur, en fonction de son statut, un compte utilisateur était créé par le service informatique dans l'annuaire LDAP du laboratoire. Il permettait d'autoriser l'accès aux différents services en fonction des demandes utilisateurs : accès aux serveurs de fichiers et de calcul, accès au service de messagerie, etc. Tant que l'utilisateur travaillait au sein ou en collaboration avec le laboratoire, son compte restait actif.
Lors de son départ, une procédure automatisée et précise permettait de fermer ce compte et de couper l'accès aux différents services.
Or, si nous maitrisions bien ce processus de durée de vie et de fin de vie du compte, notamment par l'adoption initiale d'un EOL sur ce compte (End Of Life, fin de durée de vie), les principaux problèmes étaient rencontrés lors de la création du compte.

Selon les procédures établies, tout responsable devait, pour les personnels qu'il accueillait, faire une demande de compte par l'intermédiaire d'une IHM située sur le site Intranet. Le service informatique recevait automatiquement un mail de demande d'ouverture de compte et créait le compte associé à partir des informations indiquées dans cette demande.

Ce processus semi-automatique, impliquant une réactivité forte et un respect rigoureux des consignes, était chronophage et source d'erreurs. Le cas le plus fréquent était un retard de disponibilité du compte dû, soit à un retard ou une erreur de saisie, soit à un manque de disponibilité du service informatique.

Nous avons donc mené une réflexion pour optimiser ce processus de déclaration et d'enregistrement d'un nouvel utilisateur dans le laboratoire.

Un autre point d'effort recensé était **la cohérence des données dans le système**. Comme nous venons de le voir, les données du système étaient présentes dans plusieurs sources réparties : annuaire LDAP, progiciel GLPI et fichiers de configuration des services DNS et DHCP.
La gestion manuelle de ces services et la redondance des informations pour chacun d'eux pouvaient induire des incohérences de données.

Le service informatique a donc souhaité réfléchir à une solution permettant de maintenir la cohérence des données du système d'information.

Le maintien de l'exactitude des données utilisateur est un problème fréquemment rencontré au sein des organisations. Dans toute structure, les évolutions des utilisateurs au sein de l'entreprise ont un impact sur le traitement de leurs données : mobilité géographique, changement de bureau, de projet ou d'équipe, accès à certaines ressources, connexions internes ou externes, etc.

Administrateurs des outils de gestion de ces données, le service informatique se trouvait, *de facto*, chargé de leur mise à jour. Les échanges imposés entre les différents acteurs de ce processus (secrétariat, ressources humaines, commission des locaux et service informatique) pouvaient être sources d'erreurs.

Une réflexion a donc été engagée pour décentraliser et modérer cette gestion des données au plus près de la source d'information, *i.e.* par exemple les responsables de la commission des locaux pour la gestion des bureaux et téléphones des utilisateurs.

Le service informatique souhaitait plutôt recentrer ses activités sur son cœur de métier, à savoir le maintien et la mise à disposition des données, et déléguer leurs mises à jour vers les acteurs en charge de la gestion administrative.

La gestion et la sécurisation des accès au réseau du LATT était réalisée selon un fonctionnement établi et hérité de son historique : environ 280 machines, une trentaine de serveurs, un environnement hétérogène et la grande majorité des machines sur le réseau disposant d'une adresse IP publique (une classe C IPv4 réservée au laboratoire). Cette caractéristique n'allait pas sans poser de problèmes d'adressage, de sécurité, et de disponibilité d'adresses pour les nouveaux matériels.

La gestion des stations de travail et le suivi de la durée de vie des machines étaient assez difficiles à effectuer. Si le service informatique participait à la majorité des achats de stations de travail, il était fréquemment confronté à des demandes spontanées d'adresses IP formulées par les utilisateurs ayant acheté un nouveau matériel de façon autonome.

De plus, nous maîtrisions mal le cycle de vie des matériels de notre parc. D'une part, les chercheurs qui effectuaient des mutations professionnelles conservaient leurs postes de travail et, d'autre part, les matériels obsolètes étaient parfois remplacés par les utilisateurs, sans en informer le service informatique.

Le service informatique satisfaisait donc, avec difficulté, aux demandes de nouvelles adresses IP, à la gestion de l'inventaire et ne pratiquait pas de contrôle actif d'accès au réseau.

Ces éléments nous ont incités à réfléchir à des processus d'inventaire des stations de travail et de contrôle de leurs connexions sur le réseau du laboratoire.

Le renforcement de la disponibilité du système d'information était une des actions prioritaires à mener. La disponibilité du système d'information du LATT dépendait principalement de la disponibilité de son service d'annuaire.
De la disponibilité du service LDAP dépendait le bon fonctionnement des principaux services rendus aux utilisateurs. L'annuaire LDAP constituait donc le principal point individuel de défaillance (SPOF) du système d'information.
Le service informatique a ainsi cherché une solution permettant de diminuer la criticité de ce service et d'en augmenter la disponibilité.

Le recensement de ces points d'effort a rendu évidente la nécessité d'une refonte du système d'information au sein d'un projet global regroupant l'ensemble des aspects systèmes et réseaux.

Ce projet devait permettre d'évoluer vers un système plus sécurisé, apportant des garanties en termes de disponibilité et permettant d'offrir de nouveaux services à nos utilisateurs.

3. Mon stage d'ingénieur CNAM

Fort de cette analyse, le service informatique et la direction du laboratoire m'ont confié une mission d'étude et de mise en œuvre d'un nouveau système d'information.

Mon premier axe de proposition a été dirigé vers les données ; j'ai proposé d'identifier, collecter, réunir et rationnaliser l'ensemble des données dans un référentiel unique, dont le contenu serait redistribuable et amènerait une cohérence des données dans le nouveau système d'information.

Le deuxième axe de réflexion était orienté vers la mise en place d'une délégation, avec modération, de la gestion des données administratives des utilisateurs. Cette délégation de saisie, de modification et de suppression, réalisée auprès des services compétents, devrait accroitre l'efficacité du traitement et améliorer la qualité de ces données.

Le troisième axe de réflexion concernait l'exploitation des technologies intégrées aux nouveaux matériels réseaux implantés en 2009. L'utilisation de ces fonctionnalités nouvelles permettrait d'authentifier et de sécuriser les accès aux ressources informatiques du laboratoire. La gestion de ce contrôle pourrait être simplifiée pour le service informatique grâce à l'utilisation d'un protocole d'authentification et des données d'identification associées, liées aux machines, intégrées dans le référentiel.

Le quatrième axe de réflexion était orienté vers la disponibilité du système. La mise en place des technologies de virtualisation récentes, la création de plans de reprise et de continuité d'activité, adaptés aux nouvelles technologies déployées, permettraient d'augmenter la disponibilité et la pérennité du nouveau système d'information.

A ces quatre axes principaux s'est ajoutée la volonté du service informatique de proposer de nouveaux services aux utilisateurs. Un cinquième point de réflexion a donc orienté vers la détection et la mise en place de nouveaux services pour les personnels du laboratoire d'Astrophysique.

Ces propositions, portées par le service informatique, ont été validées par la direction du laboratoire et la commission informatique. Il m'a été confié la charge de proposer une architecture pouvant apporter une plus-value en terme de qualité du service et pouvant améliorer la disponibilité du système d'information.

La direction du laboratoire, la commission informatique et mon responsable hiérarchique ont alors validé l'exécution de cette action et sa diffusion dans le cadre de mon mémoire d'ingénieur CNAM.

II. Evolution du système d'information

Nous venons d'évoquer le contexte et l'environnement informatique du LATT ainsi que les différents points d'effort sur lesquels le service devait agir en vue de l'amélioration du système d'information.

Je vais présenter la réponse technologique qui a été mise en œuvre, fournir les explications et détailler les nouveaux concepts d'architecture système qui ont été implantés au LATT entre septembre 2009 et octobre 2010. Je présenterai ensuite la méthode organisationnelle que nous avons appliquée pour mener à bien ce projet d'évolution.

1. Présentation des nouveaux concepts

Le principe de la mise en place de ces nouveaux concepts a été basé sur les critiques recensées et les points d'efforts de l'ancien système d'information. Conscient des failles du système tel qu'il existait en 2009, le service informatique a pris le parti d'entamer la réorganisation de la majorité de l'architecture de ses services.
La mise en place de cette nouvelle architecture devait se faire de façon progressive et n'entrainer aucune coupure ni perte de service supérieure à quatre heures.

Les concepts que nous avons explorés s'appuyaient sur les principes suivants :

- Le regroupement des données, initialement éclatées dans le système, dans un référentiel unique et la génération de tous les services à partir de ce dernier afin d'assurer une cohérence entre les données du système ;
- La réorganisation de la gestion des données administratives et techniques ;
- La maitrise de la gestion des accès au réseau filaire ;
- La réorganisation des services afin de minimiser l'existence de points individuels de défaillance dans le système d'information et d'en optimiser sa disponibilité.

A ces concepts, s'ajoutait une réflexion sur l'extension de l'offre de services aux utilisateurs dans le cadre de cette refonte du système d'information.

La mise en place de ces nouveaux concepts impliquait une modification profonde de l'architecture du système d'information en production. Pour cette raison, la décision a été prise de faire évoluer le système d'information de façon progressive, après avoir validé chaque modification par une phase de tests ; l'ensemble de toutes ces étapes constitue le projet global.

Je vais justifier ci-après la mise en place des nouveaux concepts d'architecture du système d'information et les orientations qui ont été effectuées au lancement de ce projet.

1.1. Création d'un référentiel unique de données

Premier axe stratégique : le regroupement des données dans un référentiel unique pour obtenir une cohérence des données avant leur redéploiement.

Dans le système d'information en production au début du stage de ce mémoire (septembre 2009), les données du système étaient réparties dans plusieurs sources de données de différents types, notamment la base de données de gestion du parc (GLPI), l'annuaire LDAP et les fichiers de configuration des services DNS et DHCP. La dispersion de ces données provoquait des incohérences qui portaient préjudice à la sécurité et à la pérennité du système.

La création d'un référentiel unique de données regroupant l'ensemble des données du système permettrait d'identifier, de qualifier et de rendre cohérentes les données du système (suppression des doublons et des données obsolètes).

Une fois ce référentiel unique créé, le laboratoire disposerait d'informations à jour sur ses utilisateurs. A partir de ce moment là, le service informatique pourrait générer toutes les configurations des services qu'il rend de façon automatisée, prenant comme source de données cette base unifiée.
De cette façon, chaque mise à jour intervenant en amont et de façon indépendante serait répercutée dans tout le système, de façon automatisée et synchrone.
A tout moment le système se trouverait dans un état stable, sans risque d'obtenir de divergence entre les réponses des différents services réseau (DNS et DHCP par exemple).

De plus, la sauvegarde complète et fréquente de ce référentiel permettrait d'assurer la pérennité de toutes les données utiles et critiques du système d'information. L'unicité de la source apporterait une simplification du processus de sauvegarde et de restauration.

1.2. Délégation de la mise à jour des données administratives de l'utilisateur.

Deuxième axe stratégique : la délégation, avec modération, de la gestion des données administratives de l'utilisateur aux services compétents.

L'idée maitresse de cet axe était de déléguer la gestion des données administratives des utilisateurs au plus près de la source d'information. En effet, le service informatique du LATT maintenait jusqu'alors les données informatisées sans être le point central de regroupement de l'information. Le secrétariat de direction ou la commission des locaux fournissaient les données au service informatique qui était en charge de la saisie des mises à jour.

Le développement d'interfaces légères et multiplateformes, permettant de mettre à jour les données des utilisateurs pour tous les services compétents (gestion administrative, commission des locaux,…) apporterait un service et une réactivité supplémentaire. De plus, la création d'interfaces de consultation et de demande de modifications orientées vers l'utilisateur final améliorerait la performance et la souplesse du processus de traitement des modifications.

Le service informatique conserverait un rôle de modérateur des modifications et poursuivrait sa mission de mise à disposition, de sauvegarde et d'archivage des données.

1.3. Maitrise de la gestion des accès au réseau filaire

Le troisième axe de réflexion concernait la sécurisation et le filtrage de l'accès au réseau physique du laboratoire.

De part la nature scientifique des stations de travail du laboratoire, une partie des utilisateurs devaient disposer des droits d'administration. Cela leur permettait d'assurer la gestion quotidienne de leur environnement applicatif métier et de configurer les connexions associées.

De plus, il n'y avait pas de contrôle *a priori* lors de la connexion physique des machines sur le réseau.
L'attribution d'adresses réseau était réalisée par le service DHCP qui attribuait une adresse IP en fonction de l'adresse matérielle de chaque carte réseau. Une fois cette adresse attribuée, elle pouvait être manuellement réutilisée sur une autre station de travail.
Par exemple, lorsqu'un visiteur était accueilli par un scientifique du laboratoire, il souhaitait se connecter, via son poste de travail portable, au réseau du laboratoire afin d'accéder à ses ressources distantes. Pour ce faire, il pouvait être amené à utiliser temporairement une adresse locale, attribuée initialement à une autre machine.
Le laboratoire devait donc se doter de solutions techniques assurant la maitrise des accès au réseau filaire. Pour cela, lors de l'appel d'offre de la rénovation des équipements réseaux réalisé en 2009, une demande spécifique avait été faite pour le support natif de protocoles d'authentification et de centralisation des connexions.

Il s'agissait alors de concevoir et d'implémenter une solution permettant de mettre en pratique ces fonctionnalités et d'assurer une gestion efficace des accès au réseau filaire.

1.4. Utilisation des nouvelles technologies pour optimiser la disponibilité du système

Quatrième axe stratégique, le déploiement et la mise en œuvre des nouvelles technologies pour augmenter la pérennité et la disponibilité du système.

Les nouvelles technologies apportent de nombreuses solutions dans différents domaines. Ce sont souvent des évolutions de systèmes en exploitation, améliorés en tenant compte des retours d'expérience.
Parmi ces nouvelles technologies, nous souhaitions explorer, d'une part, les aspects réseau autour de l'authentification et, d'autre part, l'apport de la virtualisation pour consolider l'architecture système. L'utilisation des nouvelles technologies de virtualisation pourrait permettre au laboratoire de modifier la structure même de son système d'information.
Dans le cadre de jouvence matérielle, le remplacement des serveurs physiques en production par des systèmes virtualisés pourrait probablement être envisagé.

La mise en exploitation des technologies de virtualisation nécessite néanmoins une étude et une évaluation. L'apport des deux technologies (physiques et virtuelles) devra être comparé afin d'effectuer un choix d'architecture. C'est une des actions qui sera réalisée durant ce stage.

En fonction des résultats, l'équipe système prendrait le parti de la virtualisation ou conserverait son architecture physique en la réorganisant.

L'utilisation des technologies récentes n'était qu'une piste de réflexion pour tendre vers la haute disponibilité. En parallèle, il s'agissait de repenser l'organisation de l'architecture du système d'information. Cette réorganisation passerait alors par la proposition de plans de continuité et de reprise d'activité (respectivement PCA et PRA), qui pourraient apporter des propositions de solutions organisationnelles et techniques.

La recherche et la minimisation des points individuels de défaillance, l'utilisation de technologies de sécurisation des échanges ou la redondance des services devraient aussi être évaluées.

1.5. Détection et offre de nouveaux services aux utilisateurs

Dernier point initial de réflexion : la possibilité d'améliorer et d'accroître l'offre de service à destination des utilisateurs du laboratoire.

Dans le cadre de ses activités, le service informatique a recensé des demandes de services provenant directement des utilisateurs. La réorganisation du système d'information pourrait être l'occasion de répondre aux besoins exprimés et d'offrir davantage de services aux utilisateurs du laboratoire.
L'existence d'un référentiel unique, alimenté par différentes sources de données, et la création d'interfaces orientés utilisateurs sont une base de travail intéressante pour un accroissement de l'offre de service.
L'effort serait donc orienté vers l'utilisateur final afin qu'il devienne un acteur du système d'information. Conformément à la réglementation de la CNIL, le nouveau système améliorerait la lisibilité des informations personnelles de chaque utilisateur et permettrait à ce dernier d'interagir efficacement et dynamiquement avec les services communs.

1.6. Synthèse

La mise en place de ces concepts paraissait intéressante pour l'organisation du système d'information du laboratoire. Elle permettait d'obtenir plusieurs résultats :

- De la cohérence entre les données du système d'information ;
- Une simplification et une pérennisation du processus de sauvegarde ;
- Une minimisation des points individuels de défaillance;
- Une augmentation de la qualité et de la sécurité du service ;
- Une offre de services élargie en direction des utilisateurs.

Ces nouveaux concepts, validés par la direction du laboratoire et la commission informatique, ont été le centre de ma réflexion durant mon stage d'ingénieur CNAM.

2. Choix de la méthode

Ce chapitre détaille les méthodes et l'organisation que j'ai utilisées pour mener à bien l'action qui m'a été confiée durant mon stage d'ingénieur CNAM.

2.1. Application d'une organisation projet

Afin de mener à bien le projet de migration du système d'information du laboratoire, j'ai réalisé une étude pour permettre de définir un cycle d'abstraction qui soit compatible avec l'intégration d'une telle évolution au sein du système en production.

Le mode projet m'a paru le plus adapté par sa constitution ; il permet notamment un découpage en sous-projets, puis en tâches dont la validation technique peut se faire par étape.
La Figure 3 présente le découpage du projet tel qu'il a été réalisé par l'équipe informatique.

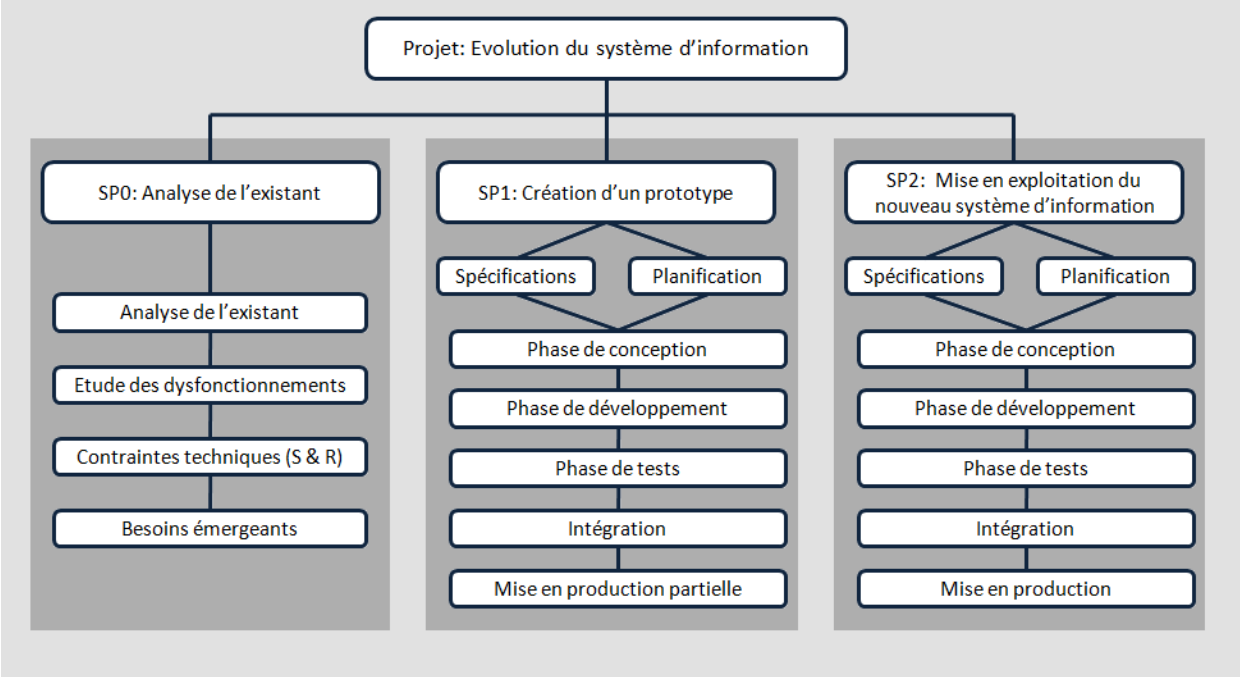


Figure 3 : Découpage du projet en sous-projets autonomes

Les trois sous-projets sont :

- **SP0 : analyse de l'existant**. Détection et recensement des besoins et des contraintes. Définition des objectifs et du prévisionnel ;

- **SP1 : création d'un prototype**. Développement et tests d'un prototype de système d'information en utilisant l'apport des nouvelles technologies. Mise en production partielle et analyse du comportement du prototype en phase de pré-production ;

- **SP2 : mise en exploitation du nouveau système d'information**. Ce sous-projet assurera l'évolution, l'amélioration et la mise en exploitation du nouveau système d'information basée sur les retours d'expérience de la phase de prototypage.
  On appliquera sur cette partie les critères de haute disponibilité inspirés des préconisations d'un plan de sécurité informatique adapté aux nouvelles technologies mises en œuvre.

Chacun de ces sous-projets a été développé de façon autonome. Le projet SP0 constitue la phase d'analyse et d'étude de l'existant. Les deux autres sous-projets (SP1 et SP2) constituent les réels développements de la nouvelle architecture.

Une fois les sous-projets identifiés, le service informatique a recensé les étapes nécessaires pour leurs développements. On trouvait pour chaque étape (des sous-projets SP1 et SP2) :

- Une phase de spécification, durant laquelle le service informatique a analysé les retours d'expérience sur le système d'information en production et établi les spécifications du nouveau système ;
- Une phase de planification où a été fixé le calendrier prévisionnel de chaque sous-projet ;
- Une phase de conception où les concepts ont été établis ;
- Une phase de développement où les nouvelles maquettes de scripts et d'IHM ont été développées ;
- Une phase de tests durant laquelle ont été testés le fonctionnement des développements et leurs procédures unitaires d'intégration dans l'architecture en production ;
- Une phase d'intégration où les nouveaux développements ont été implantés dans le système de façon progressive puis mis à l'épreuve face à plusieurs simulations d'erreurs ou de pannes ;
- Une mise en production (ou pré-production) du nouveau système développé.

2.2. Choix de la méthode de développement

De prime abord, nous avons étudié dans le cadre de notre projet la méthode de cycle de vie en V. Ce modèle permet de sécuriser l'avancée du projet en évitant les rétroactions et en assurant de ne passer à l'étape suivante que lorsque la précédente est terminée. De plus, l'accent est porté sur les phases anticipatives des développements que sont les phases d'analyse et de conception.
Par contre, la gestion dynamique des exigences liée notamment à l'émergence de nouveaux besoins ne nous permettait pas une définition exhaustive des besoins lors de la phase de spécification.
Devant le risque de voir se répéter les phases d'analyse et de spécification pour chaque sous-projet, nous avons pris conscience de la nécessité d'utiliser une autre méthode intrinsèquement adaptée à des processus de développements itératifs.
A partir de ces conclusions, l'orientation a été prise de privilégier la méthode du modèle de cycle de vie en spirale [6] [22], proposée par Barry Boehm en 1988. Chaque sous-projet, présenté dans la Figure 3, a été développé à l'aide de plusieurs itérations basées sur la méthode du cycle en V.

Modèle de cycle de vie en spirale

La démarche de la méthode Boehm était donc mieux adaptée à nos besoins puisque, pour chaque phase de développement, on retrouvait les étapes suivantes :

- La détermination des objectifs et des solutions envisageables à partir des résultats du cycle précédent et de l'analyse préliminaire des besoins ;
- L'analyse des risques et l'évaluation des possibilités à partir de phases de tests (maquettage) ;
- Le développement et la vérification de la solution retenue, sur le modèle du cycle en V ;
- L'évaluation des résultats et la planification éventuelle du cycle suivant.

Ce modèle de cycle de vie tient compte de la possibilité de réévaluer les risques en cours de développement. Il reprend les différentes phases du modèle de cycle de vie en V (troisième étape), mais permet, par l'implémentation, de versions successives d'améliorer et d'optimiser le développement global en proposant un produit de plus en plus complet. Cela permet une gestion du projet de façon plus managériale que technique.

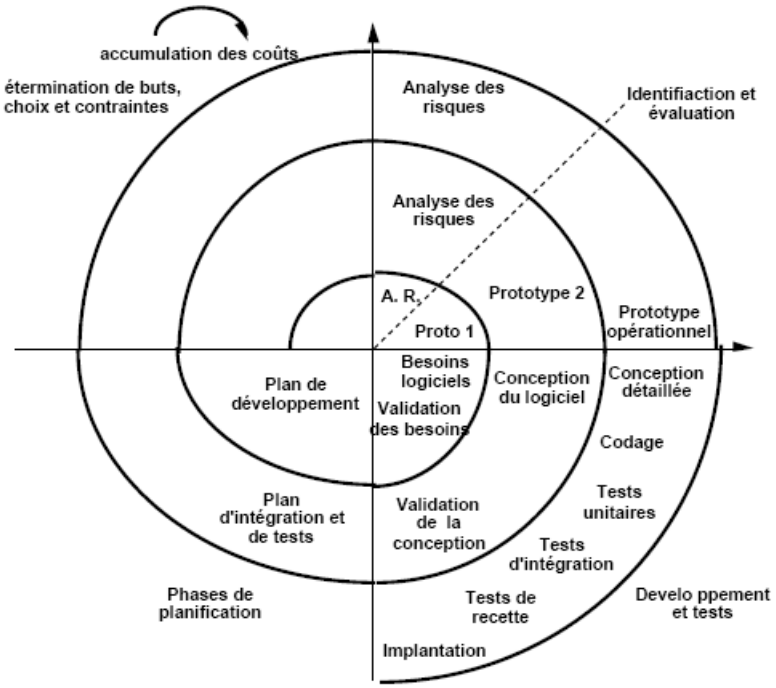


Figure 4 : Modèle de cycle de vie en spirale - Boehm (1988)

Cette méthode était conforme aux objectifs en permettant de :

- Développer une version de système d'information plus cohérente et adaptée aux besoins ;
- Améliorer la qualité du service offert et proposer de nouveaux services aux utilisateurs ;
- Augmenter la disponibilité du système d'information ;
- A plus long terme, pouvoir proposer une solution pouvant convenir à d'autres services.

La durée moyenne d'un cycle pour chaque sous-projet a été de deux mois.

L'inconvénient de la méthode Boehm est la gestion du temps. Il est en effet difficile de planifier une série de développements itératifs, sans connaitre les limites du projet initial. Le service informatique a donc fixé des jalons temporaires sur lesquels les phases de développements ont été calquées. La planification a été réalisée de manière indépendante, en amont des développements.

Modèle de cycle de vie en V

Au sein de chaque itération du cycle de vie en spirale, nous avons mis en pratique la méthode de cycle de vie en V (Figure 5).

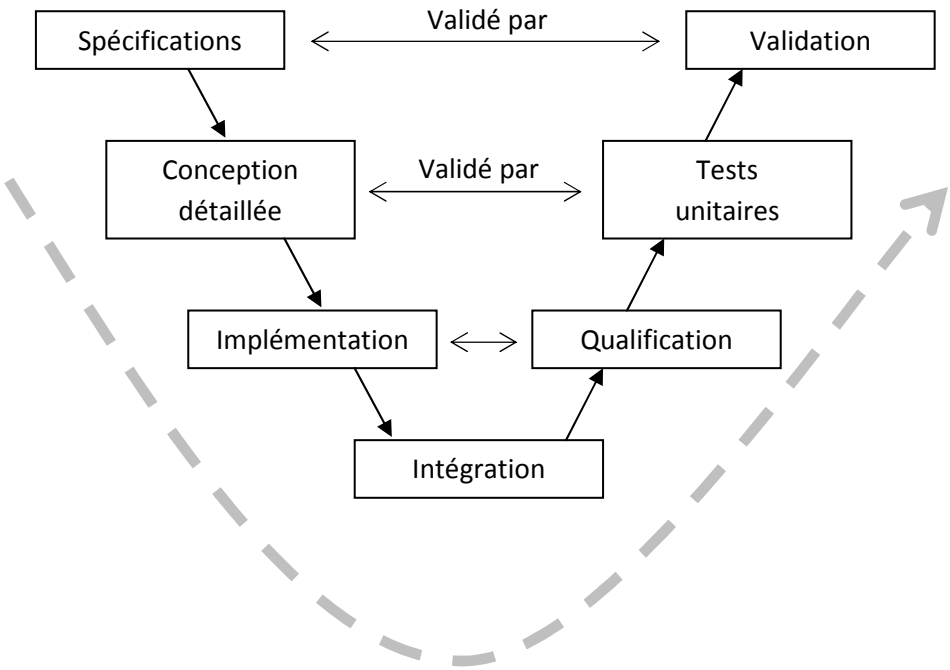


Figure 5 : Schéma du cycle de vie en V

Cette méthode de gestion de projet tient compte de la réalité de l'existant et permet de ne pas réduire le processus de développement à un enchaînement de tâches séquentielles.

En suivant le modèle de cette méthode, le service informatique a pu valider l'avancée des développements de façon progressive :

- En phase de spécification, les procédures de qualification du nouveau système ont été établies ;
- En phase de conception globale, les procédures d'intégration dans le système existant ont été testées et validées ;
- En phase de conception détaillée, les tests unitaires ont été réalisés sur des maquettes et des stations en production.

Cette méthode a notamment permis de développer les plans de tests et les documentations (technique et utilisateur) tout au long de l'avancée du projet.

Le modèle de cycle en V permet de sécuriser l'avancée d'un projet en évitant les rétroactions et en assurant de ne passer à l'étape suivante que lorsque la précédente est terminée. De plus, l'accent est porté sur les phases anticipatives des développements que sont les phases d'analyse et de conception.

3. Gestion de projet

3.1. Planification

La planification du projet d'évolution du système d'information s'est déroulée en plusieurs phases, chacune correspondant au découpage en sous-projet.

Planification du développement du prototype (sous-projet SP1)

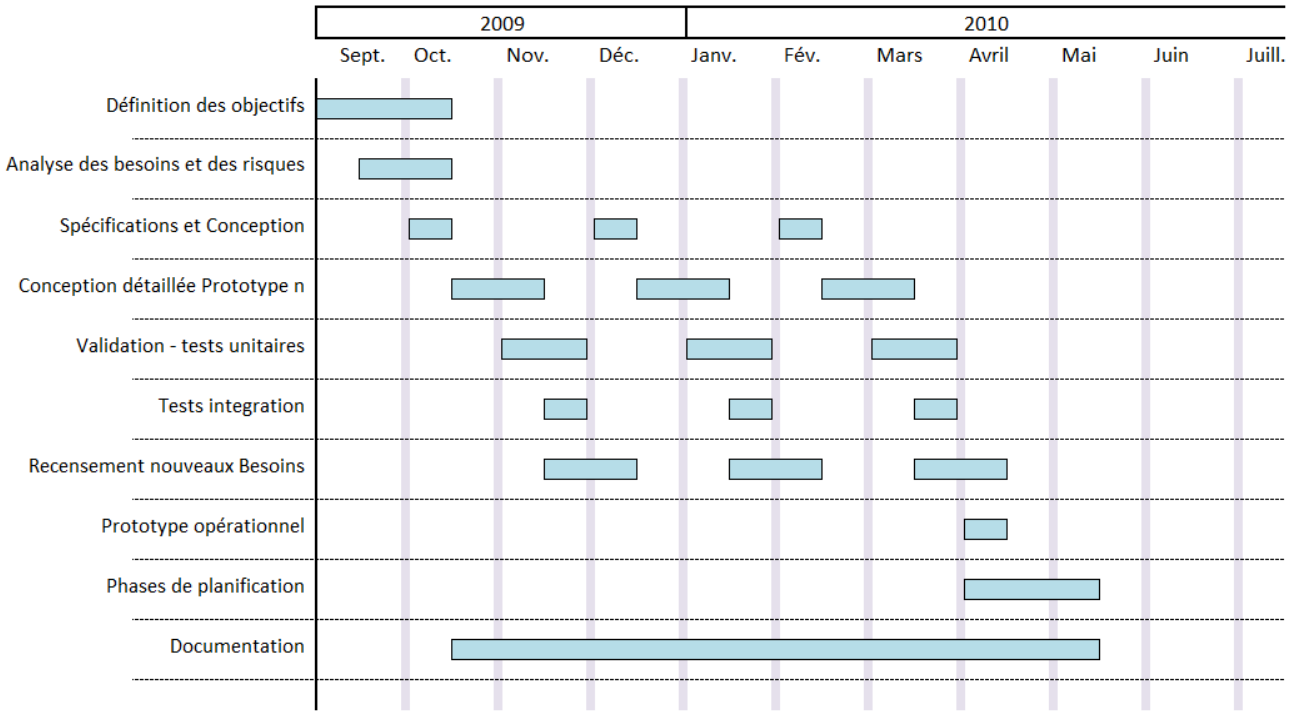


Figure 6 : Diagramme de Gantt – développement du Prototype

La première phase concerne l'élaboration et l'implantation du prototype (sous-projet SP1 : création du prototype). La durée moyenne de chaque cycle de développement de cette phase SP1 est de un à deux mois. Cette phase a débuté en septembre 2009 et s'est terminée en Avril 2010 (durée totale : 8 mois).
La répétition des phases de spécification, conception, validation, tests et recensement des nouveaux développements (phases du cycle en V) s'explique par l'application de la méthode Boehm au sous-projet.

Une phase de communication de nos résultats en a découlé, puis nous avons lancé le développement de la version finale en mai 2010. La version finale s'est nourrie des retours d'expérience du prototype, des besoins émergeants (utilisateurs et nouveaux services), et de la volonté de rendre le système d'information plus pérenne. Il a donc fallu étudier de nouvelles solutions et les mettre en place.

Planification de la mise en exploitation (sous-projet SP2)

Ce second diagramme de Gantt correspond à la dernière phase du projet, à savoir la mise en exploitation d'un nouveau système.

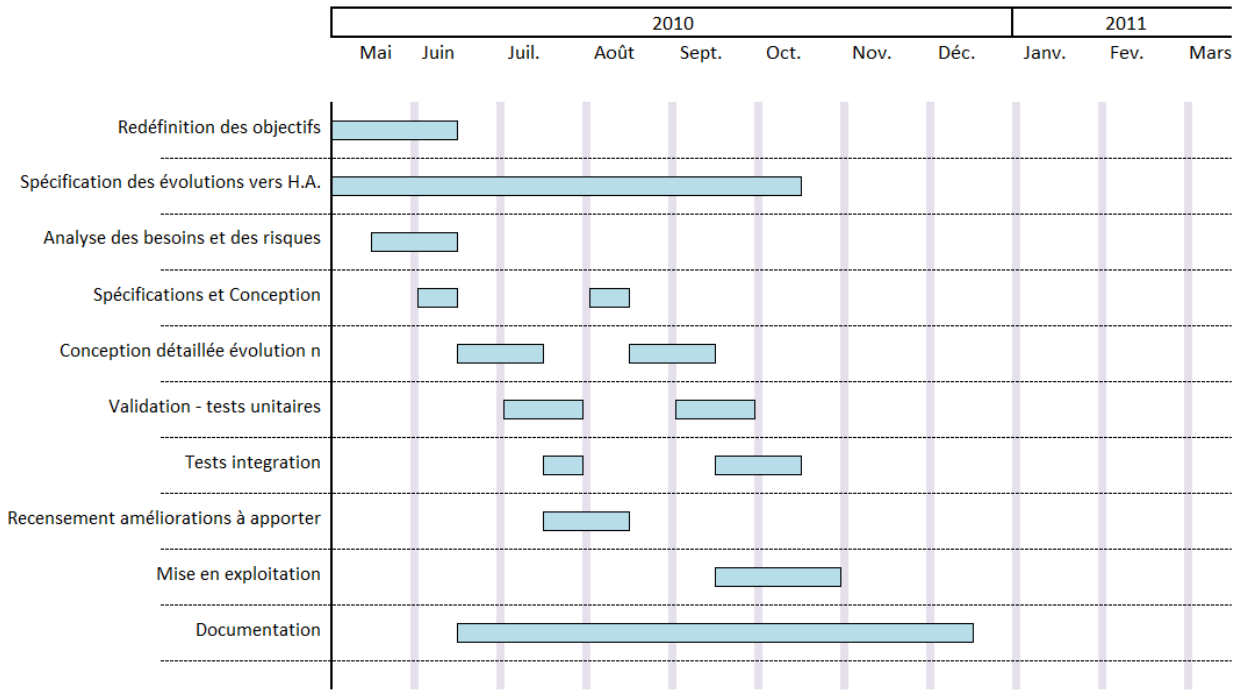


Figure 7 : Diagramme de Gantt de la mise en exploitation du nouveau système

Cette planification correspond donc à un nouveau cycle de développement, en tenant compte des retours d'expérience de la phase de prototypage et des nouveaux besoins recensés tout au long de la première phase.
Cette phase constitue donc la mise en exploitation du prototype ; elle a débuté en mai 2010 et s'est terminée, par l'implantation en production du nouveau système d'information, en octobre 2010.

Remarque : à la vue des résultats de ce projet et des demandes émergeantes de l'Observatoire Midi-Pyrénées, nous avons lancé un nouveau projet qui a débuté en janvier 2011 pour une prévision d'achèvement en novembre 2011. J'y reviendrai en conclusion de ce mémoire.

3.2. Pilotage et outils

Le pilotage et le suivi du projet ont été réalisés avec l'aide de Philippe Saby, mon responsable informatique au LATT, au travers de réunions régulières d'avancement et de validation des différentes étapes.
Durant ces réunions hebdomadaires, j'assurais la fonction de maître d'ouvrage (MOA) et réalisais :

- Un point d'avancement sur les développements en cours ;
- Le recensement des nouvelles exigences ;
- La répartition du plan de charge de chaque intervenant ;
- L'articulation du projet avec les tâches récurrentes du laboratoire.

Je pouvais alors préciser mes directives à mes collègues qui participaient au projet et occupaient davantage la fonction de maitre d'œuvre (MOE). J'ai néanmoins participé aussi à la MOE puisque la taille de l'équipe (quatre personnes) ne permettait pas disposer de suffisamment de personnel pour pouvoir dissocier les deux fonctions.

En termes d'outils de gestion de projet, j'ai principalement utilisé le logiciel de gestion de projet open source Openproj (Serena Software), puis ai redirigé mon utilisation vers Microsoft Project que je trouvais plus ergonomique.
Ces logiciels m'ont permis de réaliser la planification de mes développements et l'organisation des tâches de mon projet.
J'ai ensuite utilisé le logiciel de documentation en ligne (wiki) nommé Dokuwiki pour héberger les documentations techniques des solutions utilisées dans ce projet.
Pour plus de commodité, le logiciel Dokuwiki utilisé était celui de l'équipe informatique dans lequel nous avons développé une branche spécifique pour chaque technologie utilisée sur le projet.

III. Elaboration du prototype

Après avoir expliqué le concept d'architecture que nous souhaitions implanter, je vais détailler les moyens et méthodes que nous avons mis en œuvre pour déployer une première version. L'objectif de l'élaboration du prototype était de tester les nouveaux concepts organisationnels, d'apporter les ajustements nécessaires, d'observer leur mise en œuvre et de valider leurs fonctionnements.
Ces développements se sont déroulés entre octobre 2009 et avril 2010.

Je vais ici justifier le choix de l'architecture matérielle, des outils de développement de cette version de pré production et analyser les choix qui ont amenés l'équipe informatique à réaliser son intégration dans le système existant.

1. Nouveaux principes d'architecture

Je vais détailler ici les nouveaux concepts de réorganisation des services qui ont été développés lors de l'évolution du système d'information, en précisant les axes technologiques choisis en réponse aux problématiques posées.

Le nouveau système d'information devait être plus cohérent et moins sensible à l'environnement technologique. L'objectif était d'augmenter la qualité du service rendu et la disponibilité du système dans sa globalité. Les réponses que nous avons apportées ont été organisationnelles et techniques, et se sont structurées autour des axes suivants :

- La mise en place d'un référentiel unique de données ;
- L'augmentation de la disponibilité du système ;
- La sécurisation des accès au réseau filaire ;
- La décentralisation de la gestion des données utilisateur ;
- L'offre de nouveaux services aux utilisateurs.

1.1. La mise en place d'un référentiel unique de données

Afin de s'assurer d'une cohérence des données, le service informatique a pris la décision de mettre en place un référentiel unique. L'unicité des données à la source, et leur redéploiement dans les différents services, permet de s'assurer de la cohérence des données dans l'intégralité du système d'information.

Comme nous l'avons évoqué au chapitre précédent, le système d'information existant, bien que fonctionnel, révélait plusieurs carences, comme notamment la criticité du serveur LDAP, positionné comme point central de l'architecture (voir Figure 1).
Le service LDAP contenait toutes les données d'authentification et d'identification de tous les utilisateurs. Toutes les authentifications des serveurs du laboratoire (serveurs de fichiers, serveurs scientifiques, services de courrier électronique, de gestion des pages Internet, etc.) étaient réalisées par l'intermédiaire de ce service LDAP.

Les serveurs LDAP (deux au total) étaient dans cette organisation des machines physiques, susceptibles de subir une altération de leur fonctionnement d'un point de vue logiciel (système d'exploitation ou service défaillant, attaque informatique, etc.) ou matériel (panne d'alimentation électrique, défaillance d'un composant, surchauffe, etc.).

En cas de panne matérielle ou logicielle du serveur primaire, un service secondaire, répliqué en temps réel, assurait le service le temps que le service informatique remette en fonctionnement le serveur principal.
Malgré une sauvegarde externe des données réalisée chaque jour, les serveurs LDAP constituaient le plus important point individuel de défaillance de ce système d'information.
Un arrêt du service LDAP (serveurs primaire et secondaire) aurait entrainé un blocage de la majorité de l'informatique du laboratoire, jusqu'à ce qu'une intervention technique et manuelle du service informatique répare le dysfonctionnement.
En première intention, nous avons étudié la pérennisation du référentiel LDAP existant.

Un des objectifs du service informatique du LATT était de décentraliser la gestion des utilisateurs, et donc de multiplier les accès en écriture sur la source de données utilisateur (jusqu'alors le service LDAP). Or, par la nature même de son protocole d'accès, l'annuaire LDAP [5] apparaissait immédiatement comme inadapté pour gérer les données persistantes dans un contexte d'accès concurrentiels (cf Annexe 1 : Le protocole LDAP). La surcharge des opérations d'écriture dans le cas d'une base gérée et modifiée par plusieurs intervenants pourrait conduire à une perte partielle de données sources ou au blocage du service, et par extension au blocage du système d'information.

Pour cela, la décision a été prise de recenser l'ensemble des sources de données réparties dans le système existant et de faire remonter toutes les données dans une base de données (SGBD).
Après requalification des données, la nouvelle base de données recenserait toutes les données du système d'information existant (données utilisateur, données stations de travail, etc.) sous forme unique (suppression des doublons, etc.) et cohérente (cycle de vie du compte utilisateur maitrisé). Ce service pourrait être sauvegardé de façon simplifiée (source unique).
De plus, cette source unique de données utilisateurs et système (caractéristiques techniques des stations de travail des utilisateurs, des serveurs, matériels actifs, imprimantes, etc.) permettrait aussi de générer dynamiquement les configurations des services réseaux, assurant ainsi une parfaite cohérence des configurations des services du système d'information.

1.2. L'augmentation de la disponibilité du système

Les réponses technologiques à la problématique d'augmentation de la disponibilité du système d'information ont été d'ordre technique, mais aussi d'ordre organisationnel.

1.2.1. Minimisation des points individuels de défaillance

La minimisation des points individuels de défaillance du système et l'évaluation d'une nouvelle organisation va constituer la première évolution du système.
La suite logique de la création du référentiel unique, reposant sur un SGBD, a été la mise en place d'un système de réplication de la base de données.

Parallèlement, nous avons cherché à diminuer le niveau de criticité de l'annuaire LDAP. Pour cela, l'idée principale a été d'extraire les données du service LDAP existant et de les intégrer dans le référentiel unique. A partir de ce moment-là, nous pouvions générer dynamiquement le contenu des services LDAP du laboratoire de façon automatisée (par scripts).

De fait, l'annuaire LDAP a été repositionné comme une source de données accessible uniquement en lecture (suppression des opérations d'écriture directes) exploitant ainsi les caractéristiques du protocole LDAP sans risque pour l'intégrité des données.
De même, l'application de cette méthode aux services réseaux (DNS, DHCP, etc.) nous permettait de générer dynamiquement les fichiers de configuration des services à partir du référentiel unique. Cette méthode permettait, d'une part, de limiter la défaillance liée à la perte des fichiers de configuration et, d'autre part, de supprimer les erreurs de saisie manuelle.

En conclusion, grâce à cette nouvelle architecture de données, nous allions minimiser les principaux points individuels de défaillance et augmenter la disponibilité du système.

1.2.2. Evaluation de nouvelles technologies informatiques

L'évaluation d'une nouvelle organisation basée sur la mise en place de nouvelles technologies était un axe que le service informatique souhaitait explorer. Dans ce contexte, les technologies de virtualisation des serveurs et l'authentification des accès au réseau constituaient les deux principaux points techniques que le LATT souhaitait évaluer.

L'implantation des technologies de virtualisation permettrait de mettre en place une architecture plus souple et plus pérenne. Un des apports de la virtualisation était d'intégrer au sein de fichiers à la fois la couche d'abstraction matérielle, le système d'exploitation et les données associées. La possibilité de cloner des systèmes entiers sur disque et de pouvoir les réimplanter sans se soucier de la couche matérielle était le principal attrait de cette technologie.

La baisse permanente des coûts de stockage permettrait aussi d'envisager la possibilité de multiplier les sauvegardes complètes sur disque, autorisant une disponibilité immédiate des sauvegardes réalisées (hors archivage, toujours sur bandes). En effet, si en l'an 2000 le coût du Terra-Octets (To, soit environ 1000 Go) avoisinait 40 k€, en 2009 ce coût avait chuté à environ 500€, soit une baisse d'un facteur 80. Par exemple, en avril 2009, le LATT s'était doté de deux serveurs NAS de 20 To chacun pour un montant de 20k€.

De plus, dans l'optique de générer dynamiquement les services à partir d'une base unifiée, les technologies de virtualisation apportent un avantage supplémentaire : la possibilité de détenir des serveurs prêts à être reconfigurés et mis en production en quelques minutes seulement, permettant une reprise d'activité sensiblement accélérée. La criticité des points individuels de défaillance identifiés serait ainsi de fait diminuée.

Il restait à analyser l'adéquation entre les besoins, les solutions techniques étudiées et les moyens financiers disponibles. Pour cela, le service informatique a pris la décision d'évaluer les différentes technologies de virtualisation présentes sur le marché en 2009/2010 et de définir une stratégie sur une éventuelle mise en place de ces technologies dans son environnement.

1.3. La sécurisation des accès au réseau filaire

L'autre axe de réflexion concernait la sécurisation des accès au réseau filaire. La mise en place de la nouvelle infrastructure réseau avait doté le LATT de commutateurs capables de réaliser l'authentification à la source dès la connexion physique des matériels.

La mise en œuvre de ces technologies devait être étudiée afin de sécuriser et de pérenniser l'accès au réseau et aux services du laboratoire.

1.4. La décentralisation de la gestion des données utilisateur

L'optimisation de la gestion des données utilisateur était un pré-requis formulé autant par les utilisateurs que par les intervenants du processus de gestion administrative.
La décentralisation de la gestion de ces données au plus près de la source d'information était donc l'orientation logique choisie par le service informatique. La mise en place du référentiel unique, basé sur la base de données unifiée, permettait d'autoriser des accès concurrentiels à la source de données.

Cette modification permettait d'envisager le développement d'Interfaces Homme-Machine (IHM) d'accès et de mise à jour des données orientées vers tous les intervenants et utilisateurs du système d'information (service du personnel, commission des locaux, utilisateur).
Les utilisateurs et les intervenants pourraient consulter et demander la mise à jour de leurs informations personnelles en temps réel, sous modération de personnels autorisés et identifiés (secrétariat de direction, service de gestion des ressources humaines, direction du laboratoire, etc.).
Le délai de mise à jour des données serait ainsi optimisé et réalisé par les personnels qualifiés et informés. L'équipe informatique pourrait ainsi se focaliser sur sa mission initiale de mise à disposition et de pérennisation de ces données (sauvegarde, archivage, restauration).

1.5. L'offre de nouveaux services aux utilisateurs

La création d'IHM de gestion des données utilisateurs constituait le premier pas vers la nouvelle offre de services que l'équipe informatique désirait mettre en place pour les utilisateurs du laboratoire.
Le développement et l'intégration d'interfaces supplémentaires permettrait, après l'enrichissement de la base unifiée avec des données endogènes et exogènes aux utilisateurs, d'étoffer l'offre de services.

En effet, de façon régulière, l'équipe informatique recevait des demandes de renseignements, concernant autant des informations systèmes (utilisation de disques serveurs, disponibilité de services, etc.) que des informations administratives telles que les demandes de codes d'accès pour les ressources en libre service (salles, machines, photocopieurs, etc.).

La mise en place de ce système devait apporter une réponse aux demandes d'informations des utilisateurs, mais aussi pouvoir répondre dynamiquement, de façon régulière et automatisée, à des demandes difficiles à gérer au quotidien de façon manuelle (quotas d'espace disque par exemple).

D'autres services pourraient aussi être ajoutés ultérieurement, en fonction des nouveaux besoins que cette première offre pouvait provoquer chez les utilisateurs, conformément à la méthode de développement en spirale choisie pour le projet.

2. Conception et développement du prototype

Une fois l'ensemble des évolutions énumérées, nous avons débuté la phase de développement. Je vais détailler dans ce chapitre les choix de conception et de développement du prototype.

2.1. Plan d'action

L'objectif de la création du prototype du nouveau système d'information était de tester le fonctionnement des nouveaux concepts avant la mise en production. Il s'agissait donc de réaliser une maquette pouvant être éprouvée à échelle réelle et qui permettait de valider l'intégration des nouveaux concepts. Les principales lignes directrices pour atteindre ce premier objectif étaient les suivantes :

- **Réaliser une architecture autonome pouvant permettre de tester notre maquette** sans incidence sur le système d'information en production. Cette architecture devait être compatible avec l'architecture actuelle puisqu'elle devait, en cas de validation positive de la phase de tests, s'intégrer et faire évoluer le système d'information existant sans coupure de service ;

- **Valider le fait que l'on pouvait collecter, qualifier et rendre cohérentes les données « éparpillées »** dans le système d'information en production. L'objectif de création d'une architecture indépendante a conduit au déploiement de nouveaux serveurs, disjoints des serveurs en production. Pour simplifier la maquette, le choix a été d'utiliser le système de sauvegarde existant pour effectuer les sauvegardes journalières et hebdomadaires ;

- **Valider le fait que l'on pouvait générer des configurations de services** à partir d'une source de données. La maquette devait valider que l'on puisse générer dynamiquement des configurations de services (DNS, DHCP, ou autre) et les faire fonctionner. La génération dynamique de ces services devait se faire à partir de scripts qui utiliseraient les données du référentiel pour générer, à la demande, leurs fichiers de configuration ;

- **Sécuriser l'accès au réseau informatique** : étudier et analyser le fonctionnement et l'implantation des technologies d'identification matérielle qui pourraient permettre aux commutateurs réseau de sécuriser la connexion des stations de travail au réseau filaire ;

- **Concevoir et développer un système de modération permettant de déléguer la saisie** au plus près des intervenants détenant l'information. Cette action allait nous permettre d'améliorer la cohérence et la véracité des données du système notamment pour la gestion des ressources humaines et des locaux.

Il s'agissait donc d'une évolution majeure de la gestion des données et des services du système, mais aussi du rôle même des intervenants dans le système d'information.

2.2. Création du référentiel unique de données

Je vais présenter ici le développement et l'implantation du référentiel unique de données.

La dispersion des données utilisateur dans le système d'information était le premier point dur à résoudre dans l'optique de la mise en place les nouveaux concepts. Source de données unique et socle technique d'offre des nouveaux services, ce référentiel se devait de recenser les données du système de façon exhaustive et cohérente.

2.2.1. Identification des données existantes

Les données existantes étaient réparties, pour les utilisateurs, dans l'annuaire LDAP, et pour les stations de travail, dans la base GLPI. Il a fallu analyser les données sources de chaque application et en extraire les données utiles.
L'application GLPI, par exemple, disposait de sa propre base de données dont l'information utile concernant les stations de travail n'était contenue que dans une vingtaine de tables sur les 160 que comptait le progiciel.
La méthode que j'ai appliquée pour extraire les données de cette application de gestion de parc a été d'effectuer une retro conception (« reverse engineering ») de la base de données, d'analyser le contenu de chaque table ainsi que leurs dépendances afin d'obtenir le schéma relationnel initial. Une fois le schéma logique et conceptuel établi, j'ai pu vérifier les dépendances fonctionnelles de la base et identifier toutes les tables contenant les données utilisateurs.

La seconde partie des données, orientées utilisateur, se trouvait dans l'annuaire LDAP en production. Il s'agissait plus exactement d'inventorier les données d'identification des utilisateurs (nom, prénom, login, etc.) parmi l'ensemble des données structurelles de l'annuaire LDAP.

Ces deux actions m'ont permis d'obtenir une liste exhaustive des champs et des données détenus par le système, que j'ai recensée dans un tableau de données. J'ai ensuite, pour les types de données, supprimé les doublons et complété ce tableau avec les nouveaux types nécessaires aux nouveaux développements.
Ce processus de nettoyage et de recensement a permis de créer le dictionnaire de données qui a servi de base à l'étape de modélisation du schéma de la nouvelle base de données unifiée.

2.2.2. Modélisation du schéma de la base de données

Pour effectuer la modélisation du schéma de la nouvelle base, plusieurs réunions regroupant l'ensemble des personnels du service informatique ont été organisées afin de réaliser une modélisation optimale. Ces réunions se sont déroulées d'abord sur le modèle de réunion de type « brainstorming » (remue-méninges) afin d'essayer de compléter le dictionnaire de données, puis sur des réunions de modélisation de base de données.

Après plusieurs discussions, le service informatique a pu développer un schéma adapté à l'environnement du laboratoire d'astrophysique.

La nouvelle base de données, regroupant l'ensemble des données du système, a été développée après plusieurs raffinages successifs, en s'appuyant sur les étapes itératives de la méthode Merise (Méthode d'Etude et de Réalisation Informatique pour les Systèmes d'Entreprise). L'utilisation de cette dernière nous a permis d'établir le schéma entités-associations, de créer le modèle conceptuel des données (MCD), d'en déduire les modèles logique et physique des données (respectivement MLD et MPD).

Voici le modèle conceptuel de données du prototype, modèle représentatif du schéma de la base (hormis les contraintes d'intégrité) :

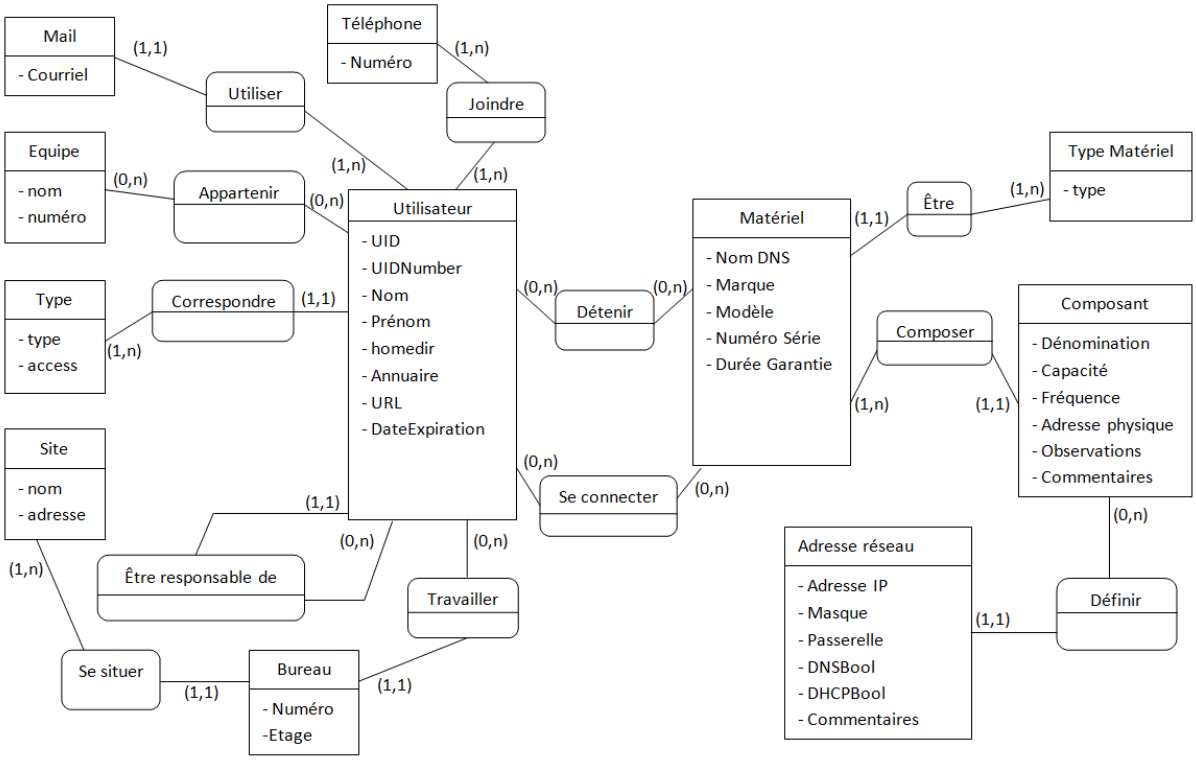


Figure 8 : Modèle conceptuel des données du prototype

Sur ce modèle conceptuel des données, il ressort visuellement qu'il existait deux familles distinctes de données utilisateur :

- Les données concernant le compte de l'utilisateur et ses caractéristiques (partie gauche) : nom, prénom, login, répertoire personnel, etc. principalement extraites des annuaires LDAP.
- Les données concernant les matériels de l'utilisateur (partie droite) : type de matériel, caractéristiques techniques du matériel, etc. principalement extraites du serveur GLPI.

Une fois ce modèle défini, nous avons généré nos modèles logiques et physiques de données.
Dans cette phase de prototypage, le choix du système de gestion de bases de Données (SGBD) s'est naturellement porté, pour des raisons de continuité et d'expertise, vers le progiciel MySQL[1].
Pour finir, j'ai installé ce progiciel sur les serveurs dédiés au prototype et créé la base et les tables correspondantes au MPD établi.

[1] *http://www.mysql.fr/*

2.2.3. Alimentation de la nouvelle base de données

A partir des actions décrites dans les paragraphes précédents, il restait à alimenter la base du prototype avec un processus d'importation des données réparties dans le système existant. Pour ce faire, j'ai procédé successivement aux étapes de création d'une base temporaire, de récupération et de traitement des données existantes et d'intégration dans le référentiel unifié.

Utilisation d'une base de données temporaire

Lors de cette première étape, pour simplifier ce processus de traitement, j'ai privilégié l'utilisation d'une base temporaire. Cette dernière avait comme structure un ensemble de tables qui reflétaient, de façon exhaustive, l'organisation de l'existant. Plus concrètement, nous avions une table pour chaque source issue du LDAP, DHCP, DNS et une table par type de données pour GLPI (utilisateur, machines,…).
A partir de scripts, développés en langage Perl, j'ai procédé à l'extraction des données des services en production et les ai intégré dans les tables temporaires correspondantes.

Qualification des données et intégration dans la base définitive

A ce stade, la base de données regroupait la collecte exhaustive des données utilisateur du système sans cohérence ni qualification, ce qui permettait de travailler sur les données sans incidence sur le système en production.
Le regroupement des données brutes a mis en évidence les problèmes identifiés par l'équipe informatique, *i .e.* des doublons d'information, des données incohérentes entre les services, des informations périmées ou manquantes, etc.

Afin de qualifier ces données, j'ai développé de nouveaux scripts (Perl) permettant de les intégrer après traitement vers la base du prototype. Ce processus automatisé de qualification et de mise en cohérence a été précédé d'une validation manuelle, notamment pour le traitement des informations obsolètes.

Le résultat obtenu était une base de données unique et cohérente, contenant l'ensemble des informations présentes de façon dispersée dans le système d'information en production.

A ce stade, nous disposions d'une source unique d'informations à jour et pouvions désormais développer les scripts d'alimentation de nos services et les IHM associées.

2.3. Développement des scripts et des IHM

L'objectif principal de la création de notre référentiel unique était de pouvoir obtenir une meilleure cohérence entre les données du système d'information. Pour cela, la génération dynamique des configurations des services réseaux à partir du référentiel et l'alimentation automatisée des données des services d'authentification à partir de ce même référentiel constituaient les pierres angulaires de la nouvelle organisation du système d'information.
De même, nous devions développer un lien privilégié avec les utilisateurs au travers d'interfaces simples, riches et intuitives.

2.3.1 Choix des langages de développement

Dans le cadre du développement du prototype, j'ai effectué moi-même les développements initiaux des scripts et des IHM. Ces développements ont été réalisés en parallèle de la charge quotidienne de mon travail d'administrateur. J'ai donc privilégié l'utilisation de technologies connues et maitrisées pouvant s'articuler facilement dans notre environnement et ne nécessitant pas d'investissement important en formation. Le choix des langages a donc été réalisé de façon réfléchie mais opérationnelle de ma part pour le développement du prototype.
Le service informatique avait décidé que, par la suite, en cas de validation des concepts et de mise en production, une étude plus approfondie des langages de scripts, de l'ergonomie des IHM et de leurs langages de développement serait réalisée.

Choix du langage de script
L'environnement de travail du laboratoire d'Astrophysique étant majoritairement composé du système Linux, je me suis logiquement dirigé vers des langages de scripts connus par la majorité des administrateurs réseaux.
Dans le cadre de ces métiers et dans des environnements « open source », les langages de script couramment utilisés sont les langages « shell » (Bash, Ksh, etc.), Python et Perl.
J'ai choisi de créer ces scripts en langage Perl car la connaissance de ce dernier était partagée au sein de l'équipe. De plus, le langage Perl est un langage disponible de façon native sur les systèmes Linux (intégrés dans les distributions), disposant des connecteurs de bases de données éprouvés pour de nombreux SGBD, et dont la compatibilité est reconnue avec les serveurs Internet (modules pour Apache).

Choix du langage de développement des Interfaces Homme-Machine
Pour le développement de nos IHM, nous avons décidé d'utiliser la technique du client léger, sur une architecture N-Tiers. Cette solution apportait plusieurs avantages :

- Une indépendance du fonctionnement de l'application vis-à-vis du système d'exploitation. L'environnement hétérogène du laboratoire amenait en permanence à proposer des solutions compatibles sur toutes les plateformes (Windows, Linux, Mac). L'application client pouvait alors interroger le serveur via un navigateur Internet en toute indépendance vis-à-vis du système d'exploitation ;
- Hormis la mise à jour du navigateur, aucune installation n'était nécessaire sur les postes de travail des utilisateurs ;
- L'application était accessible par toute personne autorisée, depuis n'importe quelle station de travail située sur le réseau du laboratoire.

Ces trois aspects étaient primordiaux puisque l'application allait notamment permettre aux utilisateurs de déclarer leurs matériels dans notre base de données. Sans cet accès facilité, ils n'auraient pas pu obtenir de façon simple leur connexion au réseau filaire. Ils devaient donc pouvoir s'authentifier et accéder à leurs services à partir de n'importe quel point d'accès du réseau, indépendamment du type de matériel ou du système d'exploitation utilisé.

J'ai donc choisi de créer les interfaces en langage PHP. Le langage PHP, comme langage Perl, disposait de connecteurs éprouvés avec de nombreux SGBD et serveurs Internet. Il était stabilisé (intégration native dans Oracle par exemple depuis Oracle 9i) et suffisamment simple d'accès pour être repris par d'autres intervenants potentiels sur le projet (évolutivité du projet).
De plus, c'était aussi un langage qui était déjà maitrisé par la majorité des intervenants du service informatique, ce qui permettait d'utiliser le temps disponible pour la recherche et la formation sur les autres technologies du projet.

2.3.2 Développement des scripts de génération des services

Les scripts développés permettaient de récupérer les informations du référentiel unique et de redistribuer les données de façon cohérente dans le système d'information. Pour cela, ces scripts ont permis la génération des configurations des services réseaux et l'alimentation dynamique des annuaires d'authentification.

Les scripts permettant la génération dynamique de la configuration des services réseaux, présentés en Figure 9, apportent plusieurs avantages.

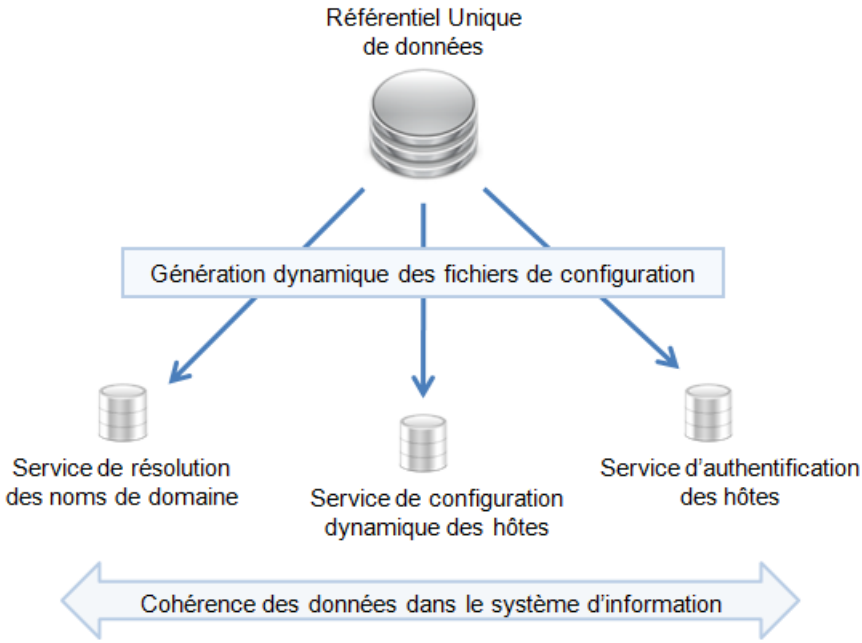


Figure 9 : Cohérence des données grâce au référentiel unique

Tout d'abord, la génération dynamique, à partir du référentiel unique, assure que chaque configuration de service est en cohérence avec la configuration des autres services du système d'information. Dès lors, la cohésion est totale entre tous les services générés de la même façon.

Ensuite, l'administration des services est simplifiée, puisque chaque modification est réalisée à l'intérieur du référentiel unique et non plus dans une configuration de service. Il n'y a donc plus de risques d'introduire manuellement des incohérences dans le système.

De plus, la rapidité d'exécution et la facilité de traitement pour propager des modifications sont largement optimisées comparativement à une intervention manuelle.

Pour finir, la possibilité de générer dynamiquement des configurations de services permet de baisser significativement la criticité des services par la possibilité de régénérer les services de façon automatisée et rapide.

Les scripts d'alimentation dynamique des annuaires d'authentification ont ainsi permis une baisse de la criticité du service.
En premier lieu, le fait de pouvoir générer le contenu du service d'authentification permet de le positionner en mode « lecture seule » sur le réseau. Cet avantage répond à la problématique des accès concurrentiels évoquée dans les évolutions souhaitées, puisque les modifications de données se font dans le référentiel unique et non plus dans l'annuaire. Le service d'authentification ne constitue plus que la mise à disposition pour l'authentification d'une extraction des données utilisateur du référentiel unique.

En second lieu, la possibilité de créer un service d'authentification de façon dynamique permet de multiplier les services d'authentification de façon modulaire. Il est dès lors possible de disposer de services secondaires ou tertiaires à moindre investissement, mais aussi de générer des services spécifiques supplémentaires à volonté (base d'authentification du pare-feu par exemple).

Au niveau configuration, il n'y a plus de nécessité de répliquer les services puisque chacun dispose de données provenant de la même base, ce qui simplifie aussi l'administration des serveurs.
Pour finir, la sauvegarde des services d'annuaires LDAP, point critique au lancement de l'action, n'est plus effectuée pour cette même raison.

Cette génération dynamique a donc permis de minimiser le plus important point individuel de défaillance, et donc d'augmenter la disponibilité du système d'information.

Il est à noter que d'autres scripts avaient été créés pour alimenter le référentiel unique et qualifier les données récupérées dans le système d'information en production (voir paragraphe 2.2.3. Alimentation de la nouvelle base de données). Ces scripts ont été utilisés à la création du référentiel et archivés ensuite.

2.3.3. Développement des Interfaces Homme-Machine

La mise en place de ce référentiel a nécessité le développement d'interfaces spécifiques de manipulation et de gestion des données. Elles ont fait l'objet de développements successifs, en fonction des nouveaux besoins rencontrés.

Deux catégories d'interfaces ont été développées : des interfaces de manipulation et de gestion des données pour les administrateurs, ainsi que des interfaces de consultation et d'offre de nouveaux services pour les utilisateurs et les intervenants.
A partir de l'intranet du laboratoire, chaque utilisateur authentifié disposait, grâce à ces IHM, d'un espace personnalisé en fonction des droits associés à son profil ; utilisateur, modérateur (responsable hiérarchique, secrétariat de direction, etc.) et administrateur (service informatique). Chaque niveau de profil permettait d'accéder à davantage de fonctionnalités dans l'interface.

IHM de consultation et d'offre de nouveaux services

La création d'IHM d'interrogation et de gestion des données utilisateurs constituait le premier pas vers la nouvelle offre de service. De plus, afin de respecter la réglementation de la CNIL, chaque utilisateur devait pouvoir accéder à ses données personnelles et soumettre des demandes de modification de ces données.

Le développement et l'intégration d'interfaces supplémentaires permettait, après avoir complété la base unifiée avec les données exogènes des utilisateurs, d'enrichir le catalogue des services :

• La création d'un espace personnel, issu de la nouvelle base de données, où l'utilisateur pouvait consulter et demander la modification de ses données personnelles;

• La mise en place d'un tableau de bord avec l'état journalier de ses comptes de messagerie ainsi que de ses espaces disques alloués;

• La mise à disposition d'informations utiles et contextuelles : codes d'accès pour les zones protégées, codes d'accès pour les matériels de reproduction (photocopieurs, imprimantes, etc.), informations confidentielles à usage spécifique ou encore réservation de matériels communs;

• L'accès à un annuaire plus complet disposant d'informations supplémentaires par rapport à l'annuaire officiel public du laboratoire et offrant un module de recherche multicritères innovant.

Ces nouveaux services proposés provenaient principalement du recensement des demandes d'informations reçues par l'équipe informatique tout au long de l'année. D'autres services pourraient ainsi être ajoutés ultérieurement, en fonction de l'émergence de nouveaux besoins.

IHM d'administration des données

Afin d'effectuer la gestion quotidienne et de faciliter le travail d'administration, des interfaces de manipulation des données ont été développées. Ces interfaces s'adressaient aux administrateurs du système (service informatique) et aux modérateurs.

Les modérateurs étaient constitués des personnels ayant une fonction administrative ou ayant autorité pour effectuer des modifications sur le référentiel des données (changements de bureau, de numéros de téléphone ou validation des demandes de prolongation de comptes). Il s'agissait principalement du responsable de la commission des locaux, de la direction et de son secrétariat.

Les interfaces, réservées uniquement aux administrateurs, facilitaient l'exécution des tâches quotidiennes de gestion et d'administration des données relatives à l'utilisateur et à ses postes de travail.

Le recensement des principales fonctionnalités apportées par ces IHM était le suivant :

• Un tableau de bord de gestion permettant la visualisation de l'ensemble du parc, la validation de stations de travail que l'utilisateur a déclaré sur son espace personnel, l'ajout de nouveaux matériels et la gestion des droits d'accès ;

• Des IHM de gestion des services réseau : ces IHM permettaient de gérer de façon rapide l'ensemble des caractéristiques réseaux des matériels du laboratoire. Elles permettraient aussi de déclencher manuellement la génération des configurations dynamiques des services réseaux (authentification, attribution dynamiques d'adresses, résolution des noms de domaines). L'intérêt était de conserver la cohérence des informations du système en s'appuyant sur les scripts existants.

• L'accès à un annuaire exhaustif présentant la totalité des informations du référentiel : données administratives, données informatiques (systèmes) et matériels enregistrés. Ces données étaient la concaténation de toutes les données réparties auparavant dans les différents systèmes et disponibles à volonté de façon synthétisée.

Ces scripts et IHM ont été développés en collaboration avec mes collègues du service informatique. Nous nous sommes réparti les tâches en fonction du temps disponible, des compétences de chacun et des besoins journaliers du service. J'ai réalisé personnellement les scripts d'intégration de données dans le nouveau référentiel et l'ensemble des interfaces homme-machine du prototype.

Des exemples d'IHM, en phase de prototypage, sont fournis en Annexe 5.

2.4. - Choix de l'architecture matérielle

L'évolution d'un système d'information, comme l'implantation de nouveaux services, induit généralement l'implantation d'une nouvelle architecture matérielle et logicielle. L'évolution d'un système d'information est fortement corrélée avec la durée de vie du matériel qui supporte l'infrastructure technique. De même, la fin des garanties des matériels, leur obsolescence, ou l'achat de nouveaux matériels ont comme conséquence l'évolution de tout ou partie du système d'information.

En lien avec l'objectif de mon stage, la mission d'évaluer la mise en place des technologies de virtualisation m'a été confiée. La virtualisation est une technologie qui avait été déjà été testée dans une problématique de gestion de parc afin de reproduire ou de simuler des installations et tests de progiciels sur les stations de travail.
Quelques tests de virtualisation de serveurs avaient aussi été réalisés, principalement pour des serveurs de projets scientifiques qui ne nécessitaient que des services de base : service Internet (site vitrine d'un projet, administré par l'équipe projet) ou un service de transfert de fichiers de type FTP (échange de données projet entre différents laboratoires). L'application de ces technologies avait alors donné entière satisfaction.

J'ai donc réalisé, fin 2009, une étude comparative multicritères des technologies qui apparaissaient comme abouties au moment du lancement de cette action. J'ai notamment évalué les avantages et les inconvénients que pouvait apporter la virtualisation de l'infrastructure *vs* l'évolution de l'architecture physique existante et proposé un choix matériel et logiciel adapté.

2.4.1. Comparaison quantitative des architectures système physiques et virtualisées

L'intégration des technologies de virtualisation sur l'infrastructure du LATT devrait initialement servir les services d'infrastructure. Par la suite, nous pourrions étendre l'application de ces techniques aux serveurs scientifiques de diffusion de l'information (serveurs Internet, transferts de fichiers, etc.).

Pour comparer les architectures physiques et virtualisées, j'ai tenu compte du nombre de serveurs existants au laboratoire et de leurs utilisations. Le système d'information du laboratoire était composé d'une vingtaine de serveurs d'infrastructure pour les services DNS, DHCP, Mail, Antivirus centralisé, NTP, Annuaires (LDAP), service Internet, services de bases de données (MySQL), services de métrologie (NAGIOS), services de gestion des logs système (SYSLOG), etc.
Les serveurs d'infrastructure consommaient peu de ressources sur chaque matériel : le taux d'occupation moyen des serveurs constaté était d'environ 20% et aucun des serveurs du laboratoire n'était exploité à plus de 40%.
La même constatation pouvait être faite sur l'utilisation des ressources disques. L'emplacement disque utilisé était principalement occupé par le système d'exploitation, le service lui-même ne consommant qu'une faible partie de l'espace occupé. Pour des raisons de sécurité, chaque serveur d'infrastructure utilisait ses disques internes, sans mutualisation d'espace.
En moyenne, le taux d'occupation constaté de ces disques était de 30%.

A cette vingtaine de serveurs s'ajoutait le nombre de serveurs dédiés à la recherche permettant de diffuser de l'information auprès de la communauté scientifique (satellites, observations, etc.). Nous en dénombrions une vingtaine mais leur nombre était en augmentation au fil des années.

Pour comparer ces deux types d'architecture sur des critères quantitatifs, je me suis appuyé sur une analyse financière basée sur les coûts d'implantation et les coûts énergétiques induits de chacune des solutions, de la même façon que N. Rasmussen dans son livre blanc [9]. J'ai ensuite analysé les capacités techniques disponibles, à terme, pour chacun des services dans les deux infrastructures.

Cette évaluation a été réalisée pour les serveurs d'infrastructure, puis pour les serveurs d'infrastructure accompagnés des serveurs scientifiques (passage de 20 à 40 serveurs).

Critères de mesure

La consommation énergétique est notée en kilowatts (kW, 1kW=1000 Watts). Il s'agit de la consommation d'électricité maximale potentielle de chaque matériel physique, multipliée par le nombre de matériels.

La dissipation calorifique s'exprime en BTU/h. Le British Thermal Unit (BTU) est une unité d'énergie anglo-saxonne définie par la quantité de chaleur nécessaire pour élever la température d'une livre anglaise d'eau d'un degré à la pression constante d'une atmosphère. Elle est généralement utilisée pour décrire la quantité de chaleur pouvant être dégagée par une unité chauffante ou réfrigérante (climatisation par exemple). La formule de calcul est la suivante :

Dissipation calorifique (BTU/h) = Tension (Volts) x Intensité(Ampères) x 3.41

Où : 3.41 est l'unité de conversion d'un Watt en BTU par heure (1 Watt = 3.41 BTU/h) [23] [24].

Cette mesure est utilisée en informatique pour calculer le dégagement de chaleur maximal possible pour un matériel. L'application directe est de quantifier la dissipation thermique de l'ensemble du matériel installé dans une salle technique. Ce résultat est une donnée essentielle lors de la rédaction du cahier des charges pour la mise en place de l'infrastructure de refroidissement (climatisation).

En termes d'occupation des baies informatiques, l'encombrement d'un matériel s'exprime en « U ». Le « U » désigne l'unité élémentaire de hauteur de chaque élément intégrable dans la baie. Un « U » correspond à une hauteur de 1.75 pouce (soit environ 4.445 cm). Une salle technique comporte des baies informatiques qui sont des armoires pouvant accueillir des éléments empilables (« rack ») dans des emplacements de taille normalisée d'une largeur de 19 pouces (48.26 cm).
Chaque baie du LATT avait une capacité de 42U.

A titre d'exemple, les anciens serveurs occupaient généralement 5U, à comparer aux nouveaux serveurs qui n'occupent plus que 1 ou 2 U. Ce gain de place permet d'optimiser l'utilisation des baies informatiques.

Pour chacune des mesures fournies, je me suis appuyé sur les données provenant des caractéristiques techniques fournies par les constructeurs des matériels[2] (Guides techniques disponibles sur Internet).

[2] *PowerEdge_R410_Technical_Guidebook.pdf , server-poweredge-r710-tech-guidebook.pdf*

1ère Option : mise en place d'une architecture physique

Pour évaluer l'architecture physique, je me suis basé sur l'achat sur de serveurs Dell R410, modèle minimal disponible au marché (obligation d'achat par l'utilisation des marchés publics).
Le détail technique d'un serveur Dell R410 se présentait comme suit : un processeur (Intel Xeon E5520), quatre cœurs cadencé à 2,26 GHz, 8 Go de mémoire vive, deux disques durs SAS de 146 Go, une carte réseau dual port intégrée, une alimentation non redondante de 480W (1637,81 BTU/h max.) et une alimentation redondante de 500W (1706,05 BTU/h max.) pour un encombrement de 1U.

La synthèse de l'investissement nécessaire et des caractéristiques disponibles pour l'implantation d'une architecture physique (1, 20 ou 40 serveurs) est présentée dans le Tableau I :

Tableau I : Investissements nécessaires à la mise en place d'une architecture physique

	Serveur Dell R410	**Architecture physique de 20 serveurs Dell R410**	**Architecture physique de 40 serveurs Dell R410**
Investissement	2 k€	40 k€	80 k€
Puissance processeur disponible	9,04 GHz	20 x 9,04 GHz	40 x 9,04 GHz
Quantité de mémoire vive disponible	8 Go	20 x 8 Go	40 x 8 Go
Espace disque brut	292 Go	20 x 292 Go	40 x 292 Go
Espace disque net	146 Go	20 x 146 Go	40 x 146 Go
Nombre de ports réseau disponibles par service	2	20 x 2	40 x2
Consommation énergétique	0,98 kW	19,6 kW	39,2 kW
Dissipation calorifique (en BTU/h maximum)	3343,86 BTU/h	66877,2 BTU/h	133754,4 BTU/h
Occupation	1 U	20 U	40 U

2ème Option : mise en place d'une architecture virtualisée

Pour évaluer l'architecture virtualisée, je me suis appuyé sur l'achat et l'implantation d'une architecture virtuelle basée sur deux serveurs adaptés Dell R710 et une baie de stockage modèle Xyratex 5412 (compatibilité avec les matériels existants). Les serveurs et la baie seraient connectés par fibre optique via le SAN existant au laboratoire.

Les caractéristiques techniques de chacun des serveurs Dell R710 étaient les suivantes : deux processeurs (Intel X5667) à quatre cœurs cadencés à 3,06 GHz, 32 Go de mémoire vive, deux disques durs SAS de 146 Go, une carte réseau dual port intégrée, deux cartes de quatre ports réseaux supplémentaires, deux alimentations redondantes « hot Plug » de 870W (2968.6 BTU/h max.) pour un emplacement de 2U.
La baie Xyratex 5412 disposait des caractéristiques suivantes : un tiroir de douze disques SAS de 146 Go chacun, alimentations redondantes « hot Plug » de 355W (1212,39 BTU/h max.), pour un emplacement nécessaire de 2U.

La synthèse de la composition de l'architecture virtualisée (deux serveurs et une baie de stockage) est présentée dans le Tableau II :

Tableau II : Investissements nécessaires à la mise en place d'une architecture virtualisée

	Serveur Dell R710	**Baie de Stockage Xyratex**	**Architecture Virtuelle (serveurs & baie)**
Investissement	4,5 k€	10 k€	19 k€
Puissance processeur	12,24 Ghz	-	24,48 Ghz
Quantité de mémoire vive disponible	32 Go	-	64 Go
Espace disque brut	292 Go	1752 Go	2336 Go
Espace disque net	146 Go	1460 Go	1752 Go
Nombre de ports réseau disponibles par services	10	-	2 x 10
Consommation énergétique	1,74 kW	0,71 kW	4,19 kW
Dissipation calorifique (en BTU/h maximum)	5937,2 BTU/h	2424,78 BTU/h	14299,18 BTU/h
Occupation	2 U	2 U	6 U

Tableaux comparatifs des deux architectures

Pour chaque analyse, j'ai comparé les chiffres de chacune des solutions et ai évalué le gain potentiel (positif ou négatif) de la solution de virtualisation. Les résultats de ces comparaisons se trouvent pour l'aspect financier dans le Tableau III, pour la consommation énergétique dans le Tableau IV et pour la synthèse dans le Tableau V.

La relation utilisée pour calculer le gain de la virtualisation est **: (1-(V/P)) x 100 = n %**

où P et V sont les valeurs issues des différents tableaux, respectivement pour les architectures physiques et virtualisées.

Cout financier de l'investissement initial

Tableau III : Comparaison du coût d'investissement d'une solution de virtualisation (hors logiciel)

	Architecture physique (20 serveurs R410)	**Architecture virtuelle (serveurs & baie)**	**Gains pour la solution de virtualisation**
Investissement	40 k€	19 k€	52.50 %

Le cout financier de l'implantation des deux architectures est largement en faveur de la virtualisation. De plus, en considérant le fait que la même infrastructure virtualisée peut aussi supporter les machines scientifiques, le facteur d'échelle est nettement en faveur de la virtualisation.
Pour l'ensemble des serveurs d'infrastructure et scientifiques virtualisables, le gain peut s'élever à 76,25% (19 k€ *vs* 80 k€).

Le coût de consommation énergétique et d'hébergement

Tableau IV : Comparaison de la consommation énergétique d'une solution de virtualisation

	Architecture physique (20 serveurs R410)	**Architecture virtuelle**	**Gain pour la solution de virtualisation**
Consommation énergétique totale	19,6 kW	4,19 kW	78.62%
Dissipation calorifique totale	66877,2 BTU/h	14299,18 BTU/h	78.62%
Occupation	20 U	6 U	70.00 %

La consommation énergétique de l'architecture virtuelle est 78% moins importante que celle de l'architecture physique. Cet écart est principalement dû au nombre de matériels physiques composant les deux architectures.
Pour le dégagement de chaleur, le gain pour l'architecture virtuelle est identique (78%). Ce résultat est cohérent puisque l'expression de la dissipation calorifique est fonction de la consommation électrique.

En termes d'occupation des baies informatiques, le constat est aussi en faveur de la virtualisation (encombrement réduit de 70%). Ces résultats ont une incidence directe sur les coûts annexes.
La virtualisation permet de diminuer le taux d'encombrement des baies informatiques, réduisant ainsi l'achat de nouvelles baies ou la création de nouvelles salles techniques informatiques (ondulées, climatisées, équipées en câblage réseau et électrique suffisant).

Le résultat de la comparaison des coûts de consommation énergétique et d'hébergement plaide clairement en faveur d'une architecture virtualisée. Le gain de consommation énergétique passerait même à 89.31%, et le gain d'encombrement à 85%, dans le cas d'une extension de l'utilisation de la même architecture virtuelle pour héberger les services scientifiques.

Une chronique technique [8], parue dans un magazine spécialisé, évoquait le gain écologique et énergétique de la virtualisation des serveurs, et avançait par exemple un facteur 6 à 10 de gain en termes d'achat de matériel. Cette consolidation permettait de réaliser une économie, à l'échelle d'une PME (réduction de 12 à 3 serveurs physiques), d'environ 10 k€ sur trois ans sur le fonctionnement, la protection et le refroidissement.

Les capacités (puissance, taille disque) disponibles pour chaque architecture

Tableau V : Comparaison de la capacité d'une solution de virtualisation

	Architecture physique (20 serveurs R410)	**Architecture virtuelle**	**Gain pour la solution de virtualisation**
Puissance processeur disponible totale	20 x 9,04 GHz	2 x 12,24 Ghz	*Comparaison non significative*
Taille mémoire disponible totale	20 x 8 Go	2 x 32 Go	*Comparaison non significative*
Taille nette disque disponible	20 x 146 Go	1652 Go	*Comparaison non significative*
Puissance processeur théorique consolidée disponible par service	9,04 Ghz	12,24 Ghz	35,39 %
Taille mémoire théorique consolidée disponible par service	8 Go	32 Go	300 %
Taille disque nette théorique consolidée disponible par service	146 Go	1652 Go	1131,50%

Les chiffres bruts des premières lignes du Tableau V ne sont pas significatifs dans ce comparatif. En effet, une machine physique ne distribue ses ressources qu'aux systèmes et applications installés et fonctionnant localement. La machine virtuelle profite de l'ensemble des ressources du serveur physique, hébergeant le système de virtualisation, sur lequel elle fonctionne. J'ai donc privilégié une analyse pondérée sur les valeurs consolidées.

Comme nous l'avons vu, les serveurs physiques en production ne fonctionnent souvent qu'à 15 ou 20% de leur capacité, le reste de la puissance étant réservé pour faire face aux montées de charge ponctuelles. Cette réserve de ressources, perdue sur une architecture physique, est mutualisée et exploitée sur des architectures virtualisées. Certaines technologies de virtualisation permettent d'optimiser l'utilisation des ressources physiques en les répartissant au besoin entre les différents serveurs virtuels fonctionnant sur l'infrastructure.

La virtualisation permet donc de réaliser de la consolidation de serveur, *i.e.* d'optimiser le taux d'utilisation des matériels physiques.
Certains hyperviseurs permettent aussi de réaliser de la surréservation de ressources en allouant aux ressources virtuelles davantage de ressources physiques que ce dont ont dispose réellement.

La comparaison quantitative est logiquement en faveur de l'architecture virtualisée. Les résultats montrent que le virtualisation permet une diminution du nombre de matériels, qui a pour conséquence une diminution de l'encombrement, une réduction de la consommation énergétique et un retour sur investissement (ROI) amélioré. La virtualisation apporte également davantage de ressources disponibles pour chaque instance de service.

Je vais à présent détailler les résultats de l'analyse que j'ai réalisée en fonction des critères qualitatifs.

2.4.2. Comparaison qualitative des architectures système physiques et virtualisées

Les critères qualitatifs suivants ont été analysés afin d'étudier les différences entre les deux architectures :

• **La sauvegarde** : sur les serveurs physiques, plusieurs solutions existent pour la sauvegarde. Le laboratoire dispose d'un robot de sauvegarde, ainsi que de plusieurs solutions de stockage sur disques. La sauvegarde des machines physiques revient alors à sauvegarder les données et configurations spécifiques de chaque service critique afin de pouvoir les restaurer en cas de problème.
La virtualisation apporte plusieurs avantages par le fait que les machines virtuelles sont vues par le système comme un simple fichier (ou répertoire). La sauvegarde peut donc se faire beaucoup plus facilement, et on peut ainsi sauvegarder l'ensemble de la machine, ce qui assure une restauration et une remise en service plus rapide et plus sure que sur une machine physique.

• **La pérennisation du service** : intrinsèquement aux principes de virtualisation, une machine virtuelle, qui respecte les conditions de compatibilité (type de processeur par exemple), est indépendante de son hôte physique. Cela induit qu'une panne matérielle d'un serveur physique n'a que peu d'incidence sur la disponibilité des machines virtuelles (VM) hébergées, pérennisant ainsi le service rendu à l'utilisateur. L'indépendance des VM vis-à-vis du système hôte nous assure un fonctionnement continu lors des opérations de maintenance ou de jouvence du parc de serveurs.
Ainsi, une infrastructure système virtualisée améliore l'efficience de l'équipe en charge de son administration et une disponibilité accrue pour les services offerts.

• **La maitrise de la technologie** : la virtualisation est une technologie récente qui apporte plusieurs innovations. Elle implique de repenser sur le plan organisationnel et technique la façon de travailler et de gérer les ressources informatiques. Le personnel intervenant doit être formé à cette technologie pour pouvoir en exploiter les avantages, ce qui implique un fort investissement humain, voire financier en termes de formation.

2.4.3. Conclusion et choix de l'architecture

Les comparaisons quantitative et qualitative font apparaître un gain financier et des possibilités d'évolution en faveur de l'architecture de virtualisation. Cette technologie nous est donc apparue compatible avec le développement de notre projet.

Le service informatique a donc pris la décision d'effectuer une étude comparative multicritères des différentes solutions de virtualisation disponibles sur le marché en fin d'année 2009 afin de déterminer quelle pouvait être l'architecture virtuelle la mieux adaptée aux besoins du laboratoire.

Les critères retenus ont été la maitrise des coûts (matériels, logiciels et formation), l'optimisation de l'ensemble des ressources physiques (serveurs et infrastructures) et l'adéquation à l'environnement technique en production au sein du laboratoire.

2.5. Choix de la plateforme de virtualisation

L'étude comparative multicritères a été découpée en une étude de marché, en une évaluation des solutions compatibles avec notre environnement et en une synthèse qui a conduit au choix de la solution mise en œuvre.

L'offre du marché

Si l'offre de virtualisation était assez large en début 2010, quelques acteurs du marché se détachaient par leur sérieux et la stabilité de leurs solutions dans le monde de l'entreprise.
J'ai réalisé une synthèse de l'offre existante à partir de discussions avec des experts ainsi qu'une étude bibliographique approfondie. Je me suis notamment appuyé sur la lecture de l'ouvrage de P. Gillet [2, chapitre 1] et sur les présentations techniques [10] [11] suivies lors des Journées Réseaux 2009 (JRES2009) qui restent une référence pour les professionnels des réseaux et des systèmes informatiques en France.

Trois systèmes de virtualisation étaient alors reconnus auprès des entreprises et des administrations pour leur qualité et fiabilité. Il s'agissait de :

- Microsoft Hyper-V : plus connu sous le nom de Windows Server Virtualization. Il s'agissait du système de virtualisation proposé par Microsoft qui était basé sur un hyperviseur 64 bits fonctionnant sur les bases de Windows Server 2008 ;

- Citrix Xen : logiciel libre de virtualisation sous licence GNU GPL. Il s'agissait d'un hyperviseur de machine virtuelle développé au départ par l'université de Cambridge ;

- VMware VI3 (Virtual Infrastructure 3) : ensemble logiciel reposant sur des hyperviseurs embarqués sur des machines physiques et un ordonnanceur de tâches (Vcenter). Cette solution, permettant une gestion optimale des ressources, était une solution de virtualisation plus industrielle et répandue que les deux précédentes. Le principal avantage de ce produit par rapport aux autres est qu'il s'installait directement sur la couche matérielle, et qu'il n'était donc pas nécessaire d'installer un système d'exploitation « hôte » sous-jacent.

Ces trois solutions étaient, au début de l'année 2010, les trois qui paraissent suffisamment fiables pour que nous puissions les implanter en production pour notre projet. Nous en avons donc réalisé une évaluation et défini laquelle des trois était la mieux adaptée pour nos besoins compte tenu de notre environnement de travail.

2.3.1. Evaluation des hyperviseurs dans notre environnement

Nous avons, comme pour le choix entre l'architecture matérielle physique ou virtuelle, déterminé des critères discriminants pour notre évaluation des systèmes de virtualisation. J'ai principalement analysé les critères suivants :

- L'indépendance du système d'exploitation vis-à-vis de l'hyperviseur ;
- La gestion de la mémoire ;
- Le système de fichiers ;
- La sécurité de l'hyperviseur vis-à-vis des machines virtuelles ;
- Le coût du système de virtualisation.

L'indépendance du système d'exploitation vis-à-vis de l'hyperviseur

Dans l'optique de la mise en place du concept, les machines virtuelles devaient fonctionner avec un système d'exploitation Linux. Ceci permettait de minorer les frais de licence, de conserver une gestion optimale de nos services et de garantir une compatibilité avec les systèmes en production.

VMware, de par son architecture, était totalement compatible avec l'environnement Linux. Citrix Xen Server était compatible avec la majorité des distributions Entreprise de Linux (principalement basées sur les distributions Red Hat et Suze). Cependant, sous Xen Server, les systèmes d'exploitation invités ont « conscience » du serveur Xen sous-jacent et ont besoin d'être « portés » (adaptés) pour fonctionner sur Xen (noyaux spécifiques).

Microsoft Hyper-V était principalement compatible avec les produits de la société Microsoft et deux distributions de Linux : SUSE Linux Enterprise Server 10 et 11 et Red Hat Enterprise Linux 5.2 ou supérieur en 64bits. Cela signifiait qu'il fallait utiliser exactement une de ces versions de Linux pour fonctionner, toute autre version n'étant pas validée sur cet hyperviseur.

La gestion de la mémoire

Avec Citrix Xen Server et Microsoft Hyper-V, l'allocation *a priori* de la mémoire physique à chaque VM hébergée conduit à une perte mécanique de la mémoire non utilisée pour chaque machine virtuelle.

Sur VMware, la mémoire physique du serveur était partagée entre les machines virtuelles. On pouvait de plus récupérer la mémoire inutilisée et faire de la déduplication des pages mémoire. Ces mécanismes permettaient d'optimiser la charge mémoire du serveur et d'absorber les montées en charges ponctuelles sans ralentissement des services.

Le système de fichiers

Sur Microsoft Hyper-V, le système de fichiers utilisé était celui de Microsoft Windows Server 2008 puisque l'hyperviseur fonctionnait sur ce système d'exploitation. Il n'était donc pas possible de partager le système de fichiers ni de créer de systèmes de fichiers en cluster. Pour le laboratoire d'Astrophysique, cette incompatibilité était gênante pour installer les machines virtuelles sous Linux.

Sur Citix Xen Server, le système de fichier était EXT3 (système de fichiers Linux) ou NFS (Network File System, système de fichiers en réseau). L'EXT3 ne permettait pas de gérer correctement les accès concurrents, alors que le NFS le permettait mais pêchait par des débits inférieurs à ceux qui étaient nécessaires pour un système en production. Xen Server restait cependant sur ce point compatible avec le projet.

Le système de fichier utilisé par VMware était le VMFS (VMware File System). Format propriétaire, il permettait de partager le système de fichiers entre plusieurs hôtes (gestion des accès concurrents) et ainsi d'assurer une localisation et un redémarrage plus aisé des machines virtuelles en cas de panne matérielle du serveur. Cet argument était un plus pour la sécurité de la future architecture.

La sécurité de l'hyperviseur vis-à-vis des machines virtuelles

Du fait de sa forte dépendance avec son système d'exploitation, Microsoft Hyper-V était vulnérable aux correctifs, mises à jour et failles de sécurité de Windows Server 2008.
Pour Citrix Xen Server en revanche, la fiabilité de l'hyperviseur était liée à la fiabilité de son système d'exploitation (Linux), jusqu'ici assez performante.

Enfin, pour VMware VI3, la fiabilité de son système d'exploitation était aussi basée sur la fiabilité de son système d'exploitation (une version Linux Red Hat remaniée) à laquelle VMware avait ajouté des API de sécurité (VMsafe).

Le coût du système de virtualisation

Le coût de Microsoft Hyper-V était celui du serveur Microsoft Windows Server 2008 auquel il fallait ajouter une vingtaine d'euros, par licence, pour le coût de l'hyperviseur (Hyper-V). Le coût restait relatif comparé à la solution VMware, mais non négligeable.

VMware était la solution payante la plus onéreuse du comparatif. Il fallait compter un minimum de 12,5k€ pour le déploiement d'une infrastructure virtuelle (VI3) au tarif éducation nationale avec un an de maintenance (mises à jour et support téléphonique). A noter qu'il était possible de déployer une version gratuite mais qui incluait seulement les hyperviseurs (ESXi), sans possibilité d'ordonnancement ou de haute disponibilité.

Citrix Xen Server était un progiciel libre et gratuit. Le coût nul de cet hyperviseur le plaçait en tête du classement par coût, à égalité avec VMware ESXi.

Autres critères évalués

Pour terminer, quelques autres critères ont été évalués tel que la volumétrie occupée par l'hyperviseur ou la gestion des ressources physiques et de leur présentation aux machines virtuelles. Dans ces deux cas, Citrix et VMware avaient une longueur d'avance sur Microsoft.
Les résultats n'étant pas significatifs en termes de comparaison, ils ne sont pas présentés ici.

2.3.2. Conclusion du choix de la plateforme de virtualisation

Au terme de mon analyse, j'ai obtenu le tableau de synthèse ci-dessous auquel j'ai intégré les quatre solutions de virtualisation étudiées. La compatibilité avec notre environnement et le prix de la solution pour ce prototype restaient les deux critères déterminants, avant tout autre critère technique.

Tableau VI : Tableau de synthèse des critères d'évaluation des solutions de virtualisation

Critères examinés	Microsoft Hyper-V	Citrix Xen	VMware VI3	VMware ESXi
Indépendance du système d'exploitation	B	A	A	A
Coût du système de virtualisation	B	A	C	A
Gestion de la mémoire	B	B	A	A
Système de fichiers	C	C	B	B
Sécurité de l'hyperviseur vis-à-vis des VM	C	A	A	A
Note de synthèse (avec pondération) :	5	11	8	13
Positionnement final (Rang) :	**4**	**2**	**3**	**1**

A : convient à notre environnement ; **B** : adaptable à notre environnement ; **C** : ne convient pas à notre environnement.

Cotation : A =2points ; **B** = 1 point ; **C** = 0 point ; les deux premiers items sont pondérés (note multipliée par deux).

A la vue du tableau récapitulatif des points évalués et de leurs justifications, le service informatique du LATT est arrivé aux conclusions suivantes :

- La solution de Microsoft (Hyper-V) était la solution la moins bien notée de ce comparatif. Adaptée aux environnements basés sur les produits Microsoft (Serveur 2008, Windows Active Directory, etc.), elle était trop éloignée de nos pratiques techniques basées en grande partie sur le monde libre (Linux et logiciels Open Source) ;
- Le produit porté par Citrix, bien que répondant à nos deux critères déterminants, ne satisfaisait qu'en partie nos critères techniques, principalement pour son mode de gestion du système de fichiers et de la mémoire ;
- Parmi les deux solutions portées par VMware, qui correspondaient le mieux aux critères d'évaluation, le critère discriminant fut le coût de l'achat des licences.

En conclusion, le choix du service informatique s'est porté, pour la phase de prototypage, sur la solution gratuite proposée par VMware. En cas de satisfaction du fonctionnement du prototype, l'implantation d'une version définitive pourrait être réalisée sur une infrastructure VMware VI3. Cette migration apporterait des fonctionnalités non disponibles sur ESXi en termes de haute disponibilité et de sécurité des systèmes.

2.5. Sécurisation de l'accès au réseau

L'analyse des points d'effort et les nouvelles fonctionnalités liées aux nouveaux matériels acquis lors de rénovation du réseau informatique nous ont permis d'améliorer la qualité du service offert. Dans ce cas précis, nous avons exploré plus particulièrement l'utilisation du protocole Radius [RFC 2865] qui nous a permis d'accroitre la mobilité, la sécurisation et le contrôle des postes de travail sur le réseau.

Radius (Remote Authentication Dial-In User Service) est un protocole client-serveur permettant de centraliser des données d'authentification. L'objectif initial de Radius était de permettre aux fournisseurs d'accès à Internet d'authentifier, à partir d'une base centralisée, les matériels déployés chez les utilisateurs distants (box) utilisant les connexions ADSL.

Dans le contexte du laboratoire, l'implémentation du protocole Radius a permis une authentification des postes de travail des utilisateurs grâce aux données d'identification du matériel et aux réseaux virtuels associés.

L'identification du matériel reposait sur l'adresse physique des cartes réseau des postes de travail (adresse MAC, Media Access Control). Chaque possesseur de matériel appartenait à une équipe identifiée au sein du laboratoire à laquelle est associé, pour des raisons de sécurité, un réseau virtuel (VLAN).

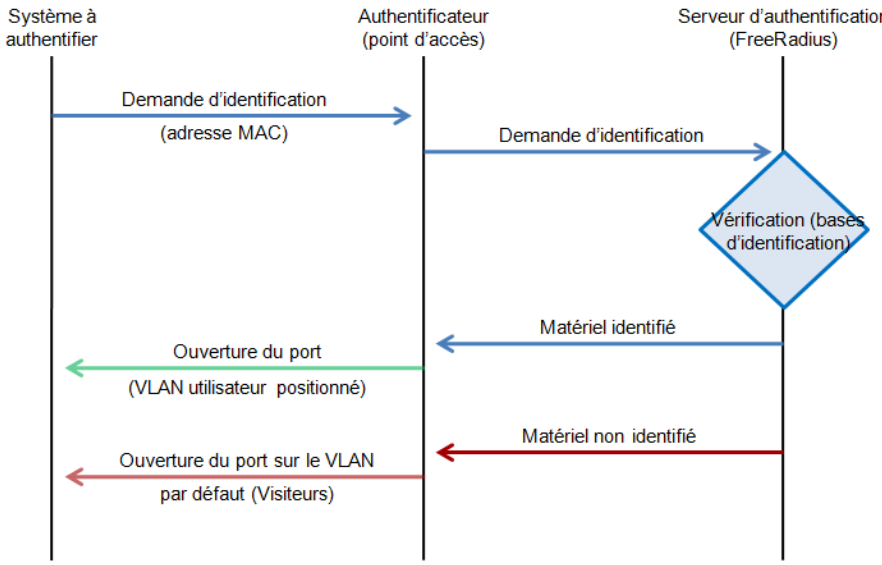


Figure 10 : Accès réseau par identification matérielle

Le fonctionnement du service d'authentification réseau, schématisé sur la Figure 10, était le suivant : tout matériel (station de travail, ordinateur portable, etc.) qui se connectait au réseau envoyait une information au répartiteur réseau sur lequel sa prise était connectée. Cette information contenait notamment l'adresse matérielle de sa carte réseau (adresse MAC) qui est unique. Le commutateur réseau récupérait cette adresse et envoyait une demande de validation de l'adresse par le service d'authentification matérielle (service Radius).

Le service Radius identifiait l'adresse physique reçue et renvoyait un numéro de VLAN associé à cette adresse dans sa base de connaissance. Dans ce cas, il y avait donc trois possibilités :

- L'adresse matérielle (MAC) était connue dans la base de connaissance du service d'authentification matérielle. L'utilisateur était positionné sur le réseau logique qui lui permettait d'accéder à tous les services qui lui sont offerts.
- L'adresse matérielle (MAC) était inconnue dans la base de connaissance du service d'authentification matérielle. L'utilisateur était positionné sur un réseau logique distinct (Figure 11), indépendant du réseau logique du laboratoire et accédait à une page d'authentification par portail captif. A partir de cette dernière, il devait s'authentifier pour accéder à une offre de service minimale avec les identifiants de son compte personnel ou un accès de type « invité ».
- L'adresse matérielle (MAC) était inconnue et l'utilisateur ne disposait pas de compte au laboratoire. Il ne pouvait pas s'authentifier sur le portail captif et ne disposait alors d'aucun accès au réseau.

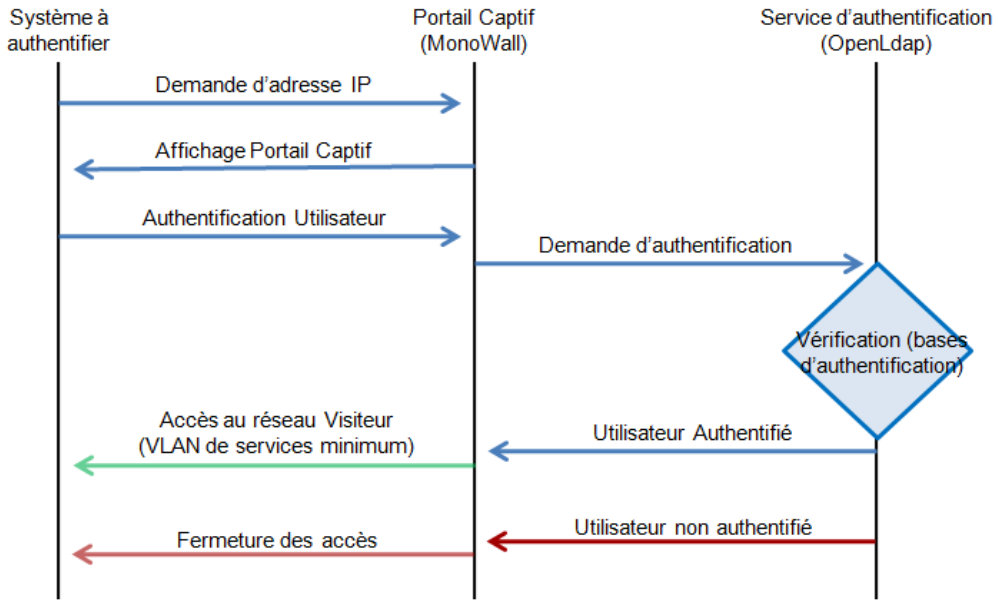


Figure 11: Authentification utilisateur pour l'accès au réseau

Un prototype d'authentification a été déployé sur le commutateur réseau irriguant une zone d'une dizaine de bureaux autour du service informatique.

En premier lieu, nous avons augmenté la sécurité et repris la maîtrise de notre réseau interne en filtrant les connexions de matériels sur nos réseaux logiques. Nous avions ainsi l'assurance que toutes les communications circulant sur ce réseau étaient réalisées par des matériels identifiés.

Le deuxième point positif est l'apport de mobilité pour les postes de travail des utilisateurs. Les utilisateurs ayant déclaré leurs postes de travail obtenaient toutes les spécificités de leurs connexions (VLAN et accès restreints) de n'importe quel point d'accès du réseau filaire. Auparavant, l'affectation de réseaux logiques (VLAN) était réalisée de façon statique, par l'intervention manuelle du service informatique sur les équipements réseau. L'arrivée de l'authentification par service Radius, conjuguée à la migration progressive des matériels utilisateurs vers des stations de travail portables, a augmenté le taux de satisfaction des utilisateurs tests. L'apport de cette mobilité a eu un écho très positif auprès de

ces derniers qui avaient désormais la possibilité de se connecter depuis tous les points d'accès ouverts du laboratoire en conservant leur environnement de travail.

Un autre avantage apporté par cette technologie a été la maitrise de l'inventaire du parc informatique. A partir du moment où tout matériel connecté était identifié, le service informatique disposait d'une base de données exhaustive et à jour des matériels du laboratoire.
En corollaire, les utilisateurs se sont inscrits dans une démarche volontaire de déclaration de leurs matériels afin d'accéder à leur services. Ce fut une amélioration sensible par rapport à la situation antérieure. Le service informatique pouvait ainsi mieux contrôler les accès des stations de travail, mieux gérer l'affectation d'adresses IP (suppression des doublons), et mieux sécuriser les stations du réseau (antivirus, pare-feux personnels, etc.) que les utilisateurs ont déclarées.
Pour finir, il est assez simple de développer des mécanismes nous permettant de vérifier les accès physiques et la présence des stations de travail dans la durée sur de simples analyses de logs des serveurs Radius. Cette possibilité n'a pas été développée dans le cadre du prototype mais sera étudiée ultérieurement afin de gérer le cycle de vie des matériels.

Le bilan de l'implantation des serveurs Radius au LATT, rendue possible par la mise en place du nouveau système d'information, a donc été positif en tous points : augmentation de la sécurité du réseau, offre de nouveaux services aux utilisateurs (mobilité), simplification de la gestion de l'inventaire du parc informatique et surtout un retour très positif sur la satisfaction des utilisateurs.

Ce bilan positif a amené le service informatique à planifier l'intégration de l'authentification matérielle sur tous les matériels actifs du laboratoire lors de la phase de mise en exploitation du prototype. L'étude de la possibilité d'intégrer la même technologie pour contrôler et sécuriser les accès au réseau sans fil du laboratoire a aussi été proposée, et devait être réalisée après la réalisation du projet global.

Bien qu'imparfaite, cette solution basée sur l'authentification MAC était adaptée dans notre environnement. Nous avons également étudié l'intégration de solutions basées sur la norme 802.1x (IEEE) [7], mais les contraintes liées à cette authentification forte ne nous ont pas permis un déploiement à l'échelle du laboratoire.

3. Mise en place du prototype

3.1. Méthode d'intégration du prototype

Pour valider ces nouveaux concepts, nous avions décidé de mettre en place une architecture qui nous permettrait d'abord de réaliser des maquettes de chacune de nos solutions de façon indépendante afin de les tester séparément.
En effet, le système d'information du LATT était un système d'information en production qui remplissait pleinement sa fonction. Il n'était donc pas question de migrer directement les systèmes en production vers un prototype en phase de tests.
La seconde étape serait de tester l'intégration de tous ces nouveaux concepts vers un prototype complet. Après ces différentes phases de développement, de tests et d'intégration des technologies, nous aurions enfin une évaluation réelle du concept global et pourrions valider ou abandonner son intégration dans le système d'information existant.

Si le prototype apportait satisfaction, nous établirions une procédure d'intégration progressive dans le système d'information en production permettant de faire cohabiter le système initial avec les nouveaux systèmes développés.

3.2. Développement de la maquette

Comme je l'ai présenté au chapitre II, la phase de développement du prototype s'est déroulée de septembre 2009 à avril 2010.
Toutes les étapes du développement de ce prototype ont suivi la méthode choisie au lancement du projet (méthode de développement en spirale). Chaque évolution du prototype a fait l'objet d'une étude tenant compte des évolutions et de l'existant, et d'une phase de tests. Ce schéma itératif a été répété pour tous les développements du prototype jusqu'à la phase de mise en exploitation. Je vais vous détailler maintenant les différentes phases du développement synthétisées sur la Figure **12**.

Le choix des technologies à employer a constitué la phase préliminaire. Cette phase de choix s'est poursuivie tout au long du développement du prototype en fonction des orientations que prenaient les développements, en accord avec la méthode de développement itérative.

La mise en place de l'architecture matérielle virtualisée, la création des machines virtuelles et leur répartition sur les serveurs ont fourni le support matériel technique pour le développement du prototype.

La création du référentiel unique, pierre angulaire du nouveau système d'information, a été la première action de la mise en place du prototype. Le recensement des données du système d'information en production, la modélisation et la création de la base de données unifiée ont constitué le socle de mise en place des nouveaux concepts.

Une fois ce référentiel unique disponible, le script de génération de la configuration du DHCP a été créé, et implanté sur la machine virtuelle qui devait servir de DHCP pour le laboratoire. Nous avons testé cette configuration en pré-production puis remis en marche l'ancien serveur DHCP après des tests positifs.

Le script de génération de la configuration du DNS a ensuite été créé sur le même schéma et implanté sur la machine virtuelle qui devait servir de DNS. Après avoir effectué plusieurs tests en pré-production, le service a aussi été rebasculé sur l'ancien serveur DNS.
Remarque : après avoir testé le fonctionnement des services DNS et DHCP, nous avons regroupé les deux services sur la même machine virtuelle puisqu'il s'avérait que chacun était peu consommateur de ressources. A titre indicatif, la machine virtuelle unique pour les deux services fonctionnait avec un disque dur (virtuel) de 8 Go et consommait environ 50 Mo de RAM (mémoire vive) en production.

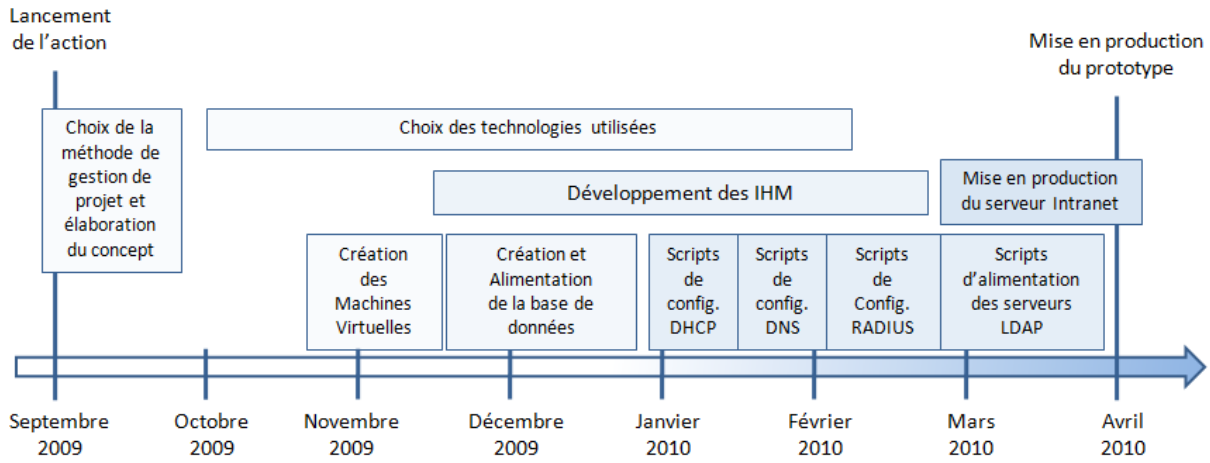


Figure 12 : Ensemble des phases du développement du prototype

Un script de génération de la configuration du service d'authentification réseau (service Radius) a de même été développé à partir des données du référentiel et implanté sur une machine virtuelle pour tester cette technologie. La mise en production du service Radius a d'abord été réalisée sur un commutateur réseau de test non relié à notre architecture réseau, puis sur le commutateur réseau en production de l'équipe informatique.
Après une phase d'observation, nous avons pris la décision de généraliser l'emploi de cette technologie à l'ensemble du réseau du laboratoire. Nous avons donc effectué une phase de mise à jour des données matérielles de nos utilisateurs afin de compléter notre référentiel et de pouvoir générer le service d'authentification réseau pour tous les matériels du laboratoire. Nous disposions ainsi de tous les pré-requis nécessaires pour déployer cette solution sécurisée à l'échelle du laboratoire.

Le script d'alimentation des annuaires LDAP est l'avant-dernière étape de la Figure 12. Le serveur LDAP était le plus important point individuel de défaillance de l'ancien système d'information. Nous avons développé les scripts de génération dynamique de ce service (basé sur l'utilisation de fichiers LDIF, format de fichiers permettant l'import et l'export des données des serveurs LDAP) et avons généré l'alimentation des fichiers de configuration des deux serveurs LDAP à partir des données de la base unifiée.
La génération dynamique a été testée pendant une semaine en positionnant ce nouveau service d'authentification comme source principale de certains serveurs de tests. Après avoir constaté un fonctionnement normal du service, nous avons logiquement remplacé les anciens serveurs LDAP par les deux serveurs LDAP virtuels générés dynamiquement (inversion des adresses IP et noms DNS).

En parallèle à ces développements et tests, le serveur Intranet permettant l'accès aux interfaces d'interrogation et de manipulation du référentiel unique a été développé. Une fois ce serveur sécurisé, nous avons développé les interfaces d'administration et d'interrogation de la base de données. Nous

avons mis au point l'interface d'interrogation client, et avons ouvert l'accès à certains chercheurs pour pouvoir effectuer des tests et un débogage avant la mise en production.
Ces IHM ont été enrichies au fur et à mesure de la remontée de nouveaux besoins par nos utilisateurs de test.

Le prototype se composait donc à ce stade de huit machines virtuelles sur lesquelles fonctionnaient les différents services (Figure 13).

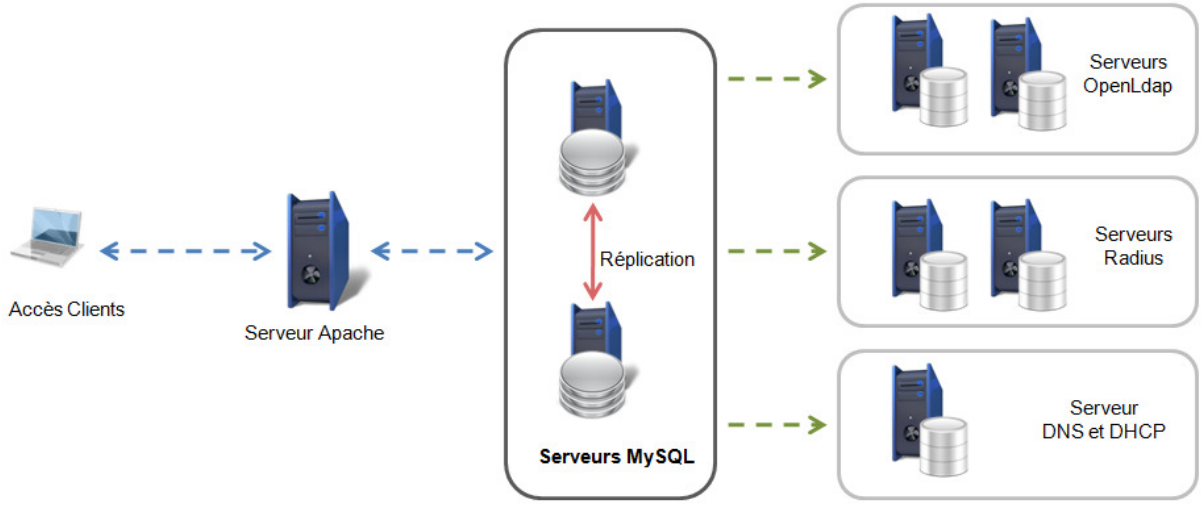


Figure 13 : Synoptique de répartition des VM du prototype

3.3. Organisation de l'implantation

Nous avions choisi d'implanter le prototype de système d'information sur une architecture virtuelle. Il fallait donc définir la répartition et l'implantation des services du prototype sur cette nouvelle architecture en vue d'obtenir une optimisation de la disponibilité et de la pérennité du système.

L'implantation matérielle du prototype a été réalisée sur deux serveurs physiques existants dans le laboratoire. Il s'agissait de machines physiques (réplication des services LDAP et DNS) qui ont été virtualisées afin de rendre les matériels disponibles. Sur ces matériels, nous avons installé la version 3.5 de l'hyperviseur ESXi de VMware. Cette solution a permis de tester le prototype sans réaliser d'investissement matériel, ce qui était un pré-requis au lancement du projet.

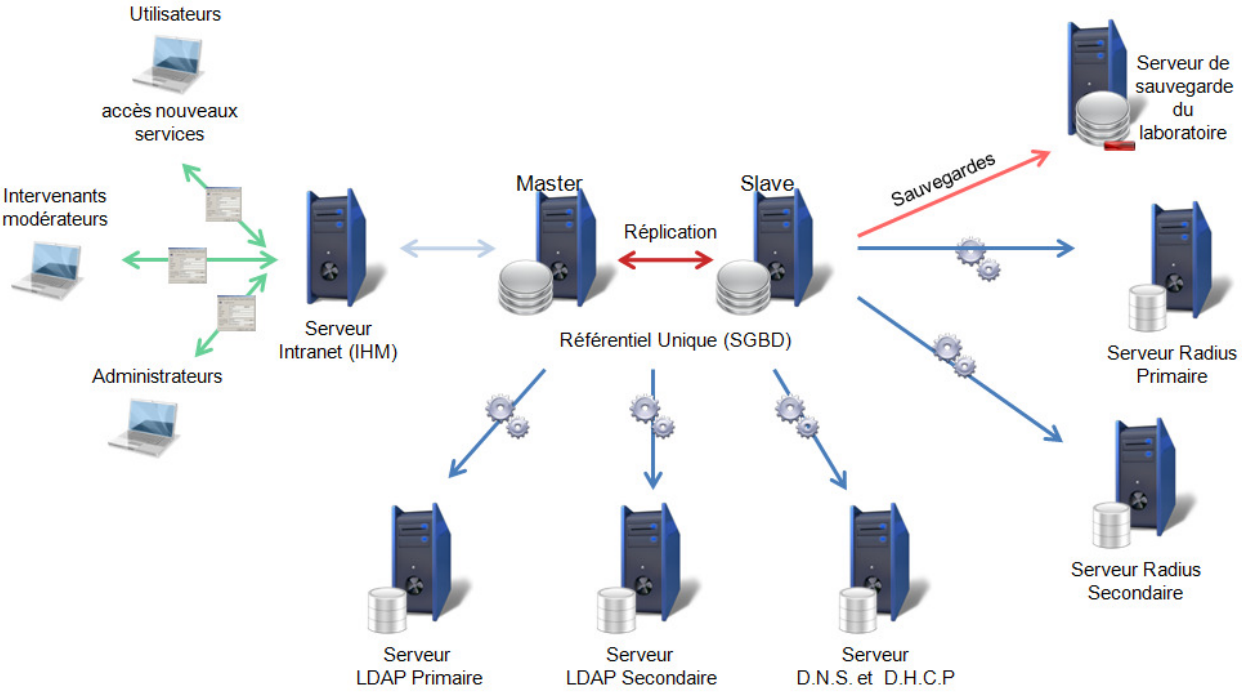


Figure 14 : Schéma du prototype du système d'information en gestion unifiée

Les services du prototype ont été implantés selon la description du schéma de la Figure 14 :

• Service de bases de données : deux serveurs linux CentOS 5 avec installation de MySQL 5 avec réplication synchrone de type Maitre-Esclave et sauvegarde quotidienne vers un serveur externe ;

• Service d'authentification : deux serveurs linux CentOS sous OpenLDAP v3, permettant de remplacer à l'identique les deux serveurs LDAP existants. Cependant, dans ce schéma, les services LDAP ne sont plus interconnectés et répliqués entre eux mais fonctionnent en parallèle puisque leur configuration est générée dynamiquement ;

• Service d'identification : deux serveurs linux CentOS sous FreeRADIUS. Le serveur primaire et le serveur secondaire étaient indépendants et redondants afin de pouvoir mettre en place l'authentification réseau par les matériels actifs de façon sécurisée ;

• Services réseau : un serveur linux CentOS permettant de rendre les services DHCP et DNS. Comme indiqué précédemment, ces deux services peu consommateurs de ressources ont été intégrés dans la même machine virtuelle sur notre prototype ;

• Service de mise à disposition des données : un serveur linux CentOS avec un service Apache et plusieurs de ses modules : modules Perl, modules PHP, connecteurs Perl et PHP vers MySql, modules de sécurité, etc. Ce serveur constituait la machine virtuelle mettant à disposition les IHM en client léger, sous authentification, pour les utilisateurs, intervenants et administrateurs.

L'installation de ces huit machines virtuelles a été réalisée sur deux serveurs physiques (Figure 15) fonctionnant avec un hyperviseur VMware ESXi. La répartition des machines virtuelles sur chaque serveur physique a été faite de façon à ce que la défaillance d'un des deux serveurs n'affecte pas la disponibilité du système.

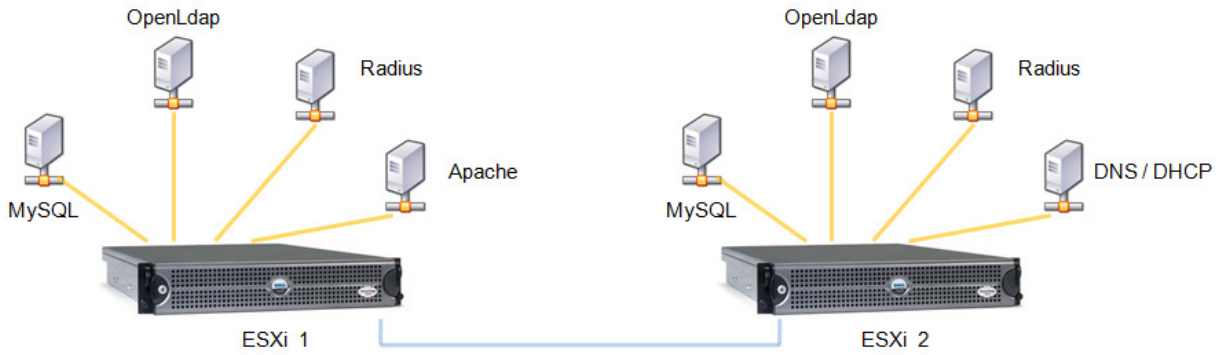


Figure 15 : Répartition physique des machines virtuelles du prototype

Chaque ESXi hébergeait effectivement un serveur de base de données, un serveur Radius et un serveur LDAP.

Pour le cas des services non redondants, nous avons utilisés le principe de la réplication « à froid ».
Les machines supportant les services DNS, DHCP et Apache avaient été copiées avant la mise en production d'un serveur vers l'autre. En cas de problème, il suffisait de quelques minutes pour redémarrer ce service manuellement.

4. Synthèse de la phase de prototypage

En avril 2010, le prototype était fonctionnel. Nous disposions donc d'une architecture virtuelle fonctionnant en parallèle avec l'architecture en production du laboratoire. Ces résultats devaient nous amener soit à abandonner ces concepts, soit à les consolider afin de pouvoir réaliser leur mise en exploitation.
La phase d'évaluation du prototype a validé la faisabilité de la majorité de nos concepts. Nous avions désormais la possibilité de :

- Créer un référentiel unique permettant de générer dynamiquement des services et d'obtenir une cohérence maximale des données du système d'information ;
- Minimiser les points individuels de défaillance du système en générant dynamiquement les services afin de minimiser leur criticité ;
- Sécuriser l'accès au réseau informatique et d'optimiser la gestion du parc au quotidien ;
- Offrir de nouveaux services à nos utilisateurs (mobilité, tableau de bord personnel).

Cette phase d'élaboration du prototype a fait apparaitre une contrainte technique liée à la nécessaire dichotomie entre l'authentification et l'identification (gestion du mot de passe), la nécessité d'implémenter un système de virtualisation professionnel et le besoin d'élaborer des plans de reprise et de continuité d'activité.

Au-delà de ce résultat technique, l'offre de services auprès des utilisateurs a connu un succès inattendu. De nouveaux besoins sont apparus et une forte pression a été exercée de la part de la communauté scientifique du laboratoire pour mettre en production le nouveau système d'information.

IV. Mise en exploitation du nouveau système d'information

Après les résultats positifs obtenus lors de la phase de prototypage, le service informatique a validé les nouveaux concepts et décidé de mettre en exploitation la nouvelle organisation de l'architecture du système d'information.
Cependant, les solutions techniques et organisationnelles choisies lors de la phase de prototypage étaient des solutions retenues sur des critères de maitrise des technologies (rapidité de développement) et de maitrise des coûts.
Pour pouvoir adapter cette nouvelle organisation du système d'information en production à l'échelle du laboratoire, il restait alors à adapter nos choix technologiques afin de pouvoir sécuriser et pérenniser les processus développés puis à réfléchir à l'organisation de leur implantation afin d'optimiser la disponibilité du système.
Nous avons donc mis en route une seconde phase de développement qui a été basée sur les retours d'expériences de la phase de prototypage, l'analyse des nouveaux besoins apparus et des nouvelles problématiques détectées afin d'élaborer une version stable, pérenne et à haute disponibilité du système d'information.

Dans un premier temps, je détaillerai succinctement le traitement des nouvelles exigences et les nouvelles évolutions que nous souhaitions apporter au système. Dans un second temps, je présenterai les évolutions technologiques et organisationnelles que nous avons mises en place. Cette phase de mise en exploitation s'est déroulée d'avril à octobre 2010.

1. Pérennisation du nouveau système d'information

Les résultats, lors de la phase de prototypage, de l'intégration des nouveaux concepts d'architecture du système d'information se sont avérés positifs. La cohérence de données et la mise en place de nouveaux services axés vers les utilisateurs ont constitué les principaux points positifs du bilan.

Traitement des nouvelles exigences

Tout d'abord, la création d'un référentiel unique de données afin de générer les services offerts a déplacé le point individuel de défaillance du système. Si le service LDAP n'était plus critique, le service de bases de données, référentiel unique, est devenu le nouveau point individuel de défaillance du système. Il fallait donc s'assurer de sa pérennité.

Ensuite, la génération dynamique du service LDAP a eu pour effet de bord de devoir stocker le mot de passe de chaque utilisateur dans la base de données. Cette solution insatisfaisante d'un point de vue de la sécurité nous a poussés à réfléchir à une solution pour extraire les données d'authentification de notre référentiel au profit des données d'identification. Il fallait donc trouver une organisation capable de résoudre cette problématique.

Apparition de nouveaux besoins utilisateurs

La mise à disposition de l'accès aux données personnelles de chaque utilisateur a engendré un effet « boule de neige » sur l'offre de services. Plusieurs utilisateurs sont venus rencontrer le service informatique pour nous faire part de leur adhésion et demander une augmentation de cette offre, par la mise en place de nouvelles fonctionnalités dans l'application mise à disposition.
Nous avons recensé ces demandes afin de réfléchir à la façon de les satisfaire.

Volonté d'optimisation des concepts du système

Le dernier point d'évolution était pour nous le plus important. Le service informatique voulait profiter de l'évolution du système d'information pour apporter davantage de sécurité et de pérennité au système en production. Nous devions, avec cette version de production, répondre aux problématiques initiales du projet.
Premièrement, sur l'aspect sécurité, nous souhaitions augmenter la sécurité du réseau, par la mise en place généralisée des technologies testées lors de la phase de prototypage, et sécuriser l'échange des authentifications sur le réseau.
Deuxièmement, il fallait repenser l'organisation et la répartition des systèmes et des services afin de pérenniser le système d'information et d'optimiser sa disponibilité.

Pour répondre à ces problématiques nouvelles, deux axes étaient clairement définis : des évolutions technologiques et des évolutions organisationnelles.

2. Evolutions technologiques

Nous allons voir dans ce chapitre quelles ont été les réponses technologiques apportées aux nouvelles problématiques. Nous verrons dans le chapitre suivant comment leur intégration a été organisée.

La réponse apportée aux nouvelles problématiques pour pouvoir mettre en exploitation les nouveaux concepts a été fixée sur deux axes de travail : un axe technologique, où la mise en place de technologies différentes pouvait répondre à certaines problématiques et un axe organisationnel grâce auquel une nouvelle organisation des services et des technologies pourrait apporter un gain en termes de disponibilité du système d'information.
Nous allons tout d'abord étudier la réponse technique que nous avons mise en œuvre.

L'amélioration du système par l'évolution technologique a été la réponse aux questions suivantes :
- La pérennisation du référentiel unique de données ;
- La séparation de l'authentification et de l'identification ;
- La sécurisation des accès réseau du laboratoire ;
- La pérennisation de l'architecture matérielle ;
- Le traitement des nouveaux besoins utilisateur (nouveaux scripts et nouvelles IHM).

2.1. Pérennisation du référentiel unique de données

Le référentiel unique de données, source de la cohérence de l'ensemble des données du système, se positionnait comme le principal point individuel de défaillance du nouveau système.
Afin d'éviter tout blocage du système, nous avons pris la décision d'étudier les différents SGBD disponibles sur le marché et de choisir celui qui pourrait le mieux satisfaire à nos exigences de

disponibilités et de pérennité. Pour cette étude, nous avons privilégié les possibilités de sauvegarde et de réplication synchrone des bases.

Lors de la phase de prototypage, le SGBD utilisé était MySQL en version 5.1. Ce SGBD avait été choisi pour des raisons de simplicité car le laboratoire disposait déjà d'un serveur MySQL qui contenait les bases du site collaboratif (site « Web »), du service de gestion de parc (GLPI) et quelques autres. Les scripts de sauvegarde étaient établis et l'équipe informatique maitrisait totalement son installation et son paramétrage. Cependant, quelques problèmes ont été recensés avec ce service durant la phase de développement et de tests du prototype :

- Un faible respect des normes SQL (libertés prises par le LMD MySQL) ;
- Un support incomplet des déclencheurs (triggers) et procédures stockées (manque d'évolutivité) ;
- L'absence de vue matérialisée, obligeant à la création de tables temporaires ;
- Le manque de solution gratuite de sauvegarde automatisée et « à chaud » ;
- Des dysfonctionnements des mécanismes de réplication liés à son implantation sur des VM.

Ces problématiques n'ont pas eu d'incidence sur les développements et les tests du prototype, mais il a fallu intervenir, à plusieurs reprises et manuellement, pour relancer la réplication entre les deux serveurs virtuels. Ces observations, conjuguées aux incertitudes concernant l'avenir de ce SGBD depuis son rachat par la société Oracle, ont motivé le service informatique du laboratoire pour évaluer d'autres solutions concurrentes.

Critères d'évaluation du nouveau SGBD

Un recensement et une évaluation des différentes technologies disponibles ont été réalisés ; les critères discriminants initiaux posés par le service informatique ont été les suivants :

- Le SGBD devait être une solution open source ;
- Il devrait avoir une compatibilité optimale avec les langages de scripts et d'IHM de façon native ;
- Il devait proposer des garanties de fiabilité, au moins sur un horizon de deux ou trois ans ;
- Le SGBD choisi devait être multiplateformes et compatible avec l'environnement technologique du laboratoire ;
- Les versions de même niveau du SGBD devaient disposer de mécanismes de réplication fiables, indépendants du système d'exploitation, et permettant des sauvegardes « à chaud » ;
- Il devait conserver un support optimal des normes du langage SQL.

De plus, le service informatique souhaitait, de préférence, privilégier un SGBD dont l'administration et l'utilisation serait réalisable par les membres du service sans nécessiter d'investissement exagéré en formation.

J'ai donc été chargé d'évaluer, en fonction de ces critères, les différents systèmes existants et de proposer le SGBD qui satisferait aux besoins de l'unité.

Pour effectuer mon choix, j'ai donc réalisé une phase de veille technologique sur les différentes bases de données du marché. Un premier raffinage selon les critères de sélection a permis rapidement de dégager une première liste de progiciels compatibles.

Recensement des solutions disponibles sur le marché

Parmi les différentes solutions présentes sur le marché, il existait plusieurs catégories de SGBD.
On trouvait d'abord des moteurs de bases de données minimalistes. Bien que très répandus et utilisés dans de nombreuses applications, ces moteurs étaient composés de bibliothèques écrites en C, d'une interface publique (API) et de connecteurs disponibles pour de nombreux langages de programmation (C++, Java, Perl, etc.).
Ensuite, dans une deuxième catégorie, j'ai recensé les SGBD orientés pour le déploiement de bases de données de moyenne importance. J'ai privilégié ceux qui correspondaient à nos critères initiaux (open source, gratuits, etc.).
Enfin, la troisième catégorie regroupait les SGBD orientés vers le déploiement d'importantes bases de données, comme Oracle ou IBM, principalement déployé sur des mini-ordinateurs (mainframes). Les éditeurs de ces SGBD proposaient des versions gratuites et allégées en fonctionnalités, mais avec des limitations de licences (éducation, personnelle, etc.).
Le tableau de synthèse ci-dessous présente l'analyse des solutions :

Tableau VII : Présentation des moteurs de bases de données étudiés

Type d'offre	Solutions Analysées	Analyse	Adéquation
Moteurs de bases de données minimalistes	• SQLite • BerkeleyDB	Simples bibliothèques écrites en langage C qui ne respectent pas le schéma habituel client-serveur mais sont directement intégrés dans les applications. Elles offrent une interface publique (API) et des connecteurs pour les langages de programmation.	Inadaptés (pas de schéma transactionnel), pas de possibilité d'installation en client-serveur.
Moteurs de bases de données de moyenne importance	• MySQL • PostgreSQL • Firebird	Support du langage SQL, respect partiel ou total des standard SQL-92, open source et gratuits. Richesse fonctionnelle, possibilités de réplication.	Systèmes en adéquation avec nos besoins et notre environnement.
Moteurs de bases de données orientés vers bases de moyenne à grande importance	• Oracle Database 10g Express Edition • IBM DB2 Express-C	Versions gratuites mais limitées des standards industriels. Préconisations techniques inadaptées : Oracle Database limitée à 4 Go, 1 CPU, 1Go de RAM et IBM DB2 Express-C nécessitant au moins 2 cœur/processeur et 2 Go RAM.	La mise en place de ces systèmes, leur administration, et le coût des licences en font des systèmes inadaptés à nos besoins et à notre environnement.

Les critères d'évaluation discriminants ont été principalement les possibilités de réplication, la compatibilité technologique avec notre environnement et l'adéquation à nos besoins. J'ai ainsi privilégié les SGBD s'adressant aux bases de données de moyenne importance. Dans cette catégorie, trois SGBD répondaient aux critères de choix : Firebird, PostgreSQL et MySQL.

Analyse des solutions choisies

Firebird était une base de données issue d'Interbase 6.0 (de Borland), reprise dans un projet open source et gérée par la Fondation FirebirdSQL.
Disposant de connecteurs disponibles pour de nombreux langages de scripts, d'un langage SQL assez proche des normes et de sauvegardes « à chaud » et incrémentales, Firebird était bien positionné pour les besoins du laboratoire malgré l'absence de vue matérialisée.
Cependant, les possibilités de réplication n'étaient pas intégrées nativement.
Plusieurs progiciels proposaient de mettre en place la réplication avec Firebird, mais l'offre pléthorique et peu lisible ne paraissait pas apporter suffisamment de garanties de pérennité, hormis par l'utilisation de solutions commerciales payantes (IBReplicator, CopyCat,…).

PostgreSQL [13] était un SGBD répandu dans le monde éducation-recherche, au même titre que MySQL. Sa communauté importante, les nombreuses ressources disponibles en littérature ou sous forme électronique permettaient de le mettre en pratique de façon assez rapide et efficace.
Open Source et gratuit, PostgreSQL supportait la majorité du standard SQL-92 et possédait aussi de nombreuses extensions comme le PL-SQL[3] (langage procédural développé par Oracle). Simple d'utilisation et d'administration, ses principaux inconvénients étaient un système de sauvegardes peu évolué, un support de bases de données de moyenne importance et une solution de réplication pas encore totalement intégrée.
Au début de l'année 2010, la version 9.0 est arrivée sur le marché en intégrant notamment comme amélioration des fonctions de réplication natives enfin stabilisées. Lors du déroulement de notre étude (été 2010), la première version stable était déjà annoncée (la release candidate 1 de PostgreSQL 9.0 est sortie le 28 aout 2010). J'ai tenu compte de ces nouvelles possibilités pour effectuer ma sélection.
La version 9.0 de PostgreSQL intégrait nativement des fonctions de réplication asynchrone, basées sur l'envoi des modifications au serveur esclave par le serveur maitre. La fonctionnalité de réplication synchrone était en développement et attendue pour la version 9.1.
Par contre, de nombreux progiciels complémentaires gratuits étaient intégrables pour diversifier les possibilités de réplication, comme la mise en place de réplication par Trigger (Slony ou Londiste) ou par duplication des requêtes envers tous les serveurs PostgreSQL du même pool (pgPool-II). J'ai pu dans ma recherche recenser plus d'une dizaine de systèmes de réplication autonomes.
Cette pluralité s'explique par le manque de solutions de réplications fiables qu'offrait PostgreSQL jusqu'à la version 8.4, ce qui a été corrigé sur la version 9.0. Cette absence de solution native a provoqué une diversité de possibilités pour la réplication des bases sous PostgreSQL, ce qui, finalement peut être considéré comme un avantage (souplesse d'utilisation).

MySQL (MySQL Community Server) [14] était le SGBD utilisé avec le prototype. Fonctionnel, son intégration dans le système du laboratoire était déjà effective. Cependant, les inconvénients cités plus haut ont poussé à sa mise en concurrence avec les autres solutions du marché.
La réplication avec MySQL était intégrée nativement depuis la version MySQL 3.23.15. Elle se réalisait par la réplication des logs binaires (réplication de type synchrone).
Dans notre environnement de tests (prototype), nous avons rencontré des problèmes de désynchronisation, problèmes que nous avons imputés au fonctionnement sous architecture virtuelle.

[3] *http://www.oracle.com/technetwork/database/features/plsql/index.html*

Ce problème n'a pas été constaté sur architecture physique sur des versions identiques d'autres bases au laboratoire.
La documentation officielle[4] de MySQL fait mention que « *Etant donné la nature non transactionnelle des tables MySQL, il est possible que le serveur esclave ne va faire qu'une partie de la modification, et retourner une erreur. (…) Si cela arrive sur le maître, l'esclave va s'arrêter et attendre que l'administrateur décide quoi faire, à moins que l'erreur ne soit légitime, et que la requête arrive à la même conclusion. (…)Si la requête sur l'esclave génère une erreur, le thread esclave s'arrête, et un message sera ajouté dans le fichier d'erreur. Vous devriez vous connecter pour corriger manuellement les données de l'esclave, puis relancer l'esclave avec la commande SLAVE START (disponible depuis la version 3.23.16. En version 3.23.15, vous devrez redémarrer le serveur.* ».
Ces possibilités recensées de désynchronisation, liées au manque de souplesse de la procédure de réactivation de la synchronisation (copie physique des bases d'un serveur sur l'autre) constituaient le point faible de ce SGBD. De plus, le rachat de MySQL par SUN en 2008, puis par Oracle (qui a racheté SUN) en 2009 laisse planer des incertitudes sur la pérennité de MySQL.

Conclusions et choix
La première conclusion de cette étude est que le marché des SGBD est très vaste et l'offre pléthorique. Le choix du SGBD doit s'appuyer sur des critères précis définis antérieurement en fonction des besoins exprimés.
Plusieurs catégories de SGBD existent, et chacun s'adresse à un marché différent : Oracle, IBM proposent des produits, principalement commerciaux, s'adressant au maintien et à l'administration d'importantes bases de données (au moins 2 Go de données). Quelques solutions spécifiques pour de petits environnements ont aussi été étudiées, comme SQLite ou Berkeley DB. Ces deux modèles étaient inadaptés à nos besoins.

L'objectif du choix du SGBD étant de pérenniser le référentiel unique de données, les possibilités de réplication et de sauvegarde ont été mes critères discriminants pour différencier ces trois systèmes.
J'ai donc proposé au service informatique de privilégier l'emploi d'une solution basée sur PostgreSQL en version 9.0. Les possibilités de réplications offertes par PostgreSQL et son environnement étaient sans commune mesure comparables avec les possibilités offertes par MySQL et Firebird.

L'option choisie a donc été l'implantation d'une version 8.4 de PostgreSQL avec mise en place de réplication par duplication des requêtes (utilisation de pgPool-II) en attendant la sortie de la version 9.0 qui allait permettre de réaliser cette réplication de façon native. Nous disposerions ainsi d'un système éprouvé tout en conservant des possibilités secondaires de réplication en cas de dysfonctionnement. Pour finir, cette approche de l'écriture en Y sur plusieurs serveurs offrait des possibilités intéressantes en termes de diversification des systèmes d'exploitation hébergeant les systèmes de gestion des bases de données.

[4] *http://dev.mysql.com/doc/refman/5.0/fr/replication-features.html*

2.2. Séparation de l'authentification et de l'identification

Comme je l'ai évoqué, la génération dynamique du service LDAP a eu pour effet de bord de devoir stocker le mot de passe de chaque utilisateur dans la base de données. Cette solution insatisfaisante d'un point de vue de la sécurité nous a poussés à réfléchir à une solution pour extraire les données d'authentification de notre référentiel au profit des données d'identification.
J'ai donc effectué une recherche sur les différentes technologies d'authentification que nous pouvions implémenter dans notre environnement et ai dirigé rapidement mon analyse vers la technologie Kerberos [RFC 4120] [12].

Kerberos est un protocole d'authentification distribué développé par le *Massachusetts Institute of Technology* (MIT). Kerberos offre une authentification sécurisée pour les échanges client/serveur reposant sur un mécanisme de clés secrètes et d'utilisation de tickets. Il permet ainsi de chiffrer les échanges entre les clients (utilisateurs, services ou machines), le serveur d'authentification Kerberos (le KDC, Kerberos Distribution Center) et les ressources auxquelles les clients souhaitent accéder.
L'avantage principal est d'éviter la circulation des mots de passe en clair sur le réseau, évitant ainsi le risque d'interception frauduleuse des mots de passe des utilisateurs.
Les informations qui sont échangées pour prouver une identité sont appelées tickets. Les informations utilisées pour chiffrer les tickets et les communications qui en découlent sont appelées clefs. Une fois l'identité d'un client vérifiée, le client reçoit un jeton qui peut être utilisé par tous les services compatibles avec Kerberos pour vérifier son identité.
Pour des raisons de sécurité, les jetons sont horodatés ainsi, ils expirent automatiquement à moins qu'ils ne soient renouvelés par un utilisateur ou un service. Les horodatages contenus dans les jetons et les tickets ne peuvent être vérifiés que si les heures et les dates sont synchronisées entre les clients et le KDC. L'authentification Kerberos échouera si les horloges se désynchronisent de plus de cinq minutes.

Plusieurs implémentations de Kerberos existent, dont par exemple l'implémentation Heimdal, créé à l'origine pour s'affranchir des restrictions à l'exportation des Etats-Unis (aujourd'hui abandonnées). Nous avons privilégié l'implémentation du MIT qui nous paraissait la plus répandue dans le monde Linux, et fournie nativement sur nos systèmes d'exploitation Red Hat Enterprise Linux.

Dans le cadre de notre projet, nous avons étudié les impacts de l'intégration de la technologie Kerberos sur notre architecture. Premièrement, nous devions extraire les mots de passe des utilisateurs de notre référentiel unique, et donc ne plus en disposer pour générer la configuration du LDAP. Cette politique de **séparation de l'authentification et de l'identification** a eu pour incidence de mettre en place une seconde base de vérité indépendante du référentiel unique d'identification (service d'authentification Kerberos).

La **sécurité de l'authentification** a également été renforcée par la **circulation cryptée des mots de passe** sur le réseau entre les clients kerbérisés et le service d'authentification. Cependant, les services non « kerbérisés » ont dû être passés en LDAPS pour éviter toute circulation de mots de passe en clair. Pour cela, nous avons supprimé la possibilité d'accéder au LDAP de façon non sécurisée (filtrage par pare-feu logiciel du serveur).

D'autre part, l'utilisation du mécanisme d'authentification **SASL** (mise en place du service *saslauthd* sur le serveur LDAP) pour interconnecter les services LDAP et Kerberos a permis de conserver l'**unicité des mots de passe** en permettant aux services non kerbérisés de s'authentifier auprès du service LDAP. LDAP utilisait le connecteur SASL pour réaliser lui-même l'authentification auprès du service Kerberos et renvoyer le résultat au client, comme s'il s'agissait d'une authentification locale.

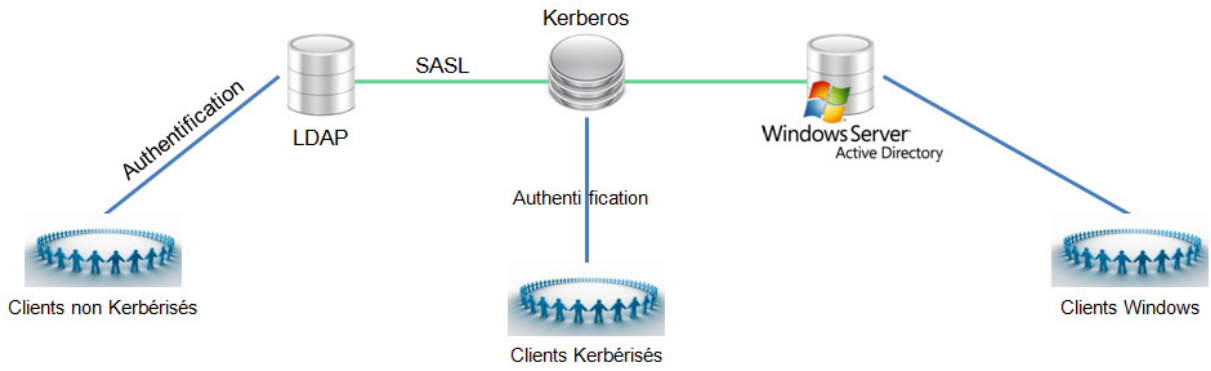


Figure 16 : Authentification en environnement hétérogène

Enfin, cette mise en œuvre a apporté des fonctionnalités nouvelles (Figure 16) que nous n'avons pas implémentées dans ce projet, comme la fédération d'identités ou l'interconnexion avec les annuaires Microsoft (Active Directory).
En effet, la mise en œuvre de notre référentiel Kerberos offrait la possibilité d'utiliser les bases d'authentification pour participer à des **fédérations d'identités** (CAS, SSO), utiles pour les échanges de services entre les laboratoires (par exemple, accès à la grappe de calcul du L.A.).
Dans le cas des annuaires Active Directory (AD), nous avions la possibilité d'offrir une authentification unique à tous les services quelque soit l'environnement de l'utilisateur et ce dès la connexion à son poste de travail.

La migration vers Kerberos a été réalisée grâce aux nouvelles IHM. Le principal problème rencontré lors de l'intégration de Kerberos au système d'information était qu'il fallait migrer l'authentification du service d'annuaire (LDAP) vers le service d'authentification (Kerberos). Cette migration a été réalisée en intégrant au code des nouvelles IHM une réécriture des mots de passe dans ces deux services.
Pour cela, lors de sa première connexion au nouveau système, l'utilisateur authentifié non enregistré dans le référentiel d'authentification se voyait proposer de changer son mot de passe. Celui-ci n'était pas enregistré dans le service LDAP mais directement dans le service Kerberos. Le service LDAP recevait la chaine de connexion permettant de déléguer l'authentification vers le service Kerberos. A la connexion suivante, l'utilisateur accédait directement à ses services.

Pour basculer l'authentification des utilisateurs du service LDAP vers le service Kerberos, nous avons utilisé un algorithme permettant de basculer les utilisateurs progressivement, au fur et à mesure de leurs connexions sur leur espace personnel. La bascule des authentifications s'est donc déroulée sans arrêt de service.
L'algorithme de mise à jour des mots de passe utilisé a été le suivant :

```
Si  Authentification Kerberos = TRUE (utilisateur) Alors
|       Accéder aux IHM
| Sinon
|       # Authentification Kerberos = FALSE, utilisateur non reconnu ou absent du KDC
|       Si Authentification LDAP = TRUE Alors
|       |       Enregistrer un nouveau mot de passe dans Kerberos
|       |       Rediriger l'authentification du LDAP vers Kerberos via SASL
|       | Sinon
|       |       Refuser l'accès aux IHM
|       FinSi
FinSi
```

Le seul point noir que nous avons rencontré avec la technologie Kerberos a été lié à la découverte d'une technologie compliquée à appréhender, implanter et maitriser. C'est le point technique sur lequel l'investissement de l'équipe informatique a été le plus important en termes de formation.

2.3. Sécurisation des accès réseau du laboratoire

La sécurisation du réseau filaire a été renforcée par la mise en place du cryptage des mots de passe, circulant sur le réseau, via Kerberos. Il fallait également s'assurer que les accès à ce réseau étaient limités aux seuls matériels autorisés.
Nous avions testé les technologies d'identification des matériels lors de leur connexion (service d'authentification Radius) sur un commutateur réseau configuré spécifiquement en phase prototypage. A la vue de ces résultats, l'équipe informatique a pris la décision de généraliser cette organisation pour l'ensemble du réseau du laboratoire.

La phase suivante d'inventaire a duré deux semaines, et a permis l'identification et la déclaration des matériels existants au laboratoire. Ensuite, une fois ces données intégrées au référentiel unique, nous avons régénéré dynamiquement les fichiers de configuration du service Radius et avons paramétré tous les commutateurs réseau pour utiliser ce type d'identification.

De cette façon, nous nous sommes assurés que toute station connectée au réseau avait été identifiée, et que tout matériel non déclaré était positionné sur un réseau logique indépendant et offrant des services réduits.

2.4. Pérennisation de l'architecture matérielle

Après les résultats obtenus par la phase de prototypage, nous avions validé les concepts et souhaitions poursuivre la voie de la virtualisation pour mettre en exploitation notre nouveau système d'information.
Ces tests ont été réalisés à un instant stratégique puisque plusieurs de nos serveurs physiques étaient hors garantie ou proche de la fin de la période de garantie. Sachant que, pour des raisons de continuité de service, le service informatique du LATT ne souhaitait pas utiliser en production de matériel hors garantie, il était temps de renouveler notre parc matériel.

Dans le cadre de cette jouvence matérielle, deux options étaient possibles : continuer à utiliser une architecture matérielle en renouvelant la garantie des serveurs, ou investir dans une infrastructure virtuelle de production. Grâce aux retours d'expérience du prototype, nous avons rapidement évalué les différences de coûts des deux solutions.

Le choix a donc été pris de poursuivre l'effort financier vers la virtualisation. Nous avons opté pour la solution payante de VMware, Virtual Infrastructure 3 (VI3), dont le coût était d'environ 15 k€.
A cet investissement, il faut ajouter l'achat de deux serveurs puissants et d'une baie de stockage équipée de contrôleurs optiques (technologie Fibre Channel).
Les deux serveurs choisis ont été sélectionné pour leur adéquation aux besoins de virtualisation : des serveurs biprocesseurs, avec pour chacun quatre Cœur par processeur (bi-quad Core), 24 Go de Mémoire (RAM) de deux disques SAS de 146 Go chacun. Le coût de ces matériels était d'environ 9 k€.
La baie de stockage compatible SAN, équipée de disques rapides (technologie SAS à 15k tr/mn) a été dédiée à l'infrastructure VMware, pour un coût d'environ 10 k€. Cette baie d'environ 3 To a été installée sur l'architecture SAN du laboratoire et était partagée par les deux serveurs dédiés à la virtualisation afin d'obtenir un espace disque commun et mutualisé pour héberger les machines virtuelles.

L'investissement matériel et logiciel pour cette architecture VMware VI3 s'est élevé a 35 k€. Comparativement à une architecture physique, le coût initial est plus élevé. Au tarif éducation, 35 K€ correspondent à l'achat d'environ douze à quinze serveurs pouvant faire fonctionner des services d'infrastructure (le coût d'un serveur scientifique est plus élevé).

Cet investissement initial a apporté malgré cela davantage de possibilités :
- En termes de serveurs, nous faisons fonctionner sur cette architecture actuellement une quarantaine de VM. Le taux d'occupation de chaque serveur est d'environ 30%, ce qui laisse la possibilité, sans investissement nouveau, d'ajouter d'autres services.
- En termes d'encombrement, deux serveurs et une baie de stockage ont remplacé avantageusement l'installation d'au moins une vingtaine de serveurs. Le gain est donc réalisé en termes d'occupation (nombre de « U »), de consommation énergétique (climatisation, onduleur et électricité) et d'infrastructure réseau (diminution du nombre de prises).
- En termes de souplesse d'utilisation : la possibilité de créer des machines virtuelles rapidement pour réaliser des maquettes de développement de services ou de répondre rapidement à la mise en place de nouveaux services à coût réduit était un net avantage sur l'architecture physique.

- En termes de continuité d'activité : l'architecture VI3 apportait des fonctions permettant de positionner notre architecture pour réaliser de la haute disponibilité : déplacement de machines virtuelles automatisé sans coupure de service (load balancing, VMware H.A.), redondance des connexions réseau (fail-over), possibilité d'organiser notre architecture sur plusieurs salles ou sites, etc.

A terme, dans notre environnement, cet investissement fut rentable car il a permis d'économiser sur plusieurs autres postes, ce qui a fait de l'opération virtualisation une opération positive en termes d'investissement informatique pour le laboratoire.

2.5. Mise à jour des IHM et des scripts d'alimentation de données

L'emploi d'une personne compétente en développement orienté vers les architectures N-tiers (clients légers) nous a permis de rendre nos interfaces plus intuitives et attrayantes. En effet, suite à la phase de prototypage, de nouvelles demandes étaient apparues de la part de nos utilisateurs, principalement en termes de gestion de leurs informations et comptes sur leurs différents serveurs d'infrastructures ou scientifiques.
Pour répondre à ces nouveaux besoins, de nouvelles IHM ont été développées pour améliorer l'ergonomie et enrichir les fonctionnalités offertes. De nouveaux scripts de récupération de données, déployés sur plusieurs serveurs scientifiques, ont eux aussi été développés.
Les langages de développement utilisés, initialement PHP et PERL, ont été complété par les technologies AJAX et JQuery (technologies Web 2.0).
Au final, nous avons obtenu des interfaces plus ergonomiques, plus intuitives, une offre de service élargie et un meilleur taux de satisfaction auprès des utilisateurs.
Quelques IHM de cette seconde version sont fournies dans l'Annexe 5.

3. Evolutions organisationnelles

Nous allons voir dans ce chapitre quelle organisation a été mise en œuvre pour réaliser l'intégration des évolutions technologies que nous venons d'énumérer.

Les solutions techniques ont été choisies et implémentées indépendamment. Il s'agissait ensuite d'organiser leur intégration et leurs interconnexions afin de pouvoir obtenir le meilleur taux de disponibilité possible.

3.1. Notion de disponibilité des systèmes

La notion de disponibilité des systèmes est un terme employé dans le monde informatique à propos d'architecture de systèmes (ou d'un service) pour désigner le fait que ces architectures (ou ce service) ont un taux de disponibilité en adéquation avec ses activités. La disponibilité est donc la possibilité d'utiliser l'ensemble des services de façon normale et continue, sans coupure de service.

L'absence de notion de haute disponibilité (HA, « high availability ») d'un système peut se traduire par des discontinuités de service à chaque défaillance (logicielle ou matérielle) et un besoin d'intervention humaine pour remettre le système dans un état stable et fonctionnel.
L'administrateur système et réseau (ASR) doit intervenir pour rechercher la cause du dysfonctionnement et prendre les mesures appropriées, comme par exemple :

- Redémarrer les processus arrêtés, en restaurant l'état du logiciel ;
- Contacter les services de support des logiciels défaillants, en cas de bug ou de limitation manifeste et reproductible ;
- Reconnecter les machines ou remplacer les pièces matérielles défectueuses.

Le plus souvent, cette analyse de problème est consolidée par des actions correctives pour éviter tout nouveau dysfonctionnement identique. Ce sont des mesures principalement curatives.
A l'inverse, l'établissement d'un plan de HA est constitué d'un ensemble de mesures préventives visant à anticiper ce type de problème, généralement par plusieurs évolutions dans l'organisation. Ces évolutions consistent essentiellement à détecter les points de défaillance du système et à les réduire par la mise en place de techniques de redondance et/ou de réplication.

Mesure et estimation de la disponibilité

La discontinuité de service peut durer jusqu'à plusieurs heures selon la gravité du dysfonctionnement et affecter le temps de disponibilité du service rendu. On peut mesurer la disponibilité d'un service ou d'un système avec un pourcentage basé sur des mesures annuelles.
Ce taux de disponibilité d'un système ou d'un service s'obtient en appliquant la relation suivante :

Disponibilité (%) = MTBF / (MTBF + MTTR)

Où : MTBF (mean time between failure) = mesure du temps estimé (ou moyen) entre deux défaillances d'un système donné (fréquence de défaillance du système considéré).
MTTR (mean time to repair) = mesure du temps estimé (ou moyen) pour réparer le système suite à la défaillance.

On obtient donc un pourcentage indiquant le temps d'indisponibilité du système. Cette règle de calcul est connue sous le terme de « règle des '9' » [17] [18], puisqu'on utilise généralement un pourcentage essentiellement composé de '9' pour l'expliquer:

- 99% désigne le fait que le service est indisponible moins de 3,65 jours par an
- 99,9%, moins de 8,75 heures par an
- 99,99%, moins de 52 minutes par an
- 99,999%, moins de 5,2 minutes par an
- 99,9999%, moins de 54,8 secondes par an
- 99,99999%, moins de 3,1 secondes par an
- etc.

Par exemple, un service qui aurait dysfonctionné pendant deux heures durant l'année écoulée (8766 heures) aurait un taux de disponibilité de :

$$8766 / (8766 + 2) = 0,999 \text{ soit } 99,9\ \%$$

Il faut cependant replacer le résultat obtenu dans le contexte de chaque entreprise et les mettre en adéquation avec les besoins de l'entité. Une interruption de service pendant quatre heures dans un laboratoire de recherche n'a pas le même impact que pour une chaine de production fonctionnant à flux tendu.

Indicateurs

Pour évaluer l'impact des dysfonctionnements dans l'environnement spécifique d'une entité, on peut s'appuyer sur deux critères : le RTO et le RPO.

• Le RTO (Recovery Time Objective) qui est la durée maximale d'interruption admissible. Cet indicateur définit le temps alloué pour effectuer, en cas de dysfonctionnement du système, un basculement vers un nouveau système et une reprise des services.

• Le RPO (Recovery Point Objective) qui est la perte de données maximale admissible. Cet indicateur définit l'état (au sens large) dans lequel doit se trouver le nouveau système après basculement et reprise des services. On fixe notamment la perte admissible d'informations que l'on peut potentiellement subir dans l'intervalle de temps où le système de secours permet de retrouver un fonctionnement normal.

Un exemple couramment donné est le cas du secteur bancaire où l'on pourra définir un RTO de 1 heure avec un RPO de 0 secondes, sans mode dégradé. Aucune transaction ne sera ainsi perdue, et le service pourra être disponible sous une heure.

Il faut pondérer ces résultats par le fait que la durée maximale d'interruption admissible varie selon les services. D'une manière générale, plus le RTO sera important et plus le risque de perte de données (RPO) sera élevé.

Cependant, la différence de criticité entre les services impliquera un ordonnancement de la relance des services afin d'obtenir le plus rapidement possible un retour à un fonctionnement normal avec le minimum de pertes d'informations.

Dernier point, mais tout aussi important, l'augmentation de la disponibilité a un impact fort sur le coût de mise en place et de maintenance d'un système en production. Mettre en place une architecture pérenne (redondée, architecture de secours, etc.) est une action qui doit impérativement être adaptée à l'activité et aux besoins de l'entreprise ou du laboratoire. En effet, chaque optimisation aura une incidence significative sur le coût de mise en place du système.

De plus, la redondance de services et la tolérance aux pannes sont des notions essentielles pour nos métiers. Cependant, il s'agit de notions difficilement justifiables d'un point de vue comptable puisque le renforcement de la disponibilité de l'architecture augmente le prix de revient de chaque service sans proposer de services supplémentaires pour l'utilisateur.
Il faut donc trouver un équilibre entre augmentation de la disponibilité et coût de revient des services offerts.

Pour cela, Il fallait estimer le taux de disponibilité cible et évaluer les actions à mettre en œuvre en tenant compte de notre environnement technologique et économique.

Mise en place de la disponibilité dans notre contexte

Pour mettre en place un système à haute disponibilité, il faut généralement établir un plan de continuité d'activité (PCA). Le PCA défini l'organisation que l'entreprise (l'entité) doit mettre en pratique afin d'assurer une disponibilité maximale de son système d'information. Il est propre à chaque entité et doit être développé, maintenu et ajusté en fonction des évolutions de l'environnement de l'entreprise : économique, technologique et politique.
C'est la solution que nous avons choisi d'adopter pour accroître la disponibilité du système d'information du LATT.

N'ayant pas initialement de compétences dans ce domaine, j'ai réalisé une période de veille technologique sur la mise en place de ce type de plan et ai proposé à ma hiérarchie des solutions à déployer dans notre environnement.

Je vais tout d'abord expliquer le cheminement que j'ai suivi et la méthode que j'ai appliquée pour établir le plan de continuité d'activité du système d'information du laboratoire. Je détaillerai ensuite les conséquences organisationnelles et technologiques de mes propositions sur l'infrastructure du système d'information et sur la mise en exploitation de la version.

3.2. Plan de Continuité d'Activité

Le plan de continuité d'activité (PCA) a pour objectif le maintien de l'activité de l'entreprise, sans coupure de service majeure, après un sinistre important survenu sur le système d'information. Le système passe dans un état dégradé qui, selon l'impact, minimise la perte de données et permet de continuer une activité en mode dégradé durant le délai de retour à un état stable du système.

Comme décrit graphiquement dans la Figure 17, il s'agit de mettre en place une organisation des systèmes permettant de maintenir la disponibilité du système en cas de dysfonctionnement (temps h) sans coupure d'activité ou de service, le temps que les intervenants puissent résoudre le dysfonctionnement (temps h+d).

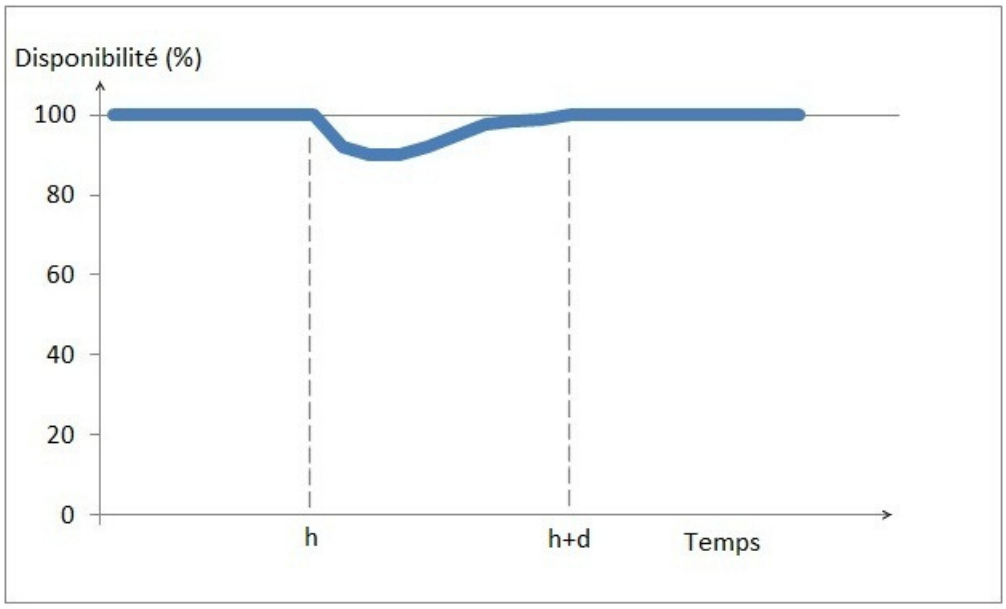


Figure 17 : Disponibilité du système avec reprise d'incident par PCA

Pour établir ce PCA, il faut tenir compte des besoins et des exigences de l'entité dans laquelle il doit être appliqué. Il doit reposer sur une phase d'analyse des risques et une phase d'analyse de l'impact des ces risques.

L'analyse de risque est l'identification des différents risques pouvant affecter l'architecture en production. L'analyse d'impact consiste à évaluer l'impact de chaque risque et à déterminer à partir de quand cet impact sera bloquant et aura une incidence sur les processus qu'il affecte.
Ces deux analyses se font par anticipation en élaborant des scénarios de risque (coupure de service, perte de matériel, etc.) et en quantifiant leur impact. Une fois ces analyses effectuées, les points faibles du système d'information, et notamment les points individuels de défaillance, vont être mis en exergue.

L'un des principaux retours d'expérience sur le sujet montre que les solutions efficaces reposent sur l'implication de tous les acteurs du système d'information : le service informatique, les services administratifs et l'ensemble des utilisateurs.

3.3. Plan de Reprise d'Activité

Aborder le PCA sans parler du plan de reprise d'activité (PRA) serait incomplet.
Le plan de reprise d'activité est un plan qui a pour objectif le redémarrage de l'activité dans un délai le plus court possible avec le minimum de perte de données (notions de RPO et de RTO).
Un plan de reprise d'activité permet d'assurer en cas de crise (majeure ou mineure) d'un centre informatique la reconstruction de son infrastructure et la remise en route des services supportant l'activité d'une organisation.
Le plan de reprise d'activité est donc l'ensemble des procédures et des actions permettant de basculer sur un système de secours capable de prendre en charge les besoins informatiques et de rendre les services nécessaires au bon fonctionnement de l'organisation, en mode normal ou dégradé, le temps de rétablir le niveau de service initial (Figure 18).

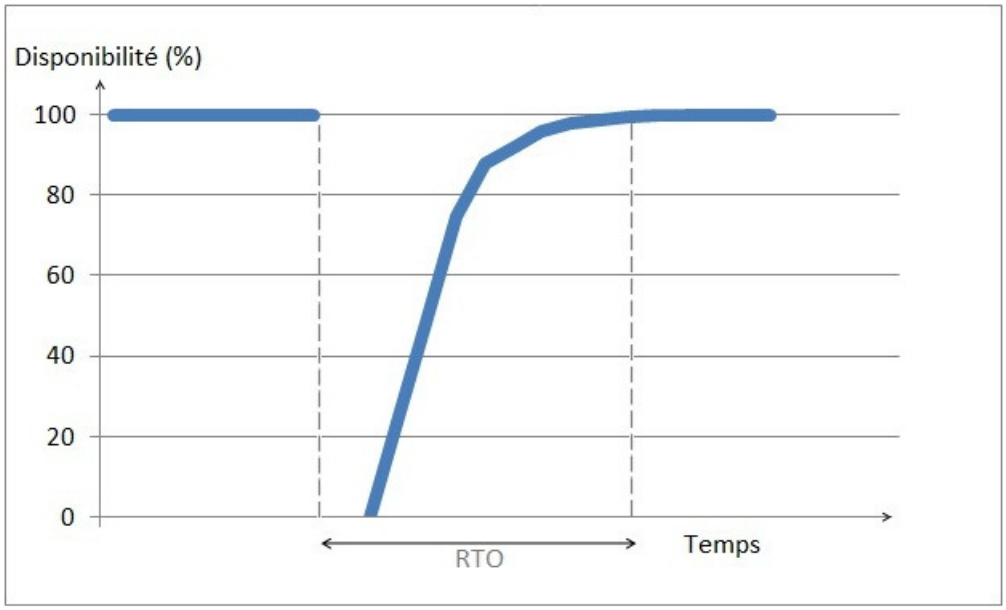


Figure 18 : Disponibilité du système avec reprise d'incident par PRA

Plusieurs niveaux de PRA sont possibles et acceptables selon les besoins exprimés par l'organisation.
C'est l'adéquation entre la nécessité de disponibilité du service et la capacité financière de l'organisation qui va permettre de fixer le système de reprise. En effet, la mise en place d'un PRA peut être une opération très onéreuse à mettre en place et à maintenir. Il est donc important de définir correctement les attentes du système de reprise en fonction des besoins réels.

L'implantation d'un PRA est une action qui devrait être considérée comme un pré-requis à toute implantation d'un nouveau système d'information. C'est le gage d'un retour à un état stable du système après toute coupure de service potentielle, qu'elle soit mineure ou majeure.

Le PRA le plus basique peut être simplement constitué de sauvegardes : certaines activités peuvent supporter un RTO (durée maximale d'interruption admissible) assez conséquent et un RPO (perte de données maximale admissible) moyen.
Le cas général sera moins laxiste et on trouvera principalement dans les entités un RPO et un RTO réduits au minimum possible ; cet intervalle de temps sera pondéré par les moyens financiers et humains mis en pratique pour ce PRA, comme évoqué au chapitre précédent.

3.4. Articulation PRA / PCA

Le choix d'application entre un PRA et/ou un PCA est un choix propre à chaque entité selon ses besoins et ses capacités humaines et financières. Si certains systèmes peuvent se contenter d'un PRA (systèmes sans criticité et risques de perte de données), la mise en place de ces deux plans est nécessaire pour améliorer le taux de disponibilité.

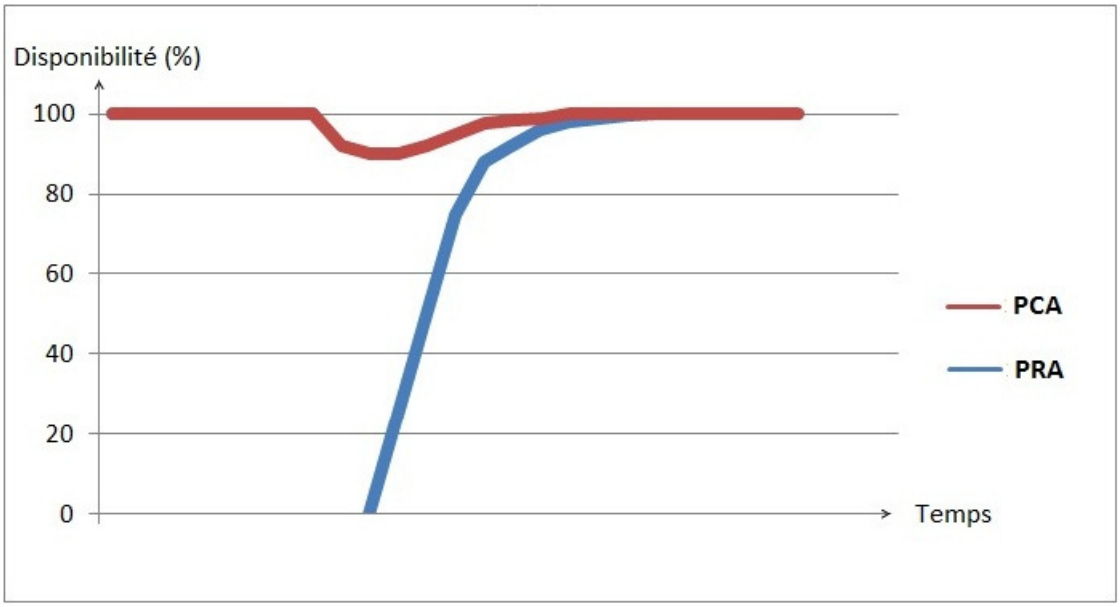


Figure 19 : Disponibilité du système - différence entre le PCA et PRA

En effet, si l'on compare les courbes de disponibilité des deux plans (Figure 19), le choix ira directement à la mise en place d'un PCA pour toute entité ayant besoin de continuité. Cependant, il n'existe pas de PCA fiable à 100% car il est simplement très difficile d'évaluer de façon exhaustive tous les scénarios de dysfonctionnement potentiels (exemple : explosion de l'usine AZF de Toulouse). La mise en place d'un PCA passe donc inévitablement par la mise en place d'un PRA, en cas de dysfonctionnement du PCA.

Le PCA est plus complexe à mettre en œuvre car il se construit en amont de l'implantation de l'architecture des systèmes. A ce titre, il ne parait pertinent que pour des structures informatiques ayant une taille critique minimale.
Le coût financier et humain étant largement supérieur à celui d'un PRA, on privilégiera le PCA pour les structures ayant un système d'information ne pouvant ou ne souhaitant pas supporter de problème de disponibilité (architectures industrielles, chaines de production, banques, etc.). Mais une fois de plus, c'est le coût financier et l'adéquation aux besoins de l'entreprise qui définira la mise en place d'un PRA seul ou d'un PCA accompagné d'un PRA.

En effet, le coût d'implantation d'un PCA est souvent supérieur à celui d'un PRA, ne serait-ce que parce qu'il impose une multiplication des services et/ou des serveurs (physiques ou virtuels) et bien souvent des technologies demandant des compétences plus spécialisées (heartbeat, failover,…).

Néanmoins, l'apport récent des technologies de virtualisation tend à diminuer l'écart existant entre les deux plans de continuité. Le coût de mise en place d'une architecture virtuelle redondante étant largement moins élevé que l'investissement nécessaire pour doubler une architecture physique, c'est au niveau de l'organisation des machines virtuelles et de l'implantation des serveurs de virtualisation qu'il faut réfléchir pour optimiser les coûts de la mise en place du PCA.

Dans le cadre de mon étude, j'ai décidé de développer un plan de continuité d'activité avant la mise en exploitation de notre nouveau prototype, et de coupler ce PCA avec l'élaboration d'un plan de reprise d'activité.

Pour cela, je me suis documenté sur les PCA et les méthodes permettant de les développer.

Je vais présenter ici mes résultats et expliquer quel a été mon choix méthodologique.

4. Mise en place de la haute disponibilité

Nous allons voir dans ce chapitre la méthodologie que j'ai appliquée pour définir l'organisation de la nouvelle architecture pour tendre vers la haute disponibilité. Nous verrons ensuite les conséquences de cette mise en œuvre sur l'architecture du système d'information du laboratoire.

4.1. Choix de la méthode pour établir le PCA

Convaincu de la nécessité de la mise en place d'un PCA et d'un PRA pour atteindre la haute disponibilité, j'ai réalisé une période de veille technologique sur la mise en place d'un plan de continuité d'activité.
Ne disposant pas d'expérience sur ce domaine, je me suis documenté auprès de diverses sources disponibles : responsables informatiques, ouvrages de référence [2] et documentations électroniques [19] [20]. J'ai recensé plusieurs méthodes d'analyse existantes et ai commencé à les étudier. J'ai éliminé les méthodes basées sur des logiciels commerciaux (comme SCORE d'Ageris Consulting) ou anciennes (comme EBIOS de la DSSI datant de 1995 ou MELISA déjà abandonnée par la DGA).
J'ai finalement orienté mes recherches sur les méthodes MARION et MEHARI proposées par le CLUSIF et citées dans de nombreuses sources que j'ai pu étudier.
Le CLUSIF (Club de la sécurité des systèmes d'information français) est une association française d'entreprises et de collectivités réunis en groupes de réflexion et d'échanges autour de différents domaines de la sécurité de l'information : gestion des risques, politiques de sécurité, cybercriminalité, intelligence économique, etc.
Cette association est notamment reconnue pour ses études sur les stratégies et solutions de secours des systèmes d'information, disponibles depuis son site Internet [15].
La méthode MARION (1980) a été abandonnée en 1995 au profit de la méthode MEHARI (MEthode Harmonisée d'Analyse de RIsques). La méthode MEHARI est une méthode de management des risques liés à la sécurité de l'information. Elle apporte des préconisations en termes de sécurité depuis de l'organisation de l'entreprise jusqu'aux spécifications techniques à mettre en œuvre.
Cette méthode est régulièrement actualisée pour être adaptée aux évolutions de l'environnement technologique et socio-économique des entités. Sa dernière mise à jour date de janvier 2010.

Ma réflexion sur l'élaboration du PCA du système d'information du LATT s'est donc basée sur l'utilisation de cette dernière méthode. Son application a permis de déceler plusieurs failles potentielles de notre système d'information et de proposer des solutions adaptée. Ces réponses apportaient des éléments permettant d'optimiser la disponibilité du système d'information en production.

Grâce aux préconisations du fournisseur, j'ai orienté la création et la mise en place des plans de continuité et de reprise d'activité de notre laboratoire.
Je vais à présent détailler la méthode, son fonctionnement, son implémentation et expliquer les répercussions de ces préconisations sur notre organisation.

Présentation de la méthode MEHARI

La méthode MEHARI est une méthode de management (analyse et de gestion) des risques liés à la sécurité de l'information. La distribution initiale de MEHARI, émanant du CLUSIF, est constituée de trois éléments : un fichier "base de connaissance", un manuel de référence des services de sécurité, et un manuel de référence des scénarios de risque.
La méthode MEHARI fournit un cadre méthodologique et des bases de connaissance pour :

- Analyser et classifier les enjeux majeurs ;
- Étudier les vulnérabilités ;
- Réduire la gravité des scénarios de risques ;
- Piloter la sécurité de l'information.

L'approche modulaire de la méthode MEHARI reprend le principe de la méthode du plan de continuité d'activité par phases successives : analyse des enjeux, analyse des vulnérabilités, identification et traitement des risques.

Dans le cadre de mon étude pour le laboratoire d'Astrophysique, je me suis principalement inspiré de la méthode MEHARI pour détecter les failles potentielles du système d'information et du manuel de référence des services de sécurité MEHARI pour trouver des réponses à nos problématiques.

Le manuel de référence a pour objectif de référencer les éléments et les méthodes de l'entité à sécuriser. Il accompagne la base de connaissance utilisée lors de la phase d'analyse des vulnérabilités et vient s'adosser à cette base :

- Par le rappel de l'objectif de chaque élément des questionnaires de la base ;
- Par l'indication des résultats attendus au niveau des services et des sous-services ;
- Par la description des mécanismes et solutions associés à chaque sous-service, y compris la distinction entre mesures organisationnelles et mesures techniques ;
- Par la clarification des éléments permettant d'établir la qualité de chaque sous-service selon trois critères d'analyse : l'efficacité, la robustesse et la mise sous contrôle.

Les questions d'audit de la méthode sont autant de points de contrôle de l'existant (technique ou organisationnel) répondant au besoin d'assurer l'objectif du service tel qu'il est décrit. La conclusion de l'analyse met en exergue les points faibles du système.

L'apport des préconisations pour augmenter la disponibilité du système d'information est pondéré par le nombre de modifications mises en œuvre en fonction de la possibilité technique et financière d'implanter les solutions proposées.

L'objectif final de l'application de la méthode MEHARI est donc l'adoption de mesures préventives visant à réduire la potentialité des risques et la mise en place de solutions assurant un niveau d'activité optimal.

4.2. Élaboration du PCA

D'une manière générale, la méthode MEHARI considère plusieurs types de services organisationnels ou techniques. J'ai mis en œuvre cette méthode pour analyser les forces et les faiblesses du système d'information existant afin d'émettre des préconisations pouvant permettre d'augmenter sa disponibilité.
Pour cela, j'ai suivi les étapes préconisées (Figure 20) par la méthode MEHARI du Clusif, présentée dans le paragraphe précédent. Le résultat de chaque étape conditionne l'étape suivante et doit faire l'objet d'une validation par le comité de pilotage (dans notre cas, le service informatique).
Ces étapes itératives correspondent aux items présentés dans la Figure 20.

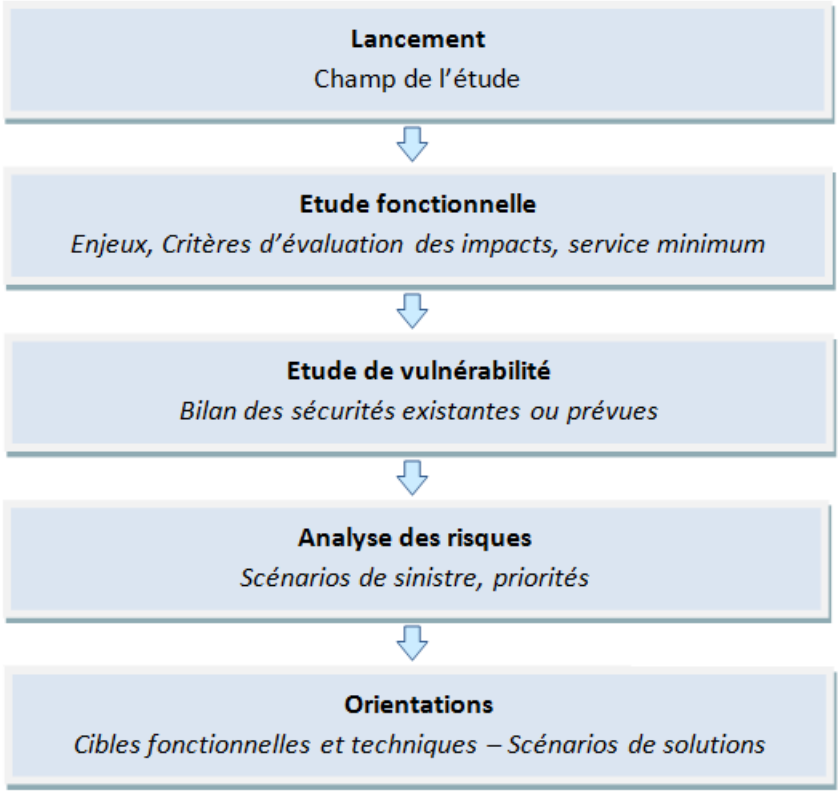


Figure 20 : Description des étapes de la méthode MEHARI

Phase 1 : Lancement

La méthode MEHARI propose une approche globale du système de l'entreprise pour l'élaboration de plans de continuité d'activité. Elle prend en compte tous les métiers de l'entreprise avant de procéder à des consolidations par cercles de métiers. Elle demande un arbitrage et une validation de la direction générale par processus itératifs. Cette méthode permet de mettre en place un PCA qui tient compte de l'organisation générale de l'entité et implique les responsables d'activité, la direction générale, les responsables sécurité et, à un moindre degré, l'ensemble des acteurs de l'entreprise.

Dans le cadre de mon stage d'ingénieur, la mission qui m'a été confiée était d'améliorer la disponibilité de la partie informatique du système d'information du laboratoire de recherche dans lequel j'exerçais mes fonctions. Ce périmètre a défini le champ de l'étude et a servi de base à l'ensemble des étapes suivantes de la méthode.

Phase 2 : Etude Fonctionnelle

La deuxième phase est la phase d'étude fonctionnelle qui a pour objectif de définir, pour chaque activité, les exigences de continuité. Cette étude a été réalisée sur l'ensemble des services du LATT. J'ai recensé les différents services existants et les ai classés selon leur criticité dans un tableau. On distinguait deux types de services :

- Les services critiques : services d'authentification, d'identification, d'accès au réseau, de bases de données, etc. Il s'agissait des services dont la coupure entrainait le blocage des autres services et donc du système (RTO et RPO faibles);
- Les services non critiques : services de messagerie, annuaire, services d'impression, etc. Il s'agissait des services dont un dysfonctionnement entrainerait une dégradation sans blocage de l'activité (RTO moyen et RPO faibles à élevés).

J'ai ensuite positionné des indicateurs sur chacun d'entre eux afin de définir le service minimum acceptable, le délai de reprise d'activité, la durée du service minimum, le niveau de dégradation de service acceptable, les conditions de retour à la normale (on retrouve les notions de RTO et le RPO) et la priorité de redémarrage du service.
En effet, l'interruption d'un service critique (authentification) était à traiter avec plus de priorité que l'arrêt d'un service non critique (impression).

J'ai aussi mené cette étude sur les matériels et salles techniques pour tenir compte au maximum de l'environnement de travail. La validation de ce recensement a été faite collégialement entre les membres du service informatique afin de tenir compte des retours d'expérience de chacun et de l'objectif à atteindre.

Cette phase a permis de déterminer la liste des applications qui constituaient le noyau stratégique du PCA. Il était constitué des référentiels unifiés (base de données et authentification) et à une moindre mesure des services primaires (DNS, DHCP, Radius et LDAP).

Phase 3 : Etude de vulnérabilité

Pour étudier les vulnérabilités de chacun des services recensés, j'ai analysé les forces et faiblesses de chacun à travers deux axes de travail :

- Les vulnérabilités intrinsèques de chaque service ;
- Les vulnérabilités dans leur implémentation au sein du laboratoire.

L'étude de vulnérabilité a également consisté à évaluer les dispositifs de sécurité existants. Mes collègues et moi avons recensés l'organisation de la sécurité générale, les moyens de secours prévus et mis en œuvre et les moyens de protections des informations stockées. Nous en avons déduit les principales vulnérabilités du système en production.
A partir de ces résultats, nous avons pu réaliser la phase d'analyse des risques en passant par les étapes d'étude des scénarios de risques et d'étude d'impact.

Phase 4 : Analyse des risques

L'objectif de la phase d'analyse des risques est, d'une part, la classification des risques d'indisponibilité totale ou partielle du système d'information et, d'autre part, la mise en évidence des priorités dans le traitement de ces risques. Cette analyse de risques peut être décomposée en deux étapes :

- Une étape technique d'étude des scénarios de sinistres

  L'étude technique s'est appuyée sur les conclusions de la phase d'étude de vulnérabilité (phase 3) et a consisté à identifier et à qualifier pour chaque objet un ou plusieurs risques significatifs.

  Cette analyse s'est donc traduite par l'identification des différents risques pouvant affecter l'architecture en production. Pour le système informatique du laboratoire, j'ai distingué différentes sources de risques potentiels :

  - Des risques informatiques : logiciels (piratage, virus, erreurs de développement, …), matériels (pannes ou dysfonctionnement), organisationnel (implantation, garanties) ;
  - Des risques et menaces humaines : malveillance, erreur de manipulation, nuisances ;
  - Des risques naturels : foudre, incendie, inondation, animal (rongeurs) ou industriel (AZF) ;
  - Des risques d'environnement : panne de climatisation, coupure électrique (EDF, onduleurs,…), coupure de l'accès réseau lié au fournisseur (université, réseaux régionaux et nationaux).

Une fois ces risques recensés, il a fallu évaluer l'impact potentiel de chacun et mettre en place des mesures d'atténuation. Le calcul des indicateurs de durée maximale d'interruption admissible (RTO) et de perte de données maximale admissible (RPO) reposait sur ces analyses de risques et d'impact.

- Une étape fonctionnelle d'étude d'impact

  L'analyse d'impact a consisté à évaluer l'impact de chaque risque et à déterminer à partir de quel moment cet impact serait bloquant et aurait une incidence sur les processus.

  Plusieurs degrés étaient possibles, pouvant aller d'un impact faible (panne d'un service de transfert de fichiers) à un impact total (incendie entrainant une destruction totale des bâtiments).

  Pour réaliser cette étape, je me suis appuyé sur le manuel de référence des services de sécurité [21] de la méthode Méhari. J'ai utilisé les fiches techniques de cas d'utilisation fournies dans cette méthode et ai relevé les préconisations qui pouvaient s'appliquer dans le champ d'application de mon étude.

  J'ai donc réalisé un document qui apportait des préconisations pour nos problématiques en fonction des réponses fournies dans ces fiches techniques. Un exemple est fourni en Annexe 4.

  Ce document a servi de base pour établir les orientations à mettre en œuvre pour l'implantation de notre nouveau système d'information.

Pour les phases 3 et 4 d'étude de vulnérabilité et d'analyse des risques, on peut se baser sur le document complet qu'est le manuel de référence des services de sécurité MEHARI. On peut ainsi s'appuyer sur la base de connaissance pour détecter les risques potentiels et positionner les réponses à ces risques détectés. C'est aussi ici l'intérêt d'utiliser une méthode qui soit régulièrement actualisée.

Phase 5 : Cible fonctionnelle et choix des solutions techniques

De cette analyse, j'ai pu déterminer le noyau fonctionnel stratégique de notre système d'information. Sans surprise, il était constitué par la base de données qui sert de source à la génération des services, et des scripts de génération de ces services. A un degré moindre se situent les principaux services générés dynamiquement.

Plusieurs conclusions ont donc été synthétisées pour orienter notre nouvelle architecture. Nous n'avons retenu que les méthodes comprises dans le périmètre de la mission du service informatique dans le respect nos critères financiers. Voici une liste non exhaustive de ce que j'ai recensé et soumis au service informatique pour une intégration intégrale ou partielle :

- **Des méthodes basées sur des règles et connaissances collectives** : sécurité d'accès aux locaux, connaissance de l'organisation informatique interne de l'entreprise. Ces préconisations ont eu pour effet l'ajout d'une règle d'usage informatique dans le règlement intérieur du laboratoire.
- **Des méthodes incitatives** : incitation au verrouillage des postes de travail en cas d'absence, utilisation raisonnée des capacités de transfert d'informations, respect de la politique de sécurité informatique, rappel au règlement intérieur, etc.
  Ces méthodes incitatives sont basées sur la participation et le comportement individuel des utilisateurs. Pour s'assurer d'une information exhaustive auprès de notre communauté scientifique, nous avons ajouté des items dans la charte informatique de l'établissement et avons communiqué sur ces préconisations par courrier électronique et sur l'intranet à l'ensemble du laboratoire.
- **Des méthodes basées sur l'utilisation de sous-traitance** (prestataires de services) : sauvegardes délocalisées chez un prestataire, mise à disposition de matériel de remplacement, assistance au dépannage, etc.
  Le LATT utilisait déjà la sous-traitance pour gérer son parc d'imprimantes ; nous n'avons donc pas étendu l'utilisation de sous-traitance pour un autre service. Nous avons alors négocié avec notre prestataire de service la fourniture de toner d'imprimante par anticipation, ce qui nous permettait de maintenir un stock local et de ne plus fonctionner à flux tendu.
- **Des méthodes techniques** : la liste des préconisations est contenue dans le Tableau VIII (page suivante).

Tableau VIII : Méthodes techniques préconisées pour le PCA du LATT

Méthodes techniques	Préconisations	Choix effectués
Choix des progiciels	• Utilisation d'un système de virtualisation professionnel (fonctionnalités de haute disponibilité) • Mise en place de Progiciels disposant de fonction de réplication « à chaud » pour le noyau fonctionnel stratégique	• VMware VI3 • PostgreSQL • Kerberos
Choix des systèmes d'exploitation	• Utilisation de systèmes d'exploitation maintenus et mis à jour en temps réel pour les systèmes critiques (licences commerciales) et de systèmes d'exploitation gratuits pour les systèmes à faible criticité	• Linux RHEL • Linux CentOS • Scientific Linux
Contrôle des accès aux salles techniques	• Mise en place de Codes d'accès pour toutes les salles techniques • Uniformisation des codes d'accès	• Un code unique pour les salles informatiques et un code pour les salles de brassage
Sécurisation des accès au réseau informatique	• Technologies d'identification des matériels et d'authentification utilisateur pour l'accès à l'infrastructure réseau filaire	• Radius • Portail captif (Monowall)
Sécurisation forte des authentifications serveurs et services	• Sécurisation par cryptage de la circulation des informations d'authentification sur le réseau	• LDAP over SSL (PAM) • Kerberos (systèmes compatibles)
Sécurisation et chiffrement des communications réseau	• Suppression de la circulation des mots de passe en clair sur le réseau (Kerberos et LDAP over SSL)	• Coupure de l'accès LDAP non sécurisé (clients de messagerie, etc.)
Systèmes distribués	• Mise en place d'une infrastructure distribuée sur plusieurs salles avec réplication des matériels et des services entre les salles	• Infrastructure VI3 distribuée • Politique de répartition des services et des espaces de stockage
Fréquence de sauvegarde des données selon le type de données	• Révision de la politique de sauvegarde en accord avec la nouvelle architecture des systèmes	• Nouvelle politique de sauvegarde et de réplication des données
Sauvegardes multi sites	• Répartition des sauvegardes et de l'archivage sur plusieurs sites afin de favoriser la reprise d'activité en cas de sinistre important (incendie, inondation,…)	• Répartition des données sur deux baies de disques dans deux salles distinctes • Sauvegarde et Archivage réalisé dans une salle indépendante dans un coffre ignifugé

Le choix et l'application de chacune de ces méthodes a été mis en adéquation avec l'entité au sens large : type de moyens humains, capacité financière, capacité technique, etc.
Afin de compléter cette étude, une description détaillée des phases de la méthode MEHARI est présentée en Annexe 3.

Dans la suite de ce document, je vais détailler les impacts de cette étude sur l'implantation de la nouvelle architecture VMware.

4.3. Impact du PCA sur l'infrastructure matérielle

Je vais, dans ce paragraphe, développer les incidences de la mise en place du PCA du laboratoire sur l'architecture matérielle. Pour cela, je présenterai les trois jalons temporels des mois d'avril, de juillet et d'octobre 2010, correspondants aux étapes majeures du développement de notre projet.

4.3.1. Architecture matérielle en Avril 2010

Comme nous l'avons vu au chapitre III, l'organisation du prototype qui a permis la validation des concepts que nous souhaitions implanter se détaillait comme schématisé dans la Figure 21.

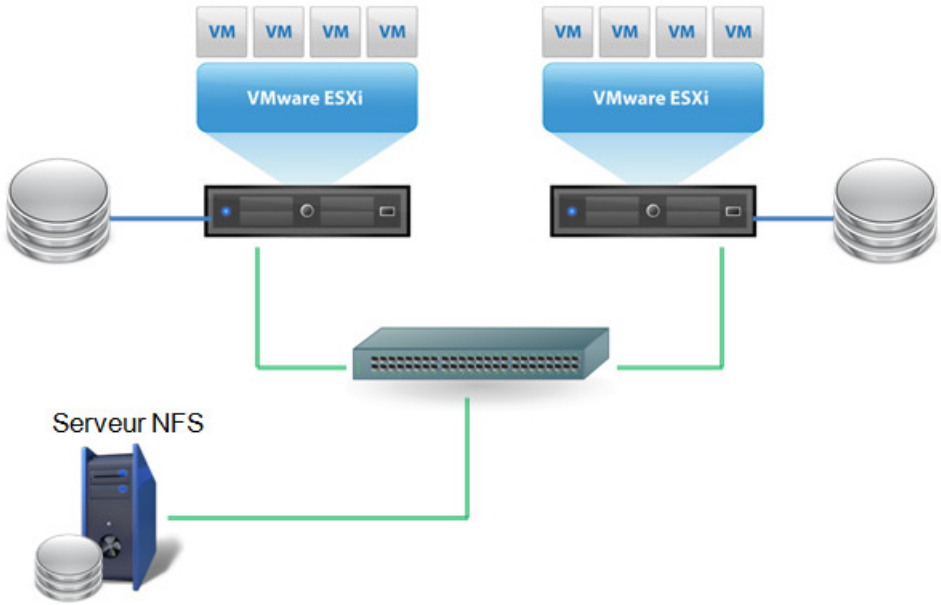


Figure 21 : Première Maquette du prototype – avril 2010

• Une architecture matérielle composée de deux serveurs ESXi (version gratuite et limitée de l'hyperviseur VMware) disposant chacun d'un espace disque local.

• Un espace disque positionné sur le serveur NFS du laboratoire partagé par les deux serveurs ESXi. Cet espace commun permettait de réaliser manuellement les sauvegardes de chacune des machines virtuelles afin de pouvoir relancer les services en cas de dysfonctionnement (PRA).
Il permettait aussi, en cas de besoin, le transfert des machines virtuelles répliquées d'un serveur à l'autre.

• Une organisation logique permettant de répartir les machines virtuelles sur les deux serveurs afin de minimiser les coupures de services lors d'une défaillance matérielle.

4.3.2. Architecture matérielle en juillet 2010

Comme précisé dans le paragraphe 2 de ce chapitre, nous avons opté, dans le cadre de notre jouvence matérielle, pour l'investissement sur une architecture virtuelle plus évoluée (VMware Infrastructure 3). Suite à une procédure d'appel d'offre adaptée, l'achat de deux serveurs et d'une baie de disques, connectés en fibre optique via l'infrastructure scientifique SAN existante, a été réalisé.
Cet investissement a permis d'augmenter la disponibilité du système d'information grâce aux avantages intrinsèques des technologies utilisées. La nouvelle architecture de production en juillet 2010 est représentée par la Figure 22.

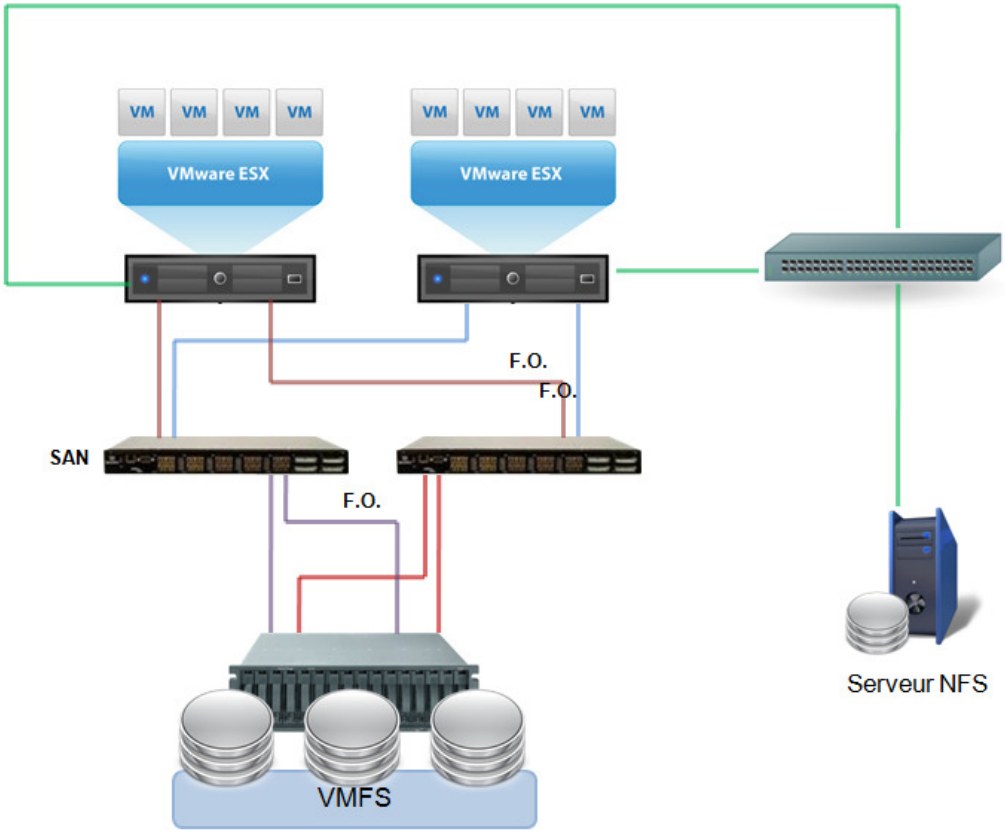


Figure 22 : Évolution du prototype après phase de tests – juillet 2010

• La mise en place de la solution de virtualisation VMware VI3 a permis d'apporter des fonctionnalités de virtualisation avancées : VMware H.A. (High Availability, haute disponibilité) et DRS (Distributed Resource Scheduler, équilibrage de charge) ou encore Vmotion (réalisation de migrations « à chaud » sans interruption de service). Cette fonctionnalité de migration est exploitée par l'ordonnanceur (DRS) pour optimiser la charge de chaque serveur et la disponibilité de chaque VM.

• L'espace de stockage des machines virtuelles a été déplacé depuis les disques internes des serveurs vers la baie de stockage dédiée. Cette baie était connectée en fibre optique, formatée en RAID 5, et comportait un système de fichiers VMFS (Virtual Machine File System). Ce système de fichiers en cluster permet d'optimiser la virtualisation du stockage et propose des taux de transfert de données supérieurs à ceux des disques internes des serveurs de l'ordre d'environ 50% (multiplicité du nombre d'axes). Ces tests ont été réalisés par copie de fichiers de 4 Go par la commande SCP. Ces résultats ont été confirmés par les mesures du progiciel embarqué d'administration du SAN.

L'avantage principal de cet espace disque mutualisé était qu'il assurait la continuité des services en cas de défaillance d'un serveur physique.

• La connexion des serveurs et de la baie par l'intermédiaire des commutateurs SAN reliés en fibre optique a été réalisée de façon à pouvoir multiplier les chemins d'accès. Nous avons donc pu mettre en pratique la notion de « multipathing ». Cette technologie permet le basculement automatique d'un équipement vers un chemin alternatif s'il existe.

• L'espace de sauvegarde dédié sur le serveur NFS a été conservé et rendu accessible pour les deux infrastructures (VI3 et ESXI). Grace à un outil de sauvegarde et de réplication dédié (Veeam Backup), nous avons automatisé la copie « à chaud » de toutes nos VM depuis le SAN vers le volume NFS.

• Les matériels ont été installés dans la salle technique informatique climatisée du laboratoire. Chaque matériel disposait de blocs d'alimentation redondants connectés, d'une part, au réseau électrique normal et, d'autre part, à une alimentation secourue et stabilisée par un onduleur.

• Les machines virtuelles du prototype ont été portées sur cette architecture physique et adaptées au nouvel environnement matériel en qualifiant et quantifiant la mémoire vive et nombre de processeurs utiles. L'infrastructure du prototype a progressivement été mise en service en exploitation pour les services DNS, DHCP, Radius et LDAP sans réaliser de coupure de services.

Suite à cette première intégration, qui a permis d'augmenter la disponibilité du système de façon significative, nous avons élaboré la version cible de notre système d'exploitation d'après les préconisations du PCA que j'avais élaboré.

4.3.3. Architecture matérielle en octobre 2010

Du point de vue de l'architecture matérielle, nous disposions de deux infrastructures VMware (production et secours), d'un réseau SAN (commutateurs et baies), d'un réseau NAS (serveurs redondants), d'une librairie de sauvegarde et d'un espace de stockage ignifugé pour l'externalisation des cartouches de sauvegarde.
La mise en pratique des préconisations du PCA imposait la répartition du matériel redondant dans deux salles machines (datacenter) disjointes.

Le LATT disposait d'une salle informatique technique, au même titre que les autres laboratoires. Nous avons mis en place un échange avec le Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS), un autre laboratoire de l'OMP, afin de réaliser un hébergement réparti. Après l'étude et la réalisation du raccordement des réseaux filaires et optiques, notre architecture physique a été scindée en implantant une infrastructure opérationnelle dans chaque salle. Chaque infrastructure était composée d'une architecture de production (serveur, SAN et baie) et d'une architecture de secours (serveur, réseau filaire et NAS).

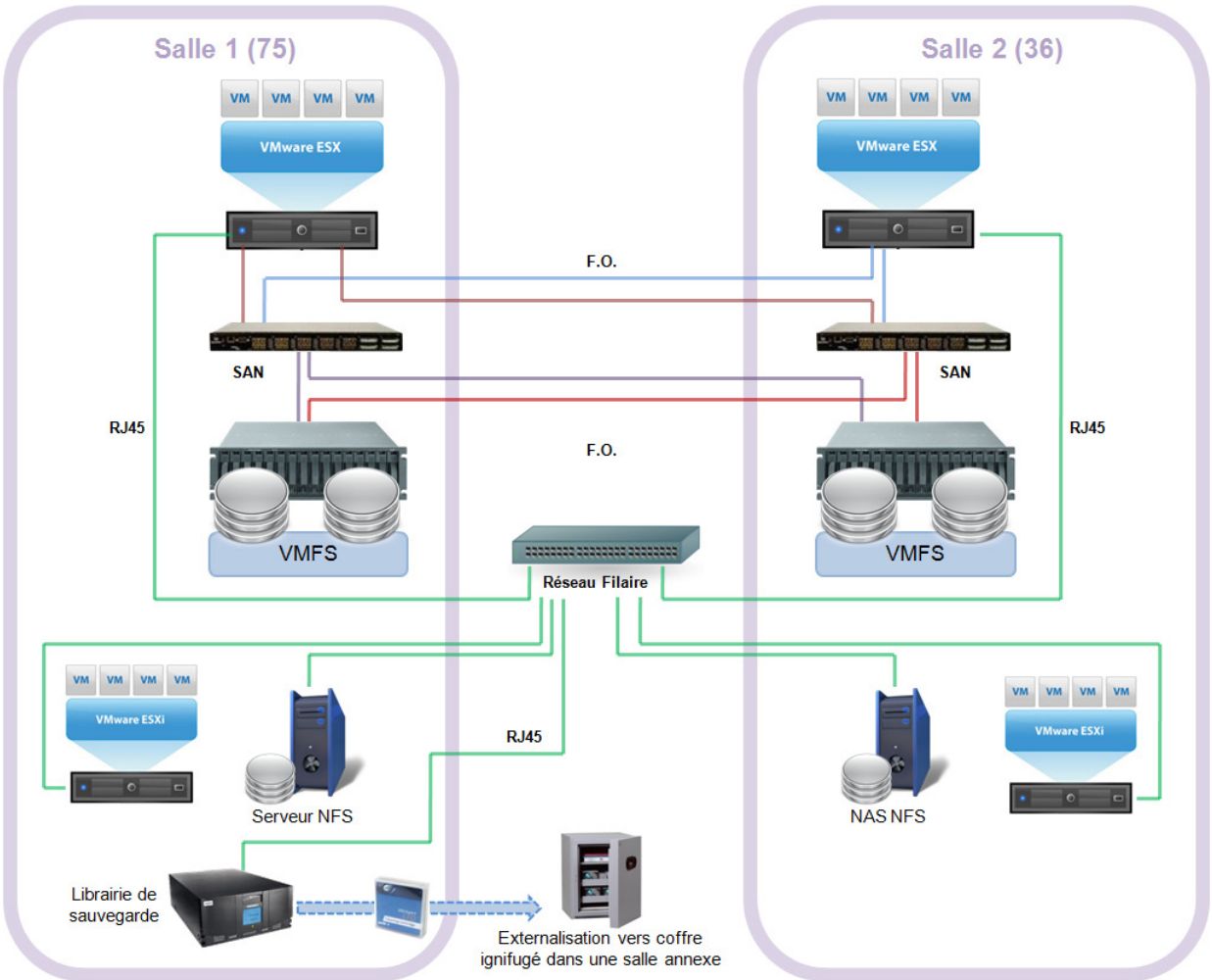


Figure 23 : Version du système d'information mise en exploitation - octobre 2010

L'interconnexion de ces deux infrastructures était assurée par les liens croisés et redondants entre chacune des salles. Le circuit d'alimentation électrique était redondant et chaque salle disposait de son arrivée électrique dédiée et d'un circuit ondulé spécifique.

De même, un système de climatisation indépendant et redondant était implanté dans chacune des salles afin de maintenir une température et une hygrométrie stable. L'accès à ces salles techniques était sécurisé par une serrure à code et réservé aux seuls agents autorisés. Le synoptique de cette infrastructure physique est présenté Figure 23.

Par anticipation, deux commutateurs réseaux ont été achetés pour pallier à la défaillance d'un matériel actif (« Cold Spare »). En effet, les matériels en production disposaient seulement d'un délai d'intervention ou de remplacement fixé au jour suivant ouvré (J+1).

Dans le cas d'une reprise d'activité après un désastre majeur, nous avions mis en place une politique de sauvegarde adaptée reposant sur une librairie de sauvegarde et sur une externalisation des supports vers un coffre ignifugé et sécurisé.

4.4. Impact du PCA sur l'infrastructure logique

A partir de l'infrastructure physique détaillée dans le paragraphe précédent, nous avons déployé une stratégie d'implantation des machines virtuelles hébergées.

4.4.1. Pérennité des référentiels d'authentification et d'identification

Le référentiel de données était devenu un point individuel de défaillance du nouveau système. Source unique pour la génération dynamique des services critiques, sa pérennité devait être assurée.
L'organisation de l'implantation des serveurs associés à ces services a fait l'objet d'une réflexion dont les résultats ont été les suivants :

• Le référentiel d'identification a été déployé en créant deux VM indépendantes. Nous avons orienté notre choix de système d'exploitation vers une version de Linux professionnelle (RHEL, Red Hat Enterprise Linux) pour sa fréquence de mise à jour. Chaque VM disposait d'une instance du SGBD (PostgreSQL) en configuration de réplication (voir détail au paragraphe IV.2.1. Pérennisation du référentiel unique de données) ;

• Le service d'authentification a, de la même façon, été implanté sur deux VM indépendantes sous Linux RHEL. Chaque serveur virtuel embarquait une instance du progiciel d'authentification Kerberos, configuré en réplication.

Contraintes de mobilité des VM

L'infrastructure VMware VI3 permettait d'améliorer la disponibilité, notamment grâce aux fonctionnalités de déplacement automatique de VM d'un serveur physique à un autre.

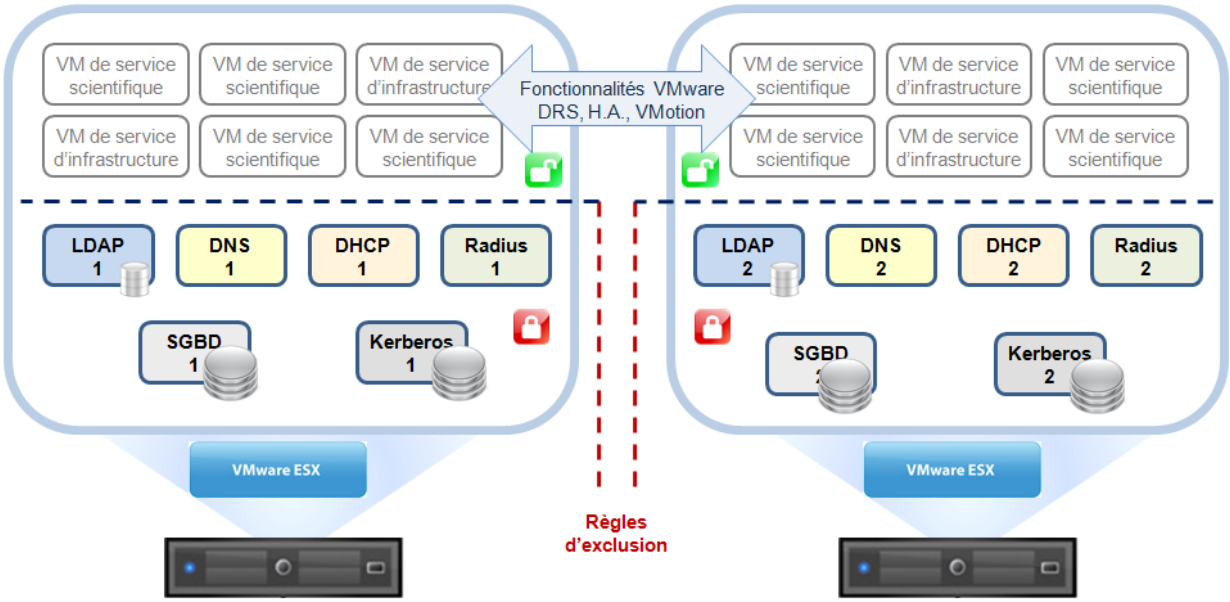


Figure 24 : Contraintes de mobilité sur les services du référentiel unique

Pour le référentiel de données comme pour le service d'authentification, nous avons positionné des clauses d'interdiction de mobilité sur chacune des VM d'infrastructure répliquées (Figure 24).

Le positionnement de ces contraintes nous permettait de nous assurer une répartition des VM critiques, en fonction des réplications de services, afin de pallier tout dysfonctionnement lié à une perte des serveurs physiques.
Les autres VM d'infrastructure et scientifiques continuaient, en parallèle, à profiter des avantages des fonctionnalités de haute disponibilité de l'hyperviseur.

Une sauvegarde complète (dump) de la base de données et du service d'authentification était réalisée deux fois par jour (à 12h et à 21h) sur un serveur physique indépendant de notre architecture virtuelle. A la vue de la périodicité de mise à jour des données de ces bases, cette fréquence a donné jusqu'à présent satisfaction.

4.4.2. Génération dynamique des services critiques

La génération dynamique des services critiques était un concept fort du nouveau système d'information. Nous avons donc généré tous nos services critiques, hors noyau stratégique, à partir du référentiel unique, comme présenté dans la Figure 25.

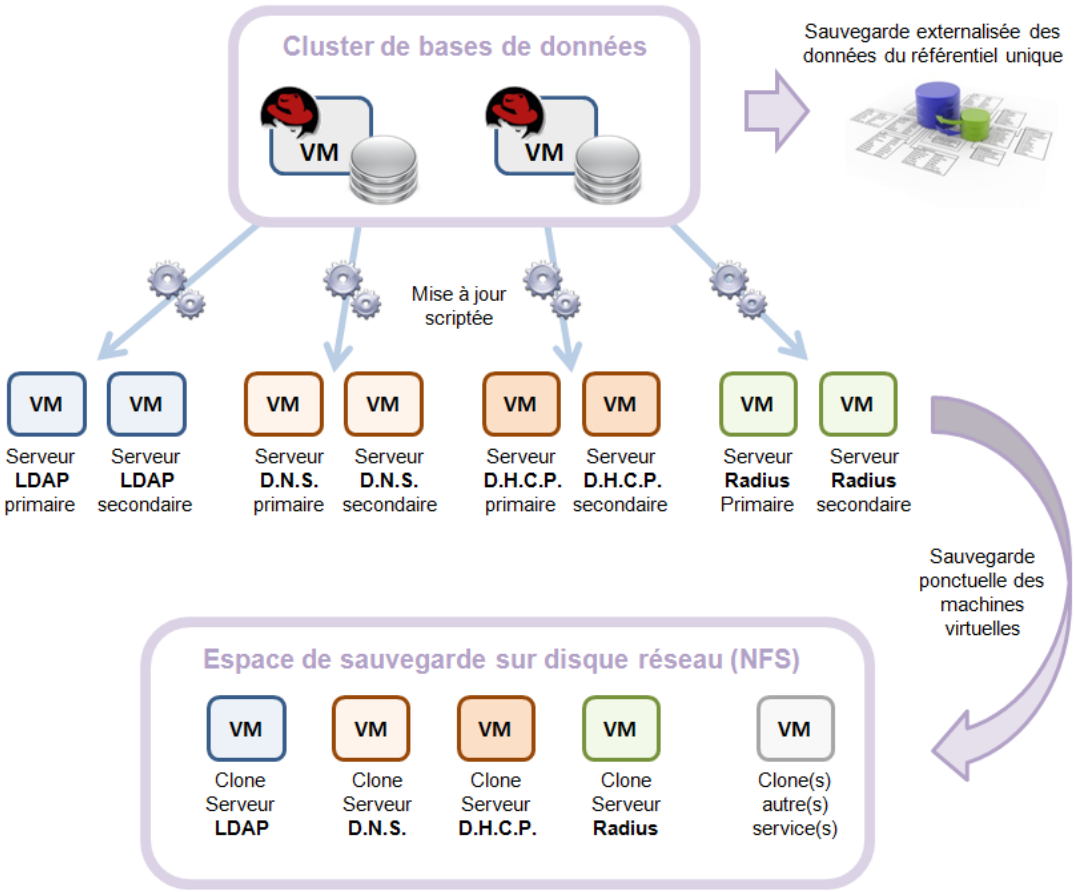


Figure 25 : Génération et sauvegarde des machines virtuelles de services

Les services LDAP, Radius, DHCP et DNS étaient donc générés à partir du référentiel unique et répartis sur l'infrastructure de production. Une copie de secours était ensuite réalisée sur l'espace disque partagé (NFS) entre l'architecture de production et l'architecture de secours.

En ce sens, en se basant sur la génération dynamique des services, nous améliorions la disponibilité par la mise en pratique des préconisations suivantes :

- Chaque service fonctionnait sur une VM (Machine Virtuelle) séparée.
- Chaque VM contenait son système d'exploitation, le démon du service et le script qui permettait de mettre à jour, à partir de la base unifiée, le fichier de configuration associé. Une VM de service était donc une machine autonome, capable de se mettre à jour toute seule dès sa mise en fonctionnement ;
- Une copie de chaque VM était sauvegardée sur un espace disque commun aux architectures de production et de secours. Une autre copie était externalisée sur bande dans un coffre ignifugé.

Cette organisation respectait ainsi les préconisations effectuées pour le PCA du système d'information et pour le PRA en utilisant l'infrastructure de secours.

4.4.3. Disponibilité des services scientifiques

Les autres services d'infrastructure ou les serveurs scientifiques virtualisés étaient relativement indépendants du noyau fonctionnel. La défaillance d'un de ces services n'entrainait pas de dysfonctionnement d'un autre service. En cas d'indisponibilité d'un de ces services, le système d'information continuait à fonctionner dans un état dégradé.

La politique choisie était de générer un clone de la VM à chaque modification.
De plus, une copie de chaque machine virtuelle en production était réalisée quotidiennement à l'aide d'un progiciel spécifique de réplication (Veeam Backup). En cas de dysfonctionnement durable, nous pouvions relancer le service en redémarrant son clone depuis l'architecture de production ou de secours.

4.5. Intégration du PRA dans le PCA

Le Plan de Continuité d'Activité (PCA) mis en œuvre garantissait une disponibilité optimale, mais non permanente. Pour éviter une coupure de service prolongée, potentiellement due à un dysfonctionnement d'un élément du PCA, nous avons développé un Plan de Reprise d'Activité (PRA) tenant compte de la nouvelle architecture.
Sachant que le PRA, comme le PCA, tient compte de l'environnement global de l'entité, j'ai proposé plusieurs axes à privilégier pour sa création :
• Privilégier une architecture multi sites qui permet au matériel sur le second site de ne pas être affecté par le sinistre ;
• Réalisation de sauvegardes synchrones et asynchrones des données, voire des systèmes, selon leur criticité et leur occupation (place disque) : duplication des données en temps réel sur des serveurs de fichiers, de messagerie ou d'infrastructure, réalisation de sauvegardes fréquentes (journalières, hebdomadaires, etc.) sur des services moins critiques. Il faut tenir compte des notions de distance pour ces réplications car l'augmentation de la distance entraine l'augmentation des coûts de la sauvegarde synchrone ;
• Externalisation journalière des sauvegardes pour les services critiques et hebdomadaires pour les services à faible criticité en utilisant des solutions adaptées (bandes LTO4 dans coffre ignifugé par exemple).

Suite à l'analyse des risques du PCA, en prévention d'un sinistre majeur, j'ai proposé une organisation de nos salles systèmes autour de trois sites : deux sites proches (campus toulousain), où les données seraient répliquées de manière synchrone, ce qui permet un basculement instantané entre les systèmes et un troisième site distant (campus tarbais) assurant une réplication asynchrone (sauvegarde et archivage).

4.4.1. Analyse des scénarios de risques

Pour développer le plan de reprise, nous avons étudié plusieurs scénarios de risques et en avons privilégié un à forte incidence, à savoir la perte totale de l'architecture VI3 et des baies de stockage.
Nous avons aussi réfléchi à l'obtention d'une granularité plus fine en envisageant la perte d'un ou de plusieurs services et les actions à réaliser pour reprendre l'activité du ou des services incriminés.
Nous avons mis en place un PRA adapté à la résolution de la majorité des scénarios de crash. Nous n'avons pas envisagé le cas de perte durable (supérieur à deux heures) d'alimentation électrique générale par exemple ou de catastrophe naturelle (incendie, inondation, etc.) à l'échelle des bâtiments. Pour les mêmes raisons qui ont orienté notre PCA vers l'instauration d'un PCA informatique (PSI), nous avons focalisé notre action sur la mise en place un PRA informatique.

4.4.2. Mise en place d'une architecture de secours

Les choix du PRA se sont portés sur la mise en place d'une architecture de secours indépendante. Tenant compte de nos contraintes financières (budgets limités), nous avons pu mettre en place ce PRA grâce à l'utilisation des matériels libérés par la virtualisation. L'intégration de cette architecture dans notre environnement est présentée au chapitre 4.3.3.

Ordonnancement de l'ordre de redémarrage des services

Nous avons créé, sur l'espace documentaire du service informatique (wiki), une documentation en ligne, imprimée à chaque mise à jour, dans laquelle nous avons défini une organisation pour la relance des services.

L'ordonnancement de relance des services a été défini en fonction du positionnement des RTO et RPO pour chacun des services. Ces critères ont permis de planifier et d'organiser les intervalles de sauvegardes et d'assurer un retour des services permettant de minimiser le RTO et d'optimiser la reprise d'activité.

Ces taux de criticité et l'interdépendance des services nous ont orientés vers l'organisation suivante :

1) Assurer la reprise sur incident des principales bases de données de services et de gestion administrative : référentiel unique, bases comptables (Xlab), bases annexes (GLPI, etc.) ;

2) Assurer l'accès au réseau physique (commutateurs réseau et services réseau : DNS, DHCP, etc.) ;

3) Assurer le bon fonctionnement du service d'authentification ;

4) Assurer l'accès aux serveurs et aux données scientifiques locales à l'établissement ;

5) Remettre en place la communication : messagerie, accès à Internet ;

6) Relancer les services non critiques : services d'impression, de gestion de jetons, etc.

La durée moyenne de relance du système d'information dépend du dysfonctionnement constaté. En cas de perte de l'infrastructure complète VMware VI3, sans perte de réseau, nous avons estimé le délai de reprise à environ deux heures, en suivant la documentation du PRA à partir du début de l'intervention du service informatique.
En cas de perte partielle ou totale du réseau, ce délai peu être augmenté de 24 à 48 heures, principalement en raison du délai d'application des garanties matérielles souscrites. La probabilité de ce scénario reste faible.

L'application de ce PRA repose sur l'intervention et l'expertise des membres de l'équipe informatique. Les choix effectués pour sa création sont en adéquation avec l'entité au sens large : type de moyens humains, capacité financière, capacité technique, etc.
Sachant que le temps de rétablissement garanti impacte le coût de la stratégie mise en œuvre, cette stratégie manuelle nous a semblé offrir le meilleur équilibre entre le coût et la rapidité de reprise de service, en adéquation avec les besoins du laboratoire.

5. Synthèse de la mise en exploitation

La mise en exploitation du nouveau système d'information nous a permis de valider les concepts, l'architecture du prototype et de consolider l'ensemble avec une infrastructure de secours.

Concrètement, la mise en place d'un référentiel unique d'identification nous a permis de répondre à la problématique de cohérence et de dispersion des données. Le déploiement des services d'authentification a apporté une simplification et une sécurisation des accès à notre système d'information et la mise en place d'une infrastructure virtualisée a offert un socle technique garantissant un haut niveau de disponibilité.

Reposant sur ce socle, l'application des préconisations du PCA et du PRA a conduit à une amélioration de l'efficience des ressources humaines et techniques.

Enfin, l'apport de nouvelles fonctionnalités de soutien et de support à la recherche induit par ce projet a reçu un accueil très favorable de la part de notre communauté scientifique (éducation - recherche).

V. Conclusion

Au lancement de ce projet, le laboratoire d'Astrophysique de Toulouse - Tarbes à l'OMP disposait d'un système informatique fonctionnel mais perfectible. Plusieurs points d'effort avaient été identifiés par le service et devaient être traités pour améliorer le fonctionnement et la pérennité du système comme la gestion des comptes utilisateurs, la gestion des redondances et des incohérences des données ou encore le contrôle des accès au réseau filaire.

En 2009, l'informatique du laboratoire arrivait à un moment charnière où le renouvellement de plusieurs serveurs d'infrastructure, en fin de garantie, devait être effectué. L'équipe a pris la décision de réfléchir à une réorganisation du système d'information afin d'améliorer l'offre de services et de d'en accroître la disponibilité.
Cette volonté a été présentée à la direction et à la commission informatique qui ont validé ce projet et m'ont confié, dans le cadre de mon mémoire d'ingénieur, sa réalisation.

Outre la gestion du travail quotidien et le maintien de l'existant, l'équipe informatique devait assurer le développement de ce nouveau système. A la vue de la charge de travail et afin de réaliser le projet dans un délai optimal, nous avons décidé de renforcer notre équipe en procédant à l'embauche d'un assistant ingénieur en CDD (Michel L'Hostis). En tant que chef de projet, l'encadrement de Michel m'a été confié pour cette action.

Au lancement du projet, nous avons organisé plusieurs réunions de type « brainstorming » afin de réfléchir à une nouvelle structure de système d'information. Plusieurs concepts ont été avancés tels que la mise en place d'un référentiel unique de données, le contrôle des accès utilisateurs au réseau, la minimisation des points individuels de défaillance mais aussi la volonté d'explorer et de mettre en pratique de nouvelles technologies à forte valeur ajoutée.

Ces nouveaux concepts ont été développés dans la phase de prototypage du projet. Cette dernière a permis de valider les concepts, de les adapter à notre environnement et d'effectuer des tests en exploitation sur un échantillon représentatif. Les résultats de l'analyse du fonctionnement de ce prototype ayant été positifs, nous avons décidés de pousser plus avant nos développements afin d'intégrer ces concepts au système en production.

Dès lors, nous avons, en avril 2010, débuté la phase de mise en exploitation des nouveaux outils. Pour cela, j'ai décidé de conserver nos concepts en étudiant la mise en place de technologies permettant de pérenniser chacun des services mis en œuvre. Afin d'obtenir un résultat optimum en accord avec les contraintes du laboratoire, j'ai réalisé une étude sur la mise en place des plans de continuité et de reprise d'activité (PCA et PRA). Les préconisations que j'ai effectuées ont été majoritairement appliquées, ce qui a engendré une modification des infrastructures physiques et logiques ainsi qu'une amélioration de l'offre de services.

En juillet 2010, nous avons poursuivi la phase de mise en exploitation du nouveau système d'information. Sans réaliser de coupure de services, nous avons pu modifier notre architecture matérielle et organisationnelle.

En octobre 2010, le projet était terminé et le nouveau système d'information était en exploitation. Depuis lors, le nouveau système d'information, issu du travail de mon stage d'ingénieur, est en production. Nous constatons jusqu'à ce jour un taux de disponibilité de 100% sur les services d'infrastructure ainsi que sur les services impactés par la refonte du système. Un retour positif a été constaté de la part des utilisateurs qui ont aujourd'hui accès à plusieurs nouveaux services, et restent demandeurs de davantage d'améliorations.
Le système en production a permis d'impliquer l'ensemble des acteurs permettant ainsi une meilleure réactivité vis-à-vis des demandes et une meilleure répartition de la charge d'administration.

La gestion des services d'infrastructure est aujourd'hui simplifiée, la satisfaction des utilisateurs constatée et la cohérence des données du système assurée.

Ce stage m'a permis d'approfondir ma connaissance sur plusieurs technologies, d'en développer les concepts et d'en assurer la mise en œuvre. J'ai découvert et identifié, au travers de la méthode MEHARI, la notion de plans de continuité et de reprise d'activité.
L'apport personnel de cette action a été aussi important du point de vue organisationnel que technique. La gestion de projet et l'encadrement de personnel sont des fonctions que je n'avais pas exercé jusque là. Ce stage d'ingénieur m'a apporté une expérience de la gestion de projet ainsi que des contraintes relationnelles et organisationnelles attenantes. D'un point de vue technique, j'ai pu aborder et pratiquer des technologies nouvelles permettant d'apporter une réelle plus-value au système d'information.

Concernant mon évolution professionnelle, ce projet m'a offert une visibilité positive auprès de l'OMP. Nous avons organisé une journée technique de présentation de nos travaux le 29 juin 2010, où nous avons pu exposer nos concepts d'architecture et nos points d'avancement à nos collègues informaticiens de l'Observatoire. Cette visibilité a eu pour incidence de recevoir une proposition de la part de la direction de l'OMP pour que je prenne la responsabilité du pôle systèmes et réseaux de l'UMS. Ce pôle coordonne les actions entre les laboratoires, fournit et gère les ressources informatiques mutualisées.
Depuis le 1er janvier 2011, j'occupe la fonction de responsable de ce service qui se compose de six informaticiens. Parallèlement à cette responsabilité, j'ai accepté la mission de chef de projet pour la mise en place d'un référentiel unifié à l'échelle de l'Observatoire. Ce référentiel, continuité de mon stage, servira de socle technique à la mise en place d'une nouvelle offre de services mutualisés.

Les compétences, l'expertise et les résultats, acquis durant ma scolarité au CNAM et mon stage d'ingénieur, m'ont permis de faire évoluer ma carrière vers un poste d'encadrement tout en restant au plus près des évolutions techniques. Mes nouvelles responsabilités me permettent aujourd'hui de m'investir sur des actions motivantes comme la gestion de projets ou le management d'équipe et d'assurer le soutien scientifique à notre communauté.

VI. Annexes

Liste des annexes

Annexe 1 : Le protocole LDAP

LDAP [RFC 1487] signifie Lightweight Directory Access Protocole, soit protocole léger d'accès aux répertoires/données. Il s'agit d'un protocole de transport standard permettant de gérer des annuaires, c'est-à-dire d'accéder à des bases d'informations sur les utilisateurs d'un réseau par l'intermédiaire de protocoles TCP/IP.
Historiquement, le protocole LDAP a été développé en 1993 par l'université du Michigan, avec pour objectif de remplacer le protocole DAP [RFC DAP ?] (servant à accéder aux services d'annuaire X.500 de l'OSI) en l'adaptant pour l'intégrer à l'environnement TCP/IP. Le service d'annuaire X.500 était un standard conçu en 1988 par les opérateurs télécoms pour interconnecter tout type d'annuaire dans un but de normalisation. X500 imposait au client et au serveur de communiquer en utilisant la pile de protocoles OSI (Open Systems Interface) ce qui impliquait une mise en place assez lourde.
C'est la raison pour laquelle en 1993 l'Université du Michigan a adapté le protocole DAP de la norme X.500 au protocole TCP/IP et mis au point LDAP (v1), qui repose sur :

- un modèle d'information définissant le type de données de l'annuaire (classe d'objets et leurs attributs) ;
- un système de nommage pour organiser et référencer les informations de définitions en organisation hiérarchique d'arbre (le DIT, Directory Information Tree) ;
- un ensemble de définitions pour l'accès à l'information contenues dans les commandes d'accès ;
- un modèle de sécurité pour définir la protection des données et des accès ;
- un processus de duplication pour définir de la répartition entre serveurs ;
- un format d'échange spécifique de données : LDIF (Interchange Format).

Dès 1995, LDAP v2 [RFC 1777] est devenu un annuaire natif, afin de ne plus servir uniquement à accéder à des annuaires de type X500, c'est-à-dire en gérant sa propre base de données. LDAP est donc une version allégée du protocole DAP prévu pour fonctionner avec les protocoles TCP/IP.
En 1997, la version 3 du protocole LDAP (Ldap v3, [RFC 2251]) proposait des mécanismes de chiffrement et d'authentification permettant de sécuriser l'accès aux informations stockées dans la base de données de l'annuaire. LDAP était devenu un protocole autonome d'accès à l'information.

Le protocole LDAP est donc prévu pour gérer l'accès aux données contenues dans les annuaires. Il permet d'effectuer plusieurs requêtes sur le serveur d'annuaire à l'aide d'une seule connexion, contrairement à ce que propose le protocole HTTP par exemple (une seule requête à chaque connexion au serveur).
Cependant, le protocole LDAP a été conçu pour être léger, ce qui signifie que s'il intègre dès la version 2 sa propre base de données, de nombreuses fonctionnalités de base de données comme le support des transactions ou les verrous d'écritures ne sont pas essentielles pour son fonctionnement.
Le protocole est donc orienté, de par sa fonction d'annuaire, pour supporter des opérations de lecture beaucoup plus fréquentes que les opérations d'écriture, contrairement aux bases de données qui reposent sur des quantités de lectures et écritures similaires.
Pour ces raisons, l'annuaire LDAP apparaissait inadapté pour gérer les données persistantes du laboratoire dans un contexte d'accès concurrentiels. La surcharge des opérations d'écriture pouvant conduire à une perte de données et par extension au blocage du système d'information.

Annexe 2 : Kerberos

Kerberos a été mis au point au MIT dans les années 1990, il est maintenant déployé et disponible pour tous les environnements (Linux, Windows, Mac). Des universités françaises (Strasbourg, Bordeaux) ont déjà migré leurs systèmes d'authentification vers Kerberos.

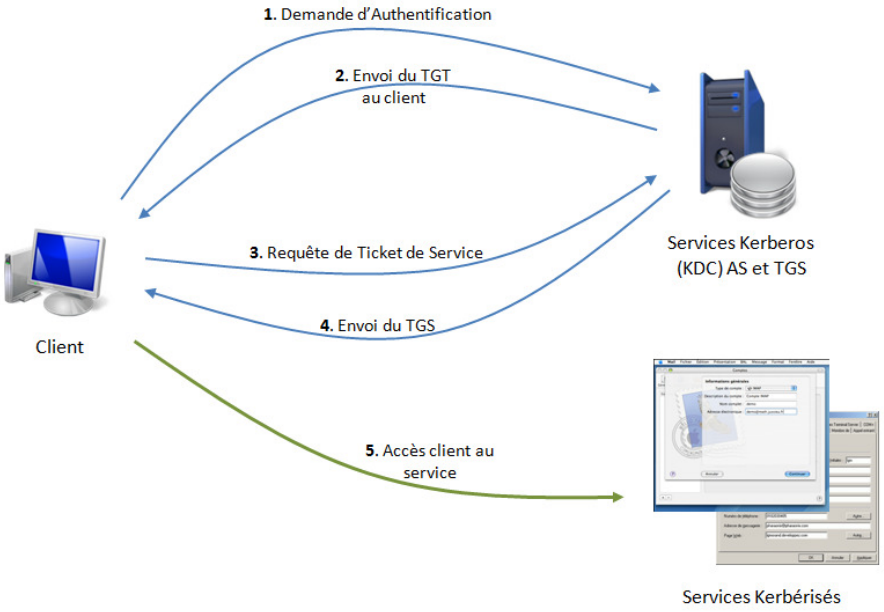


Figure 26 : Schéma d'authentification Kerberos

Le principe de fonctionnement de l'architecture Kerberos est le suivant :

1- pour s'authentifier, le client se connecte sur son application (session Windows ou Linux, accès à un service compatible,…) et entre son identifiant utilisateur (login) et son mot de passe. Le système collecte ces informations, les compile avec le nom de domaine et envoie le résultat (principal) au service d'authentification Kerberos (AS) qui fonctionne sur le serveur Kerberos appelé KDC (Key Distribution Center).

2- Le service d'authentification (AS) du KDC génère une clef de session qui sera utilisée par le client et les services distants. Il encrypte la clef de session avec le mot de passe de l'utilisateur, valide les droits du demandeur et si l'authentification du principal est conforme, le KDC renvoie un ticket appelé TGT (ticket-granting ticket) au client.

3- Si le client veut accéder à un service particulier, il envoie une seconde demande au KDC en spécifiant dans sa demande l'identifiant du service auquel il souhaite accéder. (Si le client a déjà un TGS d'accès à ce service qui n'a pas expiré, les étapes 3 et 4 ne seront pas rejouées.)

4-Le KDC renvoie un ticket de service, appelé TGS (ticket-granting service), qui donne les droits d'accès au client pour le service désiré.

5-Le client envoie sa requête au service sous forme d'un jeton Kerberos. Ce jeton est composé de la requête du client et du TGS fourni par le KDC. Le service réalise une authentification auprès du KDC avec les identifiants fournis dans le jeton du client pour valider l'exécution du service.

Annexe 3 : La méthode MEHARI (Clusif)

Voici la présentation de la méthode MEHARI du Clusif, citée dans ce document.

La méthode MEHARI est une méthode de management (analyse et de gestion) des risques liés à la sécurité de l'information. La distribution initiale de MEHARI, émanant du CLUSIF, est constituée de trois éléments : un fichier "base de connaissance", un manuel de référence des services de sécurité, et un manuel de référence des scénarios de risque.

La méthode MEHARI fournit un cadre méthodologique et des bases de connaissance pour :

- Analyser et classifier les enjeux majeurs ;
- Étudier les vulnérabilités ;
- Réduire la gravité des scénarios de risques ;
- Piloter la sécurité de l'information.

L'approche modulaire de la méthode MEHARI reprend le principe de la méthode du plan de continuité d'activité par phases successives : analyse des enjeux, analyse des vulnérabilités, identification et traitement des risques.

La méthode MEHARI permet donc d'obtenir plusieurs types d'indicateurs de sécurité globaux. Ces derniers sont associés à des services, des domaines de sécurité, des thèmes transversaux ou encore aux contrôles recommandés par la norme ISO/IEC 27002:2005. Cette Norme établit les principes généraux pour mettre en œuvre et améliorer la gestion de la sécurité de l'information.

MEHARI fournit aussi un ensemble d'outils répondant aux exigences d'un système de gestion de la sécurité de l'information (SMSI) et s'intègre donc dans une démarche de mise en conformité à la normalisation ISO/IEC 27001 ou de certification.

Les questions d'audit de la méthode sont autant de points de contrôle de l'existant (technique ou organisationnel) répondant au besoin d'assurer l'objectif du service tel qu'il est décrit. La conclusion de l'analyse met en exergue les points faibles du système.

L'apport des préconisations pour augmenter la disponibilité du système d'information est pondéré par le nombre de modifications mises en œuvre sur le système d'information en fonction de la possibilité technique et financière d'implanter les solutions proposées.

L'application de la méthode MEHARI se réalise en effectuant cinq phases successives.

Phase 1 : Lancement

Au lancement de l'élaboration d'un Plan de Continuité d'Activité, il est nécessaire de préciser le champ d'action à couvrir. Il peut s'agir de l'ensemble des activités de l'entité ou d'un focus sur un domaine stratégique. Il conviendra aussi de désigner un correspondant pour chaque activité. Ce personnel référent sera dépositaire des pratiques à mettre en œuvre en cas de dysfonctionnement.
Il faut donc faire préciser, par la direction et/ou les responsables sécurité, l'ensemble des risques que le périmètre du PCA doit prendre en compte. Selon les choix exprimés, les « objets à risques » (matériels informatiques, matériels de téléphonie, fournitures externes, personnel, locaux, etc.) et la nature des risques (risque naturel, risque social,…) seront identifiés.
Selon les risques et le périmètre retenus, le plan de secours pourra s'orienter vers un Plan de Secours Informatique (PSI) ou s'apparenter à un Plan de Continuité d'Activité global pour toute l'entreprise.
La méthode MEHARI fourni des guides d'analyse de risques qui récapitulent par grandes familles les objets à risques pouvant être pris en compte dans un PCA ainsi que les principaux risques associés.

Chaque résultat de phase conditionne la phase suivante et doit faire l'objet d'une réception formelle par le Comité de Pilotage.

Phase 2 : Etude Fonctionnelle

La deuxième phase est une phase d'étude fonctionnelle qui a pour objectif de définir pour chaque activité les exigences de continuité.

Il s'agit de déterminer les enjeux et les activités essentielles puis d'évaluer les conséquences d'une interruption ou d'une dégradation de ces activités. La comparaison de ces différentes situations de dysfonctionnement partiel ou total doit permettre d'étalonner les niveaux d'impacts qui seront utilisés lors de la phase d'analyse des risques.
L'étude fonctionnelle doit aussi permettre de préciser les conditions minimales permettant d'assurer un niveau d'activité minimal pour éviter toute coupure de l'activité en toutes circonstances.
Pour cela, il s'agira d'identifier et de définir principalement :

- Les éléments du système d'information indispensables à la poursuite de l'activité (applications, moyens de communication, données) ;
- Le service minimum acceptable par activité ;
- Les ressources humaines, les locaux et les équipements nécessaires (postes de travail, téléphones, imprimantes, réseau, etc.) ;
- Le délai de reprise d'activité, la durée du service minimum ;
- Le niveau de dégradation du service acceptable (temps de réponse, activités manuelles ou automatisées...) et les conditions de retour à la normale.

Dans le cas d'une étude à l'échelle de la structure (entreprise, laboratoire, etc.), le recensement sera réalisé en réunissant les groupes fonctionnels représentant chaque métier de l'entité. Il est nécessaire de procéder par la suite à des consolidations et de vérifier la cohérence globale des besoins exprimés, sous arbitrage parfois de la direction. Lorsque les solutions à implémenter auront été chiffrées, la notion de « service minimum acceptable » pourra être reconsidérée (processus itératif).
On déduira de cette phase la liste des applications constituant le noyau stratégique du PCA.

Phase 3 : Etude de vulnérabilité

Comme pour toute étude de sécurité, l'évaluation des dispositifs de sécurité existants ou prévus est nécessaire. Il s'agit au minimum d'analyser les items suivants :

- L'organisation de la sécurité générale (environnement, accès physiques, sécurité incendie, consignes de sécurité);
- La couverture assurance des risques informatiques ;
- Les moyens de secours existants ou planifiés (matériel informatique et réseau, alimentation électrique, climatisation, personnel, etc.) associés à leur degré de confiance (délais, documentation et tests) ;
- Les moyens de protection des informations stockées ;
- Les supports informatiques : sauvegardes, archives, procédures de restauration;
- Les supports papier (dossiers, archives, documentation, etc.) ;
- La sécurité des échanges extérieurs (protection du réseau, etc.) ;
- Les contrats de maintenance des matériels et des logiciels (degré d'engagement des prestataires).

Phase 4 : Analyse des risques

L'objectif de la phase d'analyse des risques est, d'une part, la classification des risques d'indisponibilité totale ou partielle du système d'information et, d'autre part, la mise en évidence des priorités dans le traitement de ces risques qui permettra d'orienter la mise en place du plan de continuité d'activité.

L'analyse des risques peut être décomposée en deux étapes :

- Une étape technique d'étude des scénarios de sinistres ;
  L'étude technique s'appuie sur les conclusions de la phase d'étude de vulnérabilité (phase 3) et consiste à identifier pour chaque objet un ou plusieurs risques significatifs. Pour chaque risque retenu, il conviendra ensuite de décrire les conséquences directes de sa réalisation sur le système d'information. Il ne s'agit pas pour l'instant de mesurer l'impact du sinistre mais uniquement de le qualifier. L'objectif est de réaliser un bilan des conséquences directes en termes :
  - de durée d'indisponibilité des moyens (applications, matériels, services, etc.);
  - de perte d'information (sauvegardes, etc.);
  - de potentialité du risque (soit directement attribuée, soit calculée).

- Une étape fonctionnelle d'étude d'impact.
  L'étape fonctionnelle consiste à mesurer l'impact des risques potentiels. En utilisant les résultats de l'étape technique, on détermine une mesure de gravité en combinant impact et potentialité. La classification des risques sera hiérarchisée selon leur niveau de gravité.
  Cette étape fonctionnelle permet ainsi de préciser les risques à prendre en compte dans le plan de secours et leur priorité de prise en compte.

Pour les phases 3 et 4 d'étude de vulnérabilité et d'analyse des risques, on peut se baser sur le manuel de référence des services de sécurité MEHARI. On peut ainsi s'appuyer sur la base de connaissance pour détecter les risques potentiels et positionner les réponses à ces risques détectés.

Phase 5 : Orientation : cible fonctionnelle et choix des solutions techniques

La cible fonctionnelle est déterminée par l'analyse de l'organigramme fonctionnel du système d'information et la classification des fonctions et sous fonctions selon les besoins en disponibilité estimés par la direction opérationnelle.
Cette classification permet de déterminer le noyau fonctionnel stratégique et minimum pour la survie de l'entreprise.
Il s'agira donc d'élaborer des scénarios de sinistres et d'estimer la durée d'interruption de service associée à chaque fonction vitale, en essayant de les regrouper selon des critères de gravité (4 à 5 maximum). Ces derniers, préalablement déterminés, peuvent varier d'un simple arrêt du service à une situation de « désastre ».
Tous les objets du système d'information (applications, matériels, équipements réseaux, fournitures, etc.) nécessaires au fonctionnement de ce noyau seront identifiés et regroupés en sous ensembles indissociables.
A l'analyse de ces éléments, on élaborera des scénarios de reprise et des moyens de secours pour ramener l'impact estimé à un niveau acceptable. On s'assurera aussi de conserver la cohérence d'ensemble des éléments constitutifs du secours du noyau stratégique.
En cas de dysfonctionnement total de l'activité, il faudra étudier un moyen de restauration, de reprise et de resynchronisation des données associées au noyau fonctionnel (plus proche du PRA).

Cette démarche complète aboutit à la création du cahier des charges du Plan de Continuité d'Activité.

Annexe 4 : Adaptation du manuel de référence MEHARI pour le LATT et exemple de fiche de préconisation

Tableau IX: Tableau de synthèse du manuel de référence adapté au LATT

Risque	Objectif	Résultats globaux attendus	Services de sécurité
Domaine 1 : Organisation de la sécurité			
01E Continuité de l'activité	Réduire le coût de certains sinistres en analysant au préalable les besoins fondamentaux de l'activité en termes de continuité et en mettant en place les plans correspondants (PCA & PRA)	Définir et mettre en place les plans de continuité de l'activité adaptés aux besoins de l'entité	01E01 : Prise en compte des besoins de continuité de l'activité 01E02 : Plans de continuité de l'activité 01E03 : Plans de reprise de l'environnement de travail
01C06 Enregistrement des personnes	Gérer les utilisateurs du système d'information de manière à pouvoir tracer toute action et l'imputer à une personne physique	Faire en sorte que tout utilisateur du système d'information puisse être identifié de manière sûre et sans ambiguïté et cela dès l'attribution de droits généraux..	
Domaine 2 : Sécurité des sites			
02B Protection contre les risques environnementaux divers	Minimiser les risques dus à des risques environnementaux divers, par des analyses préalables et des actions de prévention et de protection	Empêcher ces risques et quand cela n'est pas possible en limiter les conséquences	02B01 : Protection contre les risques environnementaux divers
Domaine 3 : Sécurité des locaux			
03A Services techniques	Faire en sorte que les services techniques et généraux assurés par le site ou l'établissement (EDF, FAI, ...) soient fiables et conformes aux attentes et spécifications des constructeurs	Prévenir les dégâts ou inconvénients pouvant être causés accidentellement aux systèmes d'information par une défaillance primaire des servitudes générales fournies par l'établissement	03A01 : Qualité de la fourniture de l'énergie 03A02 : Continuité de la fourniture de l'énergie 03A03 : Sécurité de la climatisation 03A04 : Qualité du câblage 03A05 : Protection contre la foudre
03B Contrôle des accès aux zones sensibles (salles informatiques, salles réseau, ...)	Faire en sorte que seules les personnes autorisées aient accès aux zones sensibles	Prévenir les actions néfastes pouvant être menées, volontairement ou non, par des personnes n'ayant pas la nécessité d'accéder dans des zones sensibles	02C02 : Contrôle d'accès physique aux zones protégées 02C03 : Gestion des autorisations d'accès aux zones protégées 02C05 : Surveillance des zones protégées
03C Sécurité contre les dégâts des eaux	Minimiser les risques de dégâts des eaux, par des actions de prévention et de protection contre l'effet des dégâts des eaux	Empêcher les dégâts des eaux et quand cela n'est pas possible en limiter les conséquences	03C01 : Prévention des risques de dégâts des eaux 03C02 : Détection des dégâts des eaux 03C03 : Évacuation de l'eau
03D Sécurité incendie	Minimiser les risques dus à un incendie, par des actions de prévention et de protection contre l'incendie	Empêcher le départ d'un incendie et quand cela n'est pas possible en limiter les conséquences	03D01 : Prévention des risques d'incendie 03D02 : Détection d'incendie 03D03 : Extinction d'incendie
Domaine 4 : Réseau étendu intersites			
04A Sécurité de l'architecture du réseau étendu et continuité du service	Mettre en place une architecture globale des réseaux intersites garantissant la facilité de communication entre les entités tout en limitant les risques d'abus ou de mauvaise utilisation de ces moyens	Garantir une continuité des communications entre entités en cas d'incident ou de panne mais assurer chacune que sa connexion à d'autres entités ne la fragilise pas. Cet aspect visent un objectif qui est d'offrir un moyen de communication sûr.	04A02 : Organisation de la maintenance des équipements 04A03 : Procédures et plans de reprise du réseau étendus 04A04 : Plans de sauvegarde des configurations 04A05 : Plan de reprise d'activité (PRA) du réseau étendu
04C Sécurité des données lors des échanges et des communications	Protéger les données transmises contre des divulgations ou des altérations illicites	Prévenir les tentatives d'accès aux informations échangées, en lecture ou en modification, par des personnes non autorisées	04C01 : Chiffrement des échanges sur le réseau étendu 04C02 : Contrôle de l'intégrité des échanges sur le réseau étendu
04D Détection et traitement des incidents du réseau étendu	Détecter les anomalies de fonctionnement ou des intrusions sur le réseau étendu pour intervenir au mieux et dans les meilleurs délais	Éviter que des actions malveillantes externes (pirates, hackers) ou internes (abus de droits d'utilisateurs), que les mesures préventives auraient pu laisser passer, perdurent et se traduisent par des dysfonctionnements qui auraient pu être détectés et interrompus	04D01 : Surveillance en temps réel du réseau étendu 04D02 : Surveillance, en temps différé, des traces, logs et journaux des événements du réseau étendu 04D03 : Traitement des incidents du réseau étendu
Domaine 5 : Réseau local			
05A Sécurité de l'architecture du réseau local	Mettre en place une architecture globale du réseau local garantissant la facilité de communication entre les unités qui en ont le besoin tout en limitant les risques d'abus ou de mauvaise utilisation	garantir une continuité des communications en cas d'incident ou de panne mais limiter les possibilités d'intrusion ou en limiter la portée par des cloisonnements internes	05A01 : Partitionnement du réseau local en domaines de sécurité 05A03 : Organisation de la maintenance des équipements du réseau local 05A04 : procédures et plans de reprise du réseau local sur incidents 05A05 : Plans de sauvegardes des configurations du réseau local 05A06 : Plans de Reprise d'Activité (PRA) du réseau local
05B Contrôle d'accès au réseau de données	Faire en sorte que seuls les utilisateurs autorisés aient accès aux réseaux internes et contrôler les accès sortants	Prévenir les tentatives d'accès, par des personnes non autorisées, aux ressources internes (serveurs en particuliers) et aux informations circulant sur le réseau	05B02 : Gestion des autorisations d'accès et privilèges 05B03-07 : Filtrage général des accès aux réseaux locaux filaires et Wifi 05B08 : Contrôle du routage des accès sortants
05D Contrôle, détection et traitement des incidents du réseau local	Détecter les anomalies de fonctionnement ou des intrusions sur les réseaux locaux pour intervenir au mieux et dans les meilleurs délais	Éviter que des actions malveillantes externes (pirates, hackers) ou internes (abus de droits d'utilisateurs), que les mesures préventives auraient pu laisser passer, perdurent et se traduisent par des dysfonctionnements qui auraient pu être détectés et interrompus	05D01 : Surveillance en temps réel du réseau local 05D02 : Analyse, en temps différé, des traces, logs et journaux d'événements sur le réseau local 05D03 : Traitement des incidents du réseau local

Domaine 7 : Sécurité des systèmes et de leur architecture			
07A Contrôle d'accès aux systèmes	Faire en sorte que seuls les utilisateurs autorisés aient accès aux systèmes	Prévenir les tentatives d'accès, par des personnes non autorisées, aux ressources internes (serveurs et applications)	07A01 : Gestion des profils d'accès 07A02 : Gestion des autorisations d'accès et privilèges 07A03 : Authentification de l'utilisateur ou de l'entité demandant un accès 07A04 : Filtrage des accès et gestion des associations 07A05 : Authentification du serveur lors des accès à des serveurs sensibles
07B Confinement des environnements	Protéger les données stockées temporairement par les systèmes contre des accès non autorisés	Prévenir les tentatives d'accès aux informations stockées temporairement, par des personnes non autorisées	07B01 : Contrôle des accès aux résidus
07C Gestion et enregistrement des traces	Permettre une analyse a posteriori des actions effectuées et ainsi le diagnostic d'actions anormales ou néfastes	Dissuader les actions néfastes de la part des utilisateurs ou du personnel d'exploitation et, à défaut, en limiter la durée dans le temps, par un diagnostic suivi de mesures visant à stopper ces actions	07C01 : Enregistrement des accès aux ressources sensibles 07C02 : Enregistrement des appels aux procédures privilégiées
07D Sécurité de l'architecture	Mettre en place une architecture globale des systèmes garantissant une continuité de fonctionnement conforme aux attentes des utilisateurs	Garantir une continuité de fonctionnement des systèmes en cas d'incident ou de panne	07D01 : Sûreté de fonctionnement des éléments d'architecture 07D02 : Isolement des systèmes sensibles
Domaine 8 : Production informatique			
08A Sécurité des procédures d'exploitation	Assurer la conformité des procédures aux exigences de sécurité spécifiées	Éviter que des erreurs, des inattentions, voire des fautes intentionnelles, dans la mise en œuvre des moyens systèmes introduisent des brèches de sécurité.	08A01 : Sécurité de l'information - relations personnel d'exploitation 08A02 : Contrôle des outils et utilitaires de l'exploitation 08A03 : Contrôle mise en production nouveaux systèmes 08A04 : Contrôle des opérations de maintenance 08A08 : Gestion des procédures opérationnelles d'exploitation
08B Contrôle des configurations matérielles et logicielles	Assurer la conformité des configurations aux exigences de sécurité spécifiées	Éviter que des erreurs, des inattentions, voire des fautes intentionnelles, dans le paramétrage des systèmes et équipements de sécurité introduisent des brèches de sécurité.	08B01 : Paramétrage des systèmes et contrôle de la conformité des configurations systèmes 08B02 : Contrôle de la conformité des configurations applicatives 08B03 : Contrôle de la conformité des programmes de référence (sources et exécutables)
08C Gestion des supports informatiques de données et programmes	Assurer avec rigueur la gestion des supports de données et programmes	Éviter que des erreurs, des inattentions, voire des fautes intentionnelles, dans la gestion des supports introduisent des brèches de sécurité.	08C01 : Administration des supports 08C03 : Sécurité physique des supports stockés sur site 08C04 : Sécurité physique des supports externalisés (site externe) 08C05 : Vérification et rotation des supports d'archivage 08C06 : Protection des réseaux de stockage
08D Continuité de fonctionnement	Assurer la continuité de fonctionnement des systèmes et applications.	Éviter que des circonstances externes ou internes, accidentelles ou malveillantes, se traduisent par des interruptions de service inacceptables (étant entendu que la recherche de ce qui est acceptable ou non fait partie des services à assurer au titre de la continuité de fonctionnement)	08D01 : Organisation de la maintenance du matériel en exploitation 08D02 : Organisation de la maintenance du logiciel (systèmes et logiciels) 08D03 : Procédures et plans de reprise des applications sur incidents 08D04 : Sauvegardes des configurations logicielles 08D05 : Sauvegardes des données applicatives 08D06 : Plans de Reprise d'Activité des services informatiques 08D07 : Protection antivirale des serveurs de production 08D08 : Gestion de la maintenance des systèmes critiques 08D09 : Sauvegardes de secours (recours) externalisées 08D10 : Maintien des comptes d'accès
08E Gestion et traitement des incidents	Détecter les anomalies de fonctionnement ou des intrusions dans les systèmes pour intervenir au mieux et dans les meilleurs délais.	Éviter que des incidents, des anomalies ou des actions malveillantes externes (pirates, hackers) ou internes (abus de droits d'utilisateurs), que les mesures préventives auraient pu laisser passer, ne soient pas détectés ni interrompus	08E01 : Détection et traitement en temps réel des anomalies et incidents d'exploitation 08E02 : Surveillance, en temps différé, des traces, logs et journaux 08E03 : Gestion et traitement des incidents systèmes et applicatifs
08F Contrôle des droits d'administration	Gérer avec rigueur l'attribution et l'utilisation de droits privilégiés sur les systèmes et applications	Éviter que les configurations sécurisées soient modifiées par des personnes abusant de droits d'administration ou que des procédures réservées à l'administration des systèmes ou des postes utilisateurs soit utilisées en dehors du contexte normalement prévu.	08F01 : Gestion des attributions de droits privilégiés sur les systèmes (droits d'administrateur) 08F02 : Authentification et contrôle des droits d'accès des administrateurs et personnels d'exploitation 08F03 : Surveillance des actions d'administration des systèmes
08G Procédures d'audit et de contrôle des systèmes de traitement de l'information	Prévenir toute activation incontrôlée d'outils d'audit technique et toute erreur ou malveillance lors des audits. Enregistrer les résultats et détecter les anomalies de fonctionnement ou modifications et intrusions éventuelles.	Éviter que des incidents, des anomalies ou des actions malveillantes externes (pirates, hackers) ou internes (abus de droits d'utilisateurs), puissent arriver du fait de ces tests d'audit et ne soient pas détectés ni interrompus.	08G01 Fonctionnement des contrôles d'audit 08G02 Protection des outils et résultats d'audit
08H Gestion des archives informatiques	Faire en sorte que les accès aux archives soient gérés de façon adéquate	Éviter que des incidents, des anomalies ou des actions malveillantes externes (pirates, hackers) ou internes (abus de droits d'utilisateurs) puissent survenir sur les supports des archives informatique et lors de leur conservation.	08H01 Politique de gestion des archives informatiques 08H02 Gestion des accès aux archives 08H03 Gestion de la sécurité des archives
Domaine 9 : Sécurité applicative			
09A Contrôle d'accès applicatif	Faire en sorte que seuls les utilisateurs autorisés aient accès aux données applicatives.	Prévenir les tentatives d'accès, par des personnes non autorisées, aux données et bases de données des applications.	09A01-02 : Gestion des profils d'accès et autorisations aux données 09A03 : Authentification de l'utilisateur ou de l'entité demandant un accès 09A04 : Filtrage des accès et gestion des associations 09A05 : Authentification des applications lors d'accès aux applications sensibles
09B Contrôle de l'intégrité des données	Protéger les données saisies, stockées ou échangées contre des altérations indues ou illicites.	Prévenir les altérations de données pouvant passer inaperçues, soit en empêchant qu'elles soient altérées, soit en détectant leur altération avant usage.	09B01 : Scellement des données sensibles 09B02 : Protection de l'intégrité des données échangées 09B03 : Contrôle de la saisie des données 09B04 : Contrôles permanents (vraisemblance, ...) sur les données 09B05 : Contrôles permanents (vraisemblance, ...) sur les traitements
09C Contrôle de la confidentialité des données	Protéger les données saisies, stockées ou échangées contre des divulgations indues ou illicites.	Prévenir les divulgations de données à des personnes non autorisées ou dans des conditions non autorisées.	09C01 : Chiffrement des échanges 09C02 : Chiffrement des données stockées 09C03 : Dispositif anti-rayonnement électromagnétique
09D Disponibilité des données	Assurer la disponibilité ou le recouvrement des données en toutes circonstances.	Éviter que des accidents ou des malveillances rendent des données totalement ou en partie indisponibles.	09D01 : Enregistrement de très haute sécurité 09D02 : Gestion des moyens d'accès aux fichiers de données

09E Continuité de fonctionnement	Assurer la continuité de fonctionnement des services applicatifs.	Éviter que des accidents ou des malveillances arrêtent les services applicatifs au delà d'une durée inacceptable par les utilisateurs.	09E01 : Reconfiguration matérielle 09E02 : Plans de continuité des processus applicatifs 09E03 : Gestion des applications critiques (vis-à-vis de la permanence de la maintenance)
09F Contrôle de l'émission et de la réception des données	Assurer certains services de sécurité lors de l'émission de messages	Éviter que des messages soient dupliqués, répudiés ou détournés sans que l'utilisateur s'en aperçoive.	09F01 : Garantie d'origine, signature électronique 09F02 : Individualisation des messages empêchant leur duplication (numérotation, séquencement) 09F03 : Accusé de réception
09G Détection et gestion des incidents et anomalies applicatifs	Détecter et gérer les incidents et anomalies applicatifs	Éviter que des incidents ou anomalies perdurent et limiter leur impact.	09G01 : Détection des anomalies applicatives.
Domaine 11 : Protection des postes de travail utilisateurs			
11A Sécurité des procédures d'exploitation du parc de postes utilisateurs	S'assurer que l'ensemble du parc des postes utilisateurs répond aux exigences contractuelles, légales et réglementaires et à la politique de l'entreprise.	Prévenir les changements de configurations non décidés par l'exploitation (du parc de postes utilisateurs)	11A01 : Contrôle de l'installation de nouvelles versions de progiciels ou systèmes sur les postes utilisateurs 11A02 : Contrôle de la conformité des configurations utilisateurs 11A03 : Contrôle des licences des logiciels et progiciels
11B Protection des postes de travail	Faire en sorte que seules les personnes autorisées aient accès aux informations contenues dans les postes de travail et protéger ceux-ci contre toute altération nuisible	Prévenir les actions néfastes pouvant être menées, volontairement ou non, par des personnes n'ayant pas la nécessité d'accéder, pour leur travail, au poste de travail des collaborateurs.	11B01 : Contrôle d'accès au poste de travail 11B02 : Travail en dehors des locaux de l'entreprise 11B03 : Utilisation d'équipements personnels ou externes (n'appartenant pas à l'entreprise)
11C Protection des données du poste de travail	Faire en sorte que seules les personnes autorisées aient accès aux informations contenues dans les postes de travail et protéger ceux-ci contre toute altération nuisible	Prévenir les actions néfastes pouvant être menées, volontairement ou non, par des personnes n'ayant pas la nécessité d'accéder, pour leur travail, au poste de travail des collaborateurs.	11C01 : Protection de la confidentialité des données contenues sur le poste de travail ou sur un serveur de données 11C03 : Confidentialité lors des opérations de maintenance des postes 11C04 : Protection de l'intégrité des fichiers 11C05 : Sécurité de la messagerie électronique (courriels) et des échanges électroniques d'information 11C06 : Protection des impressions sur imprimantes partagées
11D Continuité de service de l'environnement de travail	Assurer la continuité de service de l'environnement de travail	Faire en sorte que les utilisateurs puissent trouver un environnement de travail satisfaisant en toutes circonstances.	11D01 : Organisation de la maintenance du matériel mis à disposition 11D02 : Organisation de la maintenance du logiciel utilisateur 11D04-05 : Plans de sauvegarde des données utilisateurs 11D06 : Protection des postes utilisateurs contre des codes malveillants ou des codes exécutables non autorisés 11D07 : Plans de Reprise d'Activité des postes utilisateurs

CLUSIF Espace Méthodes	**MEHARI** **Manuel de référence des Services de sécurité**	Version 2010 décembre 2009
Domaine 5 : Réseau local		

05A02 Sûreté de fonctionnement des éléments d'architecture du réseau local

Objectif :

Assurer la fluidité des communications, à l'intérieur du réseau local, quels que soient les aléas de fonctionnement des équipements ou d'utilisation du réseau par les utilisateurs

Résultats attendus :

Éviter qu'une panne simple, voire complexe ou que des aléas dans l'utilisation du réseau local ne fassent chuter la performance du réseau avec des impact négatifs sur l'activité de l'entreprise.

Mécanismes et solutions :

Les mécanismes principaux sont du ressort de l'architecture, mais doivent s'appuyer sur des études préalables qui sont de nature organisationnelles

Mesures organisationnelles

- Les mesures organisationnelles ne sont utilisées que dans une phase préalable et consistent en une analyse des enjeux de la continuité des services du réseau local :
  - Analyse de la gravité des conséquences en fonction de la durée d'une interruption complète des services du réseau local
  - Analyse des capacités de reroutage des communications en cas de défaillance d'un équipement et du temps nécessaire à ces reconfigurations
  - Analyse des enjeux liés à une baisse de performance du réseau local

Mesures techniques

- Les mesures de base consistent en la mise en place d'une architecture à tolérance de panne telle que :
  - toute panne simple d'un équipement du réseau local puisse être soit réparée dans des délais acceptables soit contournée par des redondances d'équipement
  - toute baisse de performance significative puisse être détectée et corrigée par des reconfigurations permettant d'assurer un débit acceptable par les utilisateurs
- Les mesures techniques d'accompagnement consistent en des outils de surveillance et de monitoring du réseau local permettant effectivement de détecter toute anomalie ou baisse de performance et d'agir à distance pour allouer des ressources de contournement ou d'appoint

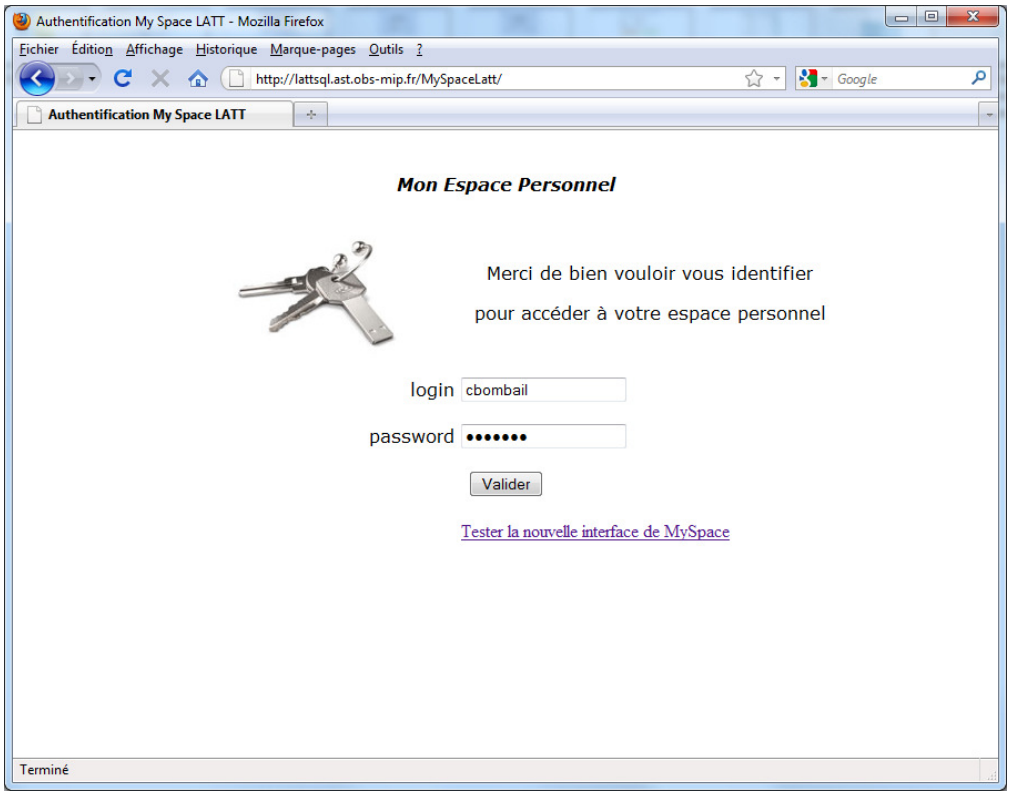
Efficacité du service :

- L'efficacité du service est directement liée à trois facteurs :
  - La qualité et la précision des analyses préliminaires
  - La rapidité et l'automaticité de détection d'une anomalie ou d'une baisse de performance
  - La rapidité et l'automaticité des reconfigurations possibles
- La robustesse des mécanismes assurant la sûreté de fonctionnement de l'architecture dépend de deux facteurs :
  - L'indépendance des équipements spéciaux (réseaux et télécoms) vis-à-vis de tout équipement de servitude (énergie, services généraux, locaux, etc.).
  - La protection, physique et logique, des équipements assurant la détection et la reconfiguration contre toute intrusion ou inhibition : contrôle d'accès physique et logique, alarme en cas d'arrêt ou de coupure d'alimentation, etc.
- La mise sous contrôle des mécanismes assurant la sûreté de fonctionnement dépend de trois facteurs :
  - Les tests de performance et de bon fonctionnement des mécanismes de détection et de reconfiguration
  - L'audit du paramétrage des équipements
  - L'audit des procédures associées à la gestion des systèmes de détection et de reconfiguration.

Annexe 5: Les interfaces homme-machine (IHM)

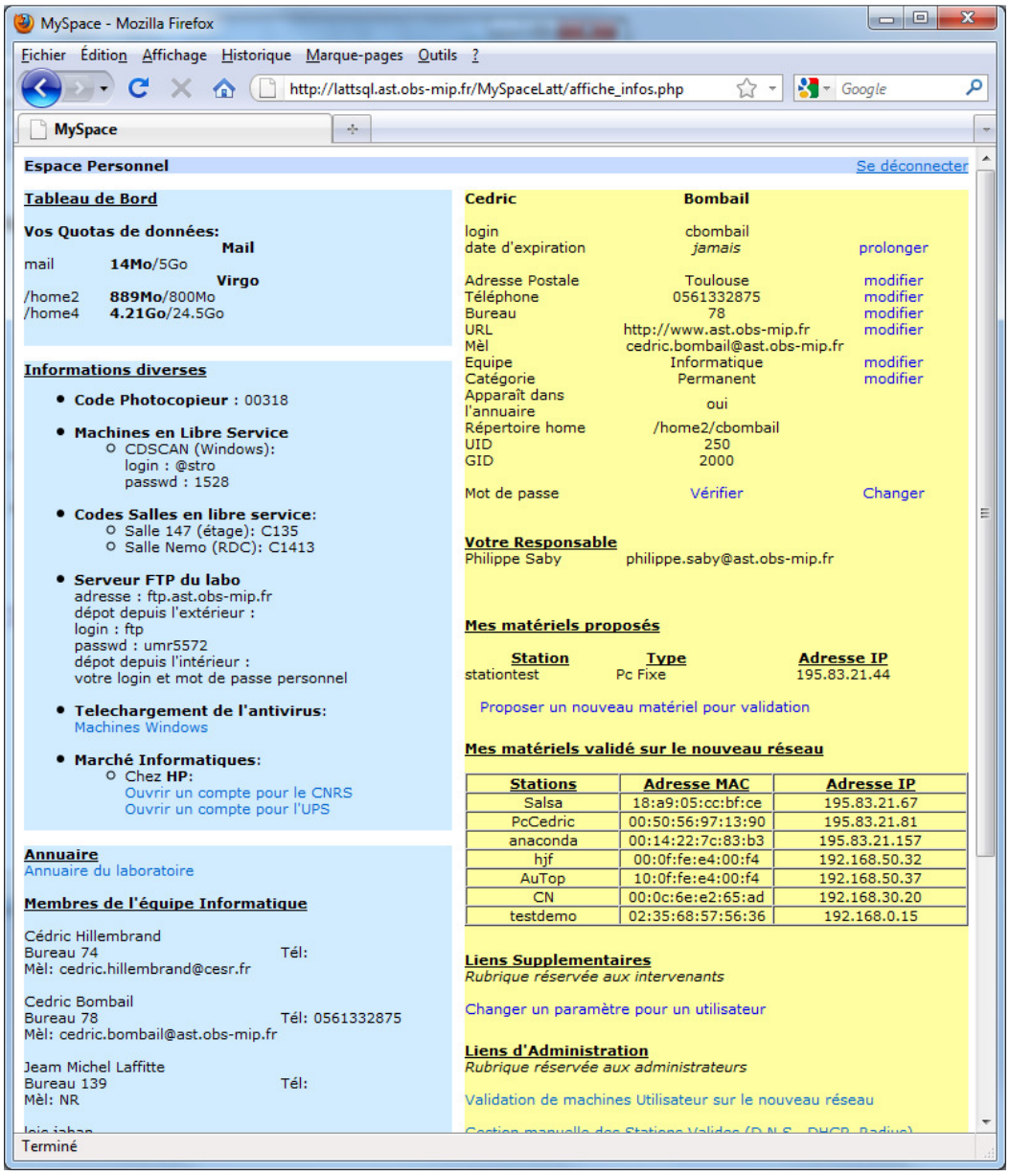
Voici quelques exemples d'IHM qui ont été développées lors de la mise en place du nouveau système d'information.
Les premières IHM ont été développées dans la version prototype et proposées aux utilisateurs de tests.
Les secondes IHM sont une évolution et un enrichissement des versions de la phase de prototypage, et sont celles qui sont en exploitation aujourd'hui.

IHM première version : phase de prototypage



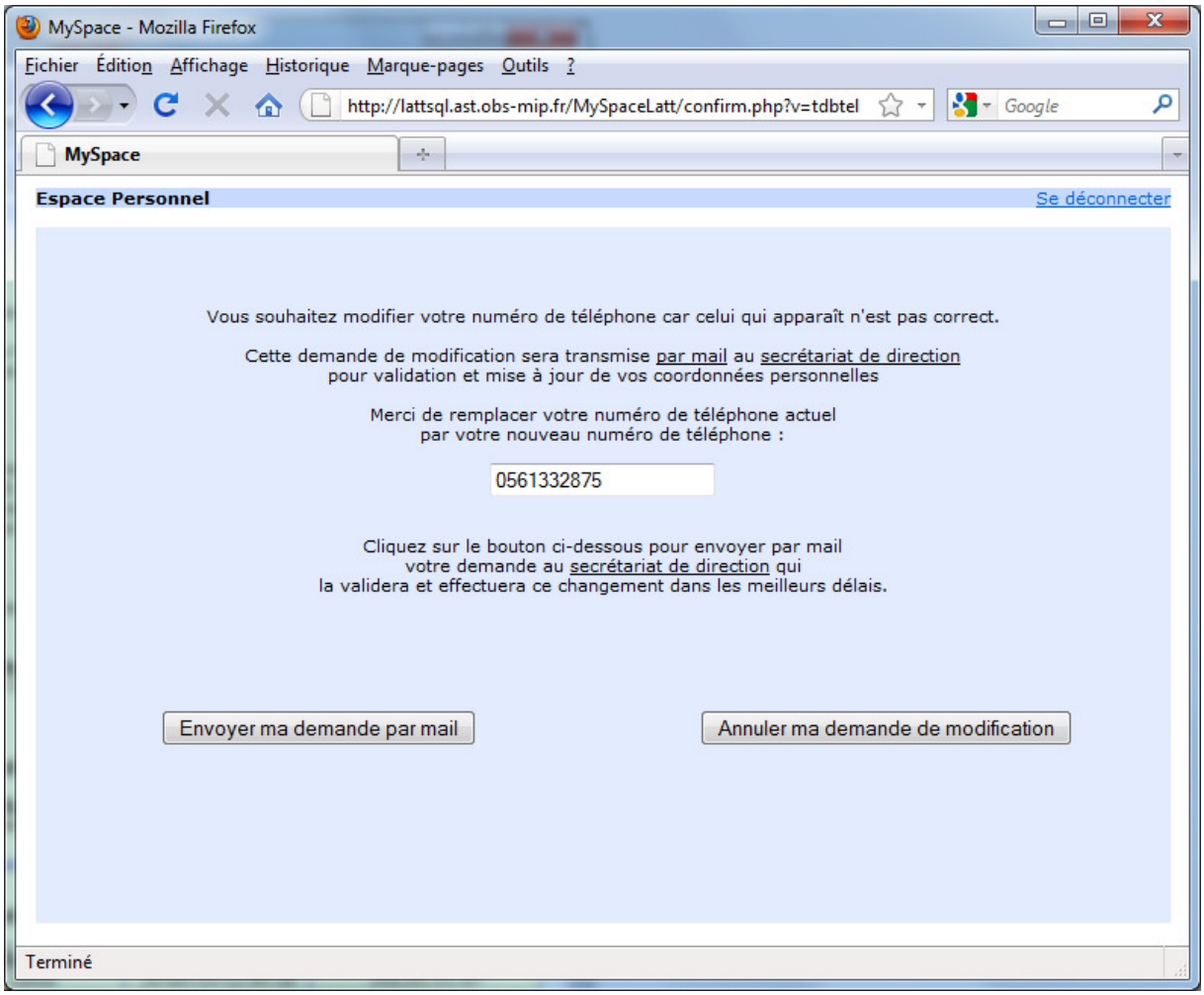
IHM 1 : Page d'authentification

La page d'authentification est la seule page qui n'ait pas évolué lors des deux phases de développement d'IHM. Son code authentifiait l'utilisateur initialement sur le service LDAP, puis en exploitation sur le service Kerberos.



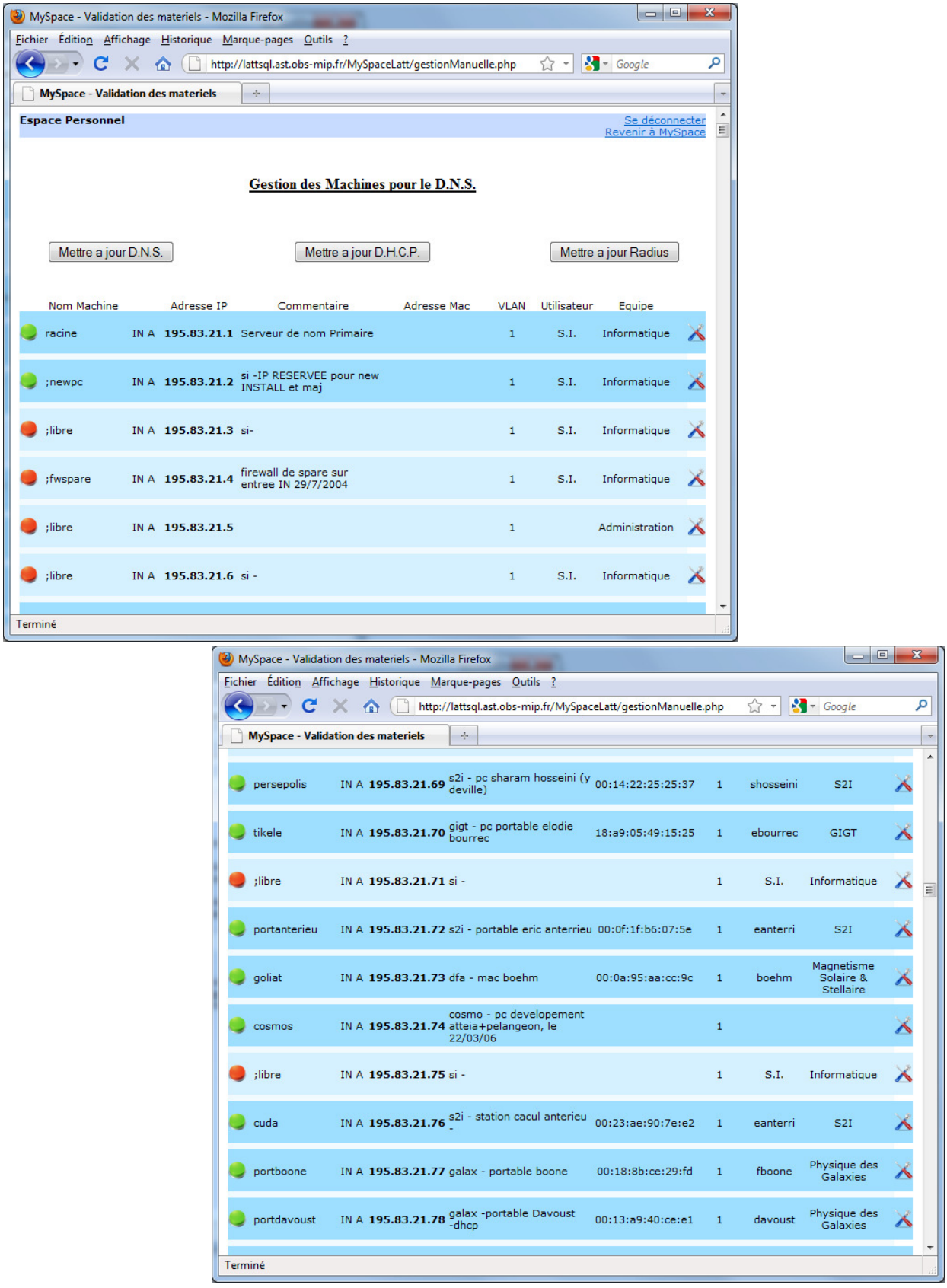
IHM 2 : Tableau de bord

Le tableau de bord contenait l'affichage des informations concernant l'utilisateur ainsi que les liens d'accès aux services en fonction du profil de chaque utilisateur.
La page présentée est celle correspondant à mon accès, *i.e.* avec un profil administrateur.

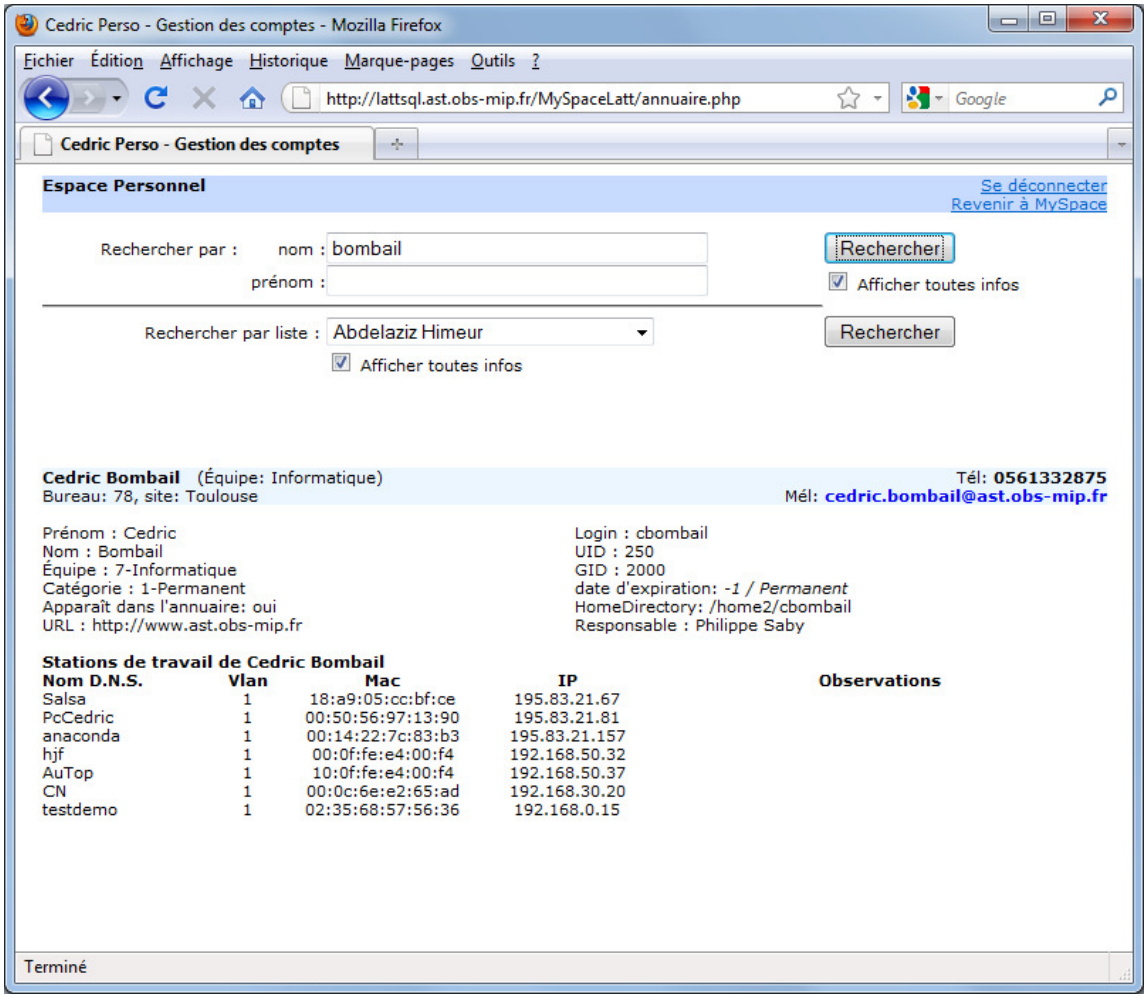


IHM 3 : IHM de demande de modification de numéro de téléphone

Cette IHM est un exemple de la nouvelle offre de services dirigée vers l'utilisateur. Elle permettait aux utilisateurs de demander une modification de leur numéro de téléphone par mail afin que les modérateurs puissent accéder à leurs requêtes dans un délai optimum.



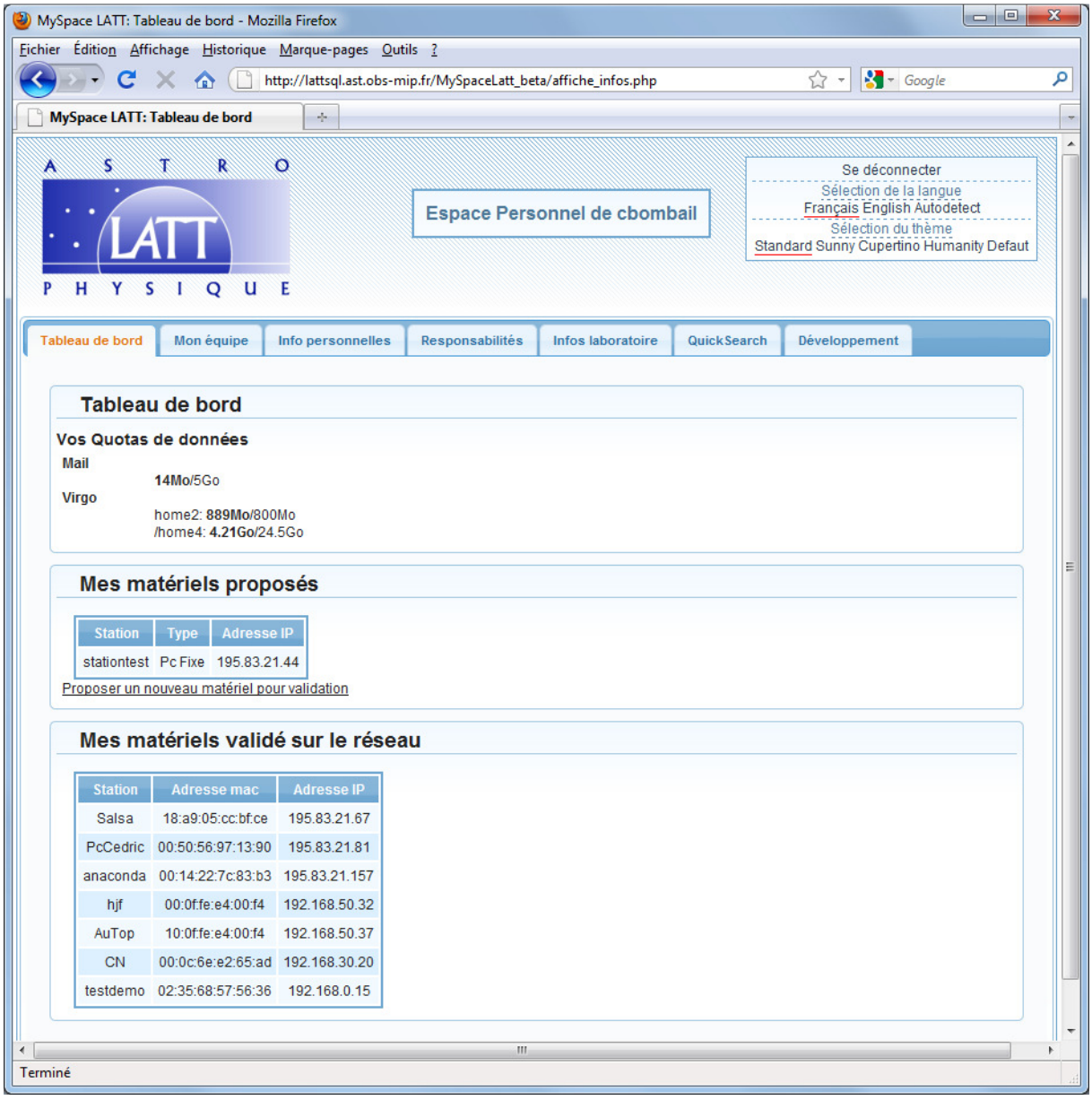
IHM 4 : Tableau de gestion des adresses IP du laboratoire (profil administrateur)



IHM 5 : Annuaire du laboratoire (profil administrateur)

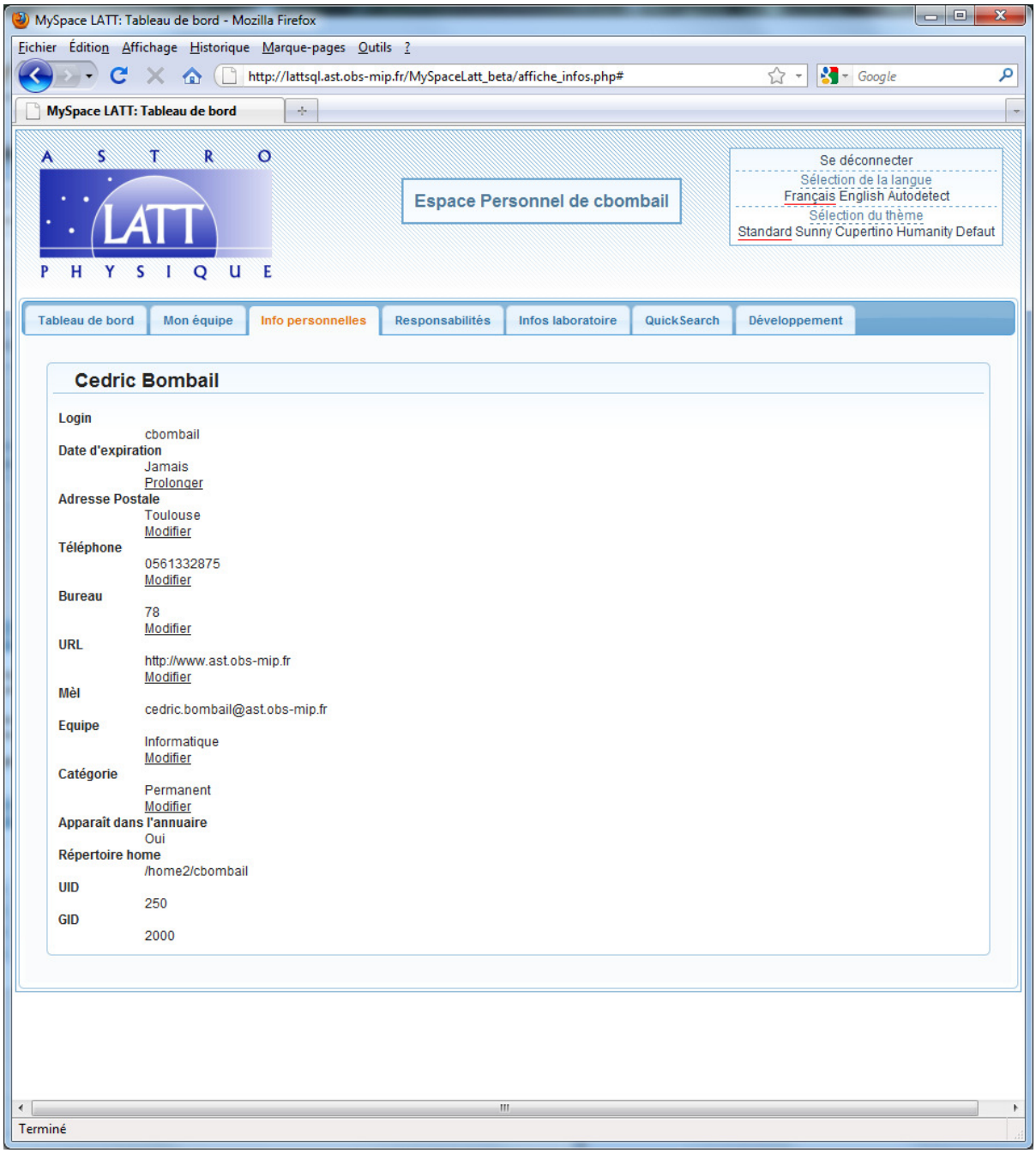
L'annuaire du laboratoire sur l'intranet authentifié, sous profil administrateur, était complet et permettait de mieux préparer les interventions quotidiennes en clientèle en récapitulant les données relatives aux utilisateurs, à leurs comptes informatiques et à leurs stations de travail.

IHM seconde version : évolutions pour la mise en exploitation

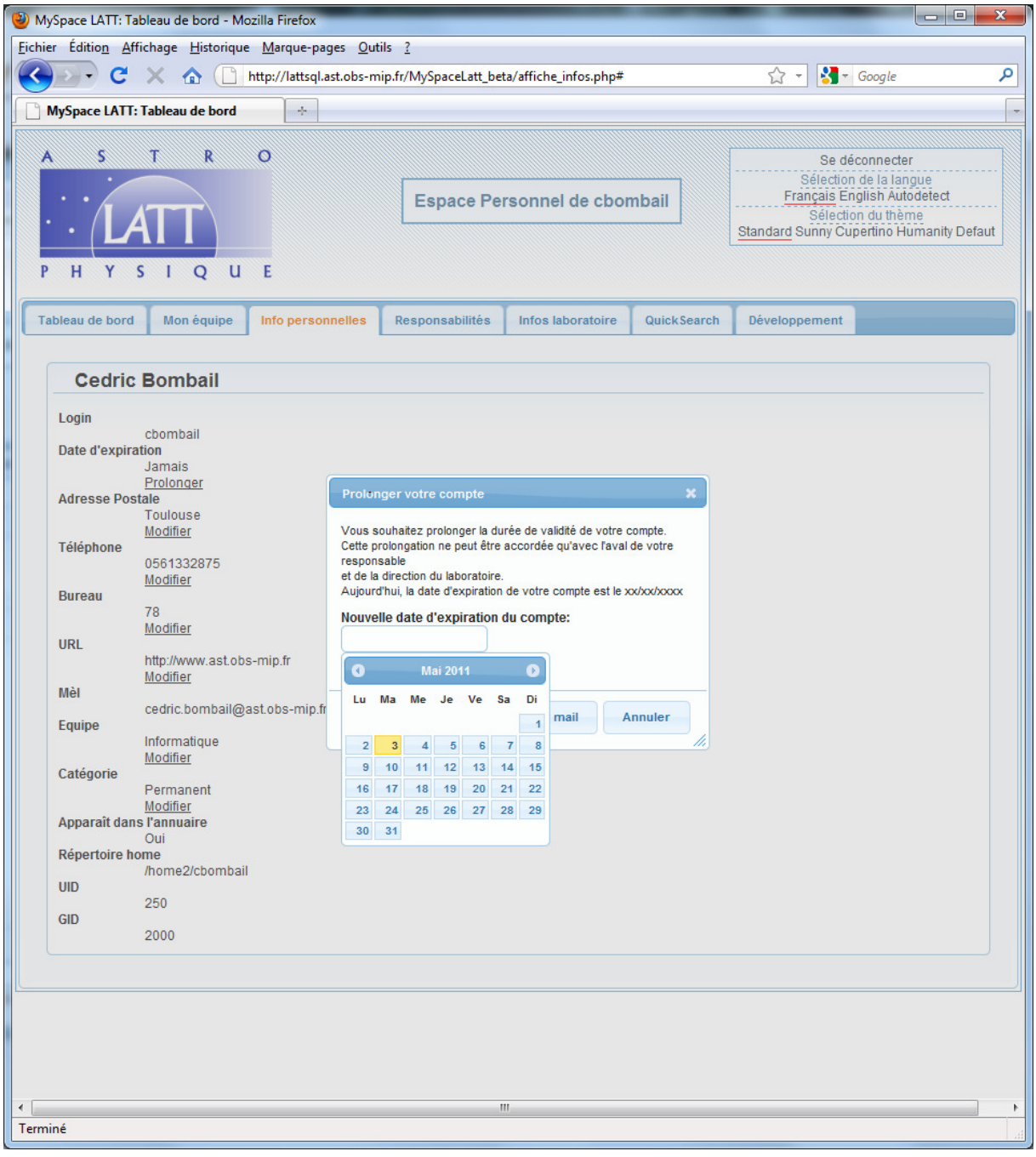


IHM 6 : Tableau de bord

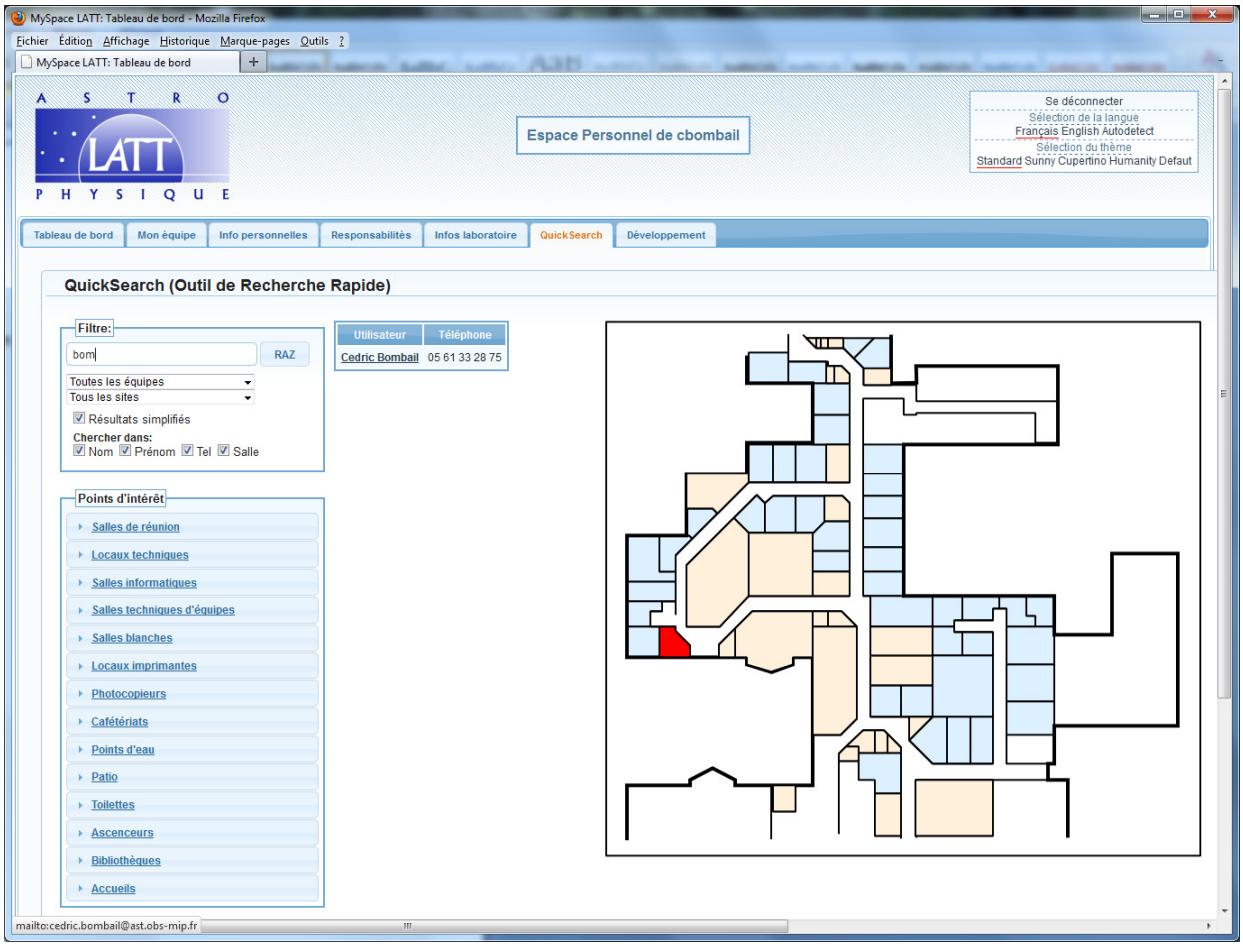
Nouvelles IHM, introduction du Web 2.0 avec les technologies Ajax et JQuery.
Nouvelle ergonomie et ajout du support en langue anglaise.



IHM 7 : Tableau de bord suite - informations personnelles



IHM 8 : Nouvelles IHM de demande de modification en AJAX



IHM 9 : Annuaire du laboratoire avec Géo localisation

L'annuaire utilisateur (accès réservé sur intranet authentifié) de la version d'exploitation permet, en plus des fonctionnalités de la version du prototype, la géo localisation des bureaux.

Annexe 6 : Tableaux de synthèse présentés en commission informatique

Ces tableaux de synthèse ont été présentés lors de la dernière commission informatique du LATT à l'ensemble des membres de la commission. Ils détaillent les principaux gains pour les utilisateurs, la direction et son secrétariat ainsi que pour le service informatique.

Utilisateurs

	AVANT	APRES
Arrivée d'une nouvelle station de travail	Mise à jour manuelle de chaque service réseau par le service informatique(SI). Risques d'incohérences de données.	Déclaration utilisateur Modération du SI. Mise à jour sur le référentiel unique et génération dynamique des configurations de services.
Modification des informations personnelles	Consultation difficile. Envoi de mail aux services administratif et SI.	Consultation : Interface personnelle Modifications : réalisées par l'utilisateur Modération optimisée répartie entre les services.
Changement d'un mot de passe	Changement mot de passe LDAP par intervention avec concours du service informatique (RDV)	Interface personnelle de modification du mot de passe
Consultation de l'état de ses comptes informatiques serveurs et Mail	Mail de demande au service informatique.	Tableau de bord personnel avec mise à jour en temps réel.
Consultation annuaire	Réponse basique (Nom, prénom, mail et téléphone)	Réponse basique plus géo localisation des bureaux. Possibilité de localiser les salles de réunion, les salles techniques, etc.
Mobilité	Demande d'intervention Brassage de prises manuel et/ou paramétrage des commutateurs (VLAN)	Automatique par identification matérielle (Radius)

Tableau X : Synthèse des apports pour les utilisateurs

Direction et secrétariat de direction

	AVANT	APRES
Prolongation de comptes	Echanges de mail avec service informatique	Modération des demandes via une IHM dédiée.
Visualisation de la composition du laboratoire	Difficile	IHM d'interrogation permettant un tri multicritères : par catégorie d'utilisateurs, par équipe, par répartition géographique, par bureaux, etc.
Création des listes électorales	Manuelle	Extraction rapide via une IHM spécialisée.

Tableau XI : Synthèse des apports pour la direction et son secrétariat

Service Informatique

	AVANT	APRES
Gestion du compte informatique	Ouverture, gestion et fermeture du compte manuelles.	Déclaration préalable saisie par l'utilisateur ou son responsable. Gestion et fermeture automatisées.
Administration des services réseaux	Configuration manuelle de chacun des services. Risques d'erreurs.	Gestion automatisée par modification du référentiel unique et génération des services (Radius, DNS, DHCP).
Maintien des données utilisateurs	Mise à jour manuelle du serveur LDAP par interface dédiée (phpLdapAdmin). Risques d'erreur de saisie.	Données modifiées par les utilisateurs et validées par les modérateurs dans le référentiel unique. Génération dynamique du LDAP.
Maintien des données matérielles	Interventions manuelles dans GLPI selon l'arrivée des informations.	Vérification des données par analyse des logs Radius.

Tableau XII : Synthèse des apports pour le service informatique

VII. Bibliographie

Ouvrages de référence

[1] DE COURCY R., 1992. Les systèmes d'information en réadaptation, In *Réseau international CIDIH et facteurs environnementaux*, no 5 vol. 1-2, Québec, 7-10.

[2] GILLET P., 2009. *Virtualisation des systèmes d'information avec VMware*, EDITION ENI, St Herblain, 292 p.

[3] HARDY D., MALLEUS G., MEREUR J., 2002. Les protocoles applicatifs In *Réseaux: Internet, téléphonie, multimédia, Convergences et complémentarités,* DE BOECK UNIVERSITE S.A., Bruxelles, 449-461.

[4] ALBITZ P., CRICKET L., tr. CARRE G., 1999. *DNS et BIND,* EDITIONS O'REILLY, Paris, 465 p.

[5] CARTER G., 2003. *LDAP System Administration*, O'REILLY & Associates, Paris, 317 p.

Articles

[6] BOEHM B., 1988. In *A Spiral Model of Software Development and Enhancement*, IEEE, 12p.

[7] SACCAVINI L., 2003. In *802.1X et la sécurisation de l'accès au réseau local*, DRSI INRIA, 6p.

[8] BARDY C., septembre 2009. *Ecologie + économies = virtualisation*, Pc Expert n°200, p. 40.

[9] RASMUSSEN N., 2003. In *Calculating Total Cooling Requirements for Data Centers*, White Paper #25, APC, 8 p.

Actes de congrès

[10] D.BENZA, D. JOIRET, *Virtualisation du service de soumission de messages*, JRES2009, Nantes, décembre 2009.

[11] H. DUARTE, *Virtualisation du poste de travail: Le cas de l'université Rennes 2*, JRES2009, Nantes, décembre 2009.

RFC (requests for comments)

Les requests for comments (RFC), littéralement « demande de commentaires », sont une série numérotée de documents officiels décrivant les aspects techniques d'Internet, ou de différents matériels informatiques.
Peu de RFC sont des standards, mais tous les standards d'Internet publiés par l'IETF sont des RFC.
Accès aux RFC : *http://www.ietf.org/rfc/rfc{numero de RFC}.txt*

• Requests for comments : LDAP

[RFC 1487] W. Yeong, T. Howes, S. Kille, *X.500 Lightweight Directory Access Protocol,* July 1993

[RFC 1777] W. Yeong, T. Howes, S. Kille, *Lightweight Directory Access Protocol,* March 1995

[RFC 2251] M. Wahl, T. Howes, S. Kille, *Lightweight Directory Access Protocol (v3),* December 1997

[RFC 3377] J. Hodges, R. Morgan J., *Lightweight Directory Access Protocol (v3): Technical Specification*, September 2002

• Requests for comments : DNS

[RFC 882] P. Mockapetris, *Domain Names - Concepts and Facilities,* November 1983

[RFC 883] P. Mockapetris, *Domain Names – Implementation and specifications,* November 1983

• Requests for comments : DHCP

[RFC 951] B. Croft, J. Gilmore, *Bootstrap Protocol (BOOTP),* September 1985

[RFC 2131] R. Droms, *Dynamic Host Configuration Protocol,* March 1997

• Requests for comments : KERBEROS

[RFC 4120] C. Neuman, T. Yu, S. Hartman, K. Raeburn, *The Kerberos Network Authentication Service (V5),* July 1995

[RFC 4537] L. Zhu, P. Leach, K. Jaganathan, *Kerberos Cryptosystem Negotiation Extension*, June 2006

[RFC 5021] S. Josefsson, *Extended Kerberos Version 5 Key Distribution Center (KDC) Exchanges over TCP,* August 2007

[RFC 5896] L. Hornquist Astrand, S. Hartman, *Generic Security Service Application Program Interface (GSS-API): Delegate if Approved by Policy,* June 2010

• Requests for comments : SASL

[RFC 2222] J. Myers, *Simple Authentication and Security Layer (SASL),* October 1997

[RFC 4422] A. Melnikov, K. Zeilenga, *Simple Authentication and Security Layer (SASL),* June 2006

Sites Internet

Références Kerberos

[12] http://web.mit.edu/kerberos/ (Juin 2010)

Références Bases de données

[13] http://www.postgresql.org/docs/ (Mai 2010)

[14] http://dev.mysql.com/doc/refman/5.0/fr/index.html (Mai 2010)

Références Haute Disponibilité

[15] http://www.clusif.asso.fr/ (Juillet 2010)

[16] http://www.haute-disponibilite.net/ (Juillet 2010)

[17] http://www.cloudconsulting.fr/index.php/en/cloud-computing/theory-in-french/mesure-et-estimation-de-la-haute-disponibilite (Juillet 2010)
[18] http://www.commentcamarche.net/contents/surete-fonctionnement/haute-disponibilite.php3 (juillet 2010)

Références Plan de Continuité d'Activité
[19] http://www.clusif.asso.fr/fr/production/ouvrages/type.asp?id=METHODES (Juillet 2010)
[20] http://cyberzoide.developpez.com/securite/methodes-analyse-risques/ (Juillet 2010)
[21] https://www.clusif.asso.fr/fr/production/ouvrages/resume.asp?id=125 (Juillet 2010)

Références Cycle de développement en Spirale (B.Boehm)
[22] http://www.cs.umd.edu/class/spring2003/cmsc838p/Process/spiral.pdf (Septembre 2009)

Références British Thermal Unit (BTU)
[23] http://h20000.www2.hp.com/bizsupport/TechSupport/Document.jsp?objectID=c00211069&lang=en&cc=us&taskId=101&contentType=SupportFAQ&prodSeriesId=1121486&prodTypeId=15351 (Novembre 2009)
[24] http://personnellement.net/btu-to-watts-calculator-pdf-1.html (Novembre 2009)

VIII. Liste des figures

IX. Liste des tableaux

Inclusion des nouvelles technologies en vue de l'évolution de la disponibilité et de la qualité du service d'un système d'information.

Mémoire d'ingénieur CNAM, Toulouse 2011

RESUME

Ce document traite de la réorganisation d'un système d'information dans un laboratoire de recherche. Il fait état d'une évolution organisationnelle et technique permettant de faire migrer un système d'information en production vers un système à haute disponibilité.

Il traite cette contrainte par la mise en cohérence de l'ensemble des données du système et la génération dynamique des services. Il détaille également les solutions apportées en termes de sécurisation des échanges et des accès au réseau.

Ces améliorations sont apportées grâce à la mise en place de technologies récentes et nouvelles, comme l'intégration d'une solution de virtualisation ou la séparation de l'authentification et de l'identification des clients (utilisateurs et machines).

Les conséquences de cette évolution ont été une simplification des tâches d'administration, une augmentation du taux de satisfaction des utilisateurs, une optimisation de la sécurité du réseau et une offre de service plus aboutie.

Mots-clés
Haute disponibilité, système d'information, virtualisation, PCA.

SUMMARY

This document discusses the reorganization of an information system in a research laboratory. It reported a technical and organizational development to migrate a system of information production to a highly available system.

It addresses this constraint by ensuring consistency of all system data and the dynamic generation of services. It also details the solutions in terms of security exchanges and network access.

These improvements are thanks to the introduction of latest technologies and news, as the integration of a virtualization solution or the separation between authentication and identification of users (users and machines).

The consequences of these developments were a simplification of administrative tasks, an increase in user satisfaction, optimization of network security and most complete service offering.

Key words
High availability, information systems, virtualization, BCP.