



HAL
open science

Contribution à la conception d'un moteur de ciblage marketing

Emmanuel Barillot

► **To cite this version:**

Emmanuel Barillot. Contribution à la conception d'un moteur de ciblage marketing. Réseaux et télécommunications [cs.NI]. 2010. dumas-00529655

HAL Id: dumas-00529655

<https://dumas.ccsd.cnrs.fr/dumas-00529655v1>

Submitted on 26 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONSERVATOIRE NATIONAL DES ARTS ET METIERS

CENTRE REGIONAL ASSOCIE DE TOURS

MEMOIRE

présenté en vue d'obtenir

le DIPLOME D'INGENIEUR CNAM

SPECIALITE : Informatique

OPTION : Informatique Réseaux et Systèmes Multimédia

par

Emmanuel BARILLOT

Contribution à la conception d'un moteur de ciblage marketing

Soutenu le 17 septembre 2010

JURY

PRESIDENT :	M. Pierre CUBAUD	Professeur au Conservatoire National des Arts et Métiers à Paris
MEMBRES :	M. Hassan BENAZIZI	Responsable de la filière Informatique au Conservatoire National des Arts et Métiers de Tours
	M.Thierry BROUARD	Professeur au Conservatoire National des Arts et Métiers de Tours
	M.Jean-Luc BRIZARD	Directeur des Systèmes d'Information, Coface Services
	M.Hugues MERCIER	Chef de Projets Informatiques, Coface Services

Remerciements

Je tiens à remercier M. Jean-Luc BRIZARD, Directeur des Systèmes d'Information de Coface Services, pour avoir accepté de faire partie de mon jury. M. BRIZARD m'a recruté il y a plus de treize années maintenant, dans une société qui s'appelait encore ORT. J'y ai fait mes premiers pas d'informaticien en entreprise. Je sortais de mes longues études académiques, et il m'a fait confiance. J'ai pu acquérir de nombreuses compétences (du moins je l'espère !) au cours des projets diversifiés auxquels j'ai été affecté. Ainsi, sous sa direction, j'ai acquis une certaine polyvalence ce dont je suis très satisfait; je l'en remercie très sincèrement.

Je remercie également les collègues avec qui j'ai eu beaucoup de plaisir à travailler chez Coface Services, depuis toutes ces années. Je pense particulièrement à notre « DBA historique », Francis, maintenant en retraite, qui m'a patiemment apporté une grande partie de ses connaissances sur Oracle, sur les systèmes et sur les « bitmaps ». Je pense à Hugues, l'inénarrable compagnon des projets difficiles mais aussi de quelques belles réussites ! Je souhaiterais remercier chaque collègue individuellement, mais un mémoire entier n'y suffirait pas. Néanmoins, j'ai une pensée très amicale pour chacune et chacun en remerciement de leur bonne humeur.

Je souhaite remercier chaque membre du jury pour avoir accepté de juger mon travail, tout imparfait qu'il soit.

Je remercie tout particulièrement Hassan BENAZIZI pour m'avoir accueilli il y a bientôt cinq années dans la filière informatique du CNAM de Tours. Sa grande humanité et ses encouragements amicaux ont été un grand soutien pour moi, notamment dans les moments difficiles durant cette dernière année.

Enfin, une pensée toute particulière pour Maryline, qui bon gré mal gré, n'a cessé de m'encourager à donner le coup de rein final pour boucler ce mémoire et tourner cette page de ma vie...

Abréviations

BODACC : Bulletin officiel des annonces civiles et commerciales

DBA : database administrator, administrateur de bases de données

DSI : direction des systèmes d'information

RNCS : registre national du commerce et des sociétés

SI : système d'information

SIREN : système d'identification national des entreprises

SIRET : système d'identification national des établissements

Table des matières

1 INTRODUCTION.....	9
2 CONTEXTE DU PROJET ET DE L'ENTREPRISE.....	10
2.1 Information entreprise (première approche).....	10
2.2 Ciblage marketing (première approche).....	13
2.2.1 Concepts.....	13
2.2.2 Architecture globale.....	14
2.2.2.1 Bases de données.....	14
2.2.2.2 Moteur de ciblage.....	15
2.2.2.3 Application web.....	15
2.3 Structure du groupe.....	15
2.3.1 Organigrammes financiers.....	15
2.3.2 BPCE.....	17
2.3.3 Natixis.....	17
2.3.4 Coface.....	18
2.3.5 Coface Services.....	20
2.3.6 Kompass International.....	21
2.4 Métiers du « Poste client ».....	23
2.4.1 Concepts.....	23
2.4.2 Zoom sur l'information entreprise.....	25
2.4.3 Zoom sur l'offre marketing.....	29
2.5 Organisation DSI.....	35
2.5.1 Production France.....	36
2.5.2 Études informatiques.....	36
2.5.3 Cycle de vie du logiciel.....	38
2.5.3.1 Mode par composants : le standard Coface.....	38
2.5.3.2 Mode projet.....	42
2.5.4 SI « Information Entreprise ».....	43
2.5.4.1 Introduction.....	43
2.5.4.2 Architecture globale.....	43
2.5.4.3 Surveillance.....	45
2.5.4.4 Diffusion.....	46
2.5.4.5 Aspects techniques.....	47
3 ANALYSE DES BESOINS, SPÉCIFICATION.....	49
3.1 Définitions - concepts.....	49
3.2 Acteurs, processus, activités.....	52
3.2.1 Acteurs.....	52
3.2.2 Processus.....	53
3.3 Caractéristiques fonctionnelles	53
3.3.1 Responsabilités du moteur de ciblage	53
3.3.2 Principe général du moteur de ciblage	54
3.3.3 Typologie des données	56
3.3.4 Modalité des requêtes.....	60
3.3.4.1 Modalité des requêtes sur des champs codes.....	60
3.3.4.2 Modalité des requêtes sur des champs textes.....	60

3.3.4.3	Modalité des requêtes sur des champs de type « nombre ».....	60
3.3.4.4	Modalité des requêtes sur des champs de type « date ».....	60
3.3.4.5	Récapitulatif des modalités des requêtes.....	61
3.3.5	Combinaison de requêtes élémentaires.....	61
3.3.6	Pré-comptage.....	61
3.4	Typologie des requêtes.....	62
3.4.1	Combinaison avec périmètre importé – Fichier repoussoir ou d’inclusion.....	62
3.4.2	Présence des fonctions de dirigeants.....	63
3.4.3	Recherche avec critère de type texte.....	63
3.4.4	Recherche avec critères définis dynamiquement (ratio, formule).....	64
3.4.5	Recherche avec critères privés créés en dynamique.....	65
3.4.6	Séquence de requêtes (une démarche de ciblage).....	66
3.4.7	Liens financiers.....	67
3.4.8	Requêtes en parallèle - Jointures.....	67
3.4.9	Liste de résultats attendues du moteur.....	68
3.4.10	Tris à plat, tris croisés.....	69
3.4.11	Mise à jour des données et des index.....	69
3.4.12	Précisions sur les fonctions de calculs.....	69
3.4.12.1	Contexte.....	69
3.4.12.1.1	Ciblage à partir de critères quantitatifs.....	69
3.4.12.1.2	Organisation des données - Dimension temporelle.....	70
3.4.12.2	Opérations.....	71
3.4.12.2.1	Utilisation directe des valeurs.....	71
3.4.12.2.2	Utilisation d’opérations.....	71
3.4.12.2.3	Utilisation de variations.....	72
3.5	Autres besoins liés au moteur.....	72
3.5.1	Session utilisateur, sécurité d’accès aux données.....	72
3.5.2	Espace de travail utilisateur.....	73
3.5.3	Stratégie.....	73
3.5.4	Combinaison de stratégies.....	73
3.5.5	Surveillance, veille.....	74
3.5.6	Export des données.....	74
3.6	Performances attendues, charge attendue.....	74
3.7	Spécifications liées à l’exploitation et à l’administration du moteur.....	75
3.7.1	Disponibilité.....	75
3.7.2	Besoins de suivi d’activité du moteur.....	75
3.7.2.1	Indicateurs.....	75
3.7.2.2	Journalisation.....	75
3.7.2.3	Statistiques.....	75

4 CONCEPTION DU MOTEUR DE CIBLAGE.....76

4.1	Synthèse des besoins.....	76
4.2	Solution actuelle : Cofacibles.....	76
4.2.1	Présentation.....	77
4.2.2	Limites de la solution « tout Oracle ».....	81
4.3	Offres des éditeurs.....	82
4.4	Proposition pour une solution composite.....	84
4.4.1	Principe général.....	84
4.4.2	Principe de la représentation « bitmap ».....	84
4.4.3	Principe de l’index.....	85

4.4.4 Exemple de construction d'un index.....	86
4.4.5 Volumes.....	88
4.4.6 Avantage du bitmap.....	88
4.4.7 Flux.....	89
4.4.8 Structures de données en mémoire.....	90
4.4.9 Techniques particulières.....	90
4.4.9.1 De la position du bit à l'adresse de la clé.....	90
4.4.9.2 Comptage des bits à 1 dans un bitmap.....	90
4.4.10 Calcul de formules.....	91
4.4.11 Jointure.....	92
4.4.12 Parallélisme.....	92
5 CONCLUSION.....	95
6 ANNEXES.....	96
7 BIBLIOGRAPHIE.....	99

1 Introduction

Pour une entreprise, cibler ses futurs clients, c'est déjà se développer. Cependant, les moyens financiers pour la prospection commerciale, sont toujours limités. Il lui faut donc obtenir l'information la plus pertinente sur les prospects les plus prometteurs.

D'un autre côté, une société comme Coface Services a développé son activité autour de l'information qu'il est possible (et permis) de récolter sur les entreprises françaises. L'idée de fournir un outil qui réponde à la demande de « ciblage des prospects », a germé il y a des années chez Coface Services. Le projet informatique « marketing direct » est né ainsi.

Nous sommes entrés chez Coface Services il y a déjà plus de treize années et nous avons participé à plusieurs phases du développement du projet informatique « marketing direct ». Notre propos est de présenter une synthèse de notre travail sur un projet qui a suivi la vie de l'entreprise. Dès 2003 nous avons développé en tant que programmeur, certains aspects du « marketing financier ». A la fin de l'année 2005, nous avons développé le module « import » de l'application Cofacibles. En 2007-2008, nous avons pris part à l'analyse fonctionnelle approfondie du projet KIM-IS, défini en commun avec Kompass International. La fin d'année 2010, sera sans doute un nouveau départ pour les applications de marketing direct chez Coface Services. Dans un premier temps, nous y prenons part comme « consultant interne » pour apporter une expérience et, nous l'espérons, des pistes techniques intéressantes.

Dans une première partie, nous présentons le contexte du projet et de l'entreprise. Le travail que nous présentons est intimement lié au métier et à leurs besoins en terme de systèmes d'information. Ce travail n'est pas un projet seulement technique, mais demande une compréhension avancée du métier.

Ensuite, nous présentons plus précisément les fonctionnalités attendues pour un moteur de ciblage marketing qui permette à Coface Services de se démarquer de ses concurrents.

Enfin, dans une troisième partie, nous abordons les aspects techniques, d'abord en présentant la solution qui existe aujourd'hui, ensuite en proposant des pistes pour dépasser certaines limites actuelles.

2 Contexte du projet et de l'entreprise

Il serait difficile de présenter un seul aspect du système d'informations (SI) de Coface Services sans présenter de façon assez précise ses métiers, dont le SI est le support. Les métiers de Coface Services étant assez complexes, nous allons faire une première approche des concepts de façon à « planter le décor » pour la suite.

2.1 Information entreprise (première approche)

Concept d' « information entreprise »

Sous cette expression, sont regroupées toutes les informations qu'il est possible d'obtenir sur une entreprise, en tant qu'objet juridique ou acteur réel de la vie économique.

Informations officielles

En France, les entreprises ont des obligations légales de publier un certain nombre d'informations (sur leur statut, leur structure, leurs comptes), auprès des Tribunaux de Commerce, éventuellement par l'intermédiaire d'un centre de formalité des entreprises. Les informations à publier, qui deviennent donc des informations officielles, dépendent globalement de la taille de l'entreprise et de sa structure juridique.

Les informations sur toutes les entreprises françaises sont regroupées dans le « Registre National du Commerce et des Sociétés » dont l'Institut National de la Propriété Industrielle » (INPI) est le dépositaire légal. Ce registre comprend toutes les informations légales et financières sur plus de 3,5 millions d'entreprises immatriculées.

L'INPI a, plus globalement, pour mission légale de diffuser toute information nécessaire à la protection des innovations et de tenir le Registre national du commerce et des sociétés. L'institut diffuse ces ressources informationnelles dans ses bases de données en ligne et les met à disposition dans le cadre de licences, pour un usage interne ou de distribution.

Il existe une autre base de données officielle, la base SIRENE gérée par l'Institut National des Statistiques et des Études Économiques (INSEE). En 1973, un système national d'identification des entreprises et un répertoire national des entreprises et de leurs établissements ont été créés pour généraliser l'utilisation d'un identifiant unique dans toutes les relations entre l'administration et les entreprises. Ce répertoire regroupe, en plus des entreprises du RNCS, les artisans (non commerçants), les exploitations agricoles, les professions libérales, les établissements publics, etc.

L'existence de ces deux répertoires nous permet déjà d'introduire les concepts suivants :

- entreprise : entité juridique,
- établissement : site d'activité d'une entreprise.

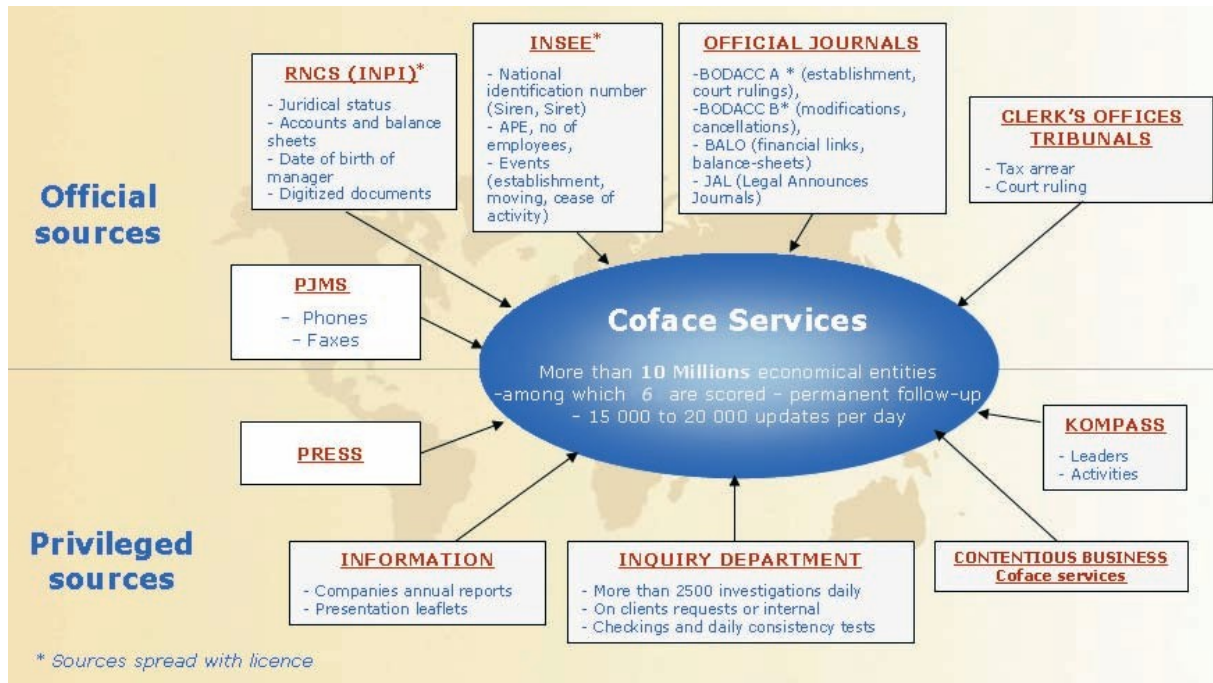
Une entreprise, qui a une existence juridique, dispose concrètement d'un ou plusieurs établissements « physiques » pour exercer son activité. Parmi ceux-ci, l'entreprise choisit celui qui est son établissement principal (son siège), et s'il y en a plusieurs, les autres sont qualifiés d'établissements secondaires.

Cette distinction est très importante pour la structure de nos bases de données; nous y reviendrons.

Informations privilégiées et privées

La loi française (comme la plupart des pays du monde ouverts à l'économie de marché)

permet la collecte d'informations, de renseignements sur les entreprises. Coface Services a développé une activité dans ce sens depuis bientôt 40 ans. Les renseignements obtenus auprès de diverses sources, dont les entreprises elles-mêmes, permettent d'enrichir considérablement et de rafraîchir les informations sans dépendre des échéances légales. Ces informations sont analysées, recoupées et constituent l'une des activités à forte plus-value pour Coface Services.



Informations légales durant la vie d'une entreprise

Le cycle de vie des entreprises est grosso modo similaire pour toutes les entreprises du monde, mais la législation n'apporte pas toujours une structuration aussi précise des informations que dans l'Union Européenne et en particulier en France.

Ainsi, la loi française impose aux entreprises de :

- s'immatriculer lors de leur création;
- signaler les modifications au RNCS :
 - dépôt d'actes,
 - modification d'enseigne,
 - Modification du Capital,
 - transfert de siège,
 - etc;
- signaler les radiations
 - mise en sommeil d'une personne physique ou morale,
 - radiation d'une personne physique ou morale,
 - fermeture d'un établissement secondaire,
 - transmission universelle du patrimoine,
 - etc;

Procédure collectives

Les entreprises étant soumises aux lois du marché, elles peuvent rencontrer des difficultés économiques et dans les cas les plus graves faire l'objet d'une « procédure collective » qui est une procédure judiciaire ouverte au Tribunal de Commerce, parmi lesquelles :

- la procédure de sauvegarde : elle permet à une société de se protéger des

- créanciers qui réclameraient le paiement de leur créance; le but est de permettre une réorganisation de l'entreprise et donc assurer sa pérennité;
- le redressement judiciaire qui intervient dans le cas plus grave de la situation de cessation de paiement, et dont le but est de déterminer dans quelle mesure l'entreprise peut survivre; un administrateur judiciaire est nommé pour la diriger;
 - la liquidation judiciaire : en résumé, la mort juridique de l'entreprise, intervient sur décision du tribunal de commerce qui constate d'une part la cessation de paiements (passif exigible supérieur à l'actif disponible), et l'impossibilité de son redressement.

Comptes annuels

Périodiquement, en général tous les 12 mois, une partie des entreprises, en fonction de leur statut et de leur taille (chiffre d'affaires en particulier) sont tenues de publier leurs Comptes Annuels, couramment connus sous le nom de « bilan ».

L'ordre de grandeur de l'ensemble des entreprises françaises assujetties à la publication des Comptes Annuels est de 900 000 chaque année.

Bases de données des entreprises de Coface Services

Coface Services a élaboré au fil des années un système d'informations dont le cœur est constitué par un ensemble de bases de données contenant de l'information entreprise.

Sans dévoiler le détail de leur structure, nous comprenons comment elles s'organisent :

- Données primaires de deux types, résultat d'un travail de recueil
 - informations qualitatives : actes de la vie d'une entreprise,
 - informations quantitatives : comptes annuels, éléments financiers.
- Informations dérivées à valeur ajoutée pour les métiers du « Poste client » (présentés plus loin)
 - solvabilité, montant du crédit conseillé, etc.

La vie des économies de marché repose, notamment, sur les échanges commerciaux entre entreprises commerciales et particuliers (le commerçant du coin par exemple) et sur les échanges entre entreprises (le B2B ou business to business des anglo-saxons); d'où l'intérêt pour une entreprise qu'elle soit cliente ou fournisseur, de disposer de ces informations le plus tôt possible pour, respectivement, son fournisseur (qui est alors leur créancier) ou son client (qui est son débiteur).

Exemples d'utilisation de l'information entreprise chez Coface Services

De la plus simple

Coface Services produit, en ligne notamment, une fiche entreprise basique (proche du « K-bis » officiel) qui donne les informations générales les plus récentes sur une entreprise. Nous en donnerons un exemple plus loin.

A la plus élaborée

Coface Services disposant des comptes annuels des entreprises, toute une analyse financière a été développée en interne pour exploiter ces informations et les diffuser à ses clients. En l'occurrence, une analyse financière précise doit permettre au Crédit Manager de déterminer le montant du crédit que la société peut accorder à son client. Elle est également utile au Directeur des achats dans sa recherche de fournisseurs pérennes, à l'investisseur dans la sélection d'entreprises performantes ... Cette analyse doit s'effectuer à partir d'une liasse fiscale complète, des comptes annuels avec l'annexe ou des comptes consolidés si l'entreprise fait partie d'un groupe. Compte tenu des options comptables, des

choix de la direction, de la présentation standard des états financiers qui ne se prête pas à l'analyse financière, il est indispensable, au préalable, de procéder à certains retraitements.

Dans ce premier paragraphe, nous avons brièvement présenté quelles notions se cachent sous l'expression « information entreprise » et qui deviennent en réalité tout un métier dont Coface Services a fait l'une de ses activités principales.

2.2 Ciblage marketing (première approche)

Le travail que nous exposons dans les parties suivantes se rapporte plus précisément à l'activité « ciblage marketing » de Coface Services. Avant d'entrer dans les détails, nous présentons dans le paragraphe suivant, une première approche de ce métier en relation avec le métier « information-entreprise » présenté ci-dessus.

2.2.1 Concepts

Ciblage marketing

Dans le cadre du B2B (échanges commerciaux entre entreprises), le ciblage marketing intéresse une entreprise qui cherche à étendre sa clientèle, sur un marché où elle est en position de fournisseur. Il s'agit pour elle de disposer d'un outil qui lui permette de déterminer une population d'entreprise, parmi celles qui sont présentes dans les bases de données de Coface Services, qui pourraient devenir ses clients. Le ciblage peut alors simplement se formuler sous la forme d'un filtre appliqué à la base de données dont nous disposons. Les critères du filtre sont choisis par l'entreprise utilisatrice de l'outil, ce qui lui permet de cibler finement ses futurs interlocuteurs commerciaux.

Lien avec l'information entreprise

Le « fond de base » de l'outil de ciblage marketing est simplement constitué des informations sur les entreprises recueillies dans le cadre des activités « information entreprise » de Coface Services. Cependant, ces informations sont retraitées et filtrées de façon à ne proposer que des informations utiles : par exemple seules les entreprises en bonne santé financière sont jugées « marketables », c'est-à-dire dont les coordonnées sont vendables, et sont proposées au choix des utilisateurs. L'information « santé d'une entreprise » qui est issue d'analyses liées au métier information entreprise, rend possible et pertinente l'activité « ciblage marketing ». L'« information marketing » ou « marketing direct » et information entreprise sont donc fortement liés.

Importance stratégique

Le ciblage marketing permet à une entreprise-fournisseur de définir sa future clientèle. Par ailleurs, Coface Services est en mesure de lui fournir un certain nombre d'informations sur la santé de ses futurs clients, au travers de produits plus élaborés contenant de l'analyse financière. Une entreprise utilisatrice du ciblage marketing peut donc « naturellement » comprendre tout l'intérêt qu'elle aurait à souscrire à des produits fournis par le métier information entreprise.

Dans une certaine mesure, l'activité de ciblage marketing peut être vue comme une « vitrine », un accès vers des produits et services plus élaborés fournis soit directement par Coface Services, soit par Coface. Nous présentons plus loin les différents métiers du « Poste client ».

Comptage

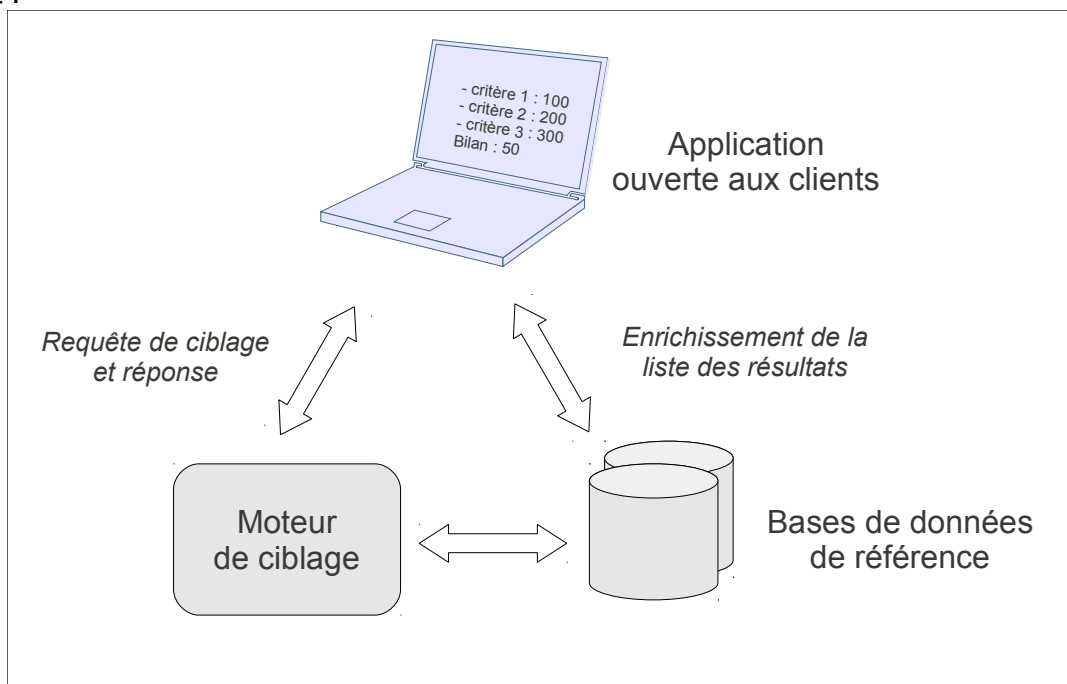
Le comptage consiste à déterminer le nombre d'entités dont les propriétés correspondent à celle qui sont spécifiées dans les critères du filtre.

Moteur de comptage et de ciblage

Le moteur de comptage et de ciblage est le « noyau » du système informatique qui permet à un utilisateur d'effectuer des requêtes de comptage et de ciblage. Ce noyau comprend le contenu (les données) et tous les mécanismes qui permettent de compter et de cibler. Il ne comprend pas les IHM (interfaces homme-machine).

2.2.2 Architecture globale

Une vue globale du système d'information du ciblage marketing est fournie par le schéma suivant :



La vision est très schématique et permet d'en voir les éléments essentiels. L'architecture plus détaillée sera présentée dans la partie technique de ce document.

2.2.2.1 Bases de données

Comme nous l'avons déjà signalé, les bases de données de référence alimentent le moteur de ciblage. Ces bases de données sont au cœur du métier information entreprise. Il y a environ 6,5 million d'entreprises enregistrées, pour 7,5 millions d'établissements. Chaque entité possède globalement plusieurs centaines d'attributs, organisés dans plusieurs tables. Les bases de données sont des instances Oracle. Les volumes sont conséquents, et nécessitent une bonne organisation de la base de donnée ainsi qu'un travail d'optimisation précis. nous y reviendrons.

Comme indiqué sur le schéma, il y a des flux de données (quotidien et hebdomadaires) entre les bases de références et le système de ciblage, de façon à tenir à jour les informations utiles au marketing.

De façon simple, un produit marketing est un fichier dont chaque enregistrement contient des informations utiles pour connaître et contacter une entreprise. Ces informations sont

issues de la base de données de référence.

2.2.2.2 Moteur de ciblage

Quelques mots sur son importance

- volumes
- performances
- sollicitation, charge, robustesse
- plusieurs milliers de requêtes par jour (pointe à 300 / heure)
- plusieurs centaines de fichiers livrés par jour

2.2.2.3 Application web

L'application web n'est que la partie visible du système. Elle se présente de prime abord sous la forme d'une page dans laquelle l'utilisateur peut construire son filtre (sa requête de ciblage en réalité) sur la population d'entreprises « marketables » disponibles. Voici une saisie d'écran de l'application Cofacibles (nom commercial de l'application de ciblage marketing développé par Coface Services) :



La suite de cette partie consiste en une mise en situation perspective du métier « ciblage marketing » dans les métiers et dans l'organisation du groupe.

2.3 Structure du groupe

Nous avons effectué une première approche du métier « information entreprise », et nous avons vu qu'il était stratégique pour le groupe auquel appartient la société Coface Services.

Dans ce paragraphe, nous présentons le groupe, ses différents métiers et nous expliciterons les liens qui existent entre eux.

2.3.1 Organigrammes financiers

L'organigramme financier simplifié suivant permet de montrer où se situe Coface Services

dans le groupe Natixis. Ces informations sont issues d'un produit développé par Coface Service pour ses clients.

Actionnaires

RCS 482 755 741 - NIC 00014	FRANCE
COFACE SERVICES	
2 actionnaire(s) direct(s)	
➤	
📁 COFACE SERVICES - RCS : 482 755 741	
└─	88,70 % - COFINPAR - RCS : 339 529 927
└─	99,99 % - COMPAGNIE FRANCAISE D ASSURANCE POUR LE COMMERCE EXTERIEUR - RCS : 552 069 791
└─	11,03 % - COMPAGNIE FRANCAISE D ASSURANCE POUR LE COMMERCE EXTERIEUR - RCS : 552 069 791
└─	100,00 % - COFACE HOLDING - RCS : 432 413 599
└─	100,00 % - NATIXIS - RCS : 542 044 524
└─	71,50 % - BPCE - RCS : 493 455 042
└─	27,40 % - PUBLIC - N° : DPU 000 714
└─	0,20 % - NATIXIS - RCS : 542 044 524
└─	0,09 % - SALARIES - N° : DSA 281 041
Nombre d'actionnaires visualisés : 9	

L'actionnaire principal de Coface Services est la société Coface, au travers d'une société de participation financières. Nous aurons l'occasion de développer les conséquences de cet actionnariat, notamment sur le système d'information.

Par ailleurs, Coface est elle même détenue par Natixis, dont l'actionnaire principal est Le groupe Banque Populaire Caisses d'Épargne (BPCE).

L'actionnariat de Coface Services est donc majoritairement constitué par des société dont l'activité concerne la banque de détail, de financement, ou l'assurance crédit. Nous verrons que les métiers de Coface Services sont fortement avec cette dernière.

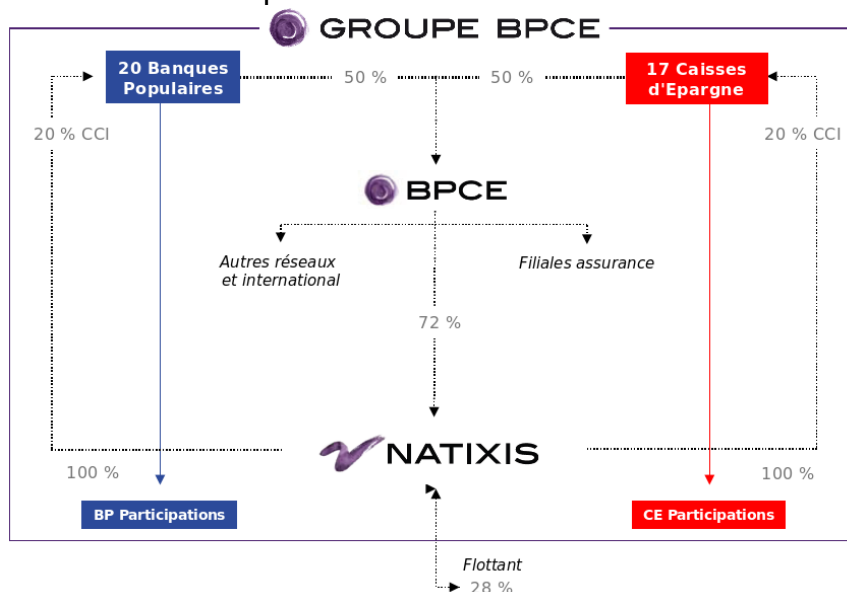
Participations

RCS 482 755 741 - NIC 00014	FRANCE
COFACE SERVICES	
7 participation(s) directe(s)	
COFACE SERVICES - RCS : 482 755 741	
100,00 % - CENTRE D ETUDES FINANCIERES - RCS : 351 875 158	
100,00 % - KOMPASS INTERNATIONAL NEUENSCHWANDER SA - RCS : 395 162 738	
100,00 % - LA LIBRAIRIE ELECTRONIQUE - RCS : 394 038 996	
100,00 % - OR INFORMATIQUE - RCS : 332 346 923	
83,00 % - EASYNUMBER COMPANY - N° : F01 758 440	
79,00 % - IJCOF - RCS : 518 528 435	
53,00 % - COFACE BELGIUM SERVICE HOLDING - N° : E01 701 770	
Nombre de participations visualisées : 7	

La société Coface Services possède un certain nombre de participations dont la plus intéressante, au regard de son activité marketing, est celle dans Kompass International. En effet, dans la partie sur l'analyse des besoins de la solution de ciblage marketing, nous auront l'occasion de mentionner l'activité marketing que Kompass International développe, et qui, dans une certaine mesure, est complémentaire de celle de Coface Services.

2.3.2 BPCE

La BPCE est une banque dont le cœur de métier est la « banque de détail ». L'importance de son réseau est stratégique pour la diffusion des produits et services du groupe auprès des clients, notamment des entreprises.



2.3.3 Natixis

Historique

Natixis est issue de la fusion des banques de financement Natexis (Banque Populaire) et

Ixis (Caisses d'Épargne) en 2006.

Chiffres clés (exercice 2009)

- 19 500 collaborateurs dans 68 pays;
- 12,7 Md€ de fond propres (tier one au sens Bâle II);
- 5,5 Md€ de produit net bancaire;
- 449 Md€ au total du bilan.

Métiers

Banque de Financement et d'Investissement (BFI)

Une banque d'investissement rassemble l'ensemble des activités de conseil, d'intermédiation et d'exécution qui concernent les opérations dites de « haut de bilan » (introduction en Bourse, émission de dette, fusion/acquisition) de grands clients (entreprises, investisseurs, mais aussi États...). Ces activités regroupent les opérations de finance d'entreprise (Corporate Finance), de marchés financiers (Global Capital Markets), et de opérations de financement structurés (Structured Finance) :

- Banque commerciale dont
 - Financements classiques
 - Conseil en fusions et acquisitions
 - Financement d'acquisitions et LBO (leverage buy out)
 - Gestion de flux
- Marchés de capitaux (investissement et gestion des risques) dont
 - Marchés actions : cash, dérivés, primaire
 - Marchés de taux, change, crédit et matières premières
- Financements structurés dont
 - Matières premières
 - Aéronautique, shipping et commerce international
 - Immobilier
 - Conseil en optimisation de bilan et transfert de risques
 - Titrisation.

La BFI de Natixis s'adresse donc à des clients-entreprises, investisseurs institutionnels et secteur public.

Épargne

Ce pôle regroupe trois métiers de gestion pour compte de tiers, la gestion d'actifs, l'assurance et la banque privée.

- la gestion d'actifs compte une vingtaine de sociétés de gestion d'actifs financiers et immobiliers présentes dans le monde entier;
- Natixis Assurances, conçoit et gère une offre complète de produits d'assurance vie individuelle et collective, d'assurance prévoyance et d'assurance dommages aux biens;
- la gestion privée assurée par Banque Privée 1818 et Natixis Private Banking International, offre des solutions patrimoniales et financières, en France et à l'étranger, adaptées aux besoins des grands investisseurs privés, notamment des entrepreneurs et des cadres dirigeants.

Services Financiers Spécialisés

Par l'intermédiaire des banques de proximité Natixis diffuse une gamme de produits et services, qui se décomposent en huit activités :

- affacturage (financement, gestion et sécurisation du Poste clients des entreprises), au travers de Natixis Factor;
- ingénierie sociale (développement de l'épargne salariale, de l'épargne retraite, de l'actionnariat salarié, de l'assurance collective et des titres spéciaux de paiement), au travers de Natixis Interépargne ;
- cautions et garanties (offre multi-marché de garanties et cautions);
- crédit-bail (financements mobilier et immobilier professionnels, LLD) ;
- crédit à la consommation (offres de crédit permanent et gestion des prêts personnels) ;
- financement du cinéma et de l'audiovisuel (crédits cash flow ou structurés, crédits moyen terme) ;
- paiements (traitements des opérations de paiement nationales, européennes et internationales, plate-forme de banque en ligne) ;
- titres (conservation de détail et de gestion privée).

2.3.4 Coface

Historique

- 1946 : création comme compagnie d'assurance spécialisée dans l'assurance-crédit à l'exportation. Elle gère ses propres produits ainsi que les garanties publiques;
- 1988 : début de la diversification métier : entre au capital de Coface SCRL (société d'information d'entreprise et de gestion de créances commerciales);
- 1992 : entame son développement international au Royaume-Uni et en Italie; création de CreditAlliance, un réseau mondial de sociétés d'assurances, d'information d'entreprises, de compagnies d'assurance et de société d'affacturage;
- 1994 : privatisation;
- 1996 : rachat de l'assureur-crédit allemand AK et autrichien OKV;
- 1999 : extension de l'offre d'affacturage en Allemagne, en créant Allgemeine Kredit Coface Finanz (AKCF);
- 2000 : lancement de la notation @rating : première notation d'entreprise mondiale et assurable;
- 2002 : Natexis Banques Populaires devient actionnaire majoritaire;
- 2004 : achat de ORT; devient, avec Coface SCRL et Coface ORT, le premier acteur français de l'information d'entreprise;
- 2005 : acquisition des assureurs-crédit sud africain CUAL et lituanien LEID; prise de participation majoritaire dans la société d'information israélienne BDI;
- 2006 : création de Coface Services (fusion de Coface SCRL et Coface ORT); acquisition d'une société américaine de recouvrement des créances (Newton & Associates); lancement de Coface Algérie; novembre : Coface devient filiale à 100% de Natixis, nouvelle banque d'affaires des Banques Populaires et des Caisses d'Epargne;
- 2007 : achat de Kompass France, sa succursale en Espagne et Kompass Belgique, via sa filiale Coface services; acquisition de Dafsaliens, le spécialiste français des liens financiers entre entreprises; la Commission Bancaire donne au Score @rating de Coface le statut d'ECAI (External Credit Assessment Institution) pour son activité de notateur en France; Coface et Creditreform lancent Easy Number, un système universel de recherche et d'identification des entreprises dans le monde entier à travers un identifiant unique;
- 2008 : création de Coface Holding qui regroupe l'ensemble des moyens du Pôle clients de Natixis, à travers Coface et Natixis Factor; lancement d'une notation mondiale sur la qualité de l'environnement des affaires; acquisition du leader de

- l'information d'entreprises en Egypte (Fiani Partners) et création de Coface Egypte; Extension de l'offre d'affacturage à la Chine, l'Australie, Israël, Tchéquie, Slovaquie et au Danemark ou Coface devient l'actionnaire majoritaire de Midt Factoring (75% du capital), le leader du secteur; lancement d'une notation d'entreprise sollicitée, alternative à celle des agences de notation traditionnelles;
- 2009 : malgré la crise, Coface continue de soutenir les entreprises en augmentant ses garanties de 14% en 2008; Coface lance la Charte Transparence qui permet aux entreprises d'avoir des informations sur leur notation et de donner des informations pour la faire évoluer; Coface s'implante dans deux nouveaux pays : Vietnam et Gabon.

Chiffres clés

- Présence dans 67 pays
- 1,5 Md€ de chiffre d'affaires
- 6 600 collaborateurs
- 135 000 clients dans le monde dont 45 % des 500 plus grands groupes mondiaux.

Métiers

Nous détaillons dans le paragraphe suivant les métiers du « Poste client », qui sont :

- Assurance-crédit : n° 3 mondial , 370 Md€ de créances garanties
- Affacturage : n° 6 mondial , 32 Md€ de créances financées
- Services
 - Notation/information d'entreprise : n° 6 mondial , 55 millions d'entreprises scorées;
 - Gestion et recouvrement de créances : n° 2 mondial, 1,5 million de dossiers gérés.

Coface propose également, en France, la gestion des garanties publiques à l'exportation pour le compte de l'État.

2.3.5 Coface Services

Historique

- 2006 : création de Coface Services, fusion de Coface ORT et de Coface SCRL
 - Coface Services obtient le marché public pour la saisie, la numérisation, l'hébergement et la diffusion des données et pièces officielles du RNCS, placé sous l'autorité de l'Institut National de la Propriété Industrielle (INPI);
 - achat de Triade (gestion de comptes clients).
 - lancement d'une nouvelle offre « rachat de créances ».
 - lancement du Pack@rating2.
- 2007 : achat de Kompass France, sa filiale en Espagne et de Kompass Belgique.
 - achat d'Idéalinfo – Dafsaliens (liens financiers).
- 2010 : création d'une joint venture avec Intrum Justitia – IJCOF, spécialisée sur le marché des créances civiles;

Historique Coface ORT

- 1973 : création de LOREA à Rocherarbon (près de Tours).
- 1977 : création de la 1ère base de données informatisée.
- 1982 : les bases de données sont consultables sur minitel.
- 1984 : LOREA devient OR Télématique.
- 1985 : Création d'une filiale à Pompey, OR Informatique, pour constituer un centre

- de saisie;
- 1988 : OR Télématique ouvre pour le compte de l'INPI l'accès au RNCS (Registre National du Commerce et des Sociétés) EURIDILE;
- 1994 : Création d'une filiale à Libourne, La Librairie Électronique, 1er centre de numérisation européen;
- 1996 : OR Télématique devient ORT;
- 2000 : ORT est achetée par le groupe Reuters;
- 2003 : Ouverture du service JURIPRO, site partenaire des professionnels du droit;
- 2004 : ORT est rachetée par Coface et devient Coface ORT;
- 2006 : création de Coface Services, fusion de Coface ORT et Coface SCRL.

Historique Coface SCRL

- 1893 : création de SCRL à Lyon, cabinet de recouvrement de créances.;
- 1950 : SCRL élargit son activité à l'information de solvabilité; SCRL invente la notion de « pré-contentieux » ; première implantation à Paris;
- 1955 : premières implantations en région, à Strasbourg et Marseille.;
- 1973 : création d'une filiale, Sogead, pour le recouvrement des créances civiles;
- 1978 : achat de SCRL par le groupe BIS;
- 1978 : informatisation des bases de données d'information.;
- 1988 : Coface entre dans le capital de SCRL à hauteur de 30 %, puis augmente sa participation à 70 % en 1992;
- 1989 : SCRL lance Diane, le premier outil d'analyse financière sur CD Rom;
- 1993 à 1996 : SCRL accélère sa croissance externe par l'acquisition de Piguet (information et recouvrement, depuis 1895), puis de Sprec en 1994 (recouvrement depuis 1964) et enfin de Bottin en 1996 (marketing direct, depuis 1796);
- 1996 : Coface devient actionnaire à 100 % du groupe SCRL;
- 1999 : création de Coface SCRL, fusion de SCRL avec ses filiales;
- 2001 : lancement du portail www.cofacerating.fr premier portail de veille concurrentielle et de gestion du risque client;
- 2003 : Coface SCRL obtient la certification ISO9001 version 2000, pour la fourniture d'informations économiques, commerciales et financières issues d'une base de données sur les entreprises françaises métropolitaines;
- 2006 : création de Coface Services, fusion de Coface Ort et Coface Scrl.

Chiffres clés

- 15 000 clients
- 730 collaborateurs en France
- 100 M€ de chiffre d'affaires.

Métiers

Coface Services est leader sur le marché français des services de « credit management » : notation, information d'entreprise et gestion de créances.

Elle permet à toutes les entreprises d'évaluer la situation financière de leurs partenaires commerciaux et leur capacité à honorer leurs engagements (information de solvabilité) ou de détecter des opportunités commerciales (information marketing). Elle propose une gamme de prestations de la gestion de factures (avant ou après échéance) au recouvrement (amiable ou judiciaire), en passant par le rachat de créances.

Ses équipes, spécialistes des métiers de l'information (collecte et analyse des données, diagnostic, statistiques sur les entreprises...), de la gestion de factures et du recouvrement de créances (analyse des risques débiteurs, procédures amiable et judiciaire...), alimentent également les autres métiers de Coface et produisent la notation

financière de Coface en France.

2.3.6 Kompass International

Historique

En juin 2008, Kompass et Kompass International fusionnent pour devenir Kompass International Neuenchwander SA, filiale à 100% de Coface Services.

Chiffres clés

- 250 salariés
- 33,9 M€ de chiffre d'affaires (2008);
- 6 millions de visiteurs par mois au niveau mondial, dont 1,7 pour la France (2009)

Métiers

Kompass International est le leader mondial de l'information marketing. Il propose aux entreprises une solution globale de développement de leur chiffre d'affaires à travers deux offres :

- la publicité qui permet à toute entreprise de présenter ses produits et services, ses décideurs et leurs emails ainsi que ses marques sur le site Kompass.com (accès libre) par le biais de rédactionnels, photos ou vidéos. À cette offre est couplé un service de réception d'appels d'offres privés et publics;
- l'information sous forme d'abonnements Internet, Extranet, DVD Rom et de fichiers marketing direct.

Le nouveau site Kompass.com rassemble une base de données exhaustive de 200 000 entreprises B2B en France et de 2,7 millions d'entreprises de 64 pays et leurs 4,2 millions de décideurs.

L'annuaire des entreprises Kompass.fr permet d'effectuer des recherches dans une base de données regroupant 2,3 millions d'établissements dans 64 pays.

- 2,7 millions d'entreprises B2B
- 23 millions de référence produits et services
- 750 000 marques dans plus de 60 pays.

Ces recherches peuvent être effectuées de 3 manières : par raison sociale, par marque et par activité grâce à une nomenclature unique de 57 000 rubriques.

Marketing Kompass / marketing Coface Services

Originellement, les outils Kompass servent à une entreprise pour trouver ses fournisseurs, la recherche la plus précise se faisant par un produit ou un service. Kompass s'est en effet historiquement spécialisée dans la collecte, le recensement et la classification de tous les produits et services offerts sur les marchés mondiaux. Ainsi, les produits et services de chaque entreprise-fournisseur sont répertoriés conformément à une nomenclature internationale homogène élaborée par Kompass.

Parallèlement, Kompass offre un « espace de visibilité » pour que chaque entreprise puisse rendre le plus visible possible ses offres de produits et services : encarts publicitaires, catalogues en ligne, etc. Kompass regroupe et classe le contenu des catalogues en les rendant disponibles à partir d'un seul point d'accès.

L'offre de recherche en ligne et les espaces publicitaires qui offrent une « visibilité » accrue aux entreprises, constitue l'activité marketing de Kompass. Cette activité constitue

la majorité de son chiffre d'affaires.

Le marketing direct tel que le propose Coface Services est complémentaire de celui de Kompass dans la mesure où les clients de Coface Services sont à la recherche de leur futurs clients, alors que ceux de Kompass sont à la recherche de leur futurs fournisseurs. Fonctionnellement et techniquement, il existe cependant des analogies notamment pour certains aspect du système d'information.

Kompass a par ailleurs élargi son offre vers l'information de solvabilité (Kompass Information Solvabilité) en bénéficiant de ses liens avec Coface Services.

Il apparaît donc clairement une forte complémentarité entre les entreprises du groupe, notamment sur le marché du marketing.

2.4 Métiers du « Poste client »

2.4.1 Concepts

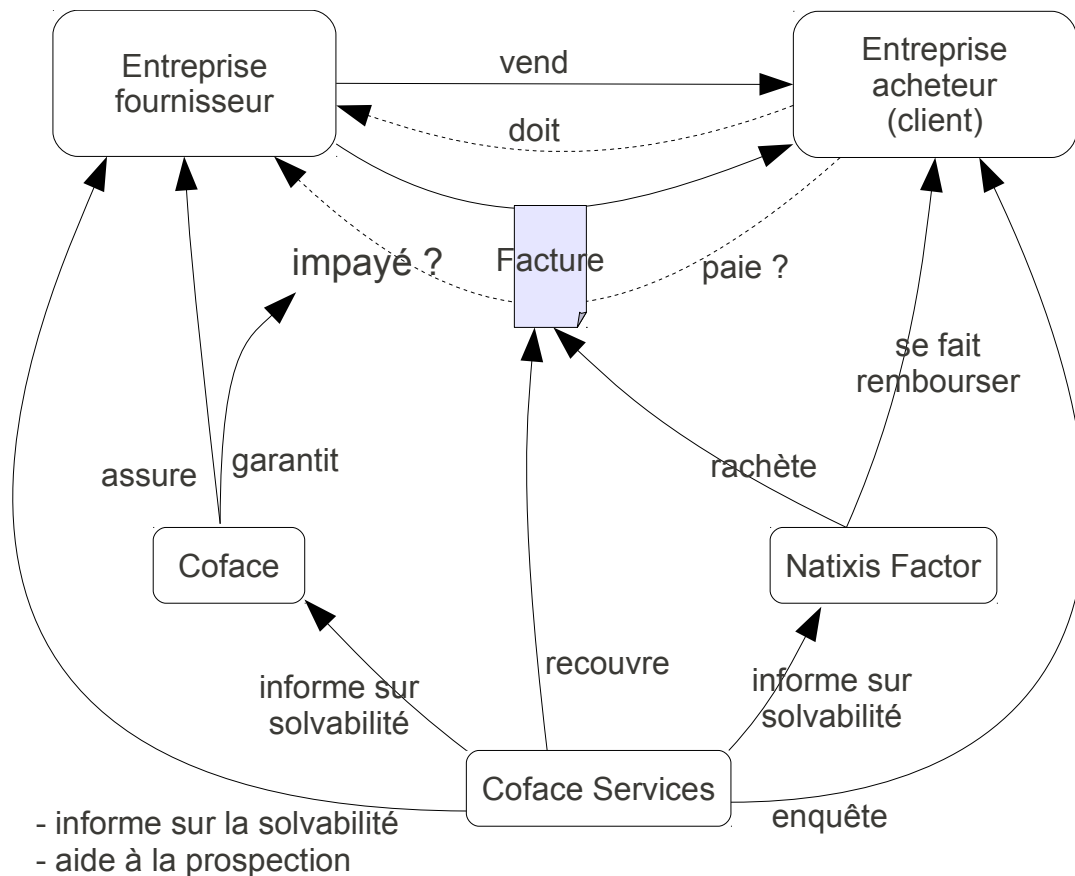
Les métiers du « Poste client » dont nous avons entrevu quelques aspect se comprennent aisément en expliquant leur fonctionnement autour du scénario basique qui met en jeu une entreprise (fournisseur) et son client (une autre entreprise).

Vendre sous-entend de vouloir être payé. Or, d'une part, l'entreprise fournisseur doit attendre pour être payée (la loi française, comme la plupart des pays, accorde un délai de paiement), d'autre part, il existe un risque de ne pas être payé.

Risque ou non, une entreprise doit donc gérer les créances qu'elle accorde, de fait, à ses clients.

D'un point de vue comptable, les créances aux clients apparaissent dans l'actif du bilan sous ce nom. Les postes (synonyme de case) du bilan correspondants sont donc les postes dédiés aux créances des clients; d'où le nom de métiers du « Poste client » pour désigner les services offerts aux entreprises par Natixis Factor, Coface, et Coface Services, de façon à leur faciliter la gestion des sommes engagées.

Le schéma suivant résume de façon simplifiée les échanges et les actions entre les acteurs principaux d'un échange commercial (l'entreprise et son client) et les sociétés de services aux entreprises qui interviennent pour aider à la gestion de ces échanges et à les sécuriser financièrement.



Ainsi la gestion du « Poste client » regroupe quatre métiers:

- l'assurance crédit,
- l'affacturage,
- la gestion de créance,
- l'information entreprise, dont le ciblage marketing.

L'assurance Crédit est le cœur de métier et le métier historique de Coface. Coface assure ses clients (assuré) contre le risque d'impayé. En cas d'impayé réel, Coface dédommage son assuré.

L'affacturage ou rachat de facture est géré par la société Natixis Factor. Les factures sont des titres négociables, car ce sont des effets de commerce. Natixis Factor rachète les factures de ses clients de façon à leur permettre de disposer de liquidités immédiatement et d'éviter un impayé à court ou moyen terme en cas de défaillance de l'entreprise débitrice.

Natixis Factor, Coface et Coface Services proposent de la gestion de créances qui est une aide au recouvrement des factures. Ce service permet aux entreprises de déléguer à Coface Service la gestion de leur factures en attente de paiement. Les entreprises n'ont plus besoin d'avoir en interne un service dédié. Parallèlement, Coface Services collecte les renseignements sur les entreprises de façon à déterminer leur solvabilité le plus précisément possible.

En effet, l'information de solvabilité est cruciale pour les autres métiers du « Poste client »: Natixis Factor n'achète des factures qu'à condition d'avoir une chance d'être remboursée;

Coface n'assure que si le risque d'impayé est faible, et en tout cas, jamais quand la défaillance est certaine !

L'efficacité et la viabilité de ces métiers reposent donc sur une information fiable et récente. D'où l'importance de l'activité de renseignement de Coface Services et le nombre important d'échanges entre Coface Services et les autres société du pôle « Poste client ».

Les anglo-saxons regroupent ces activités sous l'expression de « Trade receivables management ». Des problématiques similaires (aux normes législatives et réglementaires près) existent dans la plupart des pays du monde. Ceci explique la stratégie de développement de Natixis Factor et de Coface à l'international.

Structure d'un bilan

Une façon de distinguer les métiers du « Poste client » de ceux de la banque est de regarder la structure d'un bilan (ou comptes annuels):

Actif	Passif		
Emplois de ressources / biens	Origines de ressources / dettes		
Immobilisations	Capitaux propres	Capital	Vis à vis des propriétaires
Stocks		Bénéfices	
Créances clients	Dettes		Vis à vis des tiers
Disponibilités			

Les « créances aux clients » apparaissent dans l'actif du bilan et correspondent bien à des sommes engagées dans l'activité d'une l'entreprise. Ces sommes sont « bloquées » à court (moins d'un an) ou moyen terme (de 1 à 2 ans) et pendant ce temps, ne participent pas directement à la production de l'entreprise. D'où l'intérêt de les recouvrer.

Les « dettes » (à court, moyen ou long terme) apparaissent au passif du bilan et correspondent aux sommes empruntées par une entreprise à des tiers, parmi lesquelles bien souvent des banques.

Ce petit schéma synthétique montre que ces deux familles de métiers liés à la finance d'entreprise, ne s'adressent pas en fait aux mêmes situations et ne concourent pas de la même façon au financement de l'activité d'une entreprise. Ces activités sont complémentaires. C'est ainsi que nous comprenons la politique d'intégration des métiers du groupe Natixis.

Offre Coface pour la gestion du Poste clients

Coface propose de gérer l'ensemble du portefeuille d'une entreprise ou d'une multinationale, qu'il soit composé d'une dizaine ou de plusieurs milliers de clients. Son offre en assurance-crédit s'articule autour du contrat phare Globalliance. Celle en affacturage autour de Globalliance Finance, un contrat qui, en partenariat avec Natixis Factor, propose aux entreprises de toutes tailles un service complet intégrant la garantie, la gestion et le financement des créances commerciales :

- Globalliance Finance Intégrale : financement, protection et externalisation de la

- gestion du Poste clients,
- Globalliance Finance Liberté : formule plus souple dite en « gestion déléguée », l'entreprise conservant tout ou partie de la gestion de son Poste clients,
- Globalliance Finance Elite : le financement de la balance clients.

2.4.2 Zoom sur l'information entreprise

Dans le groupe Natixis, Coface Services a donc pour spécificité de fournir de l'information sur les entreprises, utilisée ensuite par les autres métiers en interne, ou par les services financiers des entreprises qui en dispose.

Nous avons déjà évoqué l'activité de renseignement et d'enrichissement des bases de données. Nous avons aussi évoqué l'activité de ciblage marketing. Il existe une activité supplémentaire, un peu différente du renseignement proprement dit, mais essentielle : permettre l'identification d'une entité (entreprise ou établissement) dans la vaste base de données, sans en connaître nécessairement la clé, qui est l'accès naturel le plus direct.

Identification

Aujourd'hui, l'outil dédié se nomme Easy. Nous avons évoqué plus haut, la création du « easy number » en 2007, comme le fruit d'une collaboration entre Coface et la société allemande Créditreform qui a donné naissance à la société Easy Number Company (société du groupe Coface). Cet outil est à usage externe (pour les clients du groupe) ou interne, pour les processus de traitement qui ont besoin d'identifier une entreprise ou un établissement. Cet outil contient aujourd'hui plus de 66 millions d'entités du monde entier.

Cet outil a été co-développé par les équipes de Coface et de Coface Services, sur la base d'un outil originellement développé par Coface et en bénéficiant de l'expertise de Coface Services dans le domaine de l'identification. En effet, depuis que des bases de données entreprises existent, il est indispensable de permettre à un utilisateur interne (analyste des renseignements commerciaux) ou externe (client) de retrouver rapidement dans la base de données, l'entreprise qui l'intéresse.

La DSI de Coface Services a donc développé très tôt des outils d'identification en apprenant à maîtriser différentes technologies, dont « Search Server » éditée initialement par la société canadienne Fulcrum, qui après plusieurs rachat est actuellement dans le giron de Microsoft.

La maîtrise de ces technologies est essentielle pour nos métiers dans la mesure où, par exemple, un analyste de crédit a besoin de consulter de façon interactive les informations sur une entreprise, ne connaissant que son nom et sa commune, ou son nom et le nom d'un dirigeant.

Rapprochement

A partir de l'identification proprement dite, se sont greffées d'autres activités à valeur ajoutée, comme le rapprochement / enrichissement de fichier.

Coface Services reçoit un fichier très incomplet d'un client (le SIREN peut manquer, ou l'adresse peut être incomplète) et se charge de le rectifier et de l'enrichir des informations de solvabilité par exemple. La problématique du rapprochement rejoint celle de l'identification, car dans chaque cas, il s'agit de retrouver les informations d'une ou de plusieurs entreprises en ne disposant que d'informations partielles.

Produits d'information entreprise

Exemple des prestations disponibles sur la société Coface Services, après la phase

d'identification de l'entreprise par son nom :

Recherche par les coordonnées Recherche par dirigeants Recherche libre

Vous pouvez identifier une entreprise par sa dénomination, son sigle, son enseigne, son RCS, son téléphone, le nom d'un dirigeant ou actionnaire, son domaine internet. Vous pouvez identifier une association loi 1901 par son titre.

Coface Services

OPTIONS

Dénomination, enseigne et sigle Contient le(s) terme(s) saisi(s) A l'identique

Région ou département France entière

Effacer Valider

La fiche d'identification de l'entreprise s'affiche :

RCS **482 755 741**- NIC **00014** **FRANCE**

COFACE SERVICES

Ets Siège :
1 Rue de l'Union
92500 RUEIL MALMAISON (HAUTS-DE-SEINE)
Tél. 01 47 52 40 50 - Fax 01 47 52 41 41

Forme juridique	5599 - SA à conseil d'administration
Date de création	06 juin 2005
Capital	20 294 265 Euros
N° de gestion	2005B03698
Greffe	Nanterre
Activité	8291Z - Activités des agences de recouvrement de factures et des sociétés d'information financière sur la clientèle
Nombre d'établissements	15

Votre référence dossier

Prestations disponibles

La liste des prestations proposées parmi lesquelles, on reconnaît des produits d'informations générales et des produits plus élaborés d'analyse financière ou de renseignements commerciaux (enquêtes) personnalisées.

COFACE SERVICES

Pour visualiser, cliquer sur la prestation de votre choix :

Format ⁽¹⁾Ajouter au panier⁽³⁾

Rapports de crédit actualisés au 12/08/2010

Classement du risque ORT C1	Spécimen	RTF	<input type="checkbox"/>
Cotation du risque ORT C2	Spécimen	RTF	<input type="checkbox"/>
Cotation du risque ORT C3	Spécimen	RTF	<input type="checkbox"/>

Enquêtes en différé

Avis complémentaire EN1	Spécimen		
Enquête de situation EN2	Spécimen		
Rapport approfondi EN3	Spécimen		

Contrôle

2 Actionnaire(s) direct(s)	Spécimen	RTF	<input type="checkbox"/>
7 Participation(s) directe(s)	Spécimen	RTF	<input type="checkbox"/>
Liens financiers (actionnaires et participations)	Spécimen	RTF	<input type="checkbox"/>
10 Mandataire(s) social(aux)	Spécimen	RTF	<input type="checkbox"/>
15 Etablissement(s)	Spécimen	RTF	<input type="checkbox"/>

Analyse financière

Analyse synthétique (grandes masses, principaux ratios)	Spécimen	RTF	<input type="checkbox"/>
Analyse des performances (chiffres clés, scores)	Spécimen	RTF	<input type="checkbox"/>
Analyse approfondie (grandes masses, tous ratios, flux)	Spécimen	RTF	<input type="checkbox"/>

Bilans

31/12/2009 (12 mois) - Sociaux - publiés en Euros	Spécimen	RTF	<input type="checkbox"/>
31/12/2008 (12 mois) - Sociaux - publiés en Euros	Spécimen	RTF	<input type="checkbox"/>
31/12/2007 (12 mois) - Sociaux - publiés en Euros	Spécimen	RTF	<input type="checkbox"/>
31/12/2006 (12 mois) - Sociaux - publiés en Euros	Spécimen	RTF	<input type="checkbox"/>
31/12/2005 (7 mois) - Sociaux - publiés en Franc français	Spécimen	RTF	<input type="checkbox"/>

Identité de l'entreprise

Identité de l'entreprise	Spécimen	RTF	<input type="checkbox"/>
--------------------------	----------	-----	--------------------------

Sources officielles

Copies des pièces (documents numérisés) déposées au Registre du Commerce et des sociétés⁽²⁾

- Comptes annuels
- Actes et Statuts

18 événement(s) BODACC	Spécimen		
------------------------	----------	--	--

Ajouter au panier⁽³⁾

Nous présentons en annexe un produit d'analyse financière (assez volumineux), pour montrer la richesse des informations disponibles.

Notation @rating

La notation @rating est une note qui synthétise la solvabilité d'une entreprise et son risque de défaillance de paiement. Le Score @rating est une note de risque correspondant à un découpage en tranche de probabilité de défaut .

	Risque très élevé						Risque moyennement élevé				Risque faible									
Score @rating/20	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Score @rating/10	1		2		3		4		5		6		7		8		9		10	
Taux de défaut (%) à 1 an	25%		10%		4%		2%		1,3%		0,7%		0,4%		0,15%		0,05%		0%	

La notation est d'abord calculée de façon automatique grâce à un algorithme basé sur les concepts de l'économétrie et mis au point par une équipe interne spécialisée. La notation automatique peut être reprise et modifiée par un analyste financier (interne) sur la base d'éléments particuliers, qui n'entre pas dans l'algorithme ou qui correspondent à une situation bien spécifique. a titre d'illustration, voici la répartition des 6 millions d'entreprises françaises selon leur degré de solvabilité (données pour 2008) :

Distribution du score par classe de note											
Note du score	1	2	3	4	5	6	7	8	9	10	Total
Score automatique	2,5%	7,9%	10%	21%	19%	19%	15%	4,0%	0,31%	0,04%	100%
Score@rating	2,1%	6,0%	8%	26%	19%	19%	15%	4,0%	0,35%	0,05%	100%

En parallèle du score, un « encours conseillé » est évalué : il correspond au montant maximum de la « créance client » qui devrait être accordé à une entreprise (cliente) par ses fournisseurs. Les entreprises-fournisseurs, clientes de Coface Services viennent donc consulter cet avis avant d'engager une transaction commerciale avec l'entreprise ainsi « scorée ».

L'Avis de crédit @rating (ou l'encours maximum conseillé) est indiqué comme suit :

Niveau	X	NR	R	@	@@	@@@	@@@@
Limite de crédit conseillé standard*	Aucun	7.000 euros	10.000 euros	20.000 euros	50.000 euros	100.000 euros	Plus de 100.000 euros *

Coface, avec son Score @rating, a reçu de la Commission bancaire le statut d'ECAI (External Credit Assessment Institution), qui est une reconnaissance officielle de ses notes de solvabilité des entreprises. Cet agrément, sollicité par Coface et quelques autres institutions notant les risques d'entreprise, est le premier délivré par la Commission bancaire, conformément à la réglementation Bâle II. Les normes Bâle II (le Nouvel Accord de Bâle) constituent un dispositif de prudence, destiné à mieux appréhender les risques bancaires et principalement le risque de crédit et les exigences en fonds propres des organismes financiers. Ces directives ont été préparées depuis 1988 (Bâle I) par le Comité de Bâle, sous l'égide de la « banque centrale des banques centrales » : la Banque des règlements internationaux et mis en place de 2006 à 2008.

2.4.3 Zoom sur l'offre marketing

L'offre marketing de Coface Services s'appuie sur plusieurs prestations:

- offre « on line »
 - outil Cofacibles
 - outil Astrée, développé en partenariat avec la société Bureau Van Dijk
- offre « off line »

- fourniture de fichier à façon, par une équipe dédiée qui utilise un outil spécifique développé par Coface Services aussi, mais basé sur le même contenu que Cofacibles:
- divers CD et DVD contenant une base de données marketing et un outil pour les exploiter, remis à jour périodiquement; cette formule est évidemment en perte de vitesse, à cause des commodités apportées par l'Internet.

Cofacibles

L'application Cofacibles, au développement de laquelle nous avons contribué, trouve notamment son origine dans l'application « Marketing direct » développée par ORT (qui a fusionné avec SCRL pour former Coface Services) il y a déjà plusieurs années. Cette application a évolué pour devenir Cofacibles et s'enrichir des besoins apportés par le nouveau périmètre du groupe.

L'activité « ciblage marketing » est de nouveau en train d'évoluer chez Coface Services, notamment pour rebondir à la suite de la forte baisse de chiffre d'affaires constaté sur cette activité, dont la crise financière née en 2008 est en grande partie responsable. Ce rapport s'inscrit d'ailleurs dans la perspective de la recherche de solutions aux problématiques posées par les besoins exprimés par le service commercial de Coface.

Le fond documentaire de Cofacibles est constitué des informations générales sur 6,5 millions d'établissements (rattachés à 5,5 millions d'entreprises) et des données financières sur trois années d'environ 900 000 entreprises qui publient leur comptes annuels.

Nous présentons ci après plusieurs saisies d'écran de l'application Cofacibles de façon à donner une idée plus concrète du ciblage marketing.

Accueil

The screenshot shows the Cofacibles application interface. At the top left, the logo 'cofacibles' is displayed with 'version 2.0' below it. A navigation menu on the left lists various search criteria under 'Rubriques' and 'Stratégies de recherche'. The main content area is titled 'INFORMATION MARKETING' and contains the following text:

Identifiez vos prospects à partir de notre base de données exhaustive et mise à jour en temps réel : plus de 6 000 000 entreprises en activité.

A l'aide des critères de sélection les plus fins, évaluez gratuitement votre potentiel commercial et commandez les fiches prospects adaptées à vos campagnes de marketing direct B to B.

Pour définir votre cible, il vous suffit de sélectionner dans le menu les critères de recherche de votre choix.

At the bottom of the main content area, the logos for 'coface' and 'coface SERVICES' are displayed.

Choix des critères de ciblage

Un critère qui porte sur l'activité est choisi et le comptage est affiché instantanément :

cofacibles version 2.0 Contact | Mon compte

Accès aux Rubriques

- Accueil
- Nouvelle recherche
- > Ancienneté
- > Appartenance groupe
- > Postes du bilan et ratios
- > Effectif
- > Décideurs
- > Forme juridique
- > Nombre d'établissements
- > Risque de défaillance
- > Secteur d'activité
- > Situation géographique
- > Téléphone / Télécopie
- > Tranches de C.A.
- > Stratégies de recherche
- > Importation de fichier

1 ————— 2 ————— 3
Stratégie ————— **Prestations** ————— **Commande**

Résultat du comptage : 419 entreprises 🖨️ ?

SECTEURS D'ACTIVITES

21 - Industrie pharmaceutique

TYPE D'ADRESSES

Sièges uniquement
 Sièges et établissements secondaires

Un autre critère portant sur la localisation est choisi :

Sélection cartographique

Sélection par Départements -
 Régions Départements

?

Votre sélection

CENTRE
 37 - INDRE ET LOIRE
 41 - LOIR ET CHER
 45 - LOIRET

Le comptage combiné s'affiche instantanément :

① ————— ② ————— ③
 Stratégie Prestations Commande

Résultat du comptage : 15 entreprises 🖨️ ?

👁️ Visualiser 📊 Répartir ➔ Etape ②

SITUATION GEOGRAPHIQUE Modifier

Département(s)

37 - INDRE ET LOIRE

41 - LOIR ET CHER

45 - LOIRET

SECTEURS D'ACTIVITES Modifier

📁 21 - Industrie pharmaceutique

TYPE D'ADRESSES

Sièges uniquement Sièges et établissements secondaires

👁️ Visualiser 📊 Répartir ➔ Etape ②

Une demande de visualisation affiche :

Extrait de votre sélection ?

Résultat du comptage : 15 entreprises	
Nom/Dénomination	Code postal
ORGAPHARM	45300
INDENA SAS	37000
PFIZER PGM	37530
INNOTHERA CHOUZY	41150
LES LABORATOIRES SERVIER INDUSTRIE	45520
RECIPHARM MONTS	37260
AGRO-BIO	45240
LABORATOIRES CHEMINEAU	37210
FAMAR FRANCE	45000
LABORATOIRES 3M SANTE	45300
FAMAR ORLEANS	45100
NOVEAL PHARMA	37260
ETCHELL	37210
DELPHARM TOURS	37170
LABORATOIRES PHARMACEUTIQUES POIRIER	37230

⬆️

Une demande de répartition passe d'abord par le choix des critères de répartition :

REPARTITION

1 ^{er} Critère	2 ^{ème} Critère
<input type="radio"/> Aucun	<input type="radio"/> Aucun
<input type="radio"/> Code postal	<input type="radio"/> Code postal
<input type="radio"/> Région	<input type="radio"/> Région
<input checked="" type="radio"/> Département	<input type="radio"/> Département
<input type="radio"/> Arrondissement	<input type="radio"/> Arrondissement
<input type="radio"/> Canton	<input type="radio"/> Canton
<input type="radio"/> Commune	<input type="radio"/> Commune
<input type="radio"/> Tranche de Chiffre d'Affaires	<input checked="" type="radio"/> Tranche de Chiffre d'Affaires
<input type="radio"/> Tranche de Chiffre d'Affaires export	<input type="radio"/> Tranche de Chiffre d'Affaires export
<input type="radio"/> Effectif de l'entreprise	<input type="radio"/> Effectif de l'entreprise
<input type="radio"/> Effectif des sites	<input type="radio"/> Effectif des sites
<input type="radio"/> Catégorie juridique (Code à 2 car.)	<input type="radio"/> Catégorie juridique (Code à 2 car.)
<input type="radio"/> Nombre d'établissements	<input type="radio"/> Nombre d'établissements

DECIDEURS

<input type="radio"/> Famille(s) de décideurs	
<input type="radio"/> Fonction(s) décideur(s)	

Le résultat de la répartition s'affiche :

Répartition des résultats



Résultat du comptage : 15 entreprises

Répartition de votre résultat par DEPARTEMENT / TRANCHE DE CHIFFRE D'AFFAIRES		
37 - Indre-et-Loire	8	<input checked="" type="checkbox"/>
Indéterminé	1	
Moins de 75 000 Euros	1	
De 500 000 à 1,5 millions d'Euros	1	
De 15 à 75 millions d'Euros	4	
Plus de 75 millions d'Euros	1	
41 - Loir-et-Cher	1	<input checked="" type="checkbox"/>
De 15 à 75 millions d'Euros	1	
45 - Loiret	6	<input checked="" type="checkbox"/>
De 1,5 à 3 millions d'Euros	1	
De 15 à 75 millions d'Euros	2	
Plus de 75 millions d'Euros	3	

Résultat du comptage : 15 entreprises



L'utilisateur peut revenir à tout moment sur les critères de sa sélection (ou ciblage). une fois satisfait, il passe commande du fichier marketing enrichi :

Détail de la prestation ?	
Quantité d'adresses	15
Fiche prospect sélectionnée	FP1
Mode d'expédition	<input checked="" type="radio"/> Download (vous recevrez par e-mail l'adresse url de téléchargement) <input type="radio"/> E-mail (vous recevrez par e-mail votre ou vos fichiers en pièce jointe)
Format du fichier	csv (séparateur ;)
Taille approximative du fichier	2.25 Ko

Tarifs	
TOTAL H.T.	Euros
T.V.A. 19.60 %	3.90
TOTAL T.T.C.	0.76
Mode de règlement :	
par débit de notre compte Coface Services numéro 290947	

Coordonnées d'expédition		
A l'attention de	Nom	Prénom
Monsieur <input type="button" value="v"/>	<input type="text"/>	<input type="text"/>
Société		
<input type="text" value="COFACE SERVICES COFACE SERVICES"/>		
Adresse		
<input type="text" value="CHATEAU DE SENS"/>		
Code postal	Ville	Pays
<input type="text" value="37210"/>	<input type="text" value="ROCHECORBON"/>	<input type="button" value="v"/>
Téléphone	E-mail	
<input type="text"/>	<input type="text"/>	

Après avoir choisi quelle prestation il souhaite :

Fiche Prospect 1	Fiche Prospect 2	Fiche Prospect 3	Fiche Prospect 4
Idéale pour vos campagnes de mailing.			
Tarif ?			<input type="button" value="Ajouter les contacts de votre choix"/>
Prix unitaire de la Fiche Prospect 1	Tarif HT	0.26 €	
Votre fichier de 15 adresses		3.90 €	<input type="button" value="→ Etape 3"/>
Contenu de la Fiche Prospect 1			
Dénomination (ligne 1)			
Dénomination (ligne 2)			
Adresse1			
Adresse2			
Adresse3			
Code postal			
Localité de destination			
Code Cedex et Libellé Cedex			
Code Insee Commune			
Numéro RCS			
Numéro NIC			
Enseigne			
Sigle			
Source d'information			
Civilité du dirigeant			
Prénom du dirigeant			
Nom du dirigeant			
Code Fonction du dirigeant			
Fonction du dirigeant			
Contacts au choix			

Les fiches prospects forment une gamme dont le contenu est de plus en plus riche.

Chez Coface Services, l'information entreprise regroupe de nombreuses activités, au delà de la simple compilation d'un registre des entreprises.

2.5 Organisation DSI

Coface est actionnaire de Coface Services à 100%. Les organes de décision au plus haut niveau des deux entreprises contiennent des personnes issues des deux sociétés. Par ailleurs, il y a des imbrications, des synergies évidentes entre les activités des entreprises. Un travail commun permet aussi une mutualisation des coûts et une meilleure poursuite de la stratégie commune. En particulier, dans des métiers où le traitement de l'information prédomine, la construction du système d'informations ne peut se faire qu'en cohérence. Ainsi, les Directions des Systèmes d'Information (DSI) ont mis en place une organisation qui permet le partage de décisions et la mise en commun des moyens.

Cela se traduit par la rédaction commune par les directions, du cahier des objectifs annuels, le partage des efforts pour obtenir et maintenir la certification ISO 9001 et une démarche CMMI (Capability Maturity Model Integration) conjointe.

La politique d'assurance qualité, le plan documentaire sont aussi mis en commun.

Les caractéristiques principales de la politique l'assurance qualité sont résumées dans le tableau suivant :

Characteristics	Generic levels	Interpretation
Maintainability	3	The effort needed to make specified modifications. (A good level of maintainability is normally reached by applying the CARIBOU framework and tools)
Functionality	2	A set of functions that met the customer requirements . The functions are those that satisfy stated or implied needs. The functions are qualified under the following features: suitability, accuracy interoperability, compliance with quality standards
Security	2	What may be done to protect the data and not to create infrastructure vulnerabilities . Security is a critical part of the functionality requirements and is evaluated separately.
Reliability	2	Capability of the software to maintain its level of performance under stated conditions for a stated period of time. This capability is characterised by the maturity (default frequency) fault tolerance recoverability.
Usability	1	The effort needed for use and on the individual assessment of such use by a stated or implied set of users. This effort is assessed thru the understandability; learnability and operability of the solution.
Efficiency	1	The relationship between the level of performance of the solution and the amount of resources used, under stated conditions. (time behaviour resource behaviour)
Portability	0	The ability of the solution to be transferred from one environment to another.

La DSI de Coface Services (comme celle de Coface) est divisée en deux grands services : production et études (plus un certain nombre d'autres services dont nous ne parlerons pas). Le service production a pour rôles de définir, d'installer, de maintenir, d'exploiter et de faire évoluer l'infrastructure (physique et logicielle). Le service études a lui pour rôles de concevoir, de réaliser, de maintenir les solutions logicielles créées et adaptées en réponse aux besoin en systèmes d'informations, formulés par la maîtrise d'ouvrage.

Les services des DSI de Coface et Coface Services sont répartis sur plusieurs sites physiques :

- La Défense (Coface)
- Nanterre (Coface)
- Tours (Coface Services)
- Lyon (Coface Services)
- Libourne (Coface Services)
- Rueil-Malmaison (Coface Services).

2.5.1 Production France

Le service « Production France », est un service commun aux DSI Coface et Coface Services. Les équipes intègrent en fait des personnes des deux sociétés. Les infrastructures sont partagées :

- machines : serveurs Hewlett-Packard, Fujitsu, SUN;
- réseaux, gestion de la sécurité;
- stockage des données, les sauvegardes
- logiciels pour les serveurs : Oracle, Weblogic, Tuxedo
- etc.

Les équipes disposent d'outils communs notamment pour ce qui est des outils de production organisés dans le cadre d'une démarche ITIL (Information Technology Infrastructure Library).

L'architecture standard préconisée par le groupe est de mettre les serveurs de bases de données sur des serveurs HP (avec des processeurs Itanium), et les serveurs applicatifs sur des machines sous Linux RedHat (avec des processeurs x86-64). Les données sont stockées sur des baies de disques montées en RAID5 (redundant array of inexpensive disk, mode 5), et répliquées.

Une architecture standard est indispensable au vu des nombreuses acquisitions que Coface a effectuées ces dernières années. Les systèmes d'information des entreprises acquises sont naturellement hétérogènes. De plus, le coût de maintenance de certains systèmes propriétaires un peu anciens (« mainframe » IBM par exemple), s'avère lourd à supporter à long terme.

2.5.2 Études informatiques

Présentation générale

A la différence du service « Production France », les équipes « Études informatiques » sont distinctes : chaque DSI gère ses équipes et ses projets. Cela permet par exemple aux équipes étude de rester plus proches de leurs maîtrises d'ouvrages (MOA) respectives : l'expertise du métier information entreprise est en particulier localisée chez Coface Services.

En revanche, les outils, les méthodes et les pratiques tendent vers un standard commun. Les études informatiques héritent aussi de systèmes applicatifs hétérogènes qu'il s'agit de maintenir, tout en les homogénéisant le plus possible. Nous parlons alors de projets de refonte des systèmes patrimoniaux (les legacy systems des anglo-saxons).

Pour notre part, notre travail sur la solution de ciblage marketing a été réalisée au sein du service études. Nous avons fait partie successivement de plusieurs équipes, en fonction des réorganisations des équipes projets. Nous avons donc commencé à travailler sur le projets de ciblage marketing sur le « legacy system » de ex-Coface ORT; puis dans le cadre d'un projet commun avec Kompass International, KIM, dont un certain nombre de difficultés dont la crise financière mondiale, ont imposé l'arrêt; enfin récemment, nous sommes de nouveau sollicités pour participer aux évolutions du produit Cofacibles.

L'organisation des équipes de la DSI est fonction de la stratégie et est guidée par la réactivité rendue nécessaire par les difficultés que l'entreprise peut rencontrer. La crise financière, qui a commencé en 2007-2008 par une crise du crédit, a frappé les activités

de « credit management », en particulier d'assurance crédit.

Fin 2009, la direction générale de Coface Services et les DSI de Coface et Coface Services ont décidé de faire du système développé à Tours chez ex-Coface ORT, le socle du futur système d'informations entreprise (Iris France) de Coface Services et d'abandonner progressivement tous les autres systèmes (ex-Coface ORT, ex-Coface SRCL) en les fusionnant avec celui-ci.

Cette décision a un impact sur le système de ciblage marketing, car il va devoir évoluer sur ce socle et non plus sur la refonte complète initialement envisagée, qui comprenait aussi une fusion avec les systèmes de Kompass International; la branche KIM-IS dédiée au ciblage marketing du projet KIM a été abandonnée. Néanmoins, nous pensons que l'important travail d'analyse effectué pour KIM-IS va pouvoir être réutilisé.

Plateformes techniques

Le service des études informatiques effectue son travail en respectant un cycle de développement du logiciel; à chaque étape correspond une plateforme. L'organisation des plateformes correspond au cycle de vie du logiciel, que nous décrirons plus loin, et est structurée en quatre niveaux :

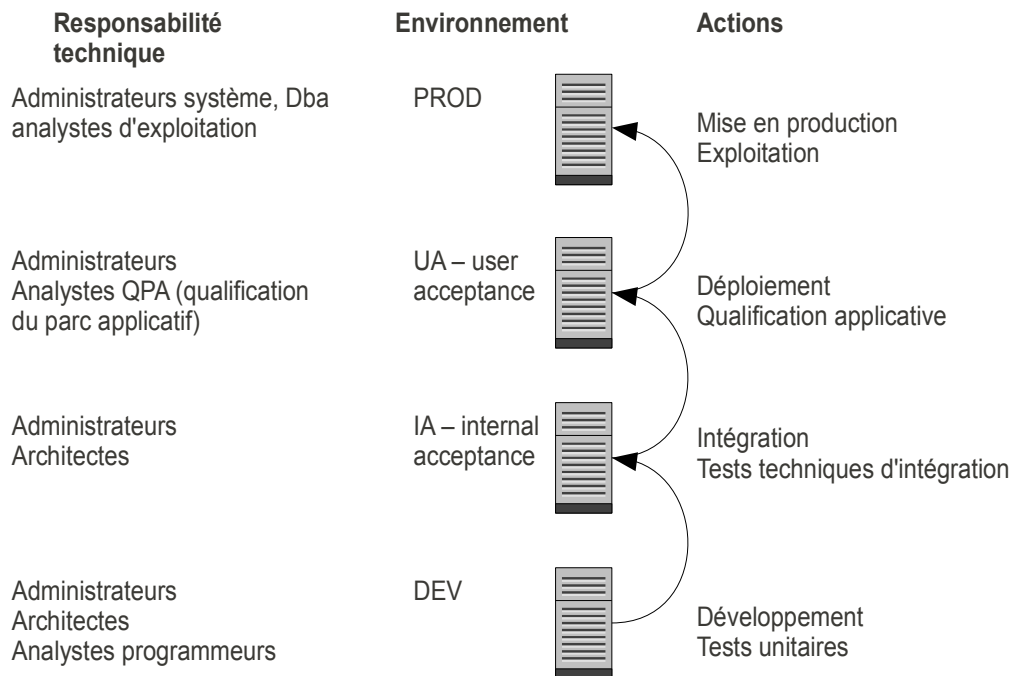
Responsabilité technique	Environnement	Utilisateurs
Administrateurs système, Db analystes d'exploitation	Environnement de production (PROD)	Clients Utilisateurs internes
Administrateurs Analystes QPA (qualification du parc applicatif)	Environnement de recette (UA – user acceptance)	Maitrise d'ouvrage Analystes testeurs Analystes fonctionnels
Administrateurs Architectes	Environnement d'intégration (IA – internal acceptance)	Architectes Analystes fonctionnels Analystes programmeurs
Administrateurs Architectes Analystes programmeurs	Environnement de développement (DEV)	Analystes fonctionnels Analystes programmeurs

Chaque niveau correspond à un état d'avancement particulier du logiciel. Le but final est évidemment de livrer en production la solution développée.

Chaque plateforme est sous la responsabilité d'une équipe technique. A chaque niveau correspond un ensemble d'utilisateurs. Plus le degré d'intégration de la solution avance, et moins les analystes développeurs sont sollicités. Cette progression permet aussi de mesurer les performances, de « stresser » chaque solution en version stable (environnement UA), avant toute mise en production. L'environnement UA présente une architecture identique et des volumes proches de ceux de la production. Il peut d'ailleurs au besoin, être « rafraîchi » avec les données récente de la production. Cet environnement permet donc de reproduire au besoin une difficulté rencontrée en production, sans la perturber.

Nous ne parlerons pas de la plateforme standard pour les postes clients qui, pour Coface, est basée sur des PC sous Windows. De plus la généralisation des interfaces homme-machine en mode web, rend l'environnement du poste client (système d'exploitation, etc.) de moins en moins perceptible au développement de solutions sur serveurs.

Procédure de développement / intégration / qualification / mise en production



Les étapes de livraison correspondent à l'organisation des plateformes :

- livraison pour intégration de, DEV à IA : le développeur livre son composant à l'intégrateur
- livraison pour déploiement, de IA à UA : l'intégrateur livre un composant complet ou une solution à QPA (qualification du parc applicatif)
- livraison pour mise en production UA à PROD : livraison d'une solution complète à la production, pour mise en exploitation.

Les outils:

- Dimension : gestion de version, de configuration; permet aussi les livraisons entre DEV et IA
- Quality center : tests et anomalies (defects); permet les échanges entre UA, IA et DEV
- Service Center : gestion des demandes de livraison entre IA et UA et UA et PROD.

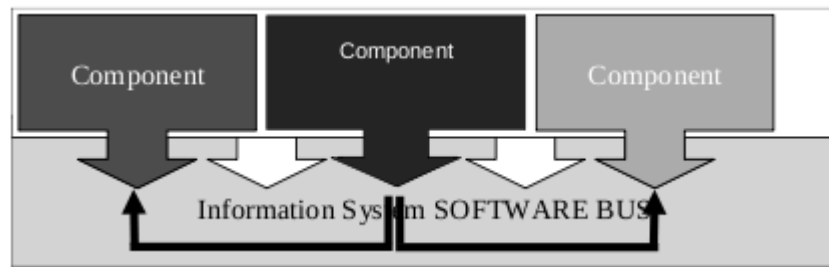
2.5.3 Cycle de vie du logiciel

2.5.3.1 Mode par composants : le standard Coface

Le rachat par Coface de plusieurs sociétés, dont chacune a un SI propre, imposait à Coface de faire converger à moyen terme les différents SI patrimoniaux hétérogènes. L'urbanisation du SI par la mise en place d'une architecture modulaire, distribuée devenait nécessaire.

Un nouveau mode d'ingénierie par composants logiciels a été adopté par Coface pour tous ses projets de taille importante : notamment le projet ELAN lancé il y a une dizaine d'années pour refondre complètement le SI Coface.

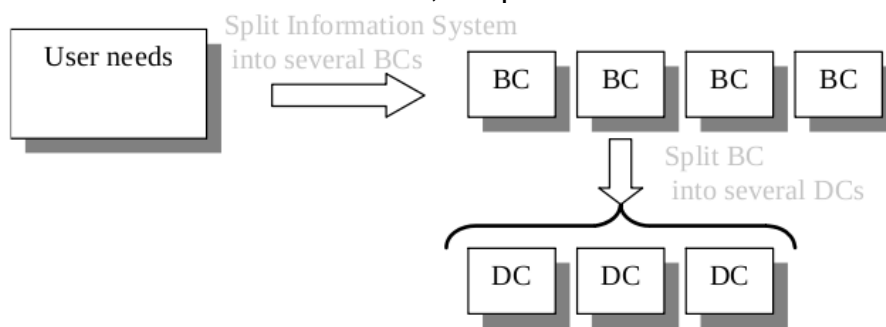
La plateforme « Caribou » a été créée. c'est est une implémentation des idées de Herzum en J2EE. Schématiquement, les composants sont des briques logicielles qui se branchent sur un bus logiciel commun standard.



Les développements par composants ont pour but

- de faciliter le contrôle de la complexité, des coûts de conception, de réalisation, d'évolution du SI;
- d'offrir un cadre standard de développement, en fournissant de large possibilité de réutilisation des composants;
- de définir une architecture souple et commode à déployer;
- de réduire les temps de mise à disposition des solutions logicielles de l'entreprise.

Un composant logiciel est au plus haut niveau un composant métier, c'est à dire qu'il répond à un ensemble cohérent de besoins, ce que résume le schéma suivant :



L'idée est de réduire la complexité initiale du problème métier. En raisonnant en terme de composants métier (BC ou business component), les analystes peuvent spécifier les fonctionnalités, indépendamment des questions techniques, dans un premier temps. Ils définissent les services rendus par le composant. Ces services seront ensuite implémentés par des éléments techniques appelés DC (distributed component). Ces composants techniques sont pris en charge par une autre équipe d'analyste-programmeurs.

Ainsi, les idées de Herzum conduisent à structurer le service des études informatiques en trois entités :

- les analystes fonctionnels, responsables de l'analyse et de la livraison de la solution logicielle auprès de la MOA (équipe CAM, coordination aide à maîtrise d'ouvrage)
- les analyste-programmeurs, responsable du développement du parc applicatif (DPA)
- les architectes-intégrateurs, responsables de la structure de la plateforme et de la cohérence globale de l'architecture du système (équipe ASI, architectes du système d'information).

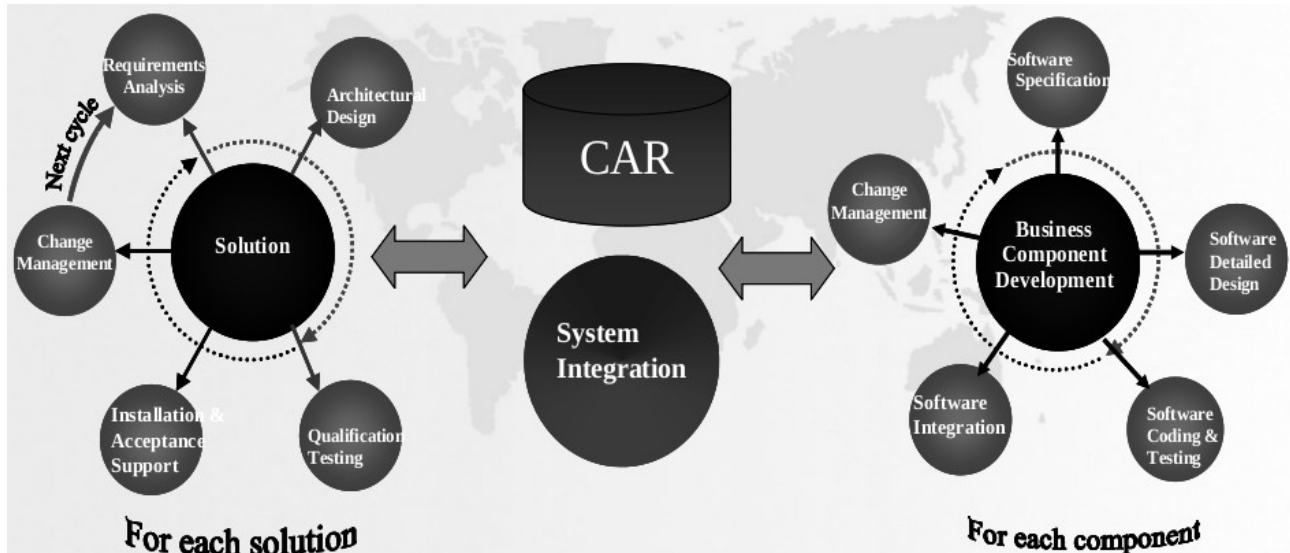
Cette organisation exige évidemment beaucoup de coordination. ainsi sont mis en place des comités de suivi, de coordination entre les équipes solution (CAM) et les équipes

composants (DPA).

Cycle de développement

Le cycle de développement d'une solution, comme conséquence des idées de Herzum, se décompose en deux sous-cycles : l'un sous la responsabilité des analystes fonctionnels (CAM), l'autre sous la responsabilité des développeurs.

Quand à la phase de « jonction », elle est assumée par les architectes intégrateurs (ASI).



CAR : component architecture repository, est le référentiel (électronique) de tous les composants du SI (sous la responsabilité de ASI).

Dans le cycle de gauche, nous reconnaissons les phases d'analyse fonctionnelle (architecture design), de qualification et de recette (environnement UA) qui sont de la responsabilité de CAM.

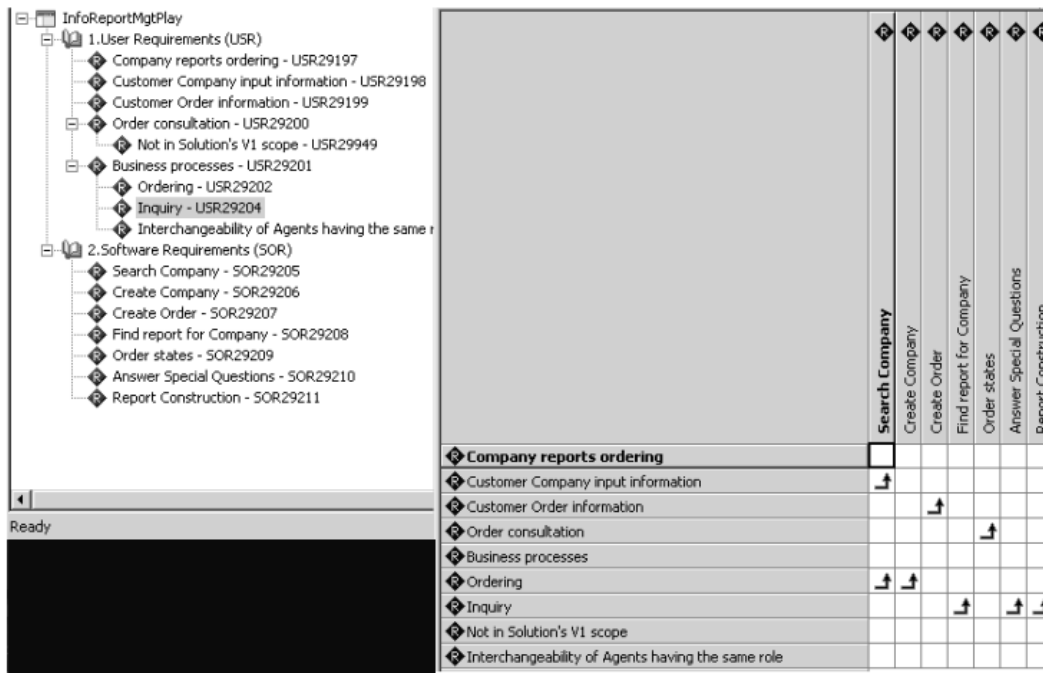
Dans le cycle de droite, les phases de spécification techniques, de programmation, de test (unitaire) et d'intégration sont sous la responsabilité de DPA.

Au centre, l'équipe ASI maintient et coordonne l'intégration des éléments développés, en rapport avec les exigences de livraison de la version attendue par CAM.

Il y a donc un lien direct entre ces cycles et les différents niveaux de plateformes physiques que nous avons décrites précédemment.

Traçabilité entre besoin et spécification technique

L'un des points forts de la méthode d'ingénierie adoptée, est l'introduction d'une traçabilité continue et rigoureuse entre les différentes étapes du développement du logiciel : du besoin spécifié au code du service implémenté. La traçabilité est gérée dans l'outil Caliber RM. Pour illustrer ce point, voici un exemple de matrice de traçabilité :

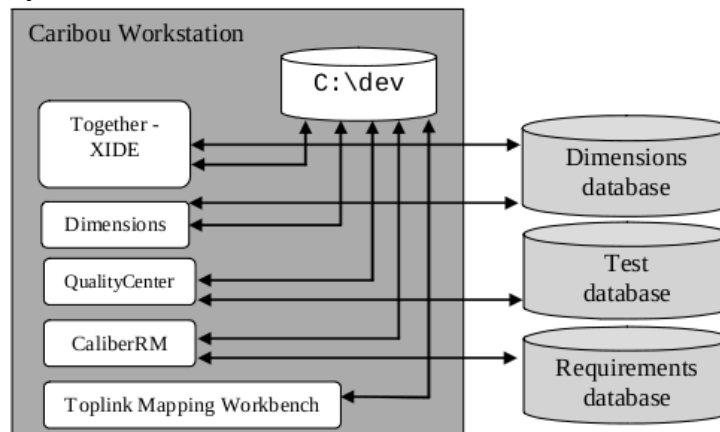


Outils pour Caribou

Pour la spécification fonctionnelle :

- Borland Together/NoMagic MagicDraw
- Borland CaliberRM (Management of requirements)
- Mercury Quality Center (Test management)
- Coface XIDE

Sur le poste de l'analyste les outils suivants sont installés :



Les outils pour la programmation et les tests :

- Eclipse avec le plugin Coface
- Coface XIDE et framework (+ libraries)
- Oracle TopLink (Object/relational mapping tool) + API
- Weblogic Workshop & Portal (GUI design)
- Weblogic Integration (Workflow)
- Serena Dimensions (configuration management database)
- Mercury Quality Center (Test management)
- Mercury QuickTest (Test execution)
- Mercury Load Runner (Load tests)

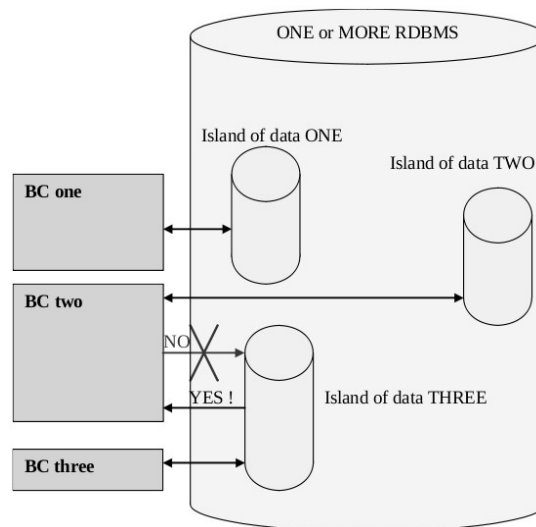
- Sitraka JPROBE (Test coverage tool)
- Junit

Quelques aspects techniques

Les composants ne communiquent pas par le biais d'objets distribués, mais par l'intermédiaire de structures de données appelées Business Data Types (BDT), qui sont décrites à l'aide d'un schéma XML.

Les composants Caribou ne sont pas des objets dans le sens où un composant n'est pas implémenté comme un objet. Un composant est une interface de services métiers.

Chaque composant gère ses propres données. On parle alors d'un îlot de données attaché à un composant. Seul le composant peut accéder à son îlot. Les autres composants ne peuvent accéder à ses données qu'au travers de services, sauf éventuellement en lecture. Ce principe est illustré sur la figure suivante:



Un îlot de données est un schéma dans une instance Oracle, ou une partie d'un schéma. Les liens entre le monde objet et les tables relationnelles sont effectués à l'aide de l'outil Oracle Toplink, qui est un outil de mapping objet-relationnel graphique.

2.5.3.2 Mode projet

Le mode projet est le mode traditionnel de gestion des études informatiques. Une même équipe, sous la responsabilité d'un chef de projets, est en charge de l'analyse fonctionnelle, de la conception, de la programmation, des tests, de l'intégration et de la livraison.

C'était le mode adopté anciennement par ORT, dont nous sommes issus. Ce mode s'est montré efficace en son temps. Les premières applications de marketing direct ont notamment été développées sur ce mode ci.

Par ailleurs, les solutions logicielles hétérogènes héritées des différentes acquisitions de Coface rendent difficile la généralisation immédiate de la méthode Caribou. Par exemple, des pans importants de ces systèmes ne sont pas développés et ne sont pas directement compatibles avec J2EE sur lequel Caribou est exclusivement implémenté. Le temps de leur « refonte », ils continuent à être maintenus en mode projet.

Pour autant, comme nous l'avons déjà signalé, le mode projet est difficile à généraliser à des projets complexes de très grande ampleur qui impliquent des équipes réparties sur plusieurs sites géographiques.

En revanche, ponctuellement, et sur un projet clairement délimité et localisé, qui peut faire l'objet d'une pression budgétaire élevée, ce mode de gestion retrouve quelque intérêt. Par pragmatisme, la DSI de Coface Services a ainsi décidé de mener le programme Iris France sur 2010 et 2011 sur la base d'un mode projet, en nommant un directeur de programme dédié à Iris. Iris est le projet de refonte du SI information entreprise, qui doit mixer en un seul système les deux systèmes hérités de Coface ORT et de Coface SCRL. Le mode d'analyse, de gestion documentaire répondent aux standards Coface. L'architecture finale, qui ne reposera pas totalement sur J2EE , sera néanmoins totalement compatible avec le SI Coface.

2.5.4 SI « Information Entreprise »

2.5.4.1 Introduction

Le SI (système d'information) information entreprise est en cours de refonte. A la fin de l'année 2009, la direction de Coface Services a décidé de fusionner les systèmes ex-Coface ORT et ex-Coface SCRL dans un nouveau système dont la base technique repose sur la structure du SI ex-Coface ORT.

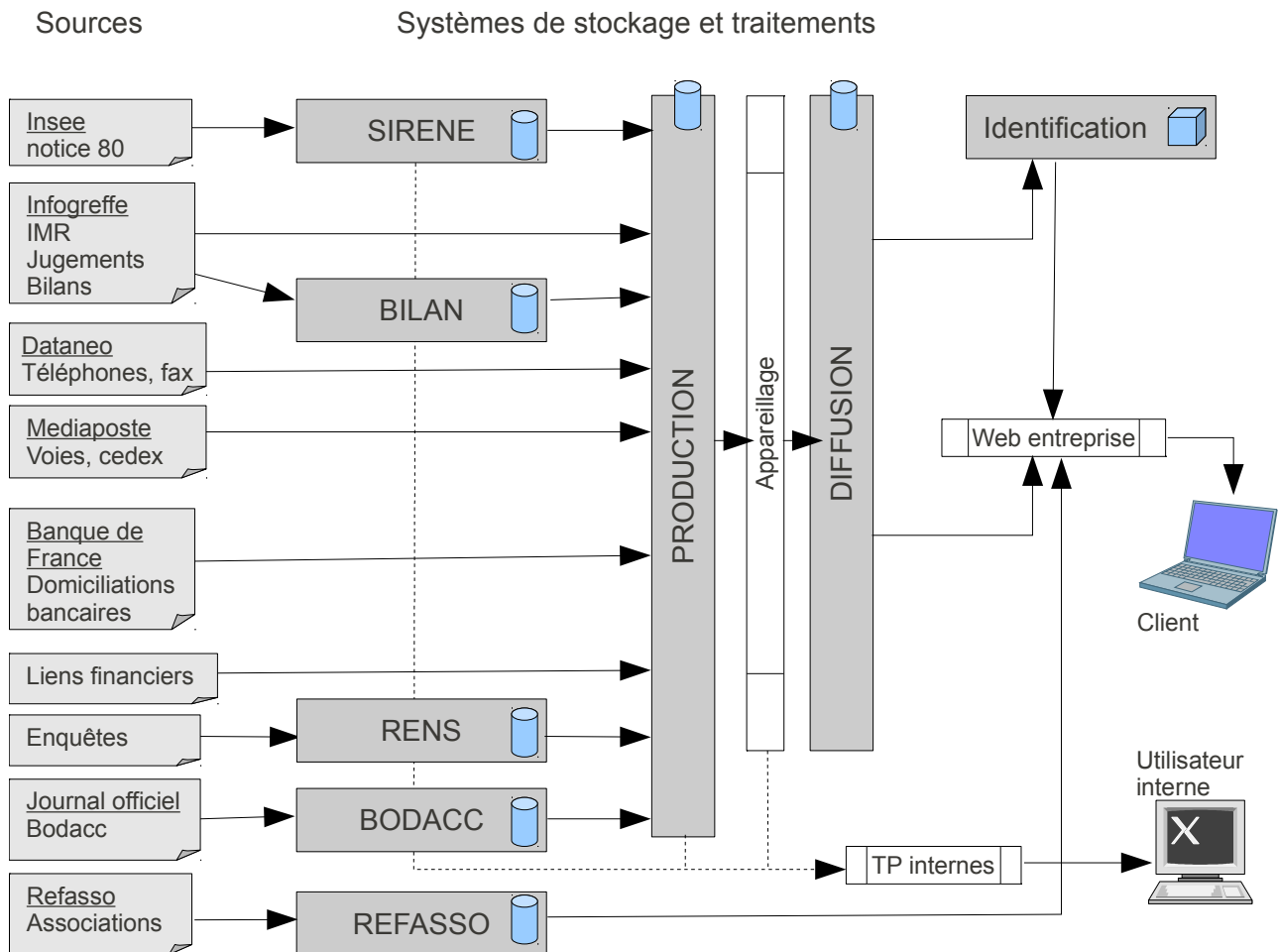
Le nouveau SI Iris n'étant pas encore finalisé, celui que nous présentons, analogue dans le principe, est celui de ex-Coface ORT. Le SI du ciblage marketing sur lequel repose la solution Cofacibles est aussi basé sur ce système.

2.5.4.2 Architecture globale

Dans ce paragraphe, nous décrivons l'architecture globale du SI dédié à l'information entreprise dans ses grandes lignes. Ce système est complexe et nous n'entrerons pas dans tous les détails.

Nous montrerons ensuite comment il est lié au système de ciblage marketing.

Le système information entreprise peut être simplement présenté comme un immense flux orienté des sources d'informations vers les clients et utilisateurs finaux. Sur ce flux, un certain nombre de processus de contrôle et de restructuration de l'information sont branchés.



Nous retrouvons les sources officielles et les sources privilégiées que nous avons mentionné dans notre première approche de l'information entreprise.

Le cœur du système est constitué de « modules » construits chacun autour d'une base de données. Certains modules sont spécialisés sur une source. Un module peut en effet avoir à gérer des données volumineuses et / ou particulières. ainsi en est-il des bilans, qui font l'objet de traitements d'analyse financière assez complexes.

Le BODACC (bulletin officiel des annonces civiles et commerciales) est aussi l'objet d'un module spécifique. Les informations arrivent en effet sous forme de textes qu'il faut analyser, découper, structurer avant de les exploiter dans la base de données « mère » que nous appelons base de « production » (dans le sens de production des informations entreprise, et non au sens de production informatique).

Un certain nombre de contrôles et de structuration préliminaires sont effectués par ces modules spécialisés. Les informations recueillies sur les entreprises françaises sont riches, diversifiées mais plus ou moins structurées selon les sources. Cela demande travail de regroupement, de classement.

La base de production agrège donc toutes les sources d'informations et les met à disposition des utilisateurs internes (analystes, enquêteurs, rédacteurs) au travers d'une application appelée « TP internes »; TP pour « Transactional Process ». Un utilisateur interne peut les consulter et les modifier en fonction de ses analyses et de ses habilitations.

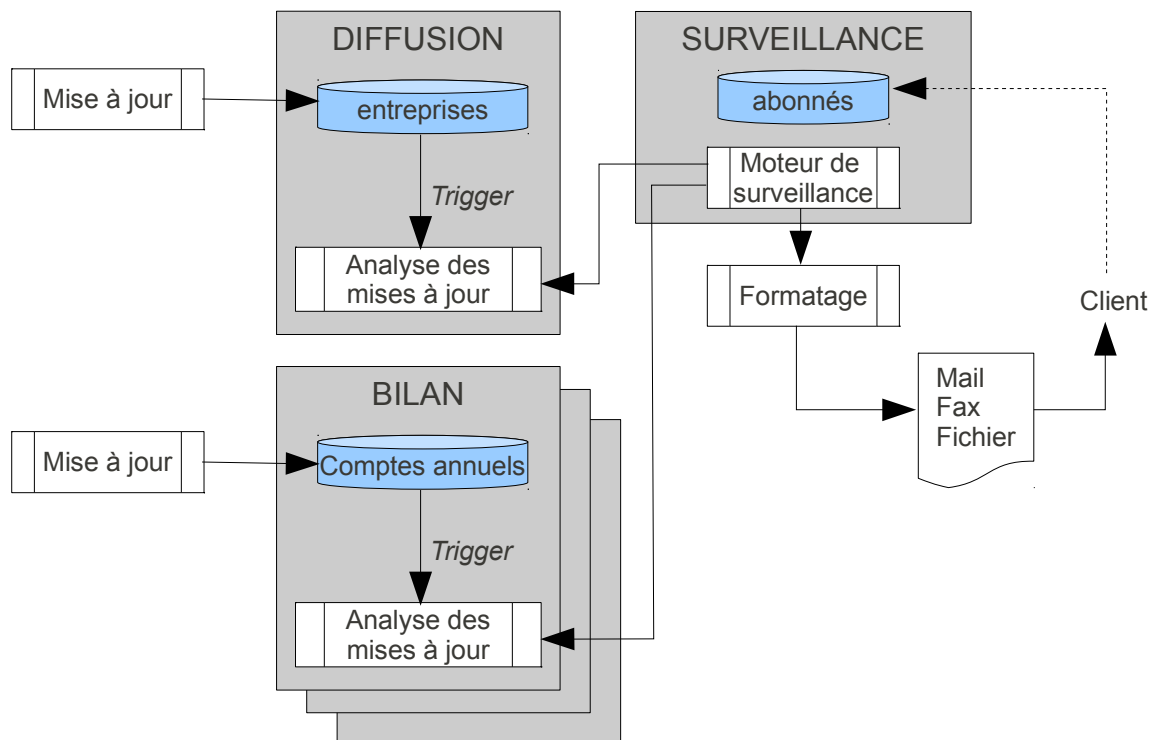
Les informations diffusées aux clients font au préalable l'objet d'un filtrage et d'une consolidation. En effet, une information de la base de production peut être fournie par

plusieurs sources (une officielle, et une source privilégiée). Il appartient au système que nous appelons « appareillage » de déterminer la meilleure informations en fonction de règles qui lui sont données (et qui sont codées). De plus, l'appareillage effectue les contrôles de cohérence globale qui n'ont pas pu être traités en amont : il vérifie par exemple avant diffusion que le capital d'une entreprise (obtenu par une source) est cohérent, au regard du droit des entreprises, avec la forme juridique de cette entreprise (qui a pu être obtenu par une autre source). Au final, le client ne doit voir que des informations à jour et cohérentes; la réputation de bonne qualité du service en dépend.

2.5.4.3 Surveillance

Le système de surveillance mis en place chez Coface Services est présenté dans le schéma suivant :

Systèmes de surveillance et de génération d'alertes



Le système de surveillance est un des modules essentiels au système d'information entreprise. Son but est d'abord, de détecter en amont toutes les modifications effectuées sur la base de données de diffusion. Chaque modification peut ainsi être traitée individuellement de façon à pouvoir alerter très précisément les clients abonnés à ce service. Le détail de chaque modification lui est transmis dans le format de son choix. Ainsi un client ne peut s'abonner qu'aux changements d'adresse, sans se soucier des procédures collectives.

Ce mécanisme, très commode, a été appliqué à d'autres bases de données qui contiennent des informations diffusibles : la base de données des comptes annuels (système BILAN) ou encore la base de données de annonces Bodacc.

Le mécanisme de surveillance est un mécanisme générique auquel est notamment abonné le système de ciblage marketing : il est mis à jour grâce aux alertes générées en amont.

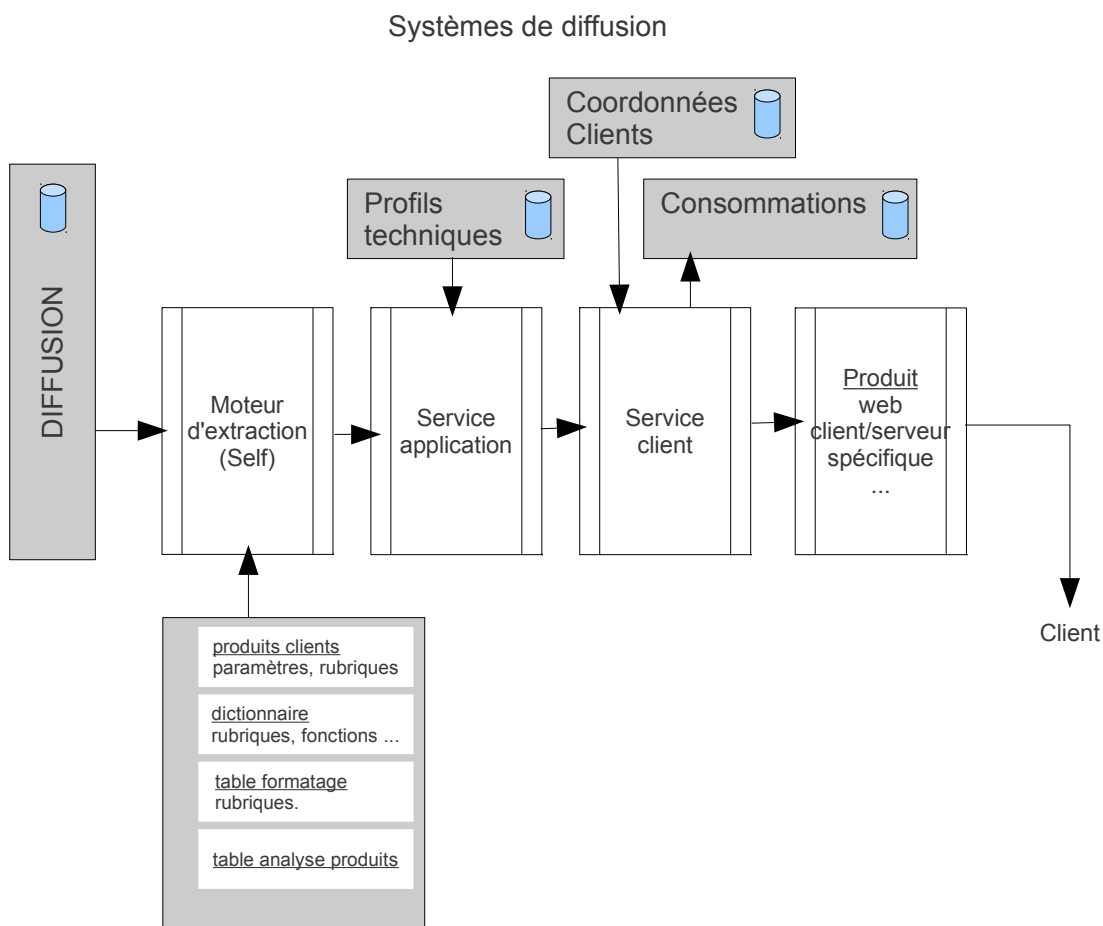
Techniquement, le mécanisme d'enregistrement des modifications est basé sur l'utilisation

de « triggers » Oracle : c'est une solution technique simple, élégante et efficace; tout se passe dans le noyau Oracle. Le système de diffusion peut « digérer » plusieurs dizaines de milliers de modifications chaque jour, sans ralentissement perceptible par les utilisateurs.

Le moteur de surveillance proprement dit, se déclenche chaque nuit (jour ouvrable) et prévient les clients abonnés. toute une gamme de produits de surveillance a été conçue : de l'alerte sur une modification de l'adresse, à une évolution du score. Ce mécanisme permet au client de ne pas avoir à faire la démarche de vérifier par lui même chaque jour si la situation financière de ses propres clients a changé.

2.5.4.4 Diffusion

Le dernier module du système d'information entreprise que nous souhaitons présenter est le système de diffusion, dans la mesure où il gère la diffusion de tous les produits qui sortent du système, parmi lesquels les produits de ciblage marketing que nous avons présentés plus haut.



Le schéma ci dessus montre qu'un produit d'information entreprise est fabriqué à partir des données contenues dans la base de données de diffusion.

Le système de diffusion est à plusieurs niveaux, chacun avec un type spécifique de service à rendre.

L'extraction est effectuée par le moteur d'extraction que nous nommons aussi « self » dans la mesure où il a été conçu de façon très souple et permet de construire par simple paramétrage un produit sur mesure pour un client, un peu comme le ferait un « self-

service ». Il repose en effet sur un système expert, qui est capable de relier chaque information demandée dans un produit à sa localisation en base de données, et ce, en minimisant le temps d'extraction global du produit.

Les autres services ont pour but de gérer les autres aspects qui concernent le client : le canal par lequel le produit lui sera envoyé, l'enregistrement de la consommation (pour facturation ultérieure), etc.

2.5.4.5 Aspects techniques

Nous souhaitons expliciter quelques éléments techniques importants du système d'information entreprise, que par ailleurs, nous avons aussi eu l'occasion de mettre en œuvre.

Technologies Oracle

Les données qui sont le cœur du système sont stockées dans plusieurs schémas Oracle eux même regroupés en plusieurs instances. Les tables sont accédées en mise à jour exclusivement au travers de fonctions PL/SQL « packagées » et compilées sur le noyau. elle contient l'implémentation des règles de gestion des données. Les fonctions qui gèrent les sources d'informations, sont notamment chargées d'effectuer un certain nombre de validation de façon à détecter très en amont les anomalies. Les anomalies sont elles même l'objet d'un suivi et d'un traitement développé en PL/SQL, au plus près de la base de données.

Par analogie avec le « monde objet », les fonctions packagées permettent une forme d'encapsulation des tables. De plus, cette technique n'impose pas de construire un « mapping objet relationnel » avant de concevoir un traitement simple; opération qui ajoute toujours de la complexité technique à une application, en imposant plusieurs couches d'objets entre la donnée et son utilisation réelle. La manipulation de la donnée est en quelque sorte « native » en PL/SQL. Cette technologie robuste et efficace, permet aussi un déploiement simple des applications.

Les traitements de masse (batch) sont développés en Pro*C. Ce langage permet d'embarquer du code SQL et PL/SQL, dans un programme écrit en langage C. Il se présente sous la forme d'un certain nombre de directives (EXEC SQL ..) à inclure dans le code C. Le code est ensuite travaillé par un pré-compilateur Pro*C qui remplace les directives par du code C qui contient les appels nécessaires à une API (application programming interface) propre à Oracle et fournie (dans la librairie libsql).

Technologies client - serveur

Les services du systèmes de diffusion, notamment, sont basés sur le moniteur transactionnel Tuxedo de l'éditeur BEA (acheté par Oracle en 2009). Tuxedo est en quelque sorte un bus logiciel, un « cadre logiciel » (le « framework » des anglo-saxons), qui soulage le développeur de l'implémentation de la gestion des communications, en les réduisant à un simple paramétrage des « domaines Tuxedo », à l'intérieur desquels les services sont publiés. La phase de développement peut donc se concentrer sur les aspects métiers.

Technologies Java

Un certain nombre de modules ont été développés en Java/J2EE, notamment sous la forme de Servlets. Elles permettent un accès simple en mode web-service pour nos clients.

Par ailleurs, il existe des ponts entre Tuxedo et Java : BEA fournit un API qui se nomme

WTC (Weblogic tuxedo connector).

La solution Iris, en cours de développement, fera vraisemblablement usage des « ponts » entre le monde Java/J2EE et le monde Oracle/Pro*C/Tuxedo.

*
* *
*

Il nous a paru important de montrer notre compréhension des différents métiers du groupe auquel appartient Coface Service, dans une première partie, avant de présenter notre contribution à la conception d'un moteur de ciblage marketing.

Nous avons poursuivi par la description de l'infrastructure, de l'organisation et des méthodes de la DSI Coface Services.

3 Analyse des besoins, spécification

Nous regroupons dans une même partie l'analyse des besoins et la spécification (générale) du système étudié. Les définitions des concepts et la présentation des principaux processus qui mettent en jeu le système de ciblage marketing, sont directement produits par l'analyse des besoins tels qu'ils ont été formulés par la maîtrise d'ouvrage (MOA) lors de nos séances de travail communes.

Dans cette partie, nous reprenons une bonne partie des éléments d'analyse que nous avons produits dans l'élaboration du « cahier des charges pour un moteur de ciblage », en tant qu'analyste assistant la MOA dans le cadre du projet KIM-IS (Kompass International Marketing – Information Service). Signalons que l'ensemble des besoins exprimés correspondent à la vision « idéale » de la MOA pour un outil qui permette à Coface Services de se démarquer de la concurrence sur un marché de plus en plus disputé. Les ambitions initiales du projet KIM-IS ont été revues depuis, pour des raisons financières que nous avons évoquées dans la partie précédente, et aussi pour des raisons techniques : les besoins sont très riches, complexes, mais les technologies disponibles sur le marchés des produits des éditeurs n'étaient et ne sont sans doute pas encore assez mûres pour répondre à tous les besoins.

Les fonctionnalités de Cofacibles (solution actuelle) ne sont aujourd'hui qu'un sous-ensemble des besoins recensés ci-après. Cependant, Cofacibles étant un produit amené à évoluer, il nous semble pertinent de présenter tous les besoins exprimés par la MOA, sachant que nous n'avons pas (encore) toutes les solutions techniques pour y répondre. La partie suivante, axée sur la conception proprement dite d'un moteur de ciblage, expose les solutions retenus pour Cofacibles et des pistes nouvelles.

3.1 Définitions - concepts

Avant d'entrer plus avant dans le détail des spécifications du moteur de ciblage marketing qui est le cœur de notre propos, nous souhaitons définir les concepts essentiels, sur lesquels reposent toutes les analyses qui vont suivre.

Sont précisés ici que les concepts utilisés dans le contexte du moteur de ciblage. Nous commençons par donner des définitions assez larges des concepts, les précisions nécessaires à la compréhension d'un paragraphe particulier, apparaîtront au fil du document.

Les lignes qui suivent donnent une première approche des définitions des notions utilisées.

Entité

C'est l'élément (l'objet) de base ciblé par les requêtes et donc par le moteur.

Dans le cadre de l'activité de ciblage marketing de Coface Services, il s'agit d'entreprises (entités juridiques) ou d'établissements (site d'activité économique).

Les entités sont caractérisées par un identifiant fonctionnel et des attributs qui dépendent fonctionnellement de cet identifiant (au sens de l'algèbre relationnelle). Un attribut peut être une donnée structurée (code, nombre, date), ou du texte.

Les entités de nature « entreprise » ont la même structure, elles sont d'ailleurs au départ

stockées dans une table relationnelle. De même pour les établissements qui ont leur propre table.

Comptage

Le comptage consiste à déterminer le nombre d'entités dont les propriétés correspondent à celle qui sont spécifiées dans les critères de la requête.

Ciblage

Le ciblage consiste à trouver les identifiants des entités dont les propriétés correspondent à celles qui sont spécifiées dans les critères de la requête.

Requête de ciblage

Une requête élémentaire de ciblage est un critère élémentaire, c'est à dire un prédicat exprimé comme une condition sur un attribut. C'est l'analogie du « select » du langage SQL. Un prédicat n'a de sens que dans le contexte d'entités de même nature (de même structure, donc issues de la même table).

Stratégie

Une stratégie de ciblage correspond à un enchaînement de requêtes élémentaires, combinées avec des opérateurs logiques. Les parenthèses sont permises. Une stratégie est donc un arbre de combinaisons logiques (et, ou, sauf) dont les nœuds sont des critères élémentaires ou des combinaisons.

Stratégie « en entonnoir »

Une stratégie « en entonnoir » est une stratégie normale. Le qualificatif « en entonnoir » concerne davantage le déroulement et la présentation des résultats du comptage.

Une stratégie en entonnoir donne le comptage de chaque résultat intermédiaire (critère élémentaire, combinaison), en plus du résultat final.

Cette idée simple au départ, exige pourtant un fonctionnement différent d'un moteur SQL classique : en SQL, seul le résultat global de la stratégie est retourné. Pour obtenir les résultats intermédiaires, il faut jouer autant de requêtes qu'il y a de nœuds dans l'arbre de la stratégie, ce qui peut s'avérer très coûteux en ressources consommées. Nous y reviendrons.

Cible

La cible est l'ensemble des entités qui ont été « cernées » par la requête de ciblage. Concrètement, pour le moteur, il s'agit de la liste des clés ou identifiants fonctionnels des entités qui ont été ciblées.

Liste des résultats

La liste de résultat est constituée de l'ensemble des entités de la cible, présentée comme une liste, une table, dont chaque ligne contient l'identifiant fonctionnel (la clé) et quelques attributs choisis d'une entité.

Cette liste permet à un utilisateur du moteur (en réalité la liste des résultats est présentée à l'utilisateur par la couche applicative au dessus du moteur) de se faire rapidement une idée de la cible qu'il a construite.

Moteur de comptage et de ciblage

Le moteur de comptage et de ciblage est le « noyau » du système informatique qui permet à un utilisateur d'effectuer des requêtes de comptage et de ciblage. Ce noyau comprend le contenu (les données) et tous les mécanismes qui permettent de compter et de cibler. Il ne

comprend pas les IHM (interfaces homme-machine).

Enrichissement

L'enrichissement d'une cible consiste à ajouter tous les attributs choisis par l'utilisateur (parmi ceux que propose Coface Services) de façon à lui retourner un fichier marketing conforme à ses souhaits.

Un « fichier enrichi » est un produit à haute valeur ajoutée pour Coface Services. Le coût de la prestation dépend de la nature et de la quantité des informations ajoutées.

Surveillances

D'un point de vue général, ce que nous appelons « surveillance » consiste à vérifier périodiquement l'état des données.

Il y a un premier type de surveillance, qui concerne chaque entité. Par exemple, surveiller une entreprise consiste à vérifier chaque jour si la valeur d'un de ses attributs a changé, et à le signaler à nos clients s'il se sont abonnés à ce service.

Un deuxième type de surveillance s'applique aux cibles, donc aux résultats des stratégies définies par les clients de Coface Services. Il s'agit dans ce cas de déterminer les éléments communs et les différences entre la cible de l'instant $t1$ et de celle de l'instant $t2$, chaque cible étant le résultat de la même requête de ciblage (même expression logique). Nous avons pris l'habitude de distinguer les « entrants » et les « sortants » pour nommer les éléments ou entités qui, respectivement, n'appartenaient pas à la cible à $t1$ mais y étaient à $t2$, ou en faisait partie à $t1$ et plus à $t2$.

Les deux types de surveillance (entités et cibles) peuvent se combiner. La surveillance des entités peut, par exemple, être demandée sur les entités qui restent dans la cible (ni entrants ni sortants), d'une vérification à l'autre. Un changement de coordonnées (adresse, téléphone, ...) intéresse évidemment un client utilisateur de fichier marketing.

Identification / ciblage

Le but de l'identification est de trouver un ensemble d'entités dont les caractéristiques sont les plus proches (pertinentes) relativement aux critères indiqués par l'utilisateur.

Dans l'esprit de l'utilisateur, il s'agit idéalement de trouver l'entité qui correspond le mieux à ses critères, à la différence du ciblage qui lui permet de cerner une population de la façon la plus exhaustive possible.

Cette différence dans le principe a des conséquences techniques : un excellent moteur d'identification n'est pas forcément un bon moteur de ciblage. Un moteur d'identification est « calibré » pour retourner très rapidement les entités les plus pertinentes au regard du critère saisi. Un moteur de ciblage doit s'assurer qu'il ne manque pas une entité pertinente dans la liste de résultat. Ce sont donc deux optiques différentes. Un moteur de ciblage doit permettre de facilement combiner des résultats intermédiaires, de les réutiliser dans autre ciblage. Un moteur de recherche (du type Google, Bing, etc) n'offre, en général, pas ces possibilités : les combinaisons de critères sont permises, mais la notion même d'ensemble d'entités (cible) issues d'une requête est assez floue; Google par exemple, n'indique qu'un nombre assez approximatif d'entités pertinentes.

Une application d'identification doit travailler sur des informations les plus fraîches possibles, ce qui exige une mise à jour en « temps réel ». Le « système d'identification » est conçu pour « absorber » ces mises à jours en continu, et les utilisateurs s'attendent à voir apparaître le plus tôt possible, dans leur moteur favori, des résultats qui portent sur

les informations les plus récentes.

A l'inverse, un système de ciblage doit annoncer explicitement l'instant de ses mises à jour, de façon à avoir un comportement prévisible et reproductible entre deux mises à jour (hebdomadaires par exemple).

L'utilisateur d'un moteur de ciblage est « confiant » si les résultats de son ciblage sont reproductibles d'une heure à l'autre, d'un jour à l'autre, car cet outil lui sert ensuite par exemple à définir sa stratégie commerciale et les tournées de ses agents commerciaux : cette activité tolère assez peu les fluctuations sur ses sources d'informations.

3.2 Acteurs, processus, activités

La présentation des concepts dans le paragraphe précédent fait apparaître un certain nombre de processus et d'acteurs qui sont engagés dans les activités de ciblage marketing chez Coface Services, et qui interagissent donc avec son système informatique de ciblage marketing.

3.2.1 Acteurs

Ainsi, nous pouvons distinguer les acteurs principaux :

- les clients
- les systèmes annexes, dont
 - le système d'enrichissement
 - les systèmes fournisseurs de données

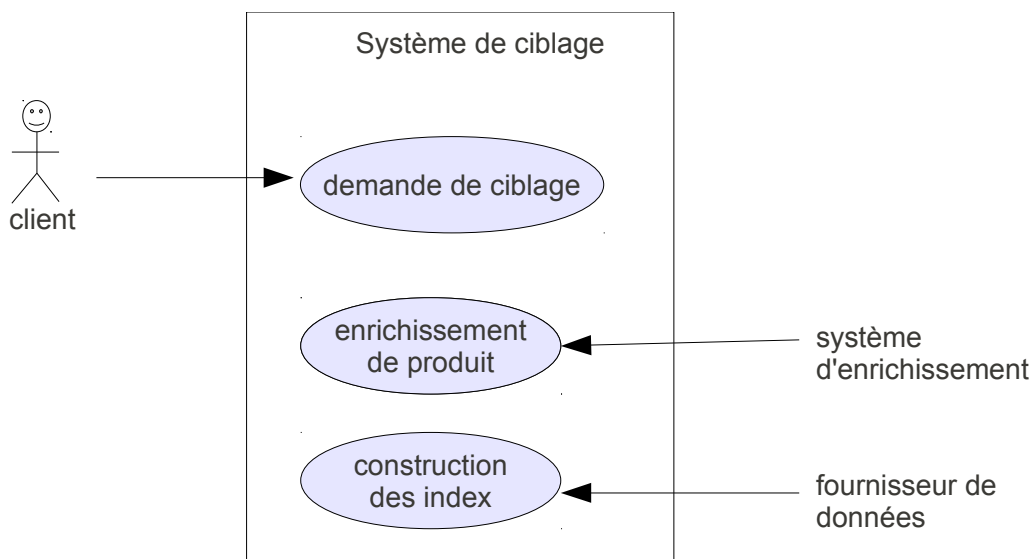


Fig. 3.4 : acteurs et scénarios

Les scénarios principaux (use-cases) sont indiqués sur la figure ci dessus. Ils permettent de « planter le décor » en présentant les services externes les plus importants, que le « système de ciblage » est amené à activer.

Le premier scénario représente la demande de ciblage du client, qui est la raison d'être du système.

Le deuxième scénario indique le processus d'enrichissement des produits, que nous ne détaillerons pas par la suite, car il pourrait faire à lui seul l'objet d'une étude complète.

C'est l'un des composants majeur du système d'informations de Coface Services et c'est l'un des plus complexes.

Le troisième scénario est représenté pour montrer l'importance du processus d'alimentation et de structuration du contenu du moteur de ciblage.

En particulier, les performances de ce processus sont essentielles pour la pertinence et la mise en place effective d'une tel système chez Coface Services.

Sans mécanisme d'alimentation / structuration performant, l'activité de ciblage marketing peut être remise en cause. Nous y reviendrons.

Nous avons omis de représenter un certain nombre de scénarios, dans la mesure où ils impliquent le système de ciblage et d'autres sous-systèmes du SI de Coface Service de façon assez complexe, sans se rapporter directement à l'activité du moteur.

Nous mentionnons plus loin ces activités complémentaires, comme la surveillance.

Nous ne nous étendrons pas plus sur la modélisation en UML du sous système d'information de Coface Service spécialisé dans le ciblage marketing.

Signalons simplement qu'une interface homme-machine fournit l'essentiel des services aux clients finaux, sous la forme d'une application web. L'ergonomie est assez simple et intuitive. Nous renvoyons à la partie 2 et à la présentation du métier « marketing » pour une présentation générale des services (en ligne) offerts aux clients de Coface Services.

3.2.2 Processus

Notre propos concerne les spécifications fonctionnelles et techniques du moteur de ciblage qui constitue le cœur du système de ciblage marketing.

Les processus principaux qui mettent en jeu le moteur sont :

- son alimentation (initiale, mises à jour) et la construction des structures de données du moteur,
- l'exécution interactive des requêtes : comptage, répartition,
- l'extraction des clés des entités ciblées,
- surveillance des stratégies.

3.3 Caractéristiques fonctionnelles

La description du moteur présente plusieurs aspects :

- les données sur lesquelles portent les requêtes (sa base de données),
- la structure des requêtes possibles,
- la gestion des requêtes ou des combinaisons de requêtes,
- les résultats attendus (comptages, tris à plat, tris croisés),
- la sécurité de l'accès aux index et aux données,
- les processus annexes, import, export, qui sont toutefois de sa responsabilité.

3.3.1 Responsabilités du moteur de ciblage

Responsabilités directes du moteur

- Cibler / Dénombrer des populations d'entités (entreprise et/ou établissements), c'est à dire cerner une liste d'entreprises partageant plusieurs critères communs tels que géographique, financiers, organisationnels, informels, ... Cette fonctionnalité est le cœur du logiciel.
- Vérifier les droits d'accès aux index / données

- données publiques
- données privées (confidentialité)
- périmètres prédéfinis
- Gérer les flux
 - Importer des informations utilisées dans les critères de recherche
 - Exporter les résultats des recherches (la liste de résultats ou result-list) indépendamment de leur présentation
- Gérer les mises à jour des données
- Surveiller :
 - les mouvements de populations (créations, suppressions, modifications)
 - les stratégies (conséquence directe du point précédent)

Performances

Les performances du moteur de ciblage (temps d'exécution des requêtes) doivent être compatibles avec la mise en place d'une application interactive avec un utilisateur; pour fixer les idées, l'ordre de grandeur pour le temps d'exécution d'une requête élémentaire (un critère) quel que soit sa complexité, ne devrait donc pas dépasser 100 millisecondes, de façon à permettre l'exécution de requêtes composées de plusieurs critères élémentaires en une seconde environ.

Responsabilités de l'application (ihm), cliente directe du moteur

- Recevoir les requêtes
- Visualiser des résultats
- Présenter des interfaces pour importer ou exporter sous divers formats et supports

Responsabilités annexes du moteur ou responsabilités partagées avec d'autres systèmes

- Gérer les comptes et les profils utilisateurs
 - définition des droits,
 - politique de gestion des accès,
 - gestion des préférences
 - gestion des profils
 - gestion des groupes
- Identifier une entreprise donnée
- Faire du rapprochement de fichier
- Faire du diagnostic de portefeuille client
- Imprimer les résultats sous divers formats
- Facturer les produits consommés

Dans la suite de cette partie, ne sont présentées que les fonctionnalités directement liées au moteur.

3.3.2 Principe général du moteur de ciblage

Le principe général du moteur est de permettre de cerner une population d'entités (entreprises, établissements) en choisissant leurs caractéristiques (ou attributs). Définir un critère de ciblage consiste, en première approche, à spécifier, parmi les valeurs possibles pour un attribut, celles que doivent avoir la population ciblée.

Un critère s'applique à une donnée structurée (code, nombre, date, ...) ou à un texte. Les types de données possibles sont explicités plus loin.

L'intérêt d'un moteur de ciblage est de permettre d'effectuer des requêtes en faisant porter les critères de ciblage sur les attributs des entités et non pas sur leur identifiant (clé), comme cela serait fait dans un accès par clé dans un SGBD relationnel. Le ciblage est un mode de requête « orthogonal » au mode d'accès par clé, habituel pour une base de données relationnelle.

Un critère de ciblage peut être envisagé comme un filtre.

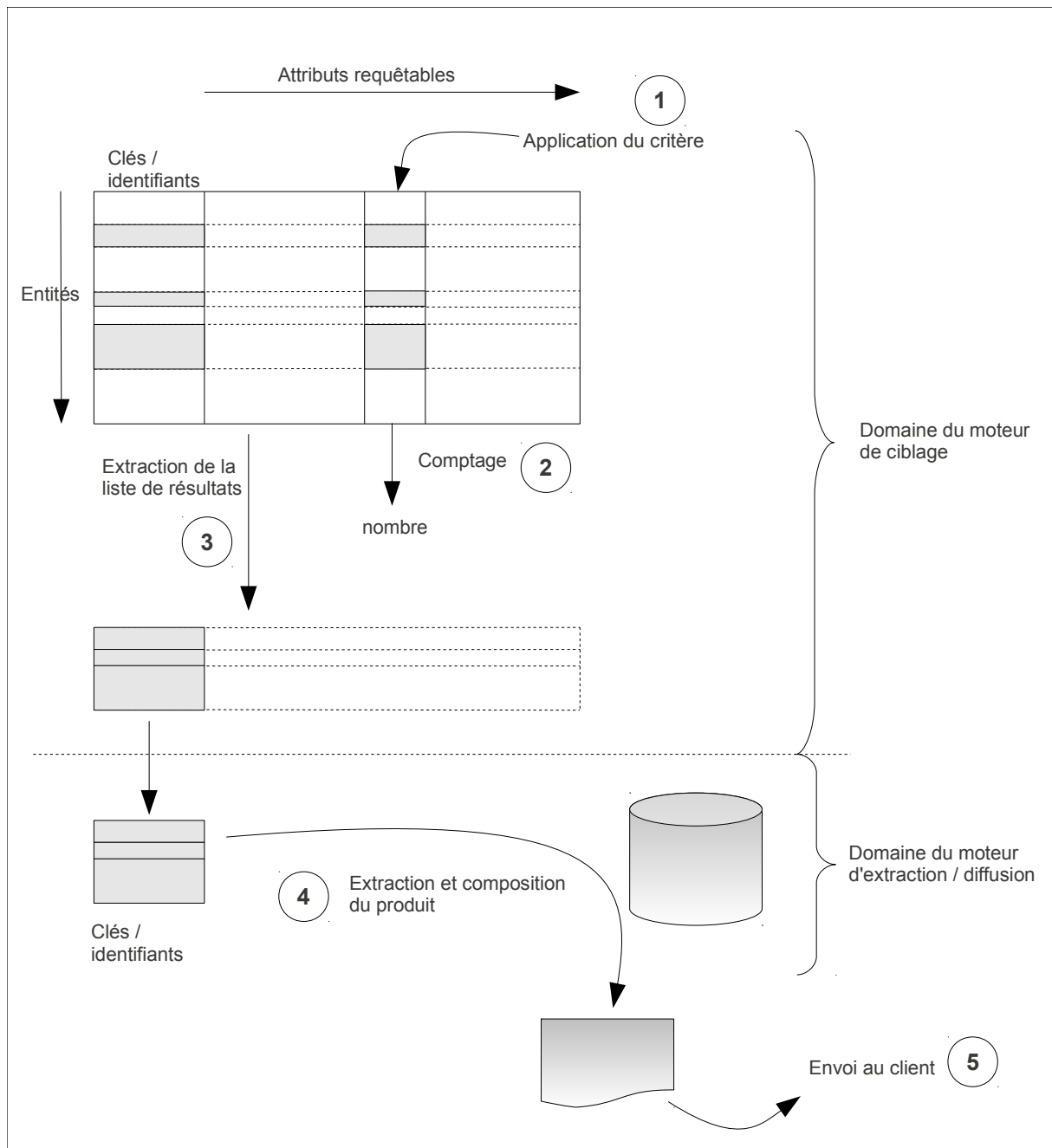
Pour simplifier, il est possible d'imaginer les données contenues dans le moteur comme une table dont les lignes sont les entités et les colonnes les attributs des entités. Chaque entité possède une clé (identifiant fonctionnel) et des attributs (données structurées ou texte).

Une requête élémentaire qui met en jeu un seul critère de ciblage porte sur un seul attribut. Une requête plus complexe met en jeu une combinaison de plusieurs critères de ciblage.

Schématiquement, voici une illustration du processus de ciblage :

Étapes

1. L'utilisateur fournit un critère qui porte sur un attribut au moteur;
2. Le moteur compte le nombre d'entités concernées et retourne le résultat à l'utilisateur;
3. Le moteur exporte la liste d'entités comptées en 2, pour une présentation éventuelle à l'utilisateur;
4. Le moteur exporte les clés des entités ciblées et les fournit au système d'enrichissement (extraction) des données pour fabriquer un produit marketing (fiche prospect Coface Services);
5. Le système de suivi des envois aux clients expédie le produit marketing par le média choisi.



3.3.3 Typologie des données

Entités et dépendances fonctionnelles

Les entités ciblées par le moteur sont des entreprises ou des établissements. Par contre, les données utilisées comme critères de ciblage sont à deux niveaux :

- les attributs rattachés aux entités elles mêmes (dépendance fonctionnelle directe, même clé fonctionnelle),
- les attributs rattachées indirectement à ces entités (dépendance fonctionnelle indirecte, clé fonctionnelle différente).

Les entités directement ciblées sont :

- les entreprises (entités juridiques),
 - dénomination,
 - attributs statutaires,

- les données financières (comptes annuels, agrégats, analyse financière),
- etc.
- les établissements (sites d'activité)
 - attributs de signalétique,
 - adresse,
 - etc.

Les informations liées et regroupées par thème sont :

- les événements légaux ou judiciaires : Bodacc
- les données privées (liées à un établissement ou une entreprise)
- les dirigeants
 - statutaires (liés à une entreprise)
 - opérationnels (liés à un établissement)
 - servent pour un ciblage des entités dans lesquelles ils ont une fonction en utilisant leur nom, prénom, date de naissance
- les liens (entre entreprises) – Nouvelle fonctionnalité (non présente dans l'existant).
 - sont représentés par la donnée de liste de couples de clés : par exemple, un lien d'actionnaire à filiale est donné par le couple (clé actionnaire, clé filiale)

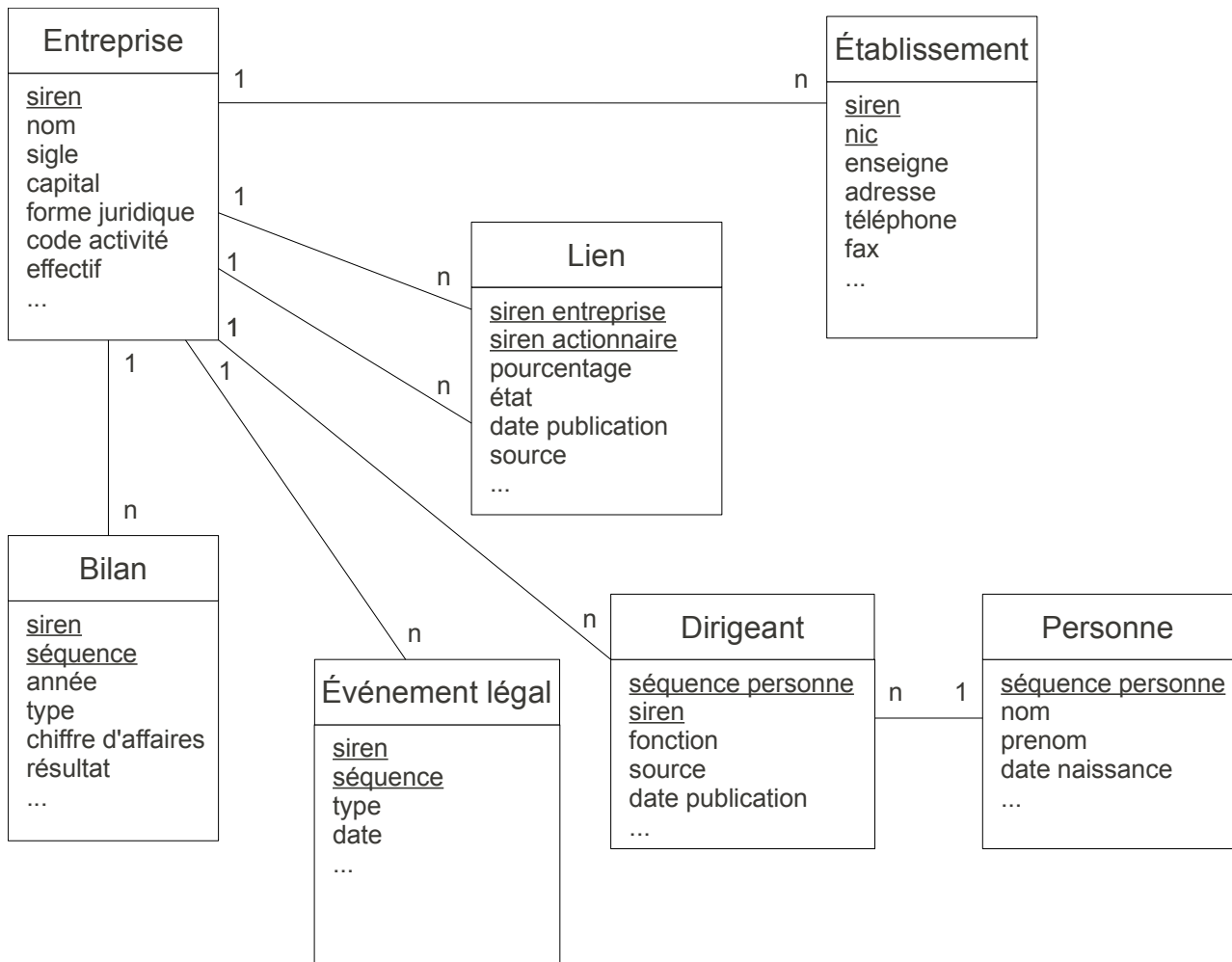
Nous ne donnons pas la liste exhaustive ni le détail de chaque champs. La description que nous venons de donner montre la diversité et la complexité des informations à gérer.

Thème

Un thème est un ensemble de données cohérentes, ayant un lien sémantique, caractérisé par une structure spécifique de clé fonctionnelle.

Il est important d'insister sur le fait que les critères utilisés n'ont pas forcément une dépendance fonctionnelle (une clé) de même structure que les entités ciblées. Par exemple, il est possible d'utiliser les attributs des établissements pour cibler des entreprises ; ou des liens entre entreprises pour cibler des établissements. Le moteur doit en tenir compte et doit donc permettre de faire du ciblage à partir de liens logiques indirects. Le moteur doit donc savoir permettre de faire des requêtes de ciblage au travers de jointures de type 1xN ou NxN.

Le schéma suivant illustre la structure logique des données, en utilisant une représentation quasi-relationnelle.



Pour fixer les idées, on peut proposer la structure (logique) suivante pour les clés :

- entreprise : numéro d'entreprise
- établissement : numéro d'entreprise + numéro d'établissement
- comptes annuels : numéro d'entreprise + type de compte + année exercice
- liens : numéro d'entreprise + numéro d'entreprise
- dirigeants : numéro de dirigeant + numéro d'entreprise.

Cartographie (macroscopique) des données (France)

Le tableau suivant décrit sous forme de pavés, la présence des données selon le type des entités. Chaque case donne une idée du volume des données disponibles pour les données sur les entités françaises.

Entreprises

	clé	Données statutaires (50 champs)	Données de localisation géographique (10 champs)	Données financières (400 champs)	Présence Fonctions des dirigeants (130 booléens)
Entreprises avec bilan (1 292 568)	■ ■ ■ ■ ■	■	■	■	■
Entreprises sans bilan (5 437 382)	■ ■ ■ ■ ■	■	■	■	■

Liens financiers (estimé minimum 160 265 entreprises pour 272 847 actionnariat)						
---	--	--	--	--	--	--

Etablissements

	clé	Données statutaires (50 champs)	Données de localisation géographique (10 champs)	Données financières (400 champs)	Présence Fonctions des dirigeants (130 booléens)
Etablissements principaux (6 715 287)					
Etablissements secondaires (818 879)					

Attributs de caractérisation des données

En plus des données qui sont des propriétés attachées aux entités, il faut ajouter des attributs qui permettent de caractériser les données, selon leur provenance, leur catégorie, etc.

Ces attributs supplémentaires sont aussi des critères de ciblage.

Visibilité, confidentialité :

- Donnée publique : donnée disponible par défaut pour tous les utilisateurs
- Donnée privée : données qui n'est accessible qu'à un certain nombre d'utilisateurs, et fait l'objet d'une gestion de confidentialité.

Provenance

- Indicateur sur la source : Insee, Rncs (IMR, immatriculations, modifications, radiations), Bodacc (bulletin des annonces civiles et commerciales)
- Comptes annuels INPI, BALO (bulletin des annonces légales officiel), autres

Périmètre (pour les données publiques)

- INSEE (SIRENE)
- RNCS (IMR)

Fiabilité, qualité des données

- indicateur sur le fait qu'une donnée soit officielle, publiée ou issue d'une enquête.

Types de données

Les données élémentaires (attributs) que le moteur devra manipuler auront les types suivants :

- codes,
- nombres entiers,
- nombres décimaux,

- booléens (indicateur de présence),
- dates,
- textes,
- liens (tables de clés représentant des liens fonctionnels (jointure) entre entités).

Les données manipulées sont essentiellement des données structurées dans le sens où elles sont au préalable analysées et organisées dans une base de données relationnelle. Cependant, il y a des informations qui sont des textes, comme la dénomination d'une entreprise, l'objet social d'une entreprise, le libellé de son activité.

Ces champs, bien que proprement découpés, sont une difficulté pour un moteur de ciblage, car un mot passé comme critère, n'est pas similaire à un code qui appartient à une liste de valeurs bine définies. Un critère bâti sur la recherche d'un mot doit en général pouvoir être étendu à quelques « variations » autour du mot : singulier / pluriel, orthographe, conjugaison, phonétique, etc. Nous verrons dans la dernière partie technique que c'est une contrainte importante pour l'architecture d'un moteur de ciblage qui sache traiter des données structurées et des données de type texte (dites semi-structurées).

3.3.4 Modalité des requêtes

3.3.4.1 Modalité des requêtes sur des champs codes

Une donnée codifiée peut être utilisée simplement dans une égalité ou dans un ensemble de valeurs.

Elle peut aussi être utilisée avec un opérateur « sinon » : si une valeur n'est pas présente alors en utiliser une autre (cascade de priorité sur les valeurs).

3.3.4.2 Modalité des requêtes sur des champs textes

Le moteur doit permettre d'effectuer des requêtes :

- à l'identique,
 - sur un ou plusieurs mots,
 - sur une expression
- à l'approchant
 - contient, débute par,
 - se termine par,
 - distance entre les mots recherchés,
- Phonétique (éventuellement multi-langue, pour l'extension de la solution à l'international).

3.3.4.3 Modalité des requêtes sur des champs de type « nombre »

Le moteur doit permettre d'utiliser des critères de type :

- quotient (ratios), formule plus complexe,
- somme, cumul,
- comparaison.

3.3.4.4 Modalité des requêtes sur des champs de type « date »

Le moteur doit permettre d'utiliser des dates en faisant :

- comparaison : inférieur, supérieur à une date donnée,
- intervalle entre deux dates,
- intervalle autour d'une date
- application de différents types de précision sur les dates : au jour, au mois, à l'année.

3.3.4.5 Récapitulatif des modalités des requêtes

Type de champ	Type de sélection						
	égal	inférieur	supérieur	intervalle	Contient	évolution temporelle	ratio formule
Codes							
Nombres							
Dates							
Textes							

3.3.5 Combinaison de requêtes élémentaires

Une des propriétés importantes, attendue du moteur est sa capacité à permettre une certaine interactivité avec l'utilisateur, c'est-à-dire sa capacité à permettre facilement (et de façon très performante) la prise en compte ou le retrait d'un ou plusieurs critères tout en retournant le comptage rapidement.

L'interactivité se concrétise dans « l'effet entonnoir ». Il signifie que l'utilisateur peut affiner son ciblage en construisant sa requête complexe pas à pas par l'assemblage de requêtes élémentaires.

L'effet entonnoir donne de plus la possibilité à l'utilisateur d'obtenir les comptages (partiels) pour chacun des critères élémentaires et les comptages progressifs obtenus après la prise en compte de chaque requête élémentaire supplémentaire. Ces comptages partiels doivent donc être présents dans le flux de retour.

Exemple :

	Requête élémentaire	Comptage (requête seule)	Comptage combiné
	Requête 1	100 000	100 000
Et	Requête 2	50 000	30 000
Et	Requête 3	80 000	20 000

Les liaisons entre les critères élémentaires d'une requête sont les opérateurs logiques : **et**, **ou**, **sauf** (analogues des opérations ensemblistes intersection, union, complémentaire).

Le détail des comptages par critères élémentaires permet à l'utilisateur d'ajuster son ciblage (compte requête seule du table).

Il faut aussi ajouter l'opérateur sinon entre critères. Cela veut dire que si un critère ne permet de cibler aucune entité alors, il faut lui substituer un autre critère.

Exemple : Recherche sur colonne A si non renseignée alors colonne B (à défaut) voir colonne C si B nulle.

3.3.6 Pré-comptage

La notion de pré-comptage concerne des requêtes avec une famille de critères qui

présentent plusieurs niveaux de granularité. C'est par exemple le cas des régions et des départements ; ou encore de la codification NAF à plusieurs niveaux des activités.

Le pré-comptage consiste donc, pour une requête qui porte sur un critère utilisé à un niveau donné de granularité, à compter aussi (lors de la même requête) à un niveau plus fin et / ou moins fin, de façon à retourner ce comptage plus fin à l'utilisateur avant qu'il ne le demande. Cette information est jugée pertinente avec sa demande initiale car elle permet à l'utilisateur d'anticiper éventuellement les étapes suivantes de son ciblage.

Exemples :

- une région sélectionnée / pré-comptage sur tous les départements de la région
- Extension coordonnées Lambert 2 (ville A avec extension aux 20Km autour).
- un département / pré-comptage sur la région et les départements de la région qui le contient

3.4 Typologie des requêtes

Une requête élémentaire est par définition un prédicat simple qui met en jeu un seul attribut ou une combinaison d'attributs (ratio entre deux grandeurs par exemple).

Un ciblage consiste donc pour l'utilisateur à définir une combinaison logique de requêtes élémentaires compréhensibles par le moteur.

Les paragraphes suivants décrivent des exemple type de combinaisons (plus ou moins complexes) de requêtes élémentaires.

Les cas types peuvent être grosso modo classés selon plusieurs critères parmi lesquels :

- type des données utilisé dans les critère (code, nombre, booléen, texte),
- pouvoir de discrimination du critère taille de la population résultante (influe notamment sur la performance),
- combinaison logique (et, ou, inclusion, exclusion, sinon),
- import,
- nombre de requêtes élémentaires combinées,
- complexité : mise en jeu d'une jointure,
- type de résultat attendu (comptage, répartition, liste).

Les exemples sont choisis de façon à couvrir le plus possible l'ensemble des cas significatifs ou jugés difficiles.

Signalons que les exemples de requêtes présentés ci dessous ont été formulées et utilisées pour « mesurer » les capacités et les performances des différents moteurs de ciblage du marché que nous souhaitons tester dans le cadre de l'évolution de l'outil Cofacibles. Dans la partie suivante, nous présenterons brièvement les offres des éditeurs que nous avons « testées ».

3.4.1 Combinaison avec périmètre importé – Fichier repoussoir ou d'inclusion

Le principe de la requête consiste à importer dans un premier temps une liste de clés (d'entités). La liste peut provenir par exemple d'un ciblage précédent, que l'utilisateur souhaite analyser ou étendre.

Cette liste peut donc servir de « liste d'inclusion » ou bien de « liste repoussoir / exclusion » dans un critère de ciblage.

Il faut noter que les clés importées peuvent être des SIREN ou des SIRET, donc concerner des entités de type entreprise ou établissement.

REQ I1 : comptage du nombre d'établissements présents dans la liste importée et étant de la région '24' (centre).

REQ I2 : comptage du nombre d'établissements des départements '37','18','28','41' et **n'étant pas** dans la liste l'importée.

REQ I3 : comptage et répartition par département du nombre d'établissements dans l'import et ayant une ancienneté de plus de 2 ans.

REQ I4 : comptage du nombre d'établissements n'étant pas dans l'import.

3.4.2 Présence des fonctions de dirigeants

REQ F1 : comptage des entreprises principaux (sièges) ayant au moins 1 famille de dirigeants renseignée dans une liste donnée de fonctions.

REQ F2 : comptage et répartition par département des entreprises (sièges ou établissements principaux) ayant

- au moins 1 dirigeant dans la famille 'INFORMATIQUE'
 - ou 1 dirigeant parmi les 'DIRECTEURS FINANCIERS'
 - ou 1 parmi les 'CREDIT MANAGERS'
 - ou 1 dirigeant parmi les 'DIRECTEURS COMMERCIAUX'
 - ou 1 dirigeant parmi les 'DIRECTEURS MARKETING'
- et étant présent dans un des départements suivants :
'75','76','77','78','92','93','94','95'

3.4.3 Recherche avec critère de type texte

Libelle code activité/BODACC et/ou texte libre (on entend par objet social le libellé code activité du BODACC potentiellement modifié par un opérateur, il s'agit alors d'un texte libre. L'idée est de cibler une population en utilisant un critère plus « approximatif » qu'une valeur numérique ou codifiée.

REQ T1 : comptage du nombre d'établissements dont l'objet social(*) contient le mot 'garage'

REQ T2 : comptage et répartition par département du nombre d'établissements dont l'objet social contient le mot 'garage'

REQ T3 : comptage du nombre d'établissements dont l'objet social contient le mot 'commerce' (*commerce est plus fréquent que garage dans l'objet social, donc ce critère cible une population plus large que la requête précédente*)

REQ T4 : comptage et répartition par département du nombre d'établissements dont l'objet social contient le mot 'commerce'

REQ T5 : comptage du nombre d'établissements dont l'objet social contient ('civile ET immobiliere') OU (,'voyage' ET 'agence')

REQ T6 : comptage et répartition par département du nombre d'établissements dont l'objet social contient ('civile' ET 'immobiliere') OU (,'voyage' ET 'agence')

REQ T7 : comptage du nombre d'établissements dont l'objet social contient le mot 'garage' et présents dans un département de la liste '75','76','77','78','92','93','94','95'

REQ T8 : comptage et répartition par département du nombre d'établissements dont l'objet social contient le mot 'garage' et présents dans un département de la liste '75','76','77','78','92','93','94','95'

REQ T9 : comptage du nombre d'établissements dont l'objet social contient le mot 'commerce' et présents dans un département de la liste '75','76','77','78','92','93','94','95'

REQ T10 : comptage et répartition par département du nombre d'établissements dont l'objet social contient le mot 'commerce' et présents dans un département de la liste '75','76','77','78','92','93','94','95'

REQ T11 : comptage du nombre d'établissements dont l'objet social contient ('civile' ET 'immobiliere') OU (,'voyage' ET 'agence') et présents dans un département de la liste '75','76','77','78','92','93','94','95'

REQ T12 : comptage et répartition par département du nombre d'établissements dont l'objet social contient ('civile' ET 'immobiliere') OU (,'voyage' ET 'agence') et présents dans un département de la liste '75','76','77','78','92','93','94','95'

Les requêtes suivantes mettent en jeu un critère de type texte et dépassent les possibilités du système actuel

*ATTENTION : Le caractère * doit remplacer 1 ou n caractères.*

REQ T13 : comptage du nombre d'établissements dont l'objet social contient les termes se suivant 'commerce équitable'

REQ T14 : comptage du nombre d'établissements dont l'objet social contient les mots 'commerc*' sauf 'commercial' (exclusion)

REQ T15 : comptage du nombre d'établissements dont l'objet social contient le terme 'cheval' avec sa lemmatisation (chevaux, etc...)

3.4.4 Recherche avec critères définis dynamiquement (ratio, formule)

Le moteur doit être capable d'effectuer des requêtes dont un critère (au moins) est basé sur des calculs à partir de formules définies par un utilisateur (donc non indexées),

- calculs mettant en jeu des champs d'un même document (enregistrement)
- calculs mettant en jeu des champs de deux documents successifs (calcul d'évolutions sur des données financières).

Voici quelques exemple :

REQ D1 : comptage du nombre d'établissements dont le chiffre d'affaires net en France de l'année en cours (poste FJ du bilan) a eu une progression de plus de 10% depuis 2 ans

REQ D2 : comptage et répartition par département et code APE (à 2 caractères) du nombre d'établissements dont le chiffre d'affaires net en France de l'année en cours (poste FJ) a eu une progression de plus de 10% depuis 2 ans

REQ D3 : comptage du nombre d'établissements dont le chiffre d'affaires net en France de l'année en cours (poste FJ) a eu une progression de plus de 10% depuis 2 ans et présents dans un département de la liste '75','76','77','78','92','93','94','95' et le code APE (2 caractères) dans la liste ('50','55')

REQ D4 : comptage et répartition par département et code APE (à 2 caractères) du nombre d'établissements dont le chiffre d'affaires net en France de l'année en cours (poste FJ) a eu une progression de plus de 10% depuis 2 ans et présents dans un département de la liste '75','76','77','78','92','93','94','95' et le code APE (2 caractères) dans la liste ('50','55')

REQ D5 : comptage du nombre d'établissements dont le chiffre d'affaires net total en France de l'année en cours (poste FL) est inférieur à celui daté de 2 ans

REQ D6 : comptage et répartition par département et code APE (à 2 caractères) du nombre d'établissements dont le chiffre d'affaires net total en France de l'année en cours (poste FL) est inférieur à celui daté de 2 ans

REQ D7 : comptage du nombre d'établissements dont le chiffre d'affaires net total en France de l'année en cours (poste FL) et celui daté de 2 ans sont renseignés

3.4.5 Recherche avec critères privés créés en dynamique

Le moteur doit permettre à un utilisateur, d'importer des données qui devront devenir disponibles pour les requêtes de l'utilisateur lui-même ou des personnes partageant ses données privées.

Précisions pour les données privées (notes)

Le format des fichiers importés doit rester le plus simple possible ou doit pouvoir provenir des logiciels de bureautique courant (format xml, format csv, format Excel par exemple).

Un fichier de données doit évidemment contenir un identifiant fonctionnel compatible avec ceux qui existent déjà dans le moteur pour les données publiques. Il est obligatoire de pouvoir rattacher le commentaire à une entité (entreprise ou établissement).

Les zones textes du style commentaire doivent avoir une longueur maximum de 2000 caractères.

Le dessin d'enregistrement du flux d'importation sera statique afin de limiter l'incidence sur les performances du ciblage mais chacun de ces éléments pourra être un objet de ciblage (critère)

Identifiant(SIREN), date (DDMMYYYY), numérique_1 ; numérique_2 ; numérique_3 ;
texte_1 ; texte_2 ; texte_3

Exemple :

Les données privées consistent par exemple en la fourniture d'un encours (de 0 à 100000) sur une population de 10000 sirets (données que nous appellerons ENCOURS → correspondance avec numerique_1).

Un autre ensemble de données privées peut aussi être une note personnelle (texte de 0 à 20 mots dont environ 50% des notes contiennent le mot CREDIT → correspondance avec texte_1) sur une population de 10000 SIRET (données que nous appellerons NOTE).

Il s'agit donc pour le moteur de permettre à un utilisateur de lancer une requête aussitôt après chargement des données privées.

Ces données doivent également apparaître dans la liste de résultat.

REQ P0a : mise « on line » de la donnée ENCOURS (prête à être requêtée).

REQ P0b : mise « on line » de la donnée NOTE (prête à être requêtée).

REQ P1 : comptage du nombre d'établissements présents dans les départements de la liste : '06','37','75','76','77','78','92','93' et dont l'encours est compris entre 0 et 1000

REQ P2 : comptage du nombre d'établissements dont le code APE (sur 4 caractères) n'est pas dans la liste :

'202Z','454J','454L','524C','527D','554C','701F','711Z','721Z','748G','752E','804A', et dont la NOTE contient le mot 'CREDIT'

REQ P3 : comptage et répartition par région et département du nombre d'établissements présent dans les départements de la liste : '06','37','75','76','77','78','92','93' et dont l'encours est compris entre 0 et 1000.

3.4.6 Séquence de requêtes (une démarche de ciblage)

REQ S1 : comptage du nombre d'établissements ayant au moins 1 dirigeant renseigné dans la famille 'DIRIGEANTS'

REQ S2 : comptage du nombre d'établissements étant présent dans un département autre que '01','07','26','38','42','69','73','74'

REQ S3 : comptage du nombre d'établissements avec un ET entre S1 et S2

REQ S4 : comptage du nombre d'établissements dont le code APE (sur 4 caractères) n'est pas dans la liste :

'202Z','454J','454L','524C','527D','554C','701F','711Z','721Z','748G','752E','804A',
'851L','853K','913C','913E','921A','921B','921C','921D'

REQ S5 : comptage du nombre d'établissements avec un ET entre S3 et S4

REQ S6 : comptage du nombre d'établissements dont l'effectif est entre 1 et 100000.

REQ S7 : comptage du nombre d'établissements avec un ET entre S5 et S6

REQ S8 : comptage du nombre d'établissements qui n'ont pas de fax renseigné.

REQ S9 : comptage du nombre d'établissements avec un ET entre S7 et S8

REQ S10 : comptage et répartition par région et département du nombre d'établissements dont l'effectif est entre 1 et 100000.

REQ S11 : comptage et répartition par département, code APE (2 caractères) et tranche de chiffre d'affaires du nombre d'établissements dont le code APE est dans la liste : '50','55' et la tranche de CA est 102 ou 106

3.4.7 Liens financiers

Le but de ces requêtes est d'illustrer l'utilisation des liens financiers comme une jointure entre deux stratégies : un ensemble de requêtes (une première stratégie) pour cibler les actionnaires et un autre ensemble de requêtes (une deuxième stratégie) pour cibler les filiales.

REQ L1 : comptage des entreprises (actionnaires) de la région Ile de France dont l'effectif est supérieur à 1000

REQ L2 : comptage des entreprises (filiales) de la région Bretagne dont l'activité est dans la liste '202Z','454J','454L','524C','527D','554C','701F','711Z','721Z','748G','752E','804A'.

REQ L3 : L1 et L2 liés par une jointure (qui n'est pas un opérateur **et**)

Version actionnaire : comptage des entreprises de la région Ile de France dont l'effectif est supérieur à 1000 et **qui sont actionnaires** d'entreprises situées dans la région Bretagne et dont l'activité est dans la liste '202Z','454J','454L','524C','527D','554C','701F','711Z','721Z','748G','752E','804A' .

REQ L4 : L1 et L2 liés par jointure

Version filiale : comptage des entreprises situées dans la région Bretagne et dont l'activité est dans la liste '202Z','454J','454L','524C','527D','554C','701F','711Z','721Z','748G','752E','804A' et **qui ont pour actionnaire** une entreprise de la région Ile de France dont l'effectif est supérieur à 1000.

3.4.8 Requêtes en parallèle - Jointures

Le moteur de ciblage doit être capable d'effectuer un ciblage au travers de la combinaison (jointure) des résultats de requêtes effectuées (en parallèle) sur différents index (un index des événement Bodacc, un des index données financières, un index des données entreprises, un index liens capitalistiques, un index des dirigeants...),

Le résultat de la requête finale peut donc nécessiter la combinaison des résultats de requêtes plus élémentaires dont les clés n'ont qu'une partie commune (jointures 1-N ou N-N); ex : une liste de SIRET à combiner avec une liste de SIREN ou d'événements Bodacc.

Exemple :

Trouver les entreprises de la région Centre dont la forme juridique est SA et qui ont été l'objet d'un procédure collective ces 5 dernières années ; cette requête met en jeu la « table » des données statutaires des entreprises et la table des événements judiciaires

(Bodacc par exemple).

3.4.9 Liste de résultats attendues du moteur

Les résultats affichés sont fonctions de l'étape à laquelle se situe l'utilisateur dans le processus de ciblage :

- comptage
 - affichage des compteurs partiels pour chaque critère
 - compteurs résultat de l'entonnoir
 - pré-comptages éventuels,
 - compteur total
- répartition (affichage du tris à plat, ou du tableau croisé),
- liste de résultats (liste limitée d'attributs personnalisable)
- produit plus riche en information sur les entités (liste de résultats également)

On distingue 2 types de flux, la partie comptage et la partie extraction des entreprises ciblées

Pour ce qui est de l'extraction, il faut évaluer les deux cas de figures suivants :

1. construction de la liste des enregistrements directement par le moteur;
2. exploitation d'une liste de clés retournée par le moteur permettant de construire la liste depuis les tables oracle existantes (ou remaniées).

1 - Comptage

La première phase consiste à retourner un flux à l'opérateur lui indiquant le détail de population ciblée par chacun de ses critères. Voir également 2.8.6 Séquence de requêtes

- affichage des compteurs partiels pour chaque critère
 - N1 entreprises correspondent au critère 1 (ex Effectifs > 250 et < 500)
 - N2 entreprises correspondent au critère 2 (Ex Region : Ile de France)
 - N3 entreprises correspondent au critère 3 (Exemple CA > 10 000 000 K€)
- pré-comptages éventuels,
 - Extension de N2 aux départements limitrophes
- compteurs résultat de l'entonnoir
 - N4 entreprises correspondant à la combinaison des critères 1 et 2
 - N5 entreprises correspondant à la combinaison des critères 1 et 2 et 3 (résultat du ciblage)
- compteur total
 - N5 entreprises correspondant à la combinaison des critères 1 et 2 et 3 (résultat du ciblage)

Le pré-comptage consiste, par exemple, pour le moteur, à pré-compter les entités pour chacun des départements qui appartiennent à une région choisie par l'utilisateur comme l'un de ses critères. Cela aide l'utilisateur à anticiper un ciblage plus fin.

2 - Extraction par clés

Le moteur doit être en mesure d'extraire les clés des entités ciblées.

Il faut un moteur capable de retourner les clés d'une population ciblée de 1 000, 10 000, 100 000, 1 000 000 entités.

3 - Extraction d'enregistrements

Le moteur doit être en mesure d'extraire des enregistrements (liste de champs à extraire) liés aux entités ciblées par une requête, dans la mesure où passer par un système

d'extraction auxiliaire ne serait pas assez réactif.

Précisions sur la liste de résultats

Les attributs qui apparaissent dans la liste de résultats sont choisis par l'utilisateur, donc imposés au moteur ou filtrés par l'application (ce qui suppose que l'ensemble des données sont retournées) au moment de la demande d'export.

3.4.10 Tris à plat, tris croisés

Le moteur doit permettre la réalisation de de tris à plat, de tris croisés, avec choix des valeurs (ou tranches de valeurs) de répartition par l'utilisateur au moment de sa demande.

Précisions sur les répartitions

Une répartition est aussi appelée segmentation.

Les dimensions sur lesquelles sont effectuées les répartitions peuvent être :

- les attributs correspondant à des données codifiées,
- des données numériques réparties sur un nombre déterminé de tranches de valeurs,
- des dates, réparties en un nombre déterminé de périodes.

Le nombre de dimensions est limité à 3 pour les répartitions.

Les tranches de valeurs sont fixées par défaut.

3.4.11 Mise à jour des données et des index

Les données sont régulièrement mises à jour : la fréquence habituelle est hebdomadaire, mais doit pouvoir devenir quotidienne au besoin. Cela impose une forte contrainte sur la durée des mises à jour. Le temps de reconstruction du ou des index supérieur à 24h poserait en effet des difficultés.

Le moteur doit donc être capable de mettre à jour ses index de façon transparente pour l'utilisateur. Et cela, sans dommage sensible pour les performances et sans dommage sensible non plus pour leur exploitabilité (volume, fragmentation, etc...).

Pour ce qui est de la partie bilan, le ciblage s'effectue sur les 5 dernières années (et toutes les entreprises ne publient pas en même temps), on devra donc gérer une « suite d'exercices glissante » (FIFO) pour chaque entreprise : le plus vieux bilan disparaît au profit du nouveau.

3.4.12 Précisions sur les fonctions de calculs

3.4.12.1 Contexte

3.4.12.1.1 Ciblage à partir de critères quantitatifs

Le but du ciblage est de cerner et de compter, de façon exhaustive une population qui répond à certains critères.

S'agissant de cibler des entreprises ou des établissements, les données quantitatives utilisables en tant que critères sont :

- soit directement des champs :
 - effectif de l'entreprise,

- effectif de l'établissement,
- postes du bilan d'une entreprise,
- ratios ;
- soit des valeurs calculées,
 - opérations complexes mettant en jeu un ou plusieurs champs.

Cette liste n'est pas exhaustive au regard du périmètre complet des données pour le projet, mais elle est une base de travail pour mettre au point les fonctionnalités attendues.

3.4.12.1.2 Organisation des données - Dimension temporelle

Contenu

Le fond de base de Cofacibles (périmètre = France) consiste en deux « tables » :

- une table ENT_MARKET (dans le langage Cofacibles)
 - contient les données des établissements
 - 6,5 millions d'enregistrements
 - 41 champs
 - peu de valeurs quantitatives :
 - le capital de l'entreprise
 - l'effectif de l'établissement
 - l'effectif de l'entreprise
- une table FIN_MARKET (dans le langage Cofacibles)
 - contient les données des entreprises qui publient leurs Comptes Annuels
 - 900 000 enregistrements
 - 273 champs
 - beaucoup plus de valeurs quantitatives
 - les mêmes que ci-dessus
 - 51 postes des Comptes Annuels (sur les 3 dernières années)
 - 18 ratios issus de l'analyse financière (sur les 3 dernières années)

Utilisations

La table FIN_MARKET est la plus propice aux tests de calculs car elle contient suffisamment de champs numériques pour permettre des calculs complexes mettant en jeu plusieurs champs numériques.

Dimension temporelle

Les données de la table FIN_MARKET présentent une dimension temporelle. En effet, chaque enregistrement contient 51 des postes des Comptes Annuels des 3 dernières années. La table, telle qu'elle est présentée aujourd'hui dans Cofacibles, est donc « dé-normalisée » au sens des normes relationnelles.

Selon la technologie utilisée, il sera peut être intéressant de re-normaliser la table FIN_MARKET de façon à disposer d'un enregistrement par année de Compte Annuel et par entreprise. La table résultante contiendra donc $3 \times 900\ 000 = 2\ 700\ 000$ lignes de 113 champs, au lieu de 900 000 lignes de 273 champs.

Cependant, dans la base de données « cible », il est d'ores et déjà prévu de prendre en compte de 5 à 10 années de Comptes Annuels. La disposition « normale », comprenant un enregistrement par année, paraît a priori, plus commode à construire et à mettre à jour. Il est resté à étudier si les calculs sont plus faciles ou non.

Valeurs non renseignées

Il faut noter que toutes les valeurs de tous les champs ne sont pas renseignées. Il en est de même pour les champs numériques.

D'où la problématique :

En cas de valeur absente, quel doit être le comportement du moteur ?

- Ignorer l'enregistrement,
- Calculer une interpolation ou extrapolation à la volée,
- Prévoir de faire les interpolations ou extrapolations en amont (au chargement).

Dans une première approche, pour simplifier, les tests de calculs peuvent donc être effectués en ignorant, au cours de chaque requête, les enregistrements dans lesquels au moins une valeur impliquée dans un calcul est absente.

3.4.12.2 Opérations

Les exemples cités sont des indications. Tout test de complexité équivalente est acceptable.

Les exemples suivants s'appliquent essentiellement aux données de la table qui contient les données financières (Comptes annuels).

3.4.12.2.1 Utilisation directe des valeurs

Cette famille de tests consiste à utiliser un critère dans lequel une ou plusieurs valeurs quantitative (d'un même enregistrement) sont mises en jeu directement. La (les) valeur(s) en question est (sont) un (des) champ(s) choisis parmi ceux cités plus haut.

- la valeur est **inférieure** ou **supérieure** à une **constante**
 - ex : compter les entreprises dont le dernier CA (poste FL du bilan) est supérieur à 500 000 euros
- la valeur est **inférieure** ou **supérieure** à une autre **valeur**
 - ex : compter les entreprises dont le CA export (poste FK) est supérieur au CA France (poste FJ) pour les Comptes annuels de la dernière année.
- la valeur est comprise dans un **intervalle absolu**
 - ex : compter les entreprises dont le dernier CA est compris entre 500 000 et 1 000 000 euros
- intervalle relatif
 - ex : compter les entreprises dont le dernier CA est de 1 000 000 euros à plus ou moins 10%

3.4.12.2.2 Utilisation d'opérations

Il s'agit de tester les capacités du moteur de comptage à utiliser un critère défini, **dynamiquement**, au moment de la requête.

- **somme** de deux (ou plus) valeurs
 - ex : compter les entreprises dont la somme des prêts (poste BF) et des immobilisations (poste BH) est supérieure aux fonds propres (poste DR)
- **rapport** entre deux valeurs d'une même année
 - ex : compter les entreprises dont le rapport entre le CA export (poste FK) et le CA France (poste FJ) est supérieur à 20 %
- **formule plus complexe**

- ex : compter les entreprises dont la somme des dettes (DU+DV+DX+DY) moins le capital social (DA), divisé par le résultat d'exploitation (GG) est supérieur à 5 %

3.4.12.2.3 *Utilisation de variations*

Il s'agit d'utiliser un critère d'évolution, en exploitant la dimension temporelle des données.

- **évolutions qualitatives** dans le temps
 - ex : compter les entreprises qui sont passées de pertes à bénéfice (HN) entre les deux derniers exercices
- évolutions quantitatives
 - ex : compter les entreprises dont le chiffre d'affaires (FL) a progressé d'au moins 5 % entre les deux derniers exercices

Comme proposé plus haut, les calculs d'évolutions temporelles peuvent être effectués :

- sur des enregistrements qui contiennent les valeurs sur plusieurs années (disposition par défaut de la table FIN_MARKET) ; les évolutions sont donc effectuées entre champs d'un même enregistrement ;
- ou sur des enregistrements différents qui contiennent seulement une année de valeurs ; les évolutions sont donc calculées entre champs d'enregistrements différents (par exemple, le CA est pris dans l'enregistrement de l'année N et dans l'enregistrement de l'année N-1 pour faire le calcul).

Le test du deuxième cas nécessite de reconstruire une nouvelle table à partir de la table FIN_MARKET fournie. Le principe est de « découper » chaque enregistrement contenant 3 années de Comptes Annuels en 3 enregistrements ne contenant que les postes des Comptes Annuels d'une année, en reportant dans chaque nouvel enregistrement tous les champs d'entête caractérisant l'entreprise.

3.5 **Autres besoins liés au moteur**

Les besoins décrits ici sont liés au moteur, mais ne sont pas de sa responsabilité directe.

3.5.1 **Session utilisateur, sécurité d'accès aux données**

Deux aspects fonctionnels

- la gestion des droits (utilisateurs, profils, groupes) :
 - Le moteur doit utiliser un système de gestion des droits compatible avec le système appelant.
- l'utilisation, la vérification des droits au moment des requêtes : c'est de la responsabilité du moteur

Caractéristiques des droits d'accès

- Granularité des droits d'accès
- Utilisateur : user/password à valider/confirmer au niveau du moteur
- Permettre un accès à des utilisateurs anonymes
- notion d'administrateur (doit exister au niveau du moteur – accès configuration et paramétrage)
- Notion de profil utilisateur (profil : un ensemble de droits partagés par un groupe d'utilisateurs)

- accès aux données privées
- périmètres
- restriction sur les données publiques
- + fonctionnalités (critères) + produits...
- Groupes d'utilisateurs, partageant les mêmes droits d'accès aux données

Droits d'accès et types de données

- restrictions sur l'accès aux données publiques (liées à un profil par exemple)
- cloisonnement des données privées (liées à un groupe d'utilisateurs, donc à un profil).

Typologie des droits d'accès

- Filtrage des critères sélectionnables ou non
- Filtrage des champs des listes de résultat, rapport, organi-groupe, etc...
- Stratégie : expressions booléennes ou non, « sauf » sur les critères, etc...
- Critères accessibles directement ou après « achat » .
- Données exportables,
- Données importables.

3.5.2 Espace de travail utilisateur

Gestion d'un espace de travail pour chaque utilisateur, sécurisé (étanche), le temps d'une session. L'espace de travail contient le résultat de chaque requête élémentaire de façon à pouvoir les combiner à volonté.

Un espace utilisateur ne peut devenir persistant entre deux sessions que s'il est associé à une stratégie.

L'espace de travail est aussi lié fortement à la notion d'interactivité et d'effet entonnoir qui consiste à pouvoir réutiliser le résultat d'une requête pour la combiner avec une suivante.

3.5.3 Stratégie

Définition : ensemble de requêtes élémentaires connectées par des opérateurs logiques.

Aspects de la gestion des stratégies :

- Plusieurs stratégies par utilisateur
- Gestion d'un stock de stratégies (gestion de l'archivage des stratégies)
- Droits d'accès aux stratégies
- Partage de stratégies
- Durée de conservation
- Gestion par le moteur ou par module auxiliaire => combinaisons de stratégies
- Exécution différée
- Surveillance.

3.5.4 Combinaison de stratégies

Il s'agit ici de généraliser le concept de combinaison entre stratégies au-delà de l'intersection et de la jointure sous tendue par la notion de lien financier. Il est permis d'imaginer une jointure de type lien commercial par exemple.

Une combinaison de stratégies est l'équivalent d'une combinaison d'ensembles des requêtes.

Il est possible de généraliser les combinaisons de stratégies :

- intersection directe (équivalent d'un « et » logique)
- union de stratégie (équivalent d'un « ou » logique)
- complément de stratégie (équivalent du concept de complémentaire d'un ensemble)
- en utilisant une table de jointure (de type lien).

3.5.5 Surveillance, veille

Le moteur faciliter la mise en place d'un mécanisme de veille permettant de détecter les entités entrantes, les entités sortantes et les entités dont les données ont été modifiées.

Typologie des surveillances

- entrants / sortants dans la population totale,
- sortants du périmètre attaché à un périmètre utilisateur (sous ensemble de la population totale),
- modifications des données,
- surveillance de stratégies.

3.5.6 Export des données

Deux types d'export :

- export de données présentes dans le moteur, ou des données déterminées à partir des données du moteur (comptages, formules, etc.)
- export de produits plus riches (qui contiennent des données liées aux entités, mais qui sont absentes du périmètre des données marketing présentes dans le moteur de ciblage).

Le premier point est de la responsabilité du moteur.

Le deuxième point est de la responsabilité d'un système auxiliaire.

Dans le premier cas, l'extraction des données est de la responsabilité du moteur.

Les champs exportés à partir du moteur sont choisis par l'utilisateur.

3.6 Performances attendues, charge attendue

Points a priori sensibles pour la performance :

- Requêtes élémentaires
- Requêtes combinées
- Requêtes sur le texte
- Requêtes sur des critères dynamiques (calcul de formule à la volée)
- Taille des périmètres repousseurs (exclusions)
- Ré-exécution de stratégies en masse (processus de surveillance)
- Volume total des données (fond de base)
- Volume des données ciblées (volume de la population ciblée, 1000, 10 000, 100 000, 1 000 000)
- Nombre de requêtes, nombre d'utilisateurs simultanés
- Jointures 1-N, N-N

3.7 Spécifications liées à l'exploitation et à l'administration du moteur

3.7.1 Disponibilité

Le moteur doit permettre une disponibilité :

- 24/24 heures,
- 7/7 jours.

Le moteur doit offrir un mécanisme pour surveiller sa disponibilité : par un « robot » doit pouvoir déterminer son état de façon à alerter un humain en cas de défaillance.

3.7.2 Besoins de suivi d'activité du moteur

Les indicateurs d'activité doivent être produits par le moteur, mais leur présentation dans une interface homme-machine, n'est pas de sa responsabilité.

3.7.2.1 Indicateurs

Possibilité d'avoir des indicateurs en temps réel (mises à jour) sur :

- le contenu de la base
 - taux de renseignement des données
- son utilisation
 - nombre de requêtes
 - nombre d'utilisateurs
 - nombre de stratégies

3.7.2.2 Journalisation

Journal avec temps CPU consommé :

- des sessions utilisateurs,
- des requêtes (utilisation des différents attributs),
- des stratégies,
- des imports,
- des exports.

3.7.2.3 Statistiques

Le moteur doit permettre d'obtenir des statistiques sur son utilisation et sur le comportement des utilisateurs :

- composition des stratégies,
- taux d'utilisation des données dans les critères,
- démarche des utilisateurs (succession des critères),
- analyse de la nature des populations ciblées (taille, caractéristique).

4 Conception du moteur de ciblage

Dans la partie précédente, nous avons présenté un ensemble très riche de besoins. A l'heure actuelle, Coface Services ne dispose pas d'une solution de ciblage qui y réponde en totalité. Y parvenir lui apporterait un avantage concurrentiel certain. Coface Services dispose aujourd'hui de la solution Cofacibles, qu'elle souhaite faire évoluer.

Dans cette partie, plus technique, nous proposons une présentation des solutions techniques retenues pour Cofacibles. Conscients des limites de ces solutions techniques, nous avons participé à la recherche d'un moteur de ciblage sur le marché. La réponse n'est pas positive. Elle n'est pas complètement négative non plus. Nous verrons pourquoi. Enfin, nous proposons un certain nombre d'idées, de pistes techniques que Coface Services pourrait développer « en interne » faute de trouver un logiciel complet sur le marché.

4.1 Synthèse des besoins

Le moteur de ciblage doit

- savoir traiter des requêtes sur des données de types
 - données structurées,
 - textes (recherches avec variations, « à l'approchant »);
- gérer des jointures entre populations de nature différentes, mais qui ont un élément de clé en commun;
- gérer des gros volumes (plusieurs million d'entités);
- être interactif (définition progressive d'une cible, « en entonnoir »);
- être performant, c'est à dire savoir répondre à une requête à un seul critère en en 100 milliseconde, typiquement ;
- supporter des mises à jour hebdomadaire, mais qui durent moins de 24h;
- savoir faire des calculs et opérations complexes ;
- ne pas avoir un coût prohibitif ;
- pouvoir être interfacé avec les composants existants du système d'information de Coface Services.

Nous pouvons déjà énumérer certaines contraintes ou difficultés :

- volumes : flux, temps de traitement (global < 24 heures),
- exigence de performances : interactivité,
- jointures,
- données de type texte,
- formules de calculs définies par l'utilisateur.

4.2 Solution actuelle : Cofacibles

Cofacibles est le nom de l'application en ligne que Coface Services propose à ses clients pour faire du ciblage marketing. Cofacibles a hérité de la (longue) expérience que notre entreprise a acquise pendant des années sur cette activité. Quelques années avant le rachat par Coface, nous avons participé à l'élaboration de la solution technique qui est devenue le socle de Cofacibles.

4.2.1 Présentation

Nous avons participé à l'élaboration de la solution actuelle dans plusieurs de ses aspects. Nous avons analysé et développé les éléments relatifs au ciblage marketing qui implique des critères financiers. Nous avons aussi développé le module de chargement en base de données d'un fichier téléchargé par l'utilisateur sur le serveur web de Coface Service (module dit « import »), au travers de l'application Cofacibles.

Architecture

L'architecture physique du sous système d'information mis en lace pour Cofacibles reprend une organisation classique en « trois tiers » :

- un niveau de serveurs web montés en « load balancing » pour gérer de façon fluide le flot des requêtes;
- un niveau de serveurs intermédiaires qui contiennent les serveurs logiciels de type « middleware »;
- enfin un serveur de bases de données, qui contient les données de référence et les tables indexées spécialement pour le ciblage.

Le schéma suivant fait apparaître aussi bien les utilisateurs-clients que les utilisateurs internes. Ils bénéficient du même moteur d ciblage, mais ces derniers ont un accès privilégié du point de vue de l'infrastructure physique.

Les communications entre les différents niveaux sont sécurisées par ce que nous avons représenté comme des firewalls (pare-feux). La réalité est un peu plus complexe, mais l'idée est bien de filtrer les accès aux différents niveaux du système.

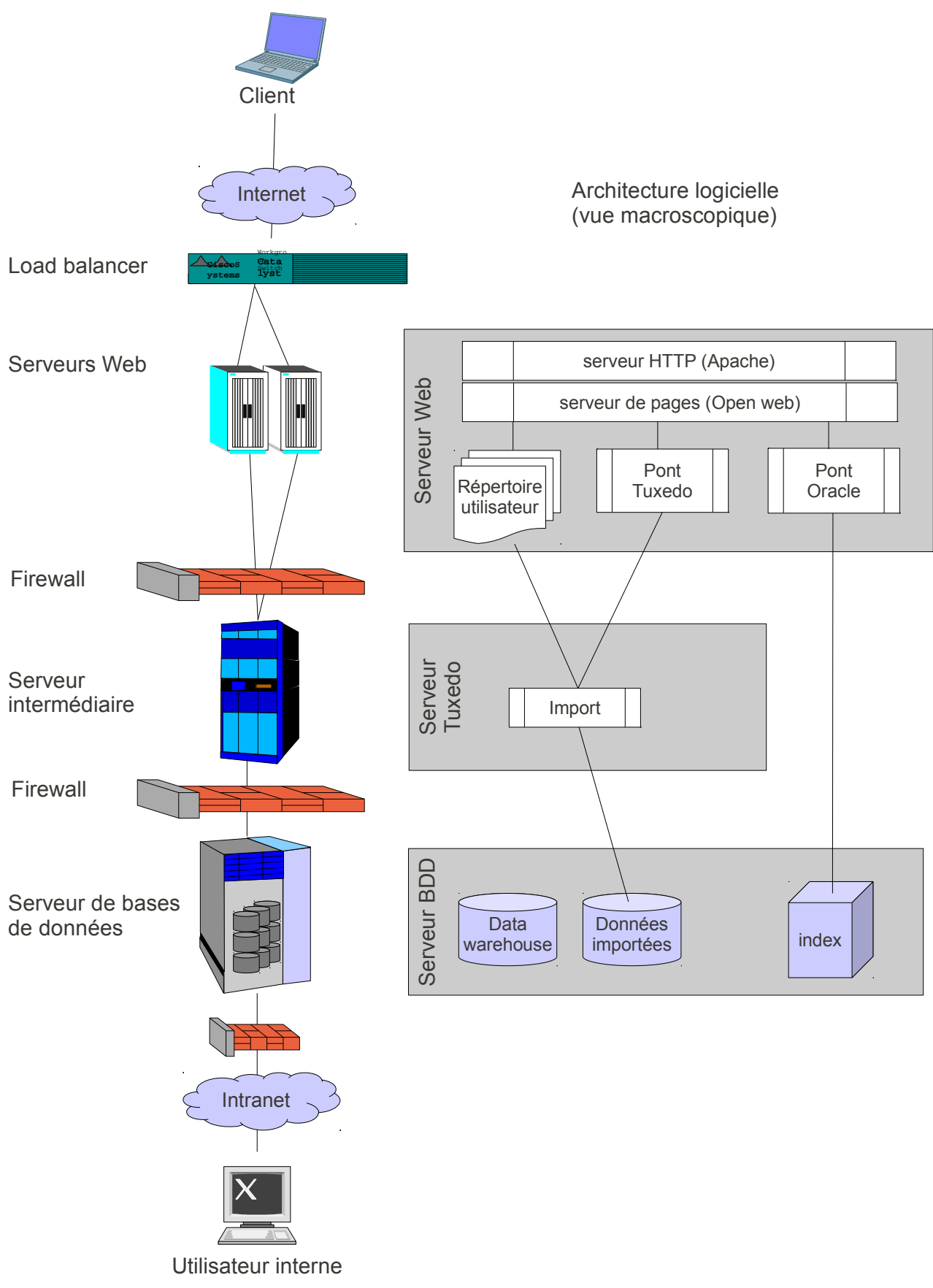
En parallèle de l'architecture, nous avons représenté l'architecture logicielle à grande échelle. Elle fait apparaître les communications (« ponts ») entre l'application web et les différentes ressources dont les index qui sont au cœur du moteur de ciblage, et les services intermédiaires. Le serveur d'application actuel est la plateforme Openweb, développée par la société Celya. Dans son principe et dans son fonctionnement, cette plateforme est analogue à une plateforme PHP, mais elle est apparue sur le marché bien avant que PHP ne soit stable. Celya proposait d'ailleurs déjà une plateforme pour le développement des applications minitel.

Openweb communique avec les ressources par des ponts logiciels : il faut un pont pour Oracle et un pont pour Tuxedo (édité par BEA). Tuxedo est l'environnement logiciel qui permet de développer des processus serveurs assez simplement : Tuxedo prend en charge tous les aspects communication inter-processus ou la gestion les communications a au travers réseau. Tuxedo facilite la conception d'applications distribuées. Tuxedo est utilisé par Coface Services pour ses services intermédiaires (middlewares), dont ceux utiles à Cofacibles :

- import de fichier utilisateur dans le moteur de ciblage,
- enrichissement pour la fabrication des produits marketing.

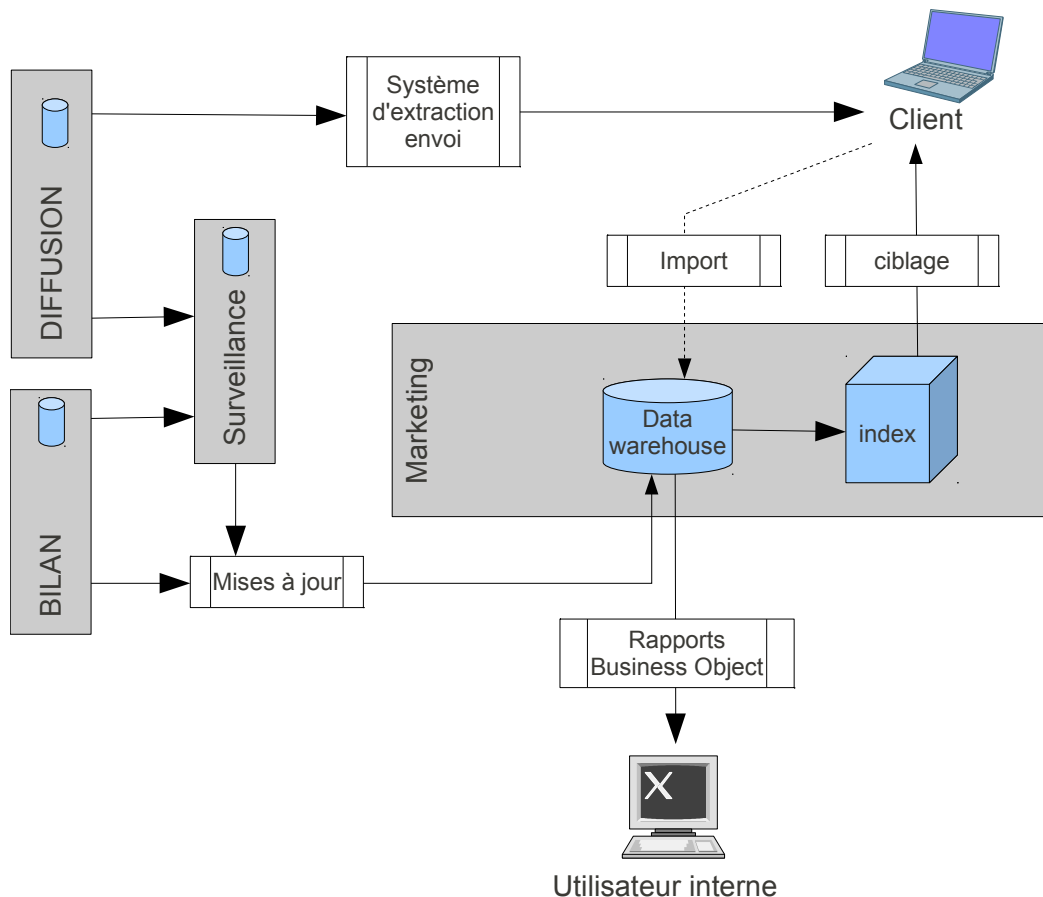
Le répertoire utilisateur héberge un fichier de clés (SIREN, SIRET) déposé en ligne, par le client, sur nos serveurs web, le temps qu'il soit importé dans le moteur de ciblage. Le contenu sert d'inclusion ou de repoussoir dans les requêtes de l'utilisateur.

Les requêtes de comptage sont communiquées directement au serveur de base de données. La solution technique que nous avons retenue est en effet basée sur une instance Oracle dont l'organisation est présentée plus en détail plus loin.



Flux

Le moteur de ciblage est alimenté chaque semaine pour les modifications qui ont été détectées, et compilées, durant la semaine écoulée, sur la base de données de référence. La partie entrepôt de données (data warehouse du schéma) est alimentée par incréments, seules les modifications sont appliquées. Les index (dont nous verrons le détail ci dessous) sont par contre reconstruits complètement à chaque fois à partir des données de l'entrepôt. Quelques heures suffisent pour la mise à jour complète du système qui est effectuée chaque nuit, du samedi au dimanche.



L'entrepôt de données sert aussi de source d'informations pour les opérateurs internes qui réalisent des rapports statistiques (outil Business Object) sur le contenu des bases : statistiques d'évolution de la répartition des scores @rating (donc de solvabilité), région par région, par exemple.

Le flux « import » du schéma correspond à la possibilité d'importer un fichier client dans le système de ciblage : cette population (« périmètre utilisateur » dans notre jargon) est ensuite utilisée comme périmètre d'inclusion ou comme repoussoir pour ses requêtes de ciblage.

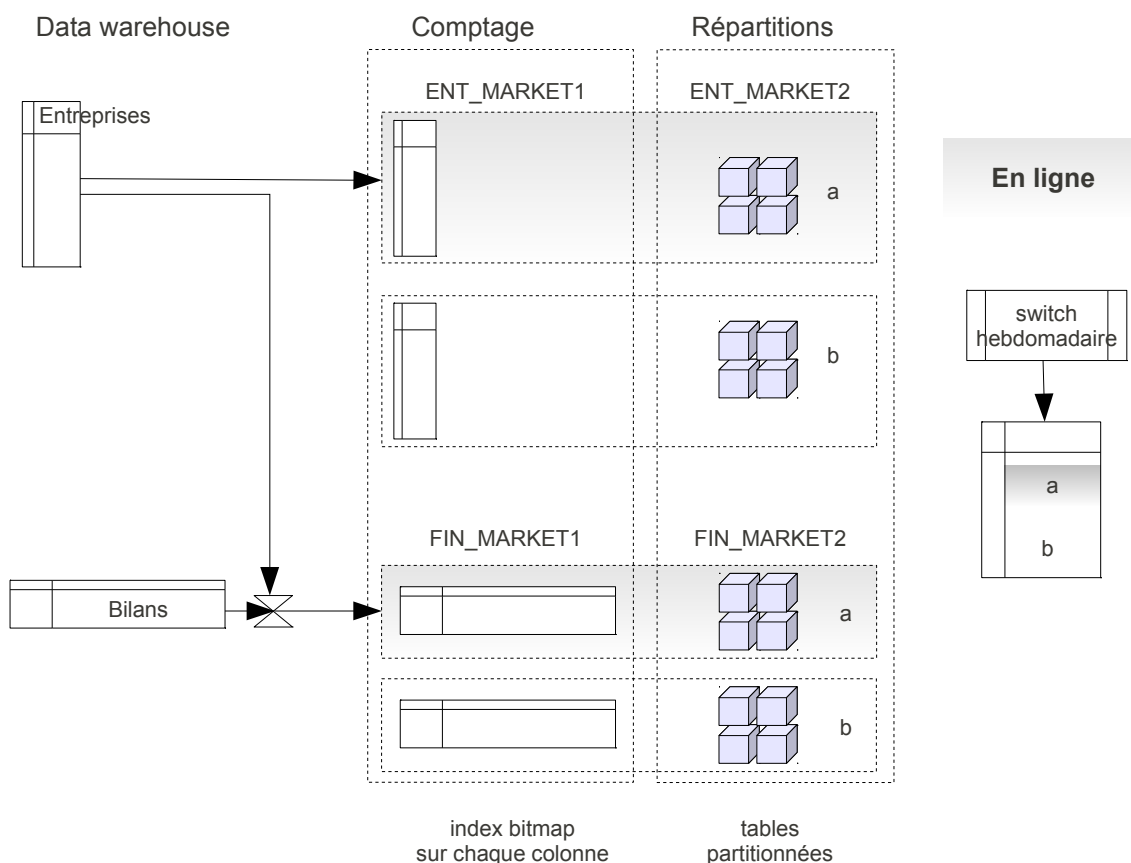
Organisation de l'instance Oracle

Le moteur de ciblage de Cofacibles est basé sur une instance Oracle, en version 8.1.7.4 au départ, et en cours de portage en version 11gR2. Nous avons défini deux schémas : un pour les données générales des entreprises, un autre pour les données financières. Dans chaque cas, il y a deux types de tables : une table pour les ciblage de type comptage, une autre pour le calcul des répartitions. En effet, il est apparu que les comptages et les répartitions (à une ou deux dimensions) ne pouvaient pas être effectués

sur la même table. La présence d'index de type bitmap Oracle, trompe l'optimiseur de requêtes (élément essentiel du noyau Oracle) et les durées des requêtes de répartition sont trop longs (plus d'une minute).

Les tables utilisées pour les comptages (ENT_MARKET1 et FIN_MARKET1) sont effet indexées de façon particulière : toutes les colonnes sont indexées avec un index de type bitmap (vecteur de valeurs binaires). Nous expliquons en détail plus loin cette technologie. Signalons simplement qu'elle simplifie énormément un comptage : les données sont regroupées en familles, dont chacune répond au critère simple du type « attribut = valeur ». Une entité appartient évidemment à autant de familles qu'il y a d'attributs. Par contre, les index bitmap implémentés par Oracle sont beaucoup moins performants pour les demandes de répartition. Pour répondre à ce besoin, nous avons été amenés à mettre en place un autre type de table.

Les tables (ENT_MARKET2 et FIN_MARKET2) sont des tables sans index (ni index B-tree classique, ni index bitmap). Mais elles sont découpées en partitions selon deux paramètres qui sont les attributs les plus souvent sollicités : le département et le code activité. Les données sont ainsi découpées en blocs plus petits (d'où le dessin en cubes sur le schéma suivant). Ils peuvent parcouru et comptés indépendamment par les processus internes du serveur Oracle. Cette organisation favorise la mise en œuvre du parallélisme dans le noyau du serveur Oracle. En procédant ainsi, une répartition sur deux critères au choix de l'utilisateur, prend une dizaine de secondes, ce qui acceptable. Par contre il faut permettre au noyau Oracle d'utiliser simultanément quatre processeurs sur un serveur Fujitsu PrimePower 2500, pour obtenir cette performance.



Sur le schéma, chaque famille de tables existe en deux exemplaires. Cette disposition permet de gérer les mises à jours des « index » sans couper l'application. Il y a toujours un jeu de tables en lignes pour recevoir les requêtes. Les mises à jour sont appliquées au

jeu de tables qui ne sont pas en ligne. Une fois celles-ci terminées et tous les index bitmap recalculés, un mécanisme de bascule (switch dans notre jargon) est déclenché pour mettre en ligne, les tables à jour.

Les tables mise hors lignes sont à leur tour mises à jour (pas besoin des index), en attendant la semaine suivante, où elles subiront les mises à jours les premières. Toute cette procédure est automatique et est gérée par l'ordonnanceur des tâches, commun à toutes les applications de l'entreprise.

L'application web ne sait interroger que la vue qui « encapsule » le nom de la table réellement en ligne. Le mécanisme de switch lui est masqué. En fonctionnant ainsi, il n'y a pas de rupture de disponibilité de l'application de ciblage.

Organisation des données

Les données sont organisées dans deux tables. Les informations générales de toutes les entreprises françaises et de leurs établissements sont dans une famille de tables (ENT_MARKETx) de structure équivalente, mais dont nous avons vu que l'usage était différent : comptage ou répartition. Pour les entreprises auxquelles sont attachées des données financières, il y a une autre famille de tables (FIN_MARKETx).

Dans la partie sur l'expression des besoins, nous avons montré une cartographie des données. Il aurait été difficile de ne faire qu'une seule table, tant il y aurait eu de valeurs non remplies.

Dans chaque table, le principe de la dé-normalisation maximum a été adopté (donc à l'inverse d'une normalisation relationnelle) : il s'agit de n'avoir qu'un enregistrement par entité dans chaque cas, en mettant « à plat » toutes les données qui lui sont attachées. Ainsi, pour les données financières, les postes des trois dernières années de comptes annuels sont-ils « concaténés ». Cette organisation, du type fichier à plat (ou comma seperated value), simplifie l'indexation bitmap et donc les comptages.

Textes

Les données de type texte comme l'objet social, sont indexées grâce à la technologie Intermedia embarquée dans le serveur Oracle. Elle autorise des critères de ciblage sur les textes avec un certain degré de variation autour des mots. Elle est cependant moins performante que d'autres solutions du marché (Exalead, Search Server par exemple). Elle a néanmoins le mérite de proposer une solution « tout Oracle » ce qui a facilité son développement et sa maintenance.

En résumé

Cette solution technique repose sur

- des index bitmap,
- des partitions,
- quelques astuces supplémentaires comme
 - utiliser des transactions autonomes qui permettent d'augmenter le degré de parallélisation de certains traitements sur le noyau Oracle,
 - utiliser les techniques d'array-fetch ou array-insert qui permettent de manipuler des blocs d'enregistrements en lecture ou en écriture dans une table, ce qui accélère sensiblement le traitements des flux de données.

4.2.2 Limites de la solution « tout Oracle »

Pour que cette solution fonctionne de façon optimale, nous avons observé qu'il valait mieux indiquer au serveur Oracle, de « verrouiller » les index en mémoire. Cette contrainte conduit à ce que les performances sont tributaires du volume de mémoire vive disponible.

Il vaut donc mieux disposer d'une machine dédiée à cette instance. Par ailleurs, le volume des index peut devenir trop important à la suite d'une mise à jour pour les « loger » complètement en mémoire. Les performances peuvent donc d'effondrer brutalement, le temps de redimensionner la machine.

Le moteur SQL d'Oracle est du type sans état (« stateless »), c'est à dire qu'il n'a pas de raison a priori de garder les résultats des opérations intermédiaires qu'il a effectuées pour un comptage ou une répartition. Ceci est un handicap pour les stratégies assez complexes qui sont construites pas à pas par un utilisateur : l'ensemble de la requête est rejoué à chaque modification ou ajout d'un critère élémentaire.

Coface Services, par la voix de son administrateur de bases de données, a demandé si Oracle envisageait de changer ce comportement, et de donner la possibilité à ses clients/utilisateurs de disposer des résultats intermédiaires et des opérations de bas niveau sur les index bitmap. Oracle, comme on pouvait s'y attendre, a refusé, car ce sont des « couches trop basses » dans le noyau et il ne leur est pas possible de les rendre publiques sous la forme d'API (application programming interface); trop difficile à maintenir et à faire évoluer ensuite.

Les possibilités de requêtes sur les textes sont assez limitées. Les caractères français accentués posent notamment problème dans certain cas à Intermedia avec Oracle en version 8.

De plus, les requêtes qui utilisent un périmètre utilisateur importé, en mode repoussoir (instruction SQL « not in »), sont parfois longues (plusieurs dizaines de secondes). Elles fonctionnent mieux en mode inclusion (jointure classique).

Ces points ont amené Coface Services à envisager d'autres solutions pour son moteur de ciblage et à étudier l'offre du marché pour les moteurs de recherche ou de comptage.

4.3 Offres des éditeurs

Comme nous l'avons déjà écrit, la recherche de nouvelles solutions s'est inscrite dans le cadre de la définition d'une offre de ciblage marketing conjointe entre Kompass International et Coface Services (projet KIM-IS). Nous avons présenté en détail les demandes en terme de ciblage. Elles conduisaient à envisager une solution différente de celle adoptée pour Cofacibles. Nous ne nous étendrons pas sur les architectures techniques envisagées pour le projet KIM-IS, nous ne ferons qu'une synthèse des informations apportées par les éditeurs rencontrés.

Les logiciels du marché peuvent être classés en grandes catégories :

- ceux qui sont plutôt orientés recherche documentaire,
- ceux qui sont orientés comptage, analyse de données financières.

Dans la première catégorie nous trouvons :

- Search Serveur,
- Fast and Search,
- Sinequa,
- Exalead.

Ces moteurs sont excellent pour trouver un mot ou une expression dans une collection de

documents. Ils sont adaptés aux gros volumes. Par contre, ils sont moins adaptés aux données structurées en tables. Il n'y a pas de concept de table, et la notion de jointure par exemple, demande quelques acrobaties de programmation.

Dans la seconde catégorie nous mettons :

- Harry Software (Cohéris),
- Isoft,
- Information Builders,
- Oracle Timesten – IMD (in memory database).

Ces moteurs ne font pas de recherche de mot dans un texte, ou alors de façon très rudimentaire (Isoft). Ces moteurs sont adaptés aux données structurées, numériques ou codifiées. Pour ce que nous avons pu en apprendre, Harry Software et Isoft font un grand usage des index de type bitmaps. Le premier intègre le langage APL pour manipuler les données et les résultats des ciblage, ce qui donne une grande homogénéité à l'ensemble. Le second réorganise les données selon la technique du « data morphing ».

Information Builders organise les données en les hiérarchisant : elles sont d'abord réparties selon les valeurs de l'attribut le plus utilisé dans les requêtes et ainsi de suite sur 5 à 10 à attributs. Cela revient à imposer une démarche de ciblage standard dans la structure des données. C'est une contrainte forte. Par contre, en suivant (au moins sur deux à trois étapes) le chemin standard, les performances en comptage et en répartition sont très bonnes.

Quant à Oracle IMD, nous n'avons pas encore d'opinion arrêtée sur les versions récentes. Il nous faut étudier de près cette possibilité, connaissant par ailleurs les limites d'une solution « tout Oracle ».

Cette présentation succincte repose sur de nombreuses réunions de travail et sur des tests, au cours desquels les requêtes présentées dans la partie précédente de ce document, ont été jouées sur les différents moteurs, qui avaient au préalable indexé les données de Cofacibles.

Les résultats obtenus sont en en demie teinte :

- soit très bons sur le texte,
- soit très bons sur les données structurées.

Aucun outil ne répondait complètement à nos besoins. Il fallait faire un choix entre richesse des fonctionnalités sur le texte et performance sur les données structurées. De plus, un assemblage ad hoc de deux moteurs (un pour les textes et un pour les données structurées) exigeait que deux éditeurs collaborent pour faire communiquer leur outil; situation que des éditeurs envisagent difficilement pour satisfaire un seul client. De plus, les coûts de licence doublent par rapport à une solution à un seul éditeur.

Il est donc apparu difficile à Coface Services d'engager l'évolution de sa solution de ciblage sur l'utilisation d'un moteur de ciblage fourni par un éditeur.

Nous avons néanmoins tiré beaucoup d'enseignements de nos contacts avec les éditeurs. Bien qu'ils soient peu bavards sur leur technologie (et on les comprend!), certaines idées font leur chemin, et sont devenues des sources d'inspiration.

4.4 Proposition pour une solution composite

4.4.1 Principe général

L'idée est de concevoir un moteur de ciblage qui tire profit des fonctionnalités d'un moteur en mode texte et des performances des moteurs basés sur la technologie bitmap, que nous savons être une excellente technique pour le ciblage sur des données structurées.

La solution solution composite que nous proposons, pourrait utiliser un logiciel fournit par un éditeur pour la partie texte de nos problématiques, et un moteur bitmap que Coface Services pourrait développer.

Un moteur pour textes est beaucoup plus difficile à développer pour Coface Services, qu'un moteur basé sur des bitmap, car ceux ci sont, par nature, proches des concepts du ciblage. Les moteurs pour textes ont atteint un très haut degré de sophistication (expérience de Google ou de Search Serveur et Fast rachetés par Microsoft par exemple). Un bitmap est plus simple à concevoir. C'est grosso modo un tableau d'indicateurs binaires positionnés à 1 ou 0 selon que l'entité concernée appartient ou non à l'ensemble (la cible) associé au bitmap.

Un développement en interne a l'avantage de pouvoir être adapté, dimensionné dès le départ pour les besoins du ciblage. Nous proposons donc de développer « en interne » la partie intégration des résultats issus de recherches hétérogènes. Le bitmap devient le format pivot de tous les résultats de recherches/ciblages partielles provenant de moteurs différents : un moteur pour les textes et un moteur pour calcul sur les données financières par exemple.

Les modules de recherche externe retournent des listes de clés, manipulables par le moteur « maison », qui se charge de les transformer en bitmap « à la volée ».

Ce module central, est une sorte « d'agrégateur » des résultats en provenance de moteurs auxiliaires spécialisés sur un type de données ou sur un fond documentaire particulier.

Nous pourrions ainsi répondre au besoin de combiner un ciblage avec un critère de recherche basé sur l'énorme fond documentaire fourni par le Bodacc (10 années de documents, un Go de données par année environ). Un tel fond documentaire ne pose, a priori pas de problème de performance aux moteurs de recherche de type « texte » que nous avons étudiés.

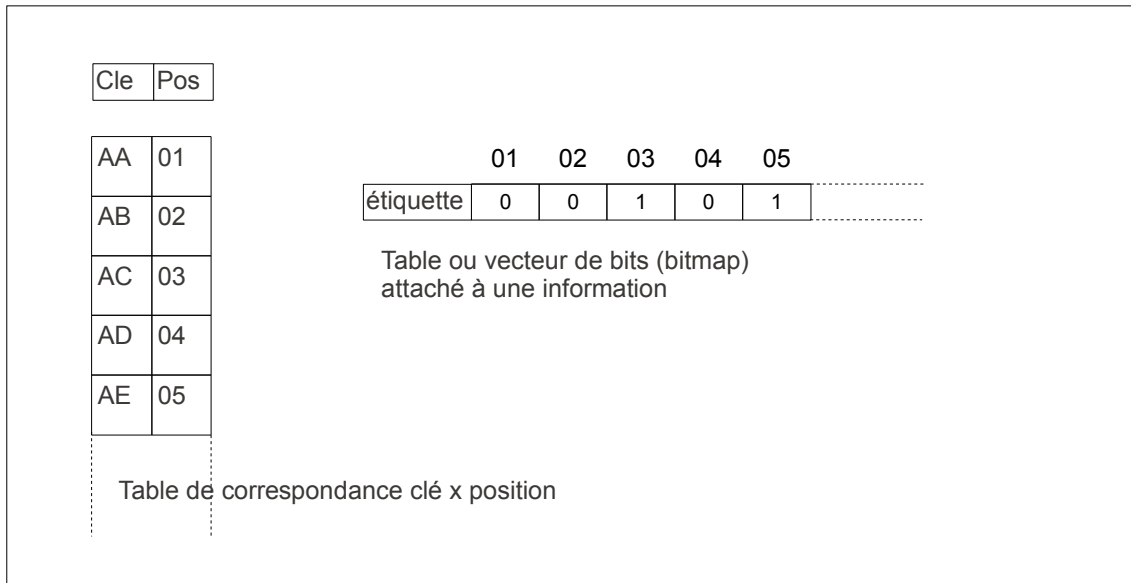
4.4.2 Principe de la représentation « bitmap »

L'ensemble des entités sur lesquelles portent le ciblage est une population figée, entre deux mises à jour, et connue, répertoriée, de façon exhaustive. Le nombre et la valeur des clés sont connus à la construction de l'index.

Pour chaque sous ensemble de la population totale, il est donc possible de dire si un élément lui appartient ou non. Cette information, « appartient ou non à tel sous ensemble » est une information binaire. En associant un tableau binaire, dont le nombre d'éléments est égal au nombre d'entités dans la population totale, à un sous ensemble, et en affectant la position de chaque élément du tableau à une entité de la population de départ, il est facile de représenter par 0 ou 1 dans la tableau si une entité appartient au sous ensemble. Nous avons ainsi construit un bitmap ou vecteur binaire.

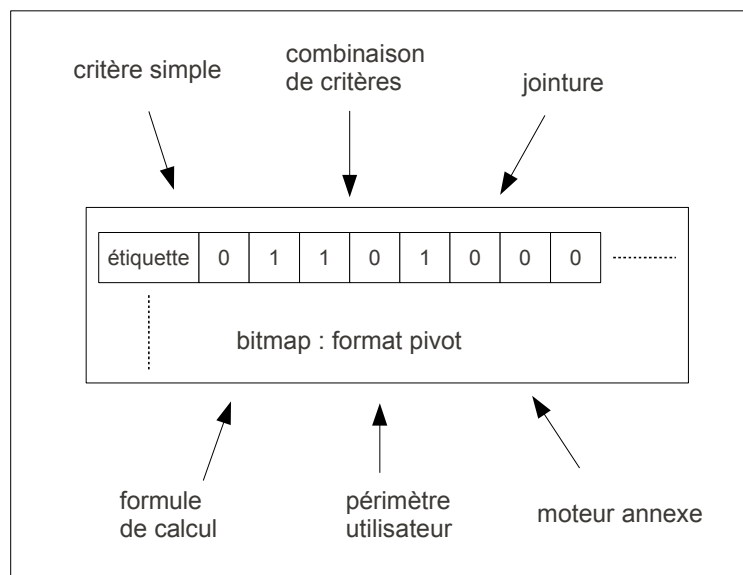
Un « index bitmap » est basé sur le principe : un bit par clé (par entité), repérée par sa

position.



L'étiquette est le nom du sous ensemble. Nous verrons comment il est associé à la valeur d'un attribut.

Bitmap comme format pivot



Tout sous ensemble, quel que soient les critères qui ont permis de le déterminer, peut être représenté sous forme d'un bitmap. Le module de combinaison des critères élémentaires de différentes nature (texte, données structurée, ...) est un agrégateur de résultats partiels.

4.4.3 Principe de l'index

Si on veut appliquer le principe du bitmap à une table relationnelle classique, il faut commencer par concevoir un index au sens « livresque du terme » : la clé joue le rôle du numéro de page, et le mot, le rôle de la valeur d'un attribut.

Le tableau suivant illustre la construction d'un index de ce genre appelé aussi parfois « table inversée ».

Table relationnelle			« Table inversée »		
Clé	Attributs		Attribut	Valeur	Clé
	attrib 1	attrib 2			
clé 1	val 11	val 22	attrib 1	val 11	clé 1
clé 2	val 12	val 21		val 12	clé 2
clé 3	val 11	val 21	attrib 2	val 21	clé 5
clé 4		val 22			clé 2
clé 5	val 12			clé 3	
				clé 1	
			val 22	clé 4	

Index (autre representation de la table inversée)			
attrib 1	<i>vide</i>	clé 4	
attrib 1	val 11	clé 1	clé 3
attrib 1	val 12	clé 2	clé 5
attrib 2	<i>vide</i>	clé 5	
attrib 2	val 21	clé 2	clé 3
attrib 2	val 22	clé 1	clé 4

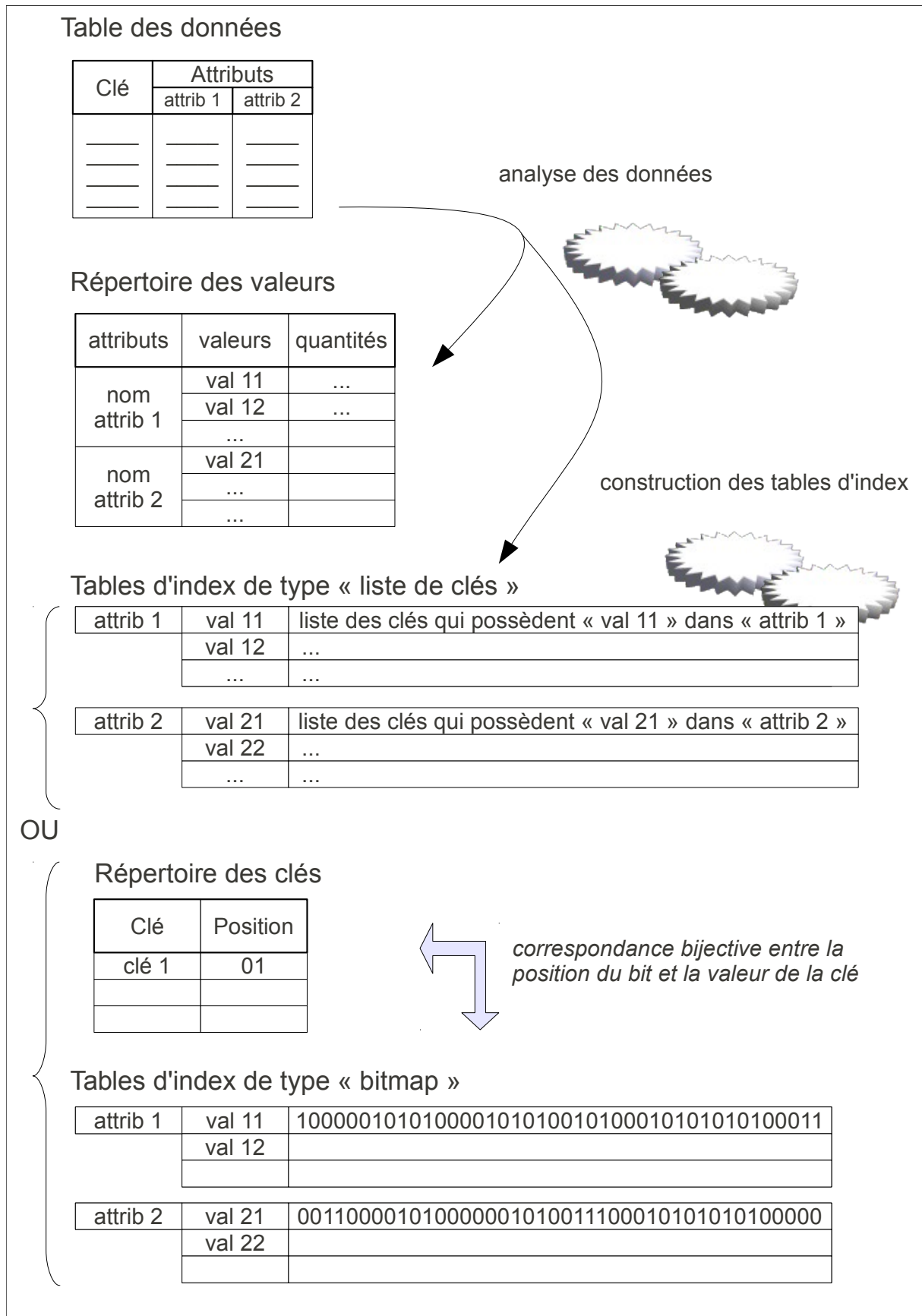
4.4.4 Exemple de construction d'un index

Le schéma suivant illustre le mécanisme complet de construction des index bitmap associés à une table relationnelle.

Nous construisons ainsi autant de sous ensembles qu'il y a de valeurs par attributs.

La représentation bitmap est d'autant plus efficace et économe en volume, qu'il y a peu de valeurs par rapport au nombre de clés, donc quand l'attribut est peu discriminant.

Constituants d'un index à base de bitmaps



Tous les bitmaps partagent la même correspondance clé-position, c'est essentiel pour

pouvoir les combiner par la suite. Sans cela, le travail du module d'agrégation n'aurait aucun sens.

4.4.5 Volumes

Estimations de volumes pour contenir les données de Cofacibles :

- la table des informations générales,
- la table des informations financières.

	Nb entités	Nb Attributs	Nb valeurs différentes	Volume données	Nb bitmaps	Volume 1 bitmap	Volume index
Infos générale	6,5 M	70	200	500 Mo	1 400	0,82 Mo	1 140 Mo
Infos financières	1,5 M	100 x 5	200	500 Mo	50 000	0,125 Mo	1 250 Mo
				1 000 Mo			2 390 Mo

Le nombre de valeurs par attribut est une moyenne estimée. Ce petit nombre de valeurs autorise une « compression » de la taille physique de chaque attribut : une représentation sur un octet est suffisante (jusqu'à 256 valeurs peuvent être encodées), au lieu d'une valeur sur plusieurs octets : la forme juridique est un code à 4 caractères, le code activité Insee (NAF 2008) est sur 5 caractères, etc. Ceci exige de disposer par ailleurs en mémoire, d'un tableau qui, pour chaque attribut, fait correspondre chaque valeur avec son code sur un octet. Il est utilisé pour au moment de la requête en traduisant la valeur en code sur un octet. Ce tableau est construit au moment de l'analyse globale des données, lors de la construction de l'index. Chaque tableau est d'un faible volume : quelques centaines à quelques milliers d'éléments.

Dans la mesure où l'ordre des valeurs a un sens pour un attribut, il faut s'arranger pour ordonner les codes dans le même sens que les valeurs. Les critères du type « attribut < valeur » sont ainsi directement transposables en remplaçant « valeur » par « code ».

Globalement, les données et les index prendraient environ 3,5 Go. Ce volume peut tout à fait loger dans la mémoire d'une machine d'un coût relativement modeste pour une entreprise (coût approximatif de 10 000 €, pour une machine serveur bi-processeur, chacun disposant de 4 cœurs cadencés à 3 Ghz et disposant de 16 Go de mémoire vive).

4.4.6 Avantage du bitmap

Un bitmap est construit pour chaque valeur de chaque attribut. Tous les critères élémentaires du type « attribut = valeur » sont directement représentés par un seul bitmap, dont le volume en mémoire est très petit : moins de 1 Mo dans notre cas. Le parcours des données en mémoire s'en trouve optimisé : pour un critère élémentaire, un seul bitmap est parcouru.

Tous les critères du type « attribut < valeur » sont calculables en appliquant une opération logique binaire « ou » (bit à bit) à l'ensemble des bitmaps qui répondent individuellement au critère.

Une requête de ciblage se résume donc à combiner des bitmaps par des opérations bit à

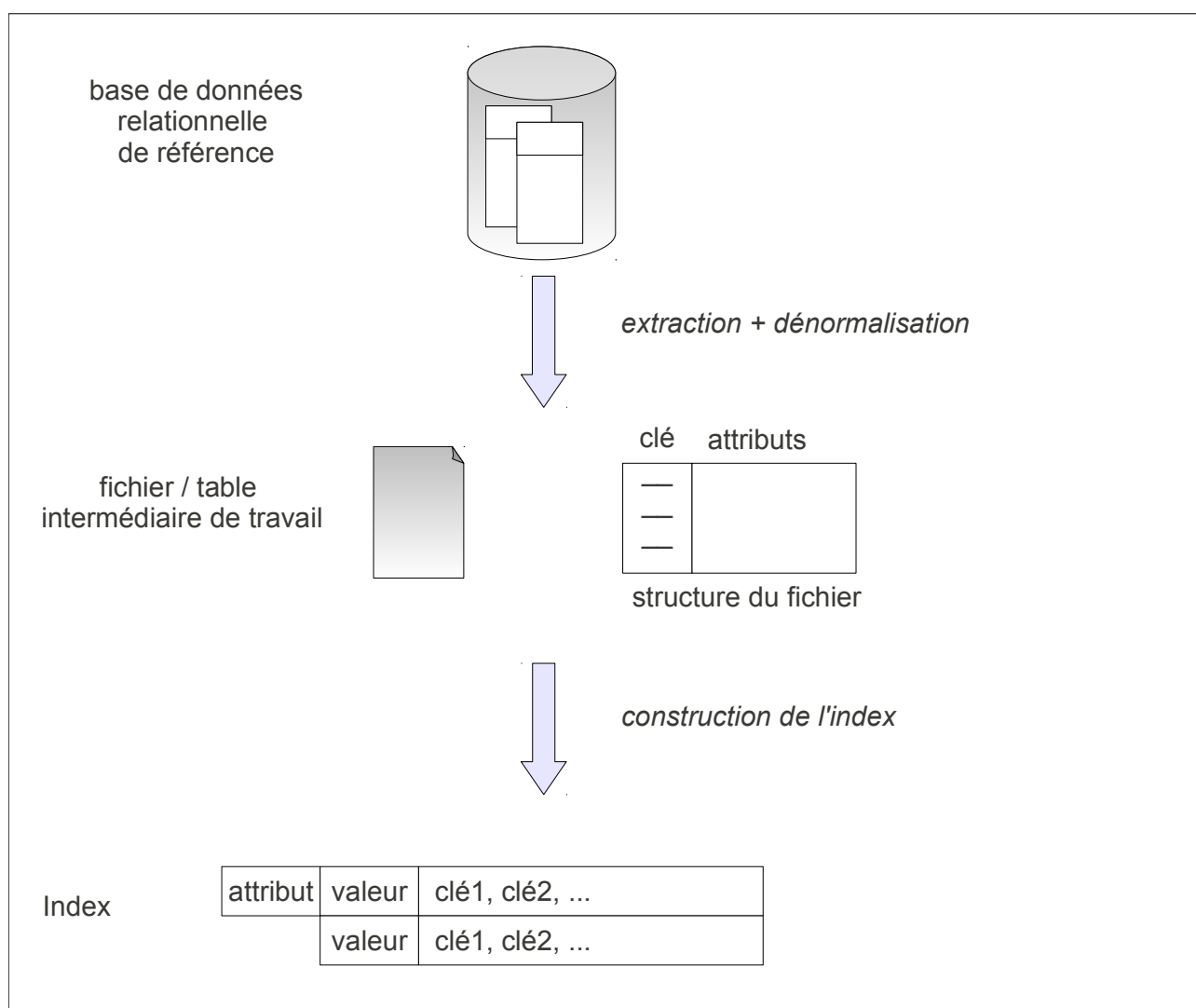
bit, ce pour quoi les processeurs sont parfaitement adaptés. Cette opération a un sens car tous les bitmaps ont la même correspondance clé-position, par construction.

Remarquons aussi que les résultats intermédiaires d'une requête qui est une combinaison « et » de critères élémentaires est très simple à obtenir, car chaque critère individuel est un bitmap, dont il suffit de compter les bits à 1.

Par rapport à la technique des bitmaps dans Oracle, nous pouvons maîtriser tous les résultats intermédiaires et maîtriser plus finement les bitmaps en mémoire pour plus de souplesse.

4.4.7 Flux

Le schéma suivant représente le flux à mettre en place pour alimenter un index à partir d'une base de données relationnelle de référence.



Mises à jour de l'index

Première version : reconstruction globale de l'index

- un nouvel index est construit pendant que l'autre est en ligne
- basculement entre index mis à jour et celui qui est en ligne

- implique de gérer la fin des requêtes en cours avant
- compatible avec une mise à jour hebdomadaire,
- présente toujours un état stable des données : pas de mise en ligne avant construction finale de l'index.

Deuxième version

Principe : mise à jour incrémentale d'une copie (hors ligne) de l'index

- Avantages :
 - moins de volume manipulé, plus rapide,
 - des mises à jours plus fréquentes sont possibles.
- inconvénient
 - plus difficile à implémenter.

4.4.8 Structures de données en mémoire

Nous proposons de garder toutes les informations du moteur de ciblage dans la mémoire du serveur qui héberge le moteur, de façon à éviter la lenteur d'accès aux données sur disque. Nous avons montré que c'est envisageable sur une machine serveur, relativement modeste.

En mémoire, il faut donc conserver :

- les données d'origine sous la forme de tableaux dont l'accès se fait par la clé;
- la table de correspondance entre les clés et leur position dans les bitmap;
- les tables de codification des valeurs des attributs;
- la collection de tous les bitmaps;
- les résultats des requêtes de ciblage des utilisateurs.

4.4.9 Techniques particulières

4.4.9.1 De la position du bit à l'adresse de la clé

A chaque mot de 64 bits on peut faire correspondre un lot de 64 adresses contiguës en mémoire : un décalage d'un bit correspond à un décalage de 8 octets pour l'adresse. Cette correspondance est indispensable pour passer des bits à 1 dans un résultat à la liste des clés qui ira alimenter le système de fabrication des fichiers marketing.

4.4.9.2 Comptage des bits à 1 dans un bitmap

L'algorithme de comptage des bits à 1 de Wegner-Lehmer-Kernighan :

```

unsigned int counting_bits_KR_ul_i_a(unsigned long int * v, unsigned int nw)
{
    unsigned int ctemp = 0; // up to 4 billion bits
    unsigned int c = 0; // up to 4 billion bits
    unsigned int inw = 0;
    for (inw=0; inw<nw; inw++) {
        for (ctemp=0; v[inw]; ctemp++) {
            v[inw] &= v[inw] - 1;
        }
        c += ctemp;
    }
    return c;
}

```

Cet algorithme basique est capable de compter sur un vecteur de 10 millions de bits (ordre de grandeur compatible nos volumes) en 0,01 secondes sur un ordinateur modeste (ordinateur portable familial).

Le nombre d'itérations dépend du nombre de bits à 1. De plus, il est parallélisable en découpant le parcours du bitmap en morceaux, qui sont pris en charge par des threads ou des processus indépendants. Nous avons commencé à développer un module de ce type en nous basant sur les techniques de parallélisation utilisées pour le calcul intensif (OpenMP en l'occurrence).

Autre algorithme cité par Anderson, à coût constant :

```
Counting bits set in a 32 bits word, in parallel

unsigned int v; // count bits set in this (32-bit value)
unsigned int c; // store the total here
static const int S[] = {1, 2, 4, 8, 16}; // Magic Binary Numbers
static const int B[] = {0x55555555, 0x33333333, 0x0F0F0F0F, 0x00FF00FF, 0x0000FFFF};

c = v - ((v >> 1) & B[0]);
c = ((c >> S[1]) & B[1]) + (c & B[1]);
c = ((c >> S[2]) + c) & B[2];
c = ((c >> S[3]) + c) & B[3];
c = ((c >> S[4]) + c) & B[4];
```

The B array, expressed as binary, is:

```
B[0] = 0x55555555 = 01010101 01010101 01010101 01010101
B[1] = 0x33333333 = 00110011 00110011 00110011 00110011
B[2] = 0x0F0F0F0F = 00001111 00001111 00001111 00001111
B[3] = 0x00FF00FF = 00000000 11111111 00000000 11111111
B[4] = 0x0000FFFF = 00000000 00000000 11111111 11111111
```

We can adjust the method for larger integer sizes by continuing with the patterns for the Binary Magic Numbers, B and S. If there are k bits, then we need the arrays S and B to be $\text{ceil}(\lg(k))$ elements long, and we must compute the same number of expressions for c as S or B are long. For a 32-bit v, 16 operations are used.

Mêmes remarques que ci dessus, sur les possibilités de parallélisation. Il existe donc plusieurs techniques éprouvées et assez simples à mettre œuvre pour calculer sur des bitmaps.

4.4.10 Calcul de formules

Nous n'avons pas trouvé de solution satisfaisante pour calculer des formules sur les valeurs financières, organisées sous la forme de bitmaps. Sur ce point, nous proposons donc d'appliquer la force brute du processeur : parcourir les données en mémoire et calculer à la volée la formule.

Ce type de calcul est parallélisable et il existe plusieurs techniques de programmation issues du calcul scientifique intensif pour les mener à bien sans programmation complexe. Signalons deux de ces techniques :

- OpenMP pour les machines à mémoire partagée (multi-coeur en l'occurrence),
- MPI (message passing interface) pour les machines à mémoire distribuée ou pour un groupe de machines reliées par un réseau et dédiées à un calcul.

Nos premiers tests sur des calculs d'une formule de la forme $(a+b)/(c+d)$, courante en analyse financière (calcul de ratio par exemple), est réalisée en environ une seconde pour une population de 10 millions de valeurs, sur une machine modeste, sans parallélisation.

Ces résultats sont encourageants.

Nous classons le calcul de formules en deux niveaux de difficulté :

- structure statique de la formule,
- structure dynamique de la formule.

Structure statique

Le calcul peut être implémenté « en dur » dès la programmation du moteur.

Au moment de l'exécution, il faut juste mettre dans la bonne variable, les valeurs des attributs, choisis par l'utilisateur.

Exemple d'une formule $(a+b)/(c+d)$: la signification de a, b, c et d est déterminée au moment de la requête, les valeurs sont recopiées dans les tableaux en mémoire qui représentent a, b, c, d.

Structure dynamique

Il faut ici imaginer un langage (simple) qui permette à l'utilisateur de définir sa formule à partir de fonctions mathématiques connues, appliquées aux attributs.

La formule doit ensuite être vérifiée, traduite en langage C, compilée dans une bibliothèque dynamique et « liée » dynamiquement au moteur, avant d'être appliquée en masse aux attributs. Cette possibilité (notamment offerte par un système Unix, Linux) est concevable, mais elle est assez complexe à réaliser. On peut aussi imaginer que cette possibilité ne soit pas complètement interactive. Le client définit sa formule, et celle-ci n'est implémentée dans le moteur que quelques minutes plus tard, le temps de la génération et de l'attachement de la bibliothèque au moteur, après tests automatiques de validité de la formule (nous pensons au problème de division par zéro, d'overflow, d'underflow, commun en calcul sur des nombres en représentation binaire).

Quelle que soit la complexité de la formule, ce module de calcul, qui ignore les bitmap dans un premier temps, peut être vu comme un moteur auxiliaire de calcul sur les informations financières, capable de fournir une liste de clés en sortie. Cette liste est ensuite transformée en bitmap (format pivot) pour être manipulée par le module de combinaison/agrégation des critères élémentaires du moteur de ciblage.

4.4.11 Jointure

Nous avons vu qu'un bitmap, ou plutôt la structure d'une famille de bitmap est directement reliée à la table (relationnelle) sur laquelle elle a été construite. Il y a une structure de bitmap par structure de clé, donc par type de table. Effectuer une jointure entre deux tables, c'est « marier » deux structures de bitmaps, structures dont les clés ont un élément commun. Le jointure peut s'effectuer par comparaison des tables de correspondance clés-position. Il nous faut valider concrètement la validité de ce principe et mesurer les performances d'un tel mécanisme.

Cependant, pour avoir par le passé développé dans un autre contexte, un mécanisme d'appariement de clés (une sorte de jointure hors base de données relationnelle, sur des millions de clés), nous sommes très confiants sur la possibilité d'obtenir de bons résultats sur ce point.

4.4.12 Parallélisme

Quasiment, plus un processeur pour serveur ne sort aujourd'hui, sans contenir 2, 4, 6 ou bientôt 8 cœurs. Le parallélisme physique est natif et disponible à volonté (ou presque)

pour le programmeur. D'autant qu'il s'agit d'un parallélisme à mémoire partagée : l'espace d'adressage d'un processus est partagé par tous les cœurs. Ce mode de parallélisme est plus facile à programmer.

Par ailleurs, disposer d'un petit nombre de cœurs (relativement aux machines massivement parallèles du calcul intensif), n'est pas forcément une grosse contrainte. en effet, la loi d'Amdahl, nous enseigne que pour « récupérer » dans un algorithme parallélisable, un degré de parallélisme égal au nombre cœurs, il faut un degré très élevé de parallélisation du code. Ainsi, en disposant d'un code parallélisable à 90%, avec 4 cœurs le facteur d'accélération maximal est de 3, avec 6 cœurs il est de 4, et avec 8 il est de 4,5 environ. Ainsi, un petit nombre de cœurs (4) apporte déjà un facteur d'accélération intéressant.

Pourquoi le parallélisme est applicable

- les bitmaps sont indépendants les uns des autres;
- les enregistrements des tableaux qui contiennent les données, sont indépendants les uns des autres.

De plus, les structures de données sont relativement simples : des tableaux. Ce genre de structure se prête bien à un mode de parallélisme par distribution des données.

Avantages

Bénéfices attendus : performances.

Inconvénients

Plus délicat à coder, à tester.

Threads

Nous pensons d'abord à mettre en place le parallélisme sur la base de threads concurrents (au standard Posix), supportés par tous les systèmes Unix ou analogues, donc portables. Les processeurs multi-cœurs étant très largement répandus, la programmation parallèle à mémoire partagée devient incontournable, et il n'est plus nécessaire de faire appel à une bibliothèque comme MPI pour mettre en œuvre du parallélisme entre plusieurs machines.

Concernant le moteur de ciblage que nous proposons de développer, nous envisageons de généraliser l'utilisation du parallélisme à toutes les opérations en mémoire : celles du module d'agrégation, celles de parcours des données financières numériques.

Parallélisme dans les clients Oracle

Le parallélisme du serveur Oracle est bien connu et fonctionne très bien. Par contre, il est beaucoup plus rare de mettre en place un traitement parallèle côté client Oracle. Nous avons récemment montré (avec l'aide d'un stagiaire) que cette technique était assez simple à mettre en place et fonctionnait parfaitement bien en utilisant une bibliothèque conforme à la norme Posix threads. Nous avons obtenu des facteurs d'accélération intéressant sur un traitement de lecture de plusieurs tables en parallèle et de chargement d'une autre table ensuite.

De plus, cette technique peut être combinée avec l'utilisation des array fetch / insert que nous avons déjà mentionné. Un client Oracle qui fait de l'extraction ou du chargement de données dans une base peut ainsi être accéléré en combinant cette technique-ci avec le

multi-threading. De plus, le gain en performance est contrôlable, en diminuant ou en augmentant le nombre de threads réels dans un processus client Oracle. Ceci donne une souplesse supplémentaire pour l'exploitation des programmes de chargement / extraction. Nous pensons notamment aux ingénieurs système et aux analystes d'exploitation qui répartissent et surveillent constamment les ressources disponibles sur les serveurs de façon à les attribuer au mieux aux différents traitements informatiques de l'entreprise.

*
* *

Dans cette partie du document, plus technique, nous avons présenté l'architecture de l'application actuelle. Nous en avons montré les limites. Nous avons ensuite proposé un certain nombre de pistes techniques dont le fil conducteur est le concept de bitmap. Cette technique, qui est présente dans la plupart des moteurs de comptage et de répartition des données structurées que nous avons étudiés, nous semble être une base sérieuse pour le développement « en interne » d'un moteur de ciblage chez Coface Services.

5 Conclusion

Nous venons de présenter un projet auquel nous avons participé, par périodes, sur plusieurs années, en fonction des besoins et de l'activité de notre employeur. Ce projet nous a amené à aborder plusieurs aspects du métier d'informaticien : analyse, conception, architecture, programmation. Nous avons eu la possibilité d'approfondir certaines problématiques techniques, notamment celle liés aux bases de données et aux techniques d'indexation.

Par ailleurs, au fil des années, nous avons pu approfondir certains aspects des métiers du « Poste client » et leur richesse fonctionnelle.

Concernant l'activité du ciblage marketing, nous avons abordé une problématique métier dans son ensemble : enjeux, besoins, recherche de solutions, mise en œuvre. Cependant, l'histoire n'est pas terminée, Coface Services est à nouveau à la recherche de solutions pour faire évoluer son offre et notamment de solutions techniques.

Nous espérons avoir convaincu de la pertinence des quelques pistes que nous nous proposons de développer. Mais elles devront « affronter » les contraintes organisationnelles et financières de notre entreprise. Puisse ce mémoire apporter quelques éléments de réflexion à nos décideurs qui auront la lourde responsabilité de choisir la « bonne voie ».

Avouons aussi que sur un plan intellectuel et personnel, passer des concepts de l'analyse financière ou des arcanes du crédit aux entreprises, au calcul intensif sur des vecteurs de bits, a quelque chose de vertigineux et de passionnant.

Sur un plan plus personnel, nous espérons aussi convaincre Coface Services de l'intérêt qu'il y a, à développer des solutions adaptées, à l'intérieur de l'entreprise; y compris pour des solutions techniques pointues, pour le développement desquelles nous sommes prêts à nous engager.

6 Annexes

Exemple de produit d'analyse financière délivré par Coface services

Analyse synthétique

RCS 482 755 741 - NIC 00014

COFACE SERVICES

LES GRANDES MASSES DU BILAN **

** Après retraitements

ACTIF

	31/12/2007		31/12/2008		31/12/2009	
	E	%	E	%	E	%
ACTIF IMMOBILISE	84 467 357	68,79	115 161 709	72,74	124 133 722	76,52
Incorporel	26 923 162	21,93	45 472 710	28,72	58 091 510	35,81
Corporel	6 155 943	5,01	5 692 337	3,6	6 949 784	4,28
Financier	51 388 252	41,85	63 996 662	40,42	59 092 428	36,43
ACTIF CIRCULANT	36 625 910	29,83	41 092 195	25,95	36 158 519	22,29
Stocks	97 736	0,08	94 394	0,06	62 817	0,04
Créances	35 783 059	29,14	40 616 534	25,65	35 917 465	22,14
Disponibilités	587 109	0,48	5 644	0	56 314	0,03
COMPTE DE REGUL. ACTIF	1 702 424	1,39	2 071 262	1,31	1 930 568	1,19
TOTAL ACTIF	122 795 691	100	158 325 166	100	162 222 807	100

PASSIF

	31/12/2007		31/12/2008		31/12/2009	
	E	%	E	%	E	%
CAPITAUX PROPRES	33 287 253	27,11	52 183 515	32,96	53 836 576	33,19
Capital social	18 000 000	14,66	20 294 265	12,82	20 294 265	12,51
PROVISIONS	7 502 865	6,11	7 448 868	4,7	7 498 109	4,62
DETTES	73 920 017	60,2	90 603 519	57,23	93 241 274	57,48
Financières	31 713 014	25,83	45 563 334	28,78	51 328 194	31,64
Fournisseurs	11 161 181	9,09	13 914 093	8,79	8 265 849	5,1
Fiscales et sociales	14 441 624	11,76	11 469 197	7,24	9 813 379	6,05
Autres Dettes	16 604 198	13,52	19 656 894	12,42	23 833 852	14,69
COMPTE DE REGUL. PASSIF	8 085 556	6,58	8 089 265	5,11	7 646 848	4,71
TOTAL PASSIF	122 795 691	100	158 325 166	100	162 222 807	100

LES RESULTATS

LES RESULTATS

	31/12/2007		31/12/2008		31/12/2009	
	E	%	E	%	E	%
CHIFFRE D'AFFAIRES	103 915 307		104 188 249		101 557 296	
Part à l'export	0	0,00	0	0,00	0	0,00
Production	114 996 013	110,66	117 253 041	112,54	116 504 083	114,72
Marge commerciale	0	0,00	0	0,00	0	0,00
Valeur ajoutée	50 963 731	49,04	46 454 512	44,59	46 544 476	45,83
Excédent brut d'exploitation	3 212 766	3,09	7 740 669	7,43	9 407 745	9,26
Résultat brut d'exploitation	13 416 710	12,91	13 877 656	13,32	10 124 061	9,97
Résultat courant avant impôts	17 032 075	16,39	15 529 959	14,91	12 379 896	12,19
Résultat net	10 457 253	10,06	10 414 606	10,00	6 618 979	6,52
Capacité d'autofinancement	11 345 605	10,92	12 120 426	11,63	15 162 810	14,93
M.B.A.	14 648 437	14,10	15 051 148	14,45	15 067 355	14,84
Frais de Personnel	46 434 790	44,69	37 547 749	36,04	35 089 118	34,55
Frais Financiers	1 407 601	1,35	2 309 037	2,22	947 873	0,93
Effectif	786		685			

COMPARAISON SECTORIELLE

	31/12/2009			
	Entreprise	Quartile inférieur	Médiane	Quartile supérieur
Production (%)	114,72	NC	NC	NC
Valeur ajoutée (%)	45,83	NC	NC	NC
Excédent brut d'exploitation (%)	9,26	NC	NC	NC
Résultat courant avant impôts (%)	12,19	NC	NC	NC
Résultat net (%)	6,52	NC	NC	NC
Capacité d'autofinancement (%)	14,93	NC	NC	NC
Frais de Personnel (%)	34,55	NC	NC	NC
Frais Financiers (%)	0,93	NC	NC	NC

LES PRINCIPAUX RATIOS

ACTIVITE ET RENTABILITE

	31/12/2007	31/12/2008	31/12/2009
Croissance du Chiffre d'Affaires (%)	0,70	0,26	-2,53
Croissance du Chiffre d'Affaires à l'export	NC	NC	NC

(%)			
Rotation d'Actif (nbre)	0,85	0,66	0,63
Rentabilité Economique (%)	4,91	8,71	9,33
Rentabilité Financière (%)	31,42	19,96	12,29

EQUILIBRE ET STRUCTURE

	31/12/2007	31/12/2008	31/12/2009
Fonds de Roulement (jrs CA)	NC	NC	NC
Besoin en Fonds de Roulement (jrs CA)	-43,48	-34,45	-40,86
Trésorerie (jrs CA)	-102,16	-151,29	-174,12
Ratio de Fonds de Roulement (nbre)	0,50	0,53	0,51
Indépendance Financière (%)	27,11	32,96	33,19
Endettement Financier (%)	25,83	28,78	31,64
Capacité de Remboursement (années)	0,14	0,15	0,14
Dettes à long terme / Capitaux propres (unités)	0,05	0,03	0,04
FFi / EBE (%)	43,81	29,83	10,08

- Confidentiel. Ne peut être divulgué -

7 Bibliographie

Droit de l'entreprise

- Daniel Fasquelle et al., « Droit de l'entreprise », Lamy (2009)

Finance d'entreprise

- Michel Lamarche, « Sympa la compta! », Éditions d'Organisation (2002)
- Roland PORTRAIT, Patricia CHARLETY, Denis DUBOIS, Philippe NOUBEL, « Les décisions financières de l'entreprise », PUF Gestion, 6ème édition (2004)

Urbanisation, architectures logicielles

- Gérard JEAN, « Urbanisation du business et des SI », Hermes Sciences (2002)
- Bernard BRULLER, « Architectures de système d'information », Vuibert (2003)
- Peter Herzum, Oliver Sims, « Business Component Factory : A Comprehensive Overview of Component-Based Development for the Enterprise », Wiley (1999)

Documentation interne Coface Services (non diffusible)

- Note sur les spécifications fonctionnelles d'une application de ciblage marketing;
- Note sur les spécifications techniques d'un moteur de ciblage;
- Rapport sur la recherche d'un moteur de ciblage sur le marché des éditeurs;
- Spécifications Cofacibles V3;
- Documentation CARIBOU;
- Documentation sur la politique qualité: SQAM;
- Notes internes d'organisation.

Bases de données

- Hector Garcia-Molina et al., « Database system implementation », Prentice Hall (2000)
- Georges GARDARIN, « Bases de données », Eyrolles (2003)

Documentation Oracle pour la version 11gR2 (www.oracle.com)

- Concepts (référence e10713)
- Data Warehousing Guide (référence e10810)
- Performance Tuning guide (référence e10821)

Data Morphing

- Richard A. Hankins, Jignesh M. Patel, "Data morphing: an adaptive, cache-conscious storage technique", Proceedings of the 29th international conference on Very large data bases - Volume 29 table of contents , Berlin, Germany, Pages: 417 - 428 (2003)

Système d'exploitation

- Berny GOODHEART, James COX, « The magic garden explained, The internals of Unix system V release 4 », Prentice Hall (1994)
- Neil Matthew, Richard Stone, « Linux programming », 4th edition, Wrox (2008)
- Andrew TANENBAUM, « Systèmes d'exploitation », 2nde édition, Pearson Education (2003)

Algorithmes sur des bitmaps

- Sean Eron Anderson, « Bit Twiddling Hacks », <http://graphics.stanford.edu/~seabder/bithacks.html>
- Peter Wegner, « A technique for counting ones in a binary computer », CACM 3 (1960), 322
- Derrick H. LEHMER, « The machine tools of combinatorics » Chapter 1 in *Applied Combinatorial Mathematics*, edited by Edwin F. Beckenbach, Wiley, New York, 1964, pp. 5-31.

Programmation parallèle

- Programming with POSIX Threads, David R. Butenhof, Addison-Wesley, (1997)
- The art of multiprocessor programming, Maurice Herlihy, Nir Shavit, Elsevier, (2008)

Ressources Internet sur les techniques de programmation parallèle pour le calcul intensif

- Parallélisme en général
 - <http://www.mcs.anl.gov/~itf/dbpp/index.html>
 - <http://aramis.obspm.fr/~semelin/enseignement.html>
- OpenMP
 - <https://computing.llnl.gov/tutorials/openMP/>
 - <http://www.openmp.org/specs/>
 - http://www.idris.fr/data/cours/parallele/openmp/OpenMP_cours.html
- MPI
 - <https://computing.llnl.gov/tutorials/mpi/>
 - http://www.idris.fr/data/cours/parallele/mpi/mpi1_cours_couleurs.pdf
 - <http://www-unix.mcs.anl.gov/mpi/tutorial/mpiexmpl/contents.html>

Résumé

Les entreprises commerciales cherchent naturellement à étendre leur clientèle. Dans le cadre du commerce entre entreprises, il s'agit de d'élargir sa clientèle en cherchant à toucher les prospects les plus pertinents. Disposer d'un annuaire d'entreprises est alors un atout.

La société Coface Services dispose d'un annuaire des entreprises françaises qu'elle enrichit constamment d'informations à jour, parmi lesquelles des informations sur leur santé et leur solvabilité, information cruciale pour un fournisseur qui cherche des clients « bons payeurs ».

Coface Services a mis en place une solution de ciblage marketing, qui permet à une entreprise de cibler ses futurs clients en les sélectionnant dans sa base de données, en fonction de filtres choisis. Les critères de sélection sont de différentes natures : informations générales, informations économiques et financières, ...

Une solution de ce type repose sur un « moteur de ciblage » puissant, de façon à offrir un outil interactif aux utilisateurs. Nous présentons les problématiques fonctionnelles et techniques d'un tel moteur de ciblage marketing. Nous décrivons la solution mise en place actuellement. Nous proposons enfin des pistes pour répondre à un cahier des charges plus complexe, qui permette à Coface Services de se démarquer de ses concurrents sur un marché très disputé.

Mots clés : information entreprise, solvabilité, ciblage marketing, sélection, moteur de recherche, bitmap

Summary

Companies naturally seek to increase the number of their customers. In the business to business context, that means seeking to reach the most relevant prospects. Having a company directory is then an asset.

Coface Services manages a directory of French companies. This directory is constantly updated with fresh information, including information on health and solvency which is a crucial information for a vendor who is looking for « good customers ».

Coface Services has implemented a targeting tool which allows a company to search for its future clients by selecting them in its database, applying suited filters. The selection criteria are of different types: general information, economic and financial information, ...

One such solution is based on a "targeting engine", powerful enough to provide an on-line tool to users. We present the functional and technical issues of such an engine. We describe the solution implemented until now. We also suggest ways to address more complex needs, which could enable Coface Services to stand out from its competitors in a highly competitive market.

Key words : information on companies, solvency, marketing targeting, selection, search engine, bitmap