

Internet et les systèmes d'information (SI) de l'entreprise

INTERNETS
ET WEBS DE DEMAIN

L'usage quotidien de l'Internet présente des possibilités assez semblables pour chacun d'entre nous : trouver les informations dont nous avons besoin, par exemple au moyen d'un moteur de recherche généraliste de type Google, trouver des sites ou des services pertinents pour nos divers besoins (voyages, culture, loisirs, informations professionnelles, réseaux sociaux, etc.) et, bien sûr, utiliser ces services aussi simplement que possible pour mener à bien les tâches que nous avons planifiées. Notre tolérance pour les imperfections de ces sites ne cesse de diminuer, notre vie en dépendant de plus en plus, comme en témoigne, par exemple, notre recours croissant au commerce électronique pour acheter des biens de toute nature ou la montée en puissance de la banque à distance.

par François BOURDONCLE*

LES SI DU MONDE INTERNET

Ce qui est assez frappant dans cette description de l'Internet au quotidien, c'est qu'il n'y est nulle part question de « systèmes d'information » (SI) et qu'en tous cas, les utilisateurs se débrouillent très bien avec Internet sans avoir aucunement recours à ce drôle de concept (car il s'agit bien d'un concept bizarre, nous y reviendrons). Seules importent aux utilisateurs les notions de « site » ou de « service », ainsi, bien sûr, que leur facilité d'utilisation.

De manière analogue, il paraîtrait incongru que les informations disponibles sur des sites marchands (par exemple, sur celui d'Amazon) ne soient pas à jour ou incomplètes. Ou encore que ces sites ne référencent pas tous les produits auxquels que l'on puisse imaginer,

aussi exotiques soient-ils. Même les moteurs de recherche grand public déploient des efforts considérables pour maintenir leurs bases d'indexation le plus à jour possible, et ce, en quasi-temps réel : c'est ce que l'on appelle le « Web temps réel ». Pourtant, indexer des dizaines de milliards de pages Web tout en mettant à jour des centaines de millions de pages chaque jour et en servant des milliards de requêtes quotidiennement représente un défi technologique dont peu de gens imaginent à quel point il est à la limite du possible.

À la toute fin du 18^e siècle, l'astronome Pierre-Simon de Laplace, qui présentait ses travaux sur la naissance du système solaire à Napoléon, entend celui-ci lui dire : « votre travail est excellent, mais il n'y a pas de trace de Dieu, dans votre ouvrage... » Et Laplace de lui

* Co-fondateur et Directeur de la Stratégie chez Exalead S.A.

répondre : « Sire, je n'ai pas eu besoin de cette hypothèse. ». Il semble bien que nous nous trouvions aujourd'hui dans une situation similaire et que l'hypothèse « système d'information » ne soit pas indispensable au monde Internet tel que nous le pratiquons, vous comme moi, tous les jours.

Pourtant, l'Internet, comme tout système électronique et informatique, utilise une quantité considérable de ce qu'il est convenu d'appeler des « systèmes d'information » (ou SI). Ces systèmes sont donc bien là, dissimulés aux yeux de l'utilisateur final, pour permettre à celui-ci de se concentrer sur ce qu'il a à faire et non pas sur la manière dont, dans l'arrière-boutique, les gestionnaires de ces services en assurent la permanence et la qualité.

Dans le monde de l'entreprise, la situation est tout autre : le SI y est placé au centre de tout : des discours, des attentions, des investissements, ainsi, le plus souvent, que des critiques d'utilisateurs qui ne comprennent pas pourquoi il est tellement compliqué de faire, au moyen de ces SI, des choses en apparence tellement simples...

Avant de revenir longuement sur les raisons (historiques, notamment) de cet état de fait, attardons-nous brièvement sur une idée très répandue et particulièrement tenace dans le monde de l'entreprise, selon laquelle le monde de l'Internet serait un monde de « cow-boys » pour lesquels perdre des informations ne serait pas si grave que cela, et pour lesquels la qualité de service et la haute disponibilité seraient des concepts inconnus (et j'en passe). Ou, pour dire les choses de manière encore plus provocante, que les « vrais informaticiens » se trouveraient dans le monde de l'entreprise alors que l'Internet, lui, serait peuplé de gamins pré-pubères tout juste bons à fabriquer des sites Web...

S'il est vrai qu'à ses débuts le monde du Web pouvait avoir un côté « FarWest » (1), cela n'a jamais été le cas pour les *infrastructures* de l'Internet (les routeurs, les *back-bones*, le protocole IP, etc.), qui ont depuis toujours été gérés avec un professionnalisme exemplaire. Actuellement, la vérité oblige à dire que le *Service Level Agreement*, le SLA, c'est-à-dire l'engagement sur la qualité de service des entreprises de l'univers Internet est considérablement plus élevé que dans tous les autres secteurs. Des sociétés de e-commerce (comme eBay) s'imposent un SLA sur leurs services dit « à trois neuf », ce qui signifie une disponibilité de 99,999 % (autrement dit, une indisponibilité maximale de 5 minutes par an). Pour y arriver, elles imposent à tous ses fournisseurs (informatiques, notamment) un SLA « à cinq

neuf », soit 99,99999 % de disponibilité (c'est-à-dire 3 secondes par an d'indisponibilité), ce qui est un niveau extrêmement difficile à atteindre pour un éditeur logiciel traditionnel. C'est sans doute l'une des raisons qui conduisent ces grandes sociétés Internet à développer elles-mêmes leurs outils logiciels (de manière assez systématique, il faut bien le dire) pour atteindre les niveaux de SLA dont elles ont besoin, car les solutions « traditionnelles » équipant les SI « traditionnels » n'y suffisent pas.

De la même manière, les solutions traditionnelles dans l'entreprise sont chroniquement totalement inadaptées au traitement de très grandes quantités de données et à la prise en compte d'une charge d'exploitation (correspondant au nombre d'accès simultanés au système) intense. Quand on parle de centaines de milliers ou de millions de documents dans l'entreprise, il n'est pas rare de parler de milliards, voire de dizaines de milliards de documents, dans le monde de l'Internet.

D'où la naissance d'une variété assez étonnante de nouvelles technologies dans de très nombreux domaines, comme le stockage distribué (Google File System (2), Amazon S3 (3)) et les bases de données (Google BigTable (4), Amazon Dynamo (5)), les systèmes de calcul distribués (Hadoop (6), Map/Reduce (7)), etc., pour pallier les insuffisances des solutions traditionnelles.

A l'évidence, ces solutions « maison » se justifient manifestement pour des services grand public où la volumétrie des données traitées et la faiblesse des marges commerciales nécessitent une approche technique entièrement nouvelle et très particulière. Ce qui est passionnant, c'est de voir que ces approches ultra-efficaces à la fois en termes de coûts, de performances et de fiabilité constituent désormais la base d'une puissante vague qui est en train de déferler sur l'informatique d'entreprise autour du concept de « *Cloud Computing* » (« l'informatique dans le nuage »). Par-delà l'aspect marketing de cette dénomination (qui est aujourd'hui au sommet du « Hype Cycle » du cabinet Gartner) l'on assiste à une profonde évolution de l'architecture des systèmes informatiques, qu'ils soient grand public ou professionnels. Les architectures client/serveur, agiles, accessibles depuis n'importe quel terminal (typiquement via un navigateur Web) et – surtout – qui « passent à l'échelle » (c'est-à-dire qui permettent de traiter des volumes de données ou un nombre d'utilisateurs arbitrairement grands sans avoir aucunement à changer d'architecture, simplement en offrant plus de ressource) s'imposent de plus en plus.

(1) J'ai en effet souvenance que la mise en production de la technologie « Cow9 » (aussi connue sous le nom de « LiveTopics » et de « Refine »), que j'avais développée pour AltaVista, se faisait en remplaçant la librairie dynamique (DLL) concernée et en redémarrant le serveur Web à la volée : c'était en 1997, autant dire « une autre époque » !

(2) <http://labs.google.com/papers/gfs.html>

(3) http://en.wikipedia.org/wiki/Amazon_S3

(4) <http://labs.google.com/papers/bigtable.html>

(5) http://en.wikipedia.org/wiki/Amazon_Web_Services

(6) <http://en.wikipedia.org/wiki/Hadoop>

(7) <http://en.wikipedia.org/wiki/MapReduce>

LES SI DANS L'ENTREPRISE

Par contraste, les SI traditionnels ont acquis au fil du temps une complexité croissante par le biais d'un empilement de couches successives, souvent sans jamais chercher à simplifier l'architecture globale. Et ces SI traditionnels sont aujourd'hui à bout de souffle, à la fois en termes de passage à l'échelle et en termes de flexibilité ou d'agilité : l'augmentation du volume des données traitées ou du nombre d'utilisateurs potentiels coûte très cher en ressources matérielles et les temps de traitement sont parfois tellement longs qu'ils dépassent la durée pendant laquelle le résultat du traitement a une valeur d'usage (emblématiquement, il faut par exemple plus de 24 heures pour traiter des données journalières !).

Il faut donc repenser entièrement l'architecture des SI actuels afin de réduire les coûts, de permettre le passage à l'échelle et d'augmenter l'agilité (c'est-à-dire la capacité opérationnelle à déployer rapidement de nouvelles applications), faute de quoi, les Directions Générales auront beau jeu de dire que l'informatique coûte de plus en plus cher, n'apporte jamais les résultats escomptés et, au final, n'est plus un réel outil de compétitivité, mais simplement un poste de dépense et un mal nécessaire, dans le meilleur des cas.

A propos de l'agilité des Systèmes d'Informations, Jean-Pierre Corniou (8), DGA de SIA conseil, ancien DSI de Renault et également ancien Président du CIGREF (le Club Informatique des Grandes Entreprises Françaises) estime que « l'informatique d'entreprise ne peut plus s'appuyer sur des cahiers des charges statiques pour envisager de nouveaux produits et services, imaginer de nouveaux processus de travail. Elle doit susciter (et accompagner en souplesse) les ruptures ».

C'est le cœur du débat actuel sur l'impact d'Internet sur les SI d'entreprises. Où faut-il investir ? Faut-il continuer à investir sur les systèmes dit « *legacy* » (les systèmes historiques, c'est-à-dire, typiquement, ces grands équipements informatiques structurants des années 80 et 90 que sont les ERP, les outils de gestion de la relation client (CRM), etc.). Ou bien faut-il plutôt considérer que ces équipements structurants doivent être maintenus, sans investissement supplémentaire, et être utilisés comme fondements d'une nouvelle classe d'applications inspirées du monde Internet et conçues pour et avec l'utilisateur ?

L'enjeu est d'importance, car le *découplage entre les développements applicatifs « productif » et « compétitif » de la gestion des infrastructures informatiques est un des éléments clés de la compétitivité future des entreprises*. De ce découplage dépend la capacité des « métiers », dans les entreprises, à se réapproprier l'usage du système d'informations, *via* leur autonomisation vis-à-vis des informaticiens 'purs et durs' des DSI. De ce découplage dépend l'agilité dans le développement de nouvelles applications, indépendamment de la gestion (mise à

jour, migration, évolutions) du « cœur » du SI et de son architecture complexe.

Ainsi, Jean-Pierre Corniou estime que les principales caractéristiques d'un système d'informations « web-centrique » sont les suivantes :

- toutes les applications doivent être accessibles à partir d'un navigateur, quel qu'il soit : propriétaire, ou libre ;
- le choix du terminal d'accès devient dès lors secondaire, mais il est clair que la plupart des applications web doivent être rendues accessibles à partir de terminaux mobiles ;
- les applications sont installées sur des serveurs distants, qui ne sont plus nécessairement exploités sous la responsabilité directe de l'entreprise utilisatrice ;
- la mise à jour et les montées en versions des applications sont assurées en continu, sans coût additionnel ;
- la sécurité est assurée par le gestionnaire du service sur des bases contractuelles ;
- le coût total de possession est exhaustif et entièrement prévisible.

A cette liste, je rajouterai le fait que sur Internet, le cœur de l'expérience utilisateur est, le plus souvent, construit autour d'un moteur de recherche. Et au-delà des moteurs de recherche de type Google, que nous utilisons tous plusieurs fois par jour, il y a aussi les moteurs locaux de chacun des sites Internet, à commencer par les sites de e-commerce, et aussi les moteurs de recherche des systèmes de messagerie, sans lesquels nous ne pourrions plus retrouver les informations dont nous avons besoin au quotidien. Le cabinet IDC estime en effet qu'environ 30 % du temps d'un collaborateur sont consacrés à rechercher l'information pertinente et que 85 % des informations d'une entreprise sont non structurées (c'est-à-dire en texte intégral). De plus, on estime qu'aux Etats-Unis, 92 % des décisions d'achat important sont précédées d'une recherche d'informations sur le web.

Il est donc très important de mettre la fonctionnalité de recherche au cœur de tout système d'information moderne ; les technologies issues du monde Internet, qui sont également au cœur des infrastructures de *cloud computing*, fournissent de nombreuses « briques » permettant de le faire.

LES SEARCH-BASED APPLICATIONS

Cette idée de construire de nouvelles applications autour de la fonctionnalité « moteur de recherche » est au centre d'un nouveau concept, que la société Exalead (que j'ai cofondée voici de cela dix ans, et qui a été acquise par Dassault Systèmes en juin 2010) a contribué à faire émerger ces dernières années. C'est le concept de *Search-Based Application* (SBA), expression que l'on pourrait traduire en français par « application fondée sur la recherche » ou, d'une manière moins exacte mais plus idiomatique, par « application d'accès à l'information ».

Au-delà de la logique exposée précédemment, à savoir celle du rôle clef de la fonctionnalité « recherche » dans

(8) <http://jeanpierrecorniou.typepad.com>

toute application « moderne » et de l'importance de découpler le cycle de vie des applications de type « infrastructure » (ERP, etc.) des applications productives, innovantes et agiles, le concept de SBA est également sous-tendu par des motivations d'ordre nettement plus technique.

Le constat que l'on peut faire dans la majorité des SI actuels, c'est que les informations utiles au travail quotidien des collaborateurs des entreprises existent presque toujours quelque part, mais que ces informations sont souvent stockées dans des « silos » étanches, complètement séparés les uns des autres, et qu'elles sont structurées de manière totalement différente d'un silo à l'autre, et accessibles chacune au moyen d'un mécanisme qui leur est propre, d'un système de sécurité propre et d'une interface utilisateur propriétaire. Imaginer que les utilisateurs vont se connecter silo par silo pour rechercher l'information dont ils ont besoin et ce, avec une interface utilisateur différente à chaque fois, relève de l'utopie : en effet, les salariés, dans les entreprises, n'ont pas le temps de se livrer à ce genre d'exercice et, par conséquent, ils doivent le plus souvent prendre des décisions à l'aveugle, sans avoir de support objectif à leurs décisions, alors que des informations critiques pour ces prises de décisions sont disponibles, mais dissimulées au sein de systèmes d'informations opaques.

Cela conduit à de nombreux phénomènes néfastes, comme la frustration devant cette informatique « qui ne marche jamais » ou la duplication de l'information, que, faute de la retrouver, on recrée à chaque fois que l'on en a besoin, ou encore comme une perte d'expertise, quand un collaborateur quitte l'entreprise, etc.

A l'inverse, une meilleure circulation de l'information accélère les prises de décisions, qui sont *a priori* mieux motivées et elle permet, également, d'identifier dans l'entreprise des personnes qui peuvent directement contribuer à des actions ou à des prises de décision. L'échange d'informations, et par conséquent la recherche d'informations, est en effet un élément clef du *networking* dans les entreprises, permettant la création de « réseaux sociaux *ad hoc* » (ou dynamiques) au sein de celles-ci.

La réalité du monde de l'entreprise actuel, c'est donc la multiplication des *systèmes de stockage et de structuration* d'informations, ainsi que celle des applications qui sont directement construites à partir desdits systèmes. Ce sont ces fameux sigles à trois lettres, qui ont fait la fortune de sociétés comme Oracle, IBM ou SAP : ERP (*Enterprise Resource Planning* ou système de gestion intégrée de l'entreprise pour la gestion des stocks, de la facturation, des ressources humaines, de la finance, etc.), le CRM (*Customer Relationship Management* ou gestion de la relation client), le PLM (*Product Lifecycle Management* ou gestion du cycle de vie du produit de sa conception à sa fabrication, en passant par ses tests), le BPM (*Business Process Management* ou gestion des processus métiers), etc.

Chacun de ces systèmes possède un modèle de données spécifique, en général très complexe (ainsi, un ERP

typique peut contenir plusieurs milliers, voire dizaines de milliers de tables aux intitulés ésotériques) et qui n'est adapté qu'à la *gestion des données* (leur structuration, leur stockage, leur mise à jour, leur archivage, etc.), mais en aucun cas à leur *accessibilité par un utilisateur final*. La raison en est, pour l'essentiel, historique, puisque les développeurs d'applications d'entreprise n'ont guère eu d'autre option, pour développer leurs produits, que de s'appuyer sur les bases de données relationnelles (RDBMS, en anglais, ou SGBD, en français) et de mettre en œuvre les fonctionnalités que celles-ci offraient, en utilisant la structuration existante des données. Le résultat en a été des produits souvent très complexes à comprendre et à utiliser pour des utilisateurs non-experts, car *la logique sous-jacente est une logique d'informaticien, et non pas une logique d'usage, celle qu'adopte un simple utilisateur « du métier »*.

Plus récemment, sont venus se rajouter des systèmes d'information comme le courrier électronique ou l'ECM (*Enterprise Content Management*) ou la GED (gestion électronique de documents). Ces systèmes ont ceci d'original que les données qu'ils manipulent sont non structurées, c'est-à-dire qu'elles sont du texte libre, comme par exemple le corps du message d'un courrier électronique ou encore un fichier Word ou PDF. En réalité, tous ces contenus sont ce que l'on appelle semi-structurés, car, en sus du contenu non structuré existe généralement une structuration, fût-elle très légère : expéditeur ou récipiendaire, date d'expédition (par exemple, pour le courrier électronique), et auteur d'un document ou dossier du disque dur dans lequel est stocké ce document (pour les systèmes d'ECM).

Les systèmes de courrier électronique ou d'ECM ont progressivement démocratisé et généralisé l'usage des moteurs de recherche, car leur contenu en texte libre se prête très mal aux systèmes de recherche rigides à base de formulaires que proposent traditionnellement les systèmes de bases de données relationnelles (RDBMS). Ces systèmes de recherche en texte libre ont généralement beaucoup de succès auprès des utilisateurs, car ils offrent le même confort d'accès à l'information que celui que l'on trouve sur des moteurs de recherche grand public de type google.com, bing.com ou exalead.com. De plus, les systèmes de moteur de recherche ont popularisé l'idée que l'accès aux informations stockées à l'intérieur des applications en silo par un mécanisme externe (*via* le moteur de recherche) offre une souplesse beaucoup plus grande, comme par exemple la possibilité d'effectuer ce que l'on appelle une *recherche fédérée*, c'est-à-dire la capacité, à partir d'un seul formulaire de recherche rempli par l'utilisateur, de rechercher dans plusieurs systèmes d'informations à la fois, sans avoir ni à se connecter à chacun de ces systèmes ni à utiliser une interface utilisateur spécifique pour effectuer sa recherche.

L'idée s'est donc progressivement fait jour de généraliser l'accès externe aux données internes des SI pour aider à dépasser la paralysie croissante de ces systèmes, une paralysie liée pour partie au poids de l'histoire et

notamment à l'incapacité chronique des entreprises à simplifier les SI lors de leurs évolutions successives (elles se contentent, dans bien des cas, de rajouter une couche à un système déjà passablement complexe).

Pour ce faire, il est nécessaire de réfléchir à la structuration qu'il convient de donner aux informations que l'on veut stocker dans le moteur de recherche. A ce stade, il convient de dire que les moteurs de recherche dont je parle ici, par exemple le moteur de recherche CloudView™ d'Exalead, n'est pas un simple moteur de recherche de type Google ne cherchant que sur des données en texte brut. Les moteurs spécialisés dans ce type de services aux entreprises savent gérer des informations semi-structurées en prenant en compte la date de création et le classement des documents dans des arborescences, appelées taxonomies ou ontologies, une gestion fine de la sécurité (droits d'accès différenciés aux documents en fonction des utilisateurs, etc.), les liens sémantiques entre divers éléments d'informations, etc. Ces moteurs spécialisés sont en fait devenus de véritables *infrastructures logicielles* destinées au *développement d'applications d'accès unifié à l'information*. Il serait d'ailleurs très intéressant de se pencher sur le type de structure (ou de semi-structure) que ces moteurs de recherche utilisent pour construire ces nouvelles applications (mais cela sortirait du cadre de cet article). Les maîtres mots sont ici de :

- structurer autant que possible les contenus au format libre, en utilisant des *technologies d'extraction sémantique (text-mining, en anglais)* ;
- déstructurer d'une manière intelligente les contenus trop fortement structurés dans les systèmes de gestion de bases de données, de manière à rendre ces contenus homogènes, en termes de structure, avec les autres contenus semi-structurés disponibles ;
- définir un ensemble *d'éléments métiers (Business Items)* faisant sens pour l'utilisateur de l'application envisagée, que cet utilisateur pourra identifier dans la liste de résultats du moteur comme étant pertinents pour sa recherche ;
- enfin, construire, à partir des informations disponibles, les *relations sémantiques* entre ces divers éléments métiers, de manière à pouvoir *contextualiser l'information* fournie à l'utilisateur.

Par exemple, pour une application de CRM (gestion de la relation client), les éléments métiers que l'on peut envisager sont : le *client*, le *contrat*, le *produit acheté*, la *panne rencontrée*, le *prix payé*, le *vendeur* ayant vendu le produit, etc. Les relations sémantiques entre ces éléments vont être, par exemple, la raison identifiée d'une panne. Les technologies de *text-mining* vont être utilisées pour analyser automatiquement, par exemple, les messages envoyés par un client sur le site Web de l'entreprise, pour reconnaître dans ces messages le nom des produits au catalogue, le nom des vendeurs, les causes de pannes, le prix du produit, la date d'achat, etc., de manière à structurer les données qui vont être ensuite indexées par le moteur de recherche.

CONCLUSION

Les applications d'accès à l'information ainsi construites par-dessus un moteur de recherche sémantique sophistiqué sont appelées *Search-Based Applications (SBA)*. Ces applications n'ont pas du tout pour but de remplacer totalement les applications d'entreprise traditionnelles. En réalité, elles ont vocation à *modifier radicalement la manière d'accéder aux applications historiques*, en utilisant à la fois des technologies et des interfaces-utilisateur issues du monde Internet. Cela étant dit, les SBA peuvent également venir parfois en concurrence de projets de type « Très Grande Base de Données » (TGBD), très appréciés des informaticiens (mais beaucoup moins des utilisateurs !) consistant à remplacer *n* systèmes d'informations hétérogènes par une base de données unique à vocation universelle.

Ces projets de « migration » (comme on les appelle) sont des projets longs, coûteux et complexes, qui conduisent parfois à de véritables catastrophes industrielles en termes de délais et de fonctionnalités. De plus, il n'est pas rare qu'une fois ces projets menés jusqu'à leur terme, de nouvelles bases de données périphériques aient fait leur apparition et doivent à nouveau être intégrées dans la TGBD. De surcroît, le Web lui-même est de plus en plus une source de données à part entière et il est illusoire de penser pouvoir gérer une base de données qui maintiendrait à jour manuellement ces informations-là.

L'approche SBA permet donc de ne pas avoir à déployer de projet de TGBD, tout en ayant l'essentiel des avantages et en permettant d'intégrer au système résultant des informations provenant du Web (par exemple, des avis de consommateurs ou l'analyse des problèmes rencontrés par des utilisateurs et relatés sur des blogs ou des forums de discussion).

Le champ d'application des SBA ne cesse de s'étendre actuellement : vision du client à 360°, gestion de la relation client multi-canal, application de suivi du ressenti client, PLM, logistique, e-commerce, e-reputation, etc. : la liste en serait longue. Et d'autres technologies issues du monde Internet, comme les technologies d'interface utilisateur dites Web 2.0, viennent en renfort d'un *mouvement de fond de réinvention de l'informatique d'entreprise* autour des technologies et des usages de l'Internet.

Il s'agit, ni plus ni moins, de *ré-enchanter l'informatique dans les entreprises* et de permettre aux utilisateurs et aux « métiers » de se réappropriier l'outil informatique, après une trop longue confiscation par les informaticiens et les Directions Informatiques.

C'est là un enjeu absolument majeur, notamment en termes de coûts et de compétitivité, et je suis persuadé qu'une révolution est en marche qui va radicalement modifier le paysage informatique dans les entreprises au cours des dix années à venir.