

Signaux

Revue de L'Association des Diplômés de l'I.S.E.P.
Novembre 2015 - N°106



du Cloud au Big Data

SOMMAIRE

■	ÉDITORIAL	4
■	LE MOT DU PRÉSIDENT	5
■	INTRODUCTION : L'ÉTAT DU CLOUD EN EUROPE	6
■	LA CONNAISSANCE DES DONNÉES, CLÉ DE LA STRATEGIE CLOUD DE L'ENTREPRISE	8
	▶ Le cloud computing	8
	▶ Conséquences pour l'entreprise	9
	▶ Focus sur les données	10
	▶ La solution : la gouvernance d'un système d'information multi-clouds pilotée par les données	12
	▶ Conclusion	13
■	QUEL AVENIR POUR LE CLOUD ?	14
	▶ Le cloud moteur de la mobilité	14
	▶ L'irrésistible ascension du cloud	14
	▶ Et apres ?	14
	▶ Pour conclure : n'ayons plus peur du cloud !	15
■	VERS UNE ARCHITECTURE BIG DATA MODERNE POUR LA BUSINESS INTELLIGENCE : DATA LAKE	17
	▶ De la business intelligence a la business discovery	17
	▶ Cloud et big data sont une opportunité pour la business intelligence	19
	▶ Le lac de données : une nouvelle architecture d'organisation des données	21
	▶ Le lac de données est une architecture technique basee sur des outils opensource	25
	▶ Une nouvelle organisation et des nouveaux enjeux humains	26
	▶ Les caractéristiques de la réussite	27
	▶ Conclusion	28

	BIG ET LINKED DATA : INTERCONNEXION DE FLUX DE DONNÉES MASSIVES	30
	▶ Big et linked data	30
	▶ L'état de l'art	31
	▶ Références	31
	LE PARADIGME DU BIG DATA ET SON APPLICATION À LA CYBERSÉCURITÉ	33
	▶ La donnée	33
	▶ Le quatrième paradigme : la corrélation des données et métadonnées	33
	▶ Du principe de causalité au principe de corrélation	34
	▶ Les données d'entreprise, une matière première inestimable	34
	▶ Le big data pour contrer les apt	35
	▶ Siem, soc et big data	35
	ISSY-LES-MOULINEAUX, UN AN APRÈS	37
	LA RECHERCHE À L'ISEP	38
	▶ Équipe Minarc	38
	▶ Équipe SITe	38
	▶ Équipe RDI	38
	LA FORMATION CONTINUE À L'ISEP	39
	▶ Pourquoi ? Pour qui ?	39
	▶ Savez-vous ?	40
	▶ Nos catalogues	41
	▶ Contrats de professionnalisation et thèses professionnelles	42
	▶ Contact	43

Autrefois... L'entreprise possédait un gros, très gros ordinateur, que nous alimentions une ou deux fois par jour avec des batchs de cartes perforées qui contenaient nos « programmes ». Non, jeunes Isépiens, ne riez pas ! Plus tard, les passages en batch firent place au time-sharing. Le gros, très gros ordinateur était toujours là, mais chacun pouvait l'utiliser (presque) en temps réel et de façon (presque) exclusive à partir de terminaux informatiques, les célèbres VT100. La génération suivante fut celle de la micro informatique et la VT100 céda la place à l'ordinateur individuel sur lequel s'exécutaient les logiciels. Le gros ordinateur de l'entreprise fut, du moins dans les grandes entreprises, remplacé par un réseau de transmission de données et des serveurs pour le stockage et le partage de l'information.

Aujourd'hui... L'ordinateur individuel, est toujours là, accompagné d'une cohorte de smartphones et de tablettes, mais il voit son rôle fortement diminué. Ceci a commencé avec le besoin de mobilité, concrétisé par le besoin d'accéder à ses données depuis n'importe où et depuis n'importe quel appareil. D'un autre côté, le volume de l'information stockée s'est mis à croître et multiplier... vous vous rappelez, Mickey, dans l'apprenti sorcier. C'est ainsi qu'est né le Cloud Computing (ou Informatique en nuage), dans un premier temps pour offrir un espace de stockage extensible et adaptable, le Cloud, accédé au travers du réseau Internet ou du réseau privé de l'entreprise. Pendant qu'on y était, les logiciels ont aussi migré dans le Cloud. Le Cloud Computing s'est alors décliné en IaaS, SaaS, PaaS et maintenant BaaS... je n'en dirai pas plus, d'éminents spécialistes vous en livrent tous les secrets dans les chapitres qui suivent.

Dans le même temps, notre société a été saisie d'une véritable boulimie en matière d'information. On ramasse tout, on stocke tout et on essaie d'en tirer la substantifique moelle. Justement, le Cloud, avec sa formidable puissance de stockage permet de ranger tout cela... et ça a donné le Big Data. Le problème se pose ensuite d'exploiter cette masse de données. Les systèmes de bases de données traditionnels ont déclaré forfait, cédant la place à un autre modèle qui est celui du Data Lake (Lac de données), assorti de traitements appropriés pour digérer, sans en perdre la moindre miette et sans s'y noyer, des flots de données hétérogènes et non structurées.

Dans ce numéro de Signaux, nous découvrirons plusieurs articles liés au Cloud Computing et au Big Data. En guise d'introduction, **Philippe Recouppé** nous livre une réflexion sur les enjeux du Cloud Computing, plus spécialement en Europe.

Après quelques rappels et définitions, **Frédérique Patron** nous explique comment les entreprises se sont appropriées le Cloud Computing et quelle gouvernance elles doivent mettre en place pour garantir l'intégrité et la confidentialité des données dans le Cloud.

Samir Koleilat nous brosse ensuite un tableau de la montée en puissance du Cloud et de son avenir.

Du Cloud au Big Data, il n'y a qu'un petit pas à franchir et **Karim Ben Mabrouk** va nous montrer comment le Data Lake va permettre aux entreprises de développer une véritable Business Intelligence en vue d'augmenter leur compétitivité.

Manipuler un grand volume de données nécessite la mise en place de traitements spécifiques comme nous l'explique **Raja Chiky**.

Éminent spécialiste de la cybercriminalité, **Gérard Peliks** nous expose, par un exemple concret, comment le Big Data peut être utilisé pour traquer les attaques informatiques.

Sortons ensuite de notre nuage pour parler de notre école préférée.

Tout d'abord, nous ferons un petit tour à Issy-les-Moulineaux avec **Emmanuelle Vatel**, pour un premier bilan de l'implantation de l'ISEP dans ses nouveaux locaux.

L'ISEP est bien sûr une école, mais aussi, un centre de recherche que nous présente **Raja Chiky**, elle-même enseignante-chercheuse à l'ISEP.

Pour terminer, **Denis Beautier** va nous faire découvrir une facette souvent méconnue de l'ISEP qui est la formation continue et comment elle s'adresse à vous, à vos collaborateurs, enfin à tout votre environnement professionnel.

Bonne lecture !

Par Michèle GERMAIN - ISEP 1968

LE MOT DU PRÉSIDENT

Quoi de plus normal pour une école comme l'ISEP, qui dorénavant se présente comme « École d'ingénieurs du Numérique », de parler de l'externalisation de l'information dans les Clouds et Big data.

Certes, pour un certain nombre de diplômés, parmi les « anciens » qui ont pratiqué les bancs de notre école il y a quelques années (pour ne pas dire dizaines d'années), l'information était une chose rare et chère et le stockage de celle-ci encore plus (seuls les 0 et les 1 étaient stockés !). De nos jours, nous « croulons » sous l'information, et les « datas » sont accessibles partout et à tout instant ...

Parlez aujourd'hui à un jeune diplômé, génération 2014 ou 2015, de Cloud et de Big Data, il saura parfaitement vous décrire cela et en connaîtra les sous-jacents techniques. Mais en perçoit-il déjà tout le potentiel et tous les risques induits ? C'est en effet un sujet qui semble arriver à maturité pour certains, mais qui me semble pour ma part n'en être qu'à ses prémices : quelles évolutions, voire révolutions, ces termes ou technologies peuvent-elles laisser présager ? Quelles transformations sociétales, organisationnelles, culturelles... vont-ils générer ? Il suffit de regarder les premiers effets déjà constatés avec l'économie collaborative, les lanceurs d'alerte, les datas scientists... qui, se mêlant, peuvent tout à la fois être créateurs (d'emplois, de valeurs, de richesses, etc.) que destructeurs (d'emplois, de sécurité, de richesses, etc.).

Il était donc évident que votre revue se mobilise une nouvelle fois pour vous faire partager quelques réflexions sur ce thème porteur d'avenir « Du Cloud au Big Data ». Cela permet de prouver, une nouvelle fois, la nécessaire interpénétration des mondes de l'enseignement, de la recherche et de l'économie.

Je tiens à remercier les écrivains, d'un jour ou réguliers, pour leur participation à ce numéro 106 de Signaux, et tout spécialement notre rédactrice en chef, Michèle Germain, pour son implication et la dynamique qu'elle a mis dans les derniers numéros qu'elle a menés de main de maître. Elle souhaite passer la main, alors n'hésitez pas...

Par Bruno DE LA PORTE - Président d'ISEP Alumni



INTRODUCTION : L'ÉTAT DU CLOUD EN EUROPE

Par Philippe Recouppé - COGICOM
Président de Forum ATENA - Enseignant à l'ISEP Formation Continue

**« I want Europe to be not only Cloud-friendly but Cloud-active » Neelie KROES, World Economic Forum, 13 janvier 2012.
Qu'en est-il trois ans après ?**

Le Cloud Computing est une révolution pour l'informatique, une révolution industrielle, organisationnelle, technologique et aussi, bien sûr et nécessairement, réglementaire et juridique.

Le Cloud Computing est pour l'informatique la concrétisation de la révolution numérique, mais il n'en révèle aujourd'hui que les premiers domaines d'applications, les premiers services pour les entreprises et les individus, se proposant dans un futur proche de réinterroger bien d'autres activités humaines.

De nouvelles technologies, comme la virtualisation ou le stockage en réseau, associées à des économies d'échelle considérables, ont permis l'apparition d'offres de services informatiques à distance ou dans le « Cloud ». Même si l'offre d'hébergement sous toutes ses formes (SaaS, IaaS, PaaS) s'est bien développée en Europe, ce sont les États-Unis qui tirent le mieux leur épingle du jeu par leur dynamisme, leur compréhension des évolutions du marché, la présence des plus grandes entreprises de l'IT, et le bénéfice d'un meilleur facteur d'échelle. C'est l'ensemble de ces facteurs qui a favorisé le développement des grands leaders mondiaux que sont Amazon, Microsoft, Google, Apple, IBM, Oracle ou Salesforce.

En contraste, peu de grands leaders émergent en Europe, même si on compte plusieurs success stories, comme OVH avec près de 200 000 serveurs.

Pourtant, de nombreuses entreprises de technologie, petites et grandes, ont compris et anticipé la transition en cours, ainsi que certaines grandes écoles, comme l'ISEP, qui accompagne les entreprises dans la transition vers le Cloud par l'organisation d'un Mastère, aux côtés de partenaires américains et français, comme Sogeti.

Il est à remarquer également le dynamisme d'entreprises européennes dans les prestations d'accompagnement et dans le développement d'outils, plus particulièrement dans le monde de l'innovation, en adoptant une stratégie « ouverte » et en recourant fortement à l'Open Source. À suivre dans l'actualité, pour mesurer la profondeur de ce mouvement, les EuroCloud Awards organisés par l'association EuroCloud, qui vont récompenser les meilleures entreprises dans six domaines liés au Cloud. Plusieurs entreprises françaises et de nombreuses entreprises européennes sont d'ores et déjà sélectionnées.

En parallèle, et par contraste, l'impact du Cloud sur l'individu et sur la société a été perçu en Europe comme un possible danger et a donné lieu à de nombreux débats sur la protection des données personnelles et sur la souveraineté des États. Les écoutes de la NSA ont favorisé la prise de conscience, provoquant un mouvement de défiance vis-à-vis des sociétés américaines, et qui nous a amenés à nous interroger.

Ce premier temps passé, des réflexions plus approfondies sont en cours, l'exploitation massive des données n'étant pas nécessairement incompatible avec la nécessaire protection des données personnelles, les premiers exemples de ce changement d'approche sont déjà visibles dans l'actualité.

L'Europe saura-t-elle concilier opportunité technologique et respect éthique ? L'enjeu est considérable, notre sensibilité continentale peut être un atout si nous savons nous montrer entreprenants, imaginatifs, rapides, pragmatiques et surtout positifs dans cette recherche d'équilibre entre protection et opportunités.

À PROPOS DE L'AUTEUR

Philippe Recouppé a mené sa carrière à la croisée des mondes informatique, télécom et formation.

Co-fondateur en 1988 de COGITEC, entreprise de formation et développement en informatique et télécommunication, il fonde en 1995 COGICOM, entreprise centrée sur la formation et le conseil télécom, intervenant en France et dans le monde entier.

En 2007, il fonde l'association Forum ATENA dont il devient le président. Forum ATENA a pour vocation de fédérer et animer le monde du numérique, des entreprises et de l'enseignement supérieur.

Enseignant à l'ISEP, Philippe Recouppé est également chargé du développement du Mastère Spécialisé Expert Cloud Computing, premier Mastère Spécialisé dans ce domaine en France. Il est également en charge du Mastère spécialisé Management et Protection des données à caractère personnelles.



LA CONNAISSANCE DES DONNÉES, CLÉ DE LA STRATEGIE CLOUD DE L'ENTREPRISE

Par Frédérique Patron

Titulaire du Mastère Spécialisé Expert Cloud Computing et SaaS de l'ISEP

Chef de projet en conduite du changement au Ministère de l'Intérieur

► LE CLOUD COMPUTING

Pour ceux qui n'ont qu'une vague idée de ce qu'est le Cloud, je propose de repartir de la définition du NIST (National Institute of Standards and Technology).

Le Cloud Computing est une nouvelle façon de délivrer les ressources et non une nouvelle technologie.

Le Cloud Computing doit présenter **les cinq caractéristiques** suivantes :

- **Un accès en libre-service à la demande** aux capacités de calcul : pas d'intermédiaire.
- **Un accès ubiquitaire au réseau** : accessible quel que soit l'endroit où se trouve l'utilisateur.
- **Une mise en commun des ressources** : partage des ressources techniques par différents services ou colocation d'un même service de façon étanche.
- **Une « Élasticité rapide »** : montée en charge transparente pour l'utilisateur, pas de ralentissement du fait d'une augmentation du nombre de connexions au service.
- **Un service mesuré en permanence** : ne payer ou facturer que les ressources utilisées : mémoire, vCPU. Location de ressources au temps d'utilisation.

Cela correspond aux trois principaux **modèles de services** ci-dessous :

- **Cloud Software as a Service (SaaS)** : location d'applications au mois, par exemple Deezer.
- **Cloud Platform as a Service (PaaS)** : location au temps passé d'environnement de développement complet.
- **Cloud Infrastructure as a Service (IaaS)** : location de temps machine : vCPU, espace disque, espace de stockage...

D'autres modèles de service cloud computing se sont développés, comme DBaaS pour DataBase ou DRaaS pour Disaster Recovery par exemple.

On distingue quatre **modèles de déploiement** :

- **Le Cloud privé** : le cloud appartient à l'entreprise. Pour respecter la cinquième caractéristique du cloud, le service cloud est mesuré et peut être refacturé à la direction utilisatrice au sein de l'entreprise.
- **Le Cloud public** : à l'opposé le cloud n'appartient pas à l'entreprise. En France, il y a deux clouds souverains (Français) Numergy et Cloudwatt qui garantissent un hébergement en France.
- **Le Cloud communautaire** : il s'agit d'un cloud public ouvert à un ensemble spécifique d'utilisateurs.
- **Le Cloud hybride** : comme nous pourrions nous en douter, il s'agit d'un cloud privé qui peut en cas de besoin de ressource s'étendre sur un cloud public.

En résumé, avec le Cloud Computing, l'informatique est dorénavant consommée comme un service.

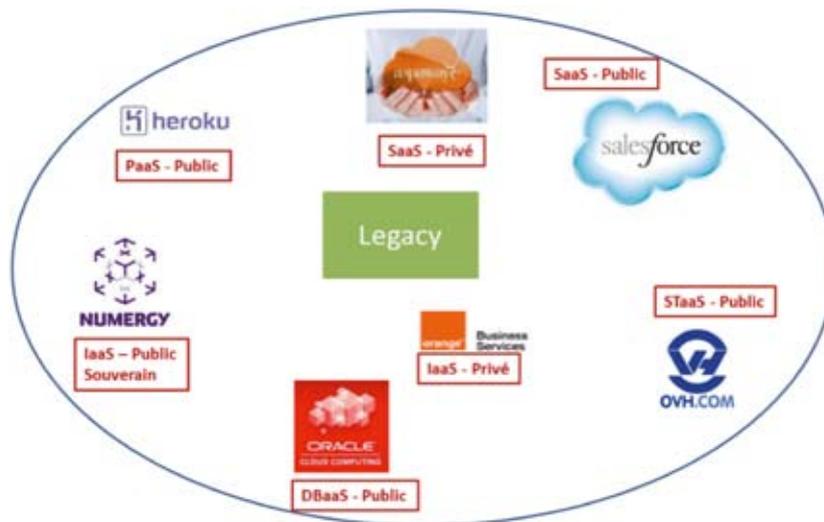
► CONSÉQUENCES POUR L'ENTREPRISE

Le multi-clouds

Nous constatons que l'entreprise a à sa disposition différentes solutions Cloud pour répondre à ses multiples besoins, et il est fort probable qu'elle en choisisse plusieurs :

- Une application CRM en mode SaaS dans un Cloud public, type Salesforce pour gérer ses clients.
- Des environnements de développements PaaS en Cloud public, comme la plateforme cloud Gandi pour les nouvelles applications qui seront déployées, soit sur un IaaS public type AWS (Amazon Web Services), soit sur un IaaS privé appartenant à l'entreprise.

Le système d'information devient alors multi-clouds et peut ressembler au schéma ci-dessous :



Le « Legacy » représente les composants historiques du Système d'Information de l'entreprise.

Pour que l'entreprise conserve son efficacité, les différents clouds doivent communiquer entre eux. Se pose alors le problème de leur intégration.

Architecture d'intégration

Avec le multi-clouds, les enjeux de la réussite de l'intégration des composants clouds au système d'information prennent une autre ampleur avec la prise en compte des items suivants :

- Les applications SaaS imposées par la maîtrise d'ouvrage sans étude préalable de leur possibilité d'intégration avec le reste du système d'information,
- Les changements de la façon de travailler du service informatique de l'entreprise,
- La dépendance du système d'information vis-à-vis d'Internet,
- La législation à respecter dans un contexte ouvert,
- La contractualisation et la définition de niveaux de service communs à tous les fournisseurs Cloud,
- La dépendance vis-à-vis d'un fournisseur (tel le fournisseur d'accès réseau par exemple),
- Le nombre d'acteurs à coordonner pour assurer la continuité de service,
- L'hétérogénéité des solutions et la difficulté à les interopérer pour maintenir la cohérence de l'ensemble
- La complexité de la sécurisation du système d'information étendu
- Le changement de relation avec les utilisateurs qui deviennent en quelque sorte clients du système d'information...

Il faut que l'entreprise définisse une architecture d'intégration des clouds pour pouvoir bénéficier pleinement dans le temps de tous les nouveaux avantages (flexibilité, rapidité, coûts optimisés, industrialisation des pro-

cessus) apportés par le Cloud Computing, sans risques, entre autres, pour la sécurité, la qualité de service ou l'intégrité des données.

Qui dit intégration de clouds, dit diffusion des données de l'intérieur vers l'extérieur de l'entreprise et inversement.

Diffusion des données

« Le multi-clouds contribue à l'entropie du systèmes d'information ».

Imaginons le multi-clouds, tel l'univers, en expansion. Il s'agit bien d'un changement de dimension pour les entreprises.



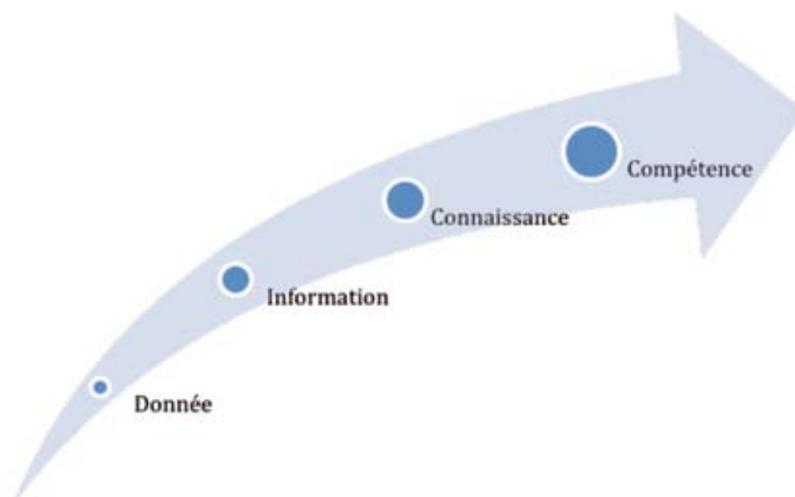
Au-delà du plaisir intellectuel que procure cette image, il ne faut pas oublier deux points :

- L'entropie mesure le degré de désordre d'un système ;
- Qui dit expansion du SI dit aussi expansion de la diffusion des données.

« Il faut donc maîtriser la diffusion des données. »

► FOCUS SUR LES DONNÉES

Les données sont le composant élémentaire de l'information, la connaissance et la compétence de l'entreprise, et à ce titre les données sont la véritable richesse de l'entreprise.



DIFFERENTES NATURES DE DONNÉES

Les données à caractère personnel

L'article 2 de la loi Informatique et libertés définit les données à caractère personnel de la façon suivante : « *Toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres* »

Il s'agit de données relatives à des personnes physiques. Les données relatives aux personnes morales sont exclues du bénéfice de la protection des données à caractère personnel.

Les données sensibles

Au sens de l'article 8 de la loi Informatique et Libertés du 6 janvier 1978 modifiée, sont considérées comme sensibles : les « *Données à caractère personnel qui font apparaître, directement ou indirectement, les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses ou l'appartenance syndicale des personnes, ou qui sont relatives à la santé ou à la vie sexuelle de celles-ci* ».

En revanche, l'adresse, la nationalité, la langue parlée et le lieu de naissance ne sont pas considérés par la CNIL comme des données « sensibles » au sens de l'article 8. **Cependant, la pertinence de leur collecte et de leur traitement doit être dûment justifiée** au cas par cas, par le responsable du traitement.

Les notions de « données sensibles » et de critères de discrimination ne coïncident pas toujours. Certains des critères de discrimination, tels que l'âge, le sexe, la situation de famille, ne correspondent pas à la notion de « données sensibles » au sens de la loi « Informatique et Libertés ».

De même, les données relatives aux infractions, condamnations et mesures de sûreté bénéficient d'un régime spécifique. Il est interdit de collecter et d'enregistrer ces données, seuls les fichiers d'État (police, RG, etc.) pouvant déroger à cette règle, sous certaines conditions.

Les données stratégiques

Il est important de ne pas confondre « les données qui sont sensibles pour une organisation » (des données financières ou un fichier clients par exemple pour une entreprise) et le régime juridique des « données sensibles au sens de la loi de 1978 ».

Nous parlerons alors de données stratégiques pour désigner les données importantes et confidentielles de l'organisation mais non couvertes par un cadre juridique.

A noter qu'une donnée stratégique peut être aussi une donnée sensible.

Les données de référence

Les données de référence, aussi appelées données maîtres ou en anglais master data, sont des données partagées entre plusieurs applications en même temps, elles soutiennent l'activité courante de l'entreprise et ses prises de décision.

Exemple : une application de facturation utilise des données de natures différentes :

- La liste des articles achetés par le client ainsi que la quantité de chaque article. Cette liste varie à chaque commande en fonction de ce dont le client a besoin. Ce sont des données qui varient à chaque transaction/opération, ce sont des données dites opérationnelles.
- Le nom, l'adresse de facturation du client, le numéro et la description des articles sont des données qui varient peu avec le temps et qui ne dépendent pas de la transaction et qui sont nécessaires à l'opération de facturation. Ce sont des données de référence.

LES EXIGENCES PAR NATURE

Nature	Personnelle	Sensible	Stratégique	Autres
De référence	Disponibilité Intégrité Cohérence Confidentialité	Disponibilité Intégrité Cohérence Confidentialité	Disponibilité Intégrité Cohérence Confidentialité	Disponibilité Intégrité Cohérence
Opérationnelle	Disponibilité Intégrité Confidentialité	Disponibilité Intégrité Confidentialité	Disponibilité Intégrité Confidentialité	Disponibilité Intégrité

► LA SOLUTION : LA GOUVERNANCE D'UN SYSTÈME D'INFORMATION MULTI-CLOUDS PILOTÉE PAR LES DONNÉES.

Un certain nombre de solutions techniques permettent l'intégration de clouds et répondent aux exigences des données en termes de disponibilité, intégrité, cohérence et confidentialité. Cependant d'autres points restent en suspens :

- Qui est le plus à même d'évaluer les problématiques juridiques du Cloud ? De la localisation des données ?
- Que deviennent les données si l'opérateur Cloud dépose le bilan ?
- La réversibilité des données et l'intégrabilité ne devraient-elles pas être étudiées avant de souscrire à toute offre ?

La mise en place d'un comité de gouvernance multi-clouds composé de personnes de la direction informatique, de la sécurité, du juridique, des achats, des métiers... veillera au respect des exigences liées aux données : disponibilité, intégrité, cohérence et confidentialité et sera organisée pour ne pas bloquer tous les projets par manque de bande passante.

Pour cela, la gouvernance doit avant tout connaître la nature des données pour définir la stratégie cloud de l'entreprise. Cette connaissance lui permettra de :

- Identifier les données et traitements qui seront dans le Cloud,
- Concevoir une architecture d'intégration,
- Définir des exigences en termes de sécurité technique et juridique,
- Analyser les risques et proposer les solutions adéquates,
- Définir le type de cloud (interne, externe) approprié aux utilisations envisagées,
- Accompagner le changement,
- Choisir le prestataire répondant aux exigences définies préalablement,
- Adapter la politique de sécurité interne aux nouvelles conditions,
- Évaluer la sécurité des données de l'entreprise :
 - identifier si des organisations tierces ont accès aux données et aux informations qui sont dans le cloud,
 - s'assurer qu'il y a une protection des données efficaces de la part du prestataire (segmentation de données, prévention d'attaques internes, techniques de protection, chiffrement...)
 - s'assurer de la disponibilité des données en cas de défaillance du prestataire cloud.

La gouvernance du SI multi-clouds doit permettre de renforcer la maîtrise de la localisation des données, la protection des données personnelles, et empêcher les transferts inappropriés d'information par rapport aux réglementations nationales.

C'est en analysant les risques et les opportunités que représentent le cloud que les entreprises pourront identifier l'intérêt ou non d'une migration de leurs données dans le cloud.

► CONCLUSION

Depuis quelques années, avec l'informatisation des entreprises et le déploiement des systèmes ouverts interconnectés, les entreprises ont de plus en plus conscience de la nécessité de gérer ou administrer des données, et avant tout les données de référence. Nous pouvons le constater par l'essor des projets MDM (Master Data Management).

En allant vers le multi-clouds, cela sera l'occasion pour l'entreprise de poursuivre ce mouvement en mettant la donnée à sa place, au centre de son organisation, car c'est la donnée qui guidera les orientations stratégiques pour le Cloud et donc par conséquent celles de l'entreprise.

Elle permettra d'établir une communication entre la technique et les métiers avec la mise en place de la gouvernance multi-clouds et facilitera ainsi le passage vers une informatique de service pérenne.

La place centrale de la donnée permettra de détecter rapidement les changements de stratégie cloud à effectuer lorsque la législation évolue ou tout simplement les règles métier entraînant un changement de nature de la donnée.

À PROPOS DE L'AUTEUR

Après avoir travaillé en société de service, **Frédérique Patron** a été responsable du domaine Échanges de données, Dématérialisation et Archivage des documents pour un des leaders du travail temporaire dans le monde. Elle est aujourd'hui chef de projet en conduite du changement au Ministère de l'Intérieur. En 2014, elle est diplômée de la 2^{ème} promotion du Master spécialisé «Expert Cloud Computing et Saas » de l'ISEP.



QUEL AVENIR POUR LE CLOUD ?

Par Samir Koleilat
Président - fondateur d'ACROPOLIS _ Vice - Président de Forum ATENA

► LE CLOUD MOTEUR DE LA MOBILITÉ

Faisons tout d'abord un bilan des apports du cloud en matière de mobilité.

S'affranchissant de la dépendance des données à leur support physique, le cloud est l'élément indispensable pour assurer l'interopérabilité et la synchronisation de bases de données distribuées sur divers supports. Vos données sont toujours avec vous et à jour, où que vous vous trouviez, sur votre ordinateur, votre smartphone ou encore votre tablette. Outre cet aspect de partage, les données de référence stockées dans le cloud assurent une sauvegarde fiable.

Toutes vos applications, tant vos applications « business » (messagerie, agenda, carnet d'adresses...) que vos applications opérationnelles sont des applications cloud, que vous utilisez en SaaS.

► L'IRRESISTIBLE ASCENSION DU CLOUD

Les statistiques (2015) nous révèlent que :

- 92% des directeurs techniques (CTO) pensent que leur infrastructure n'est pas assez flexible,
- Les services cloud augmentent de 20% tous les ans,
- 30% des entreprises ont l'intention de passer au cloud dans l'année,
- 54% des entreprises utilisent déjà les applications SaaS.

Rien qu'en France, le marché du cloud devrait augmenter de 4,9 milliards dans l'année.

Que trouve-t-on aujourd'hui dans le cloud ? Tout d'abord un espace de stockage fiable et adaptatif pour stocker des données, des images, des sites Web... Ensuite un partage d'applications centralisées dont les données sont synchronisées entre différentes machines, comme la messagerie, l'agenda, le carnet d'adresses cités ci-dessus, pour ne parler que de celles-ci. Ceci parce que les trois « Grands » que sont Microsoft, Apple et Google l'ont rendu possible (One Drive, iCloud, Drive)... et aussi les autres (IBM Cloud, SugarSync, Amazon Cloud Drive, etc.).

► ET APRES ?

L'objectif est qu'en 2018 il n'y ait plus de licences d'installation de logiciels sur des serveurs. Tout sera proposé en SaaS et les fournisseurs de SAAS vont remplacer l'ancienne génération de revendeurs à valeur ajoutée. C'est déjà le cas de Windows Office 365 office, SAGE ERP, Exchange, Zimbra, Google Apps et autres.

Les avantages

Les avantages en sont nombreux. Tout d'abord, le cloud améliore la productivité et la mobilité, comme nous l'avons vu. Les DSI n'ont plus le souci de la maintenance d'un parc à jour. L'utilisation d'applications en SaaS garantit à l'ensemble des utilisateurs qu'ils travaillent sur la même version la plus à jour d'un logiciel. L'utilisation d'un logiciel en SaaS en garantit l'authenticité et met fin aux pratiques de logiciels pirates.

Le coût d'investissement est réduit de manière drastique puisque l'achat de licences (CAPEX) est remplacé par une formule d'abonnement (OPEX). L'entreprise ne paie que ce qu'elle consomme en ajustant sa demande à ses besoins, ce qui est également facteur d'économies. En effet, elle choisit des applications en mode locatif, payées par utilisateur et par mois, autrefois trop chères à l'acquisition sous forme de licences logicielles et qui de plus nécessitaient des investissements en serveurs coûteux pour les héberger. L'utilisateur accède à des applicatifs spécifiques ou à des capacités de calcul sur une durée déterminée et ne paie que le temps et la capacité choisis.

Les données de l'entreprise ne sont plus éparpillées sur plusieurs ordinateurs personnels mais centralisées dans le Cloud commun à tous les utilisateurs d'une même entreprise.

Comment ?

Il faut oublier les vieilles habitudes, s'habituer à la dématérialisation totale... L'entreprise qui migre en SaaS doit mener une réelle conduite du changement pour glisser « en douceur » dans le mode cloud.

Tout passant par le Net, il est impératif de disposer d'une bande passante performante. La fibre et la 4G répondent parfaitement à ce besoin.

Il faut enfin être vigilant quant à l'approche législative du cloud et bien choisir son ou ses fournisseurs. En tout premier lieu, il convient de prendre en compte l'aspect légal du lieu de présence des documents, afin que les données ne soient accessibles que par les services légaux du pays et uniquement sur réquisition judiciaire légale, afin d'écartier le danger de la mise des données à la disposition des services étrangers, tel que le « Patriot Act ».

► POUR CONCLURE : N'AYONS PLUS PEUR DU CLOUD !

Voyons ce qui concerne les équipements des entreprises, et surtout le budget informatique de productivité de ses collaborateurs, tout d'abord dans le modèle traditionnel. Par temps de crise ou pas, il est le suivant : les serveurs sont amortis, les logiciels sont installés et payés et le fonctionnement est normal. Cette situation est très confortable, tant au niveau des financiers qu'au niveau des dirigeants qui voient les charges baisser et les amortissements fondre.

La question qui se pose est alors la suivante : quel est l'instrument de mesure de la productivité d'une entreprise ? Comment un dirigeant peut-il être certain que ses équipements informatiques lui apportent le bon rendement et la bonne efficacité ? A quel moment cette productivité commence-t-elle à baisser ? La routine est-elle une source de tranquillité ? Si les cadres ne demandent rien, est-ce un bon signe ?

À quel moment une société doit-elle se mettre en cause et comparer ses équipements à ce qui se fait sur le marché ?

Si la situation d'attentisme est confortable, elle est soutenue par l'absence d'investissement, mais si l'évolution n'est plus conditionnée par des investissements, ne faut-il pas chercher cette évolution afin d'être à la pointe de progrès sans toucher aux fonds propres ou à un endettement éventuel ?

Rares sont les entreprises qui sont passées de la théorie à la pratique, à savoir, comparer leur productivité par rapport au marché global en général, et à certains concurrents en particulier. C'est un exercice assez compliqué... S'il faut prendre une image, nous vivons mieux si nous ne voyons pas de médecin, car nous ne voulons pas aller voir un médecin de peur qu'il nous dise que nous sommes malades ; ignorer notre état de santé est mieux que savoir quelle maladie nous devons soigner. La comparaison est assez brutale, oui, mais voilà, si l'esprit est frappé par cette comparaison, il sera plus attentif au cloud...

Pour passer au cloud, il n'y a pas d'investissement à faire. Il y a certes le coût de la saisie de l'opportunité et de la migration, mais ici migration ne veut pas dire changer de logiciel ou d'habitudes, ni avoir besoin de formation. Ce sont en effet ces contraintes qui souvent freinent les changements car la formation est considérée comme une dépense importante et elle écarte les collaborateurs de leur poste de travail pendant leur déroulement, ralentissant de ce fait la productivité tout le temps que dure l'adaptation à un nouvel environnement de travail.

Si le sujet du Cloud est un sujet assez sensible, c'est parce qu'il a des coûts cachés, tels que la gestion du changement, mais ne perdons pas de vue que dans le modèle traditionnel, le budget informatique des entreprises est élevé, et ceci d'autant plus que le nombre des collaborateurs est élevé.

Plus le nombre des collaborateurs de l'entreprise est important, plus son budget informatique est important, et plus le besoin se justifie pour passer au CLOUD.

À PROPOS DE L'AUTEUR

Samir Rachad Koleilat est né à Beyrouth en 1954. En France depuis 1976 et après un DUT à Metz, il suit des cours d'informatique au CNAM à Metz puis à Nancy en formation continue. A Paris en 1985, il suit des cours de télécommunication en formation continue jusqu'au master à l'ENST.

Serial entrepreneur, après son succès en 1988-89 dans sa première entreprise HYQUEST, il part travailler pour Western Digital à Irvin en Californie du sud. De retour en France en 1991 il crée la société CHIP-TEC, et en parallèle, il crée la société INTERNEXT, opérateur d'opérateur pour la vente d'accès Internet aux ISP. En 2000 PRIMUS rachète INTERNEXT et le nomme président du groupe en France. En 2001 à la chute de la bulle Internet il crée ACROPOLIS TELECOM, opérateur de téléphonie sur IP, première plateforme de virtualisation des iPBX téléphoniques des entreprises, ensuite de virtualisation des serveurs et lance en 2008 des data center dédiés au stockage des données et des logiciels des entreprises.

Il est inventeur du brevet IDFIX sur l'identité numérique (2005).

De 1998 à 2008 il a été l'animateur du fond d'investissement Busines Angel, CYBOURSE. Samir Koleilat est également membre fondateur et vice-président de l'association Forum ATENA depuis sa création en 2007.



VERS UNE ARCHITECTURE BIG DATA MODERNE POUR LA BUSINESS INTELLIGENCE : DATA LAKE

par Karim Ben Mabrouk,
Titulaire du Mastère Spécialisé Expert Cloud Computing et SaaS de l'ISEP
Fondateur de KBM Consulting.

A l'âge du digital, la croissance permanente des données a donné naissance au concept du Big Data. Les entreprises sont de plus en plus conscientes que la donnée est un actif précieux dont il faut tirer de la valeur. L'exploitation de ces données, de plus en plus massives, variées et volatiles, nécessite la mise en place de solutions techniques capables d'intégrer et d'exploiter les Big Data, et la mise en place d'une gouvernance et une gestion des données internes et externes à l'entreprise.

Dans le contexte économique actuel, l'ensemble des entreprises essaient d'améliorer leurs performances économiques. Un premier levier consiste à optimiser le pilotage des différentes activités (production, vente, marketing, innovation...) par la mise en place de systèmes analytiques et d'aide à la décision permettant non seulement de mesurer les performances, mais d'anticiper, de prévoir et d'ajuster afin de les optimiser. Un deuxième levier est la maîtrise des coûts de fonctionnement, notamment ceux relatifs aux systèmes d'information. Le Cloud Computing offre en outre un modèle économique très compétitif pour consommer de l'informatique.

Le Lac de données (ou Data Lake) est une architecture moderne qui répond à ces deux enjeux. D'un côté, il offre une plateforme Business Intelligence et analytique capable de tirer valeur des données. D'un autre côté, il se déploie sur une infrastructure cloud constituée d'équipements à bas coût et dont la facture est proportionnelle à l'utilisation.

► DE LA BUSINESS INTELLIGENCE A LA BUSINESS DISCOVERY

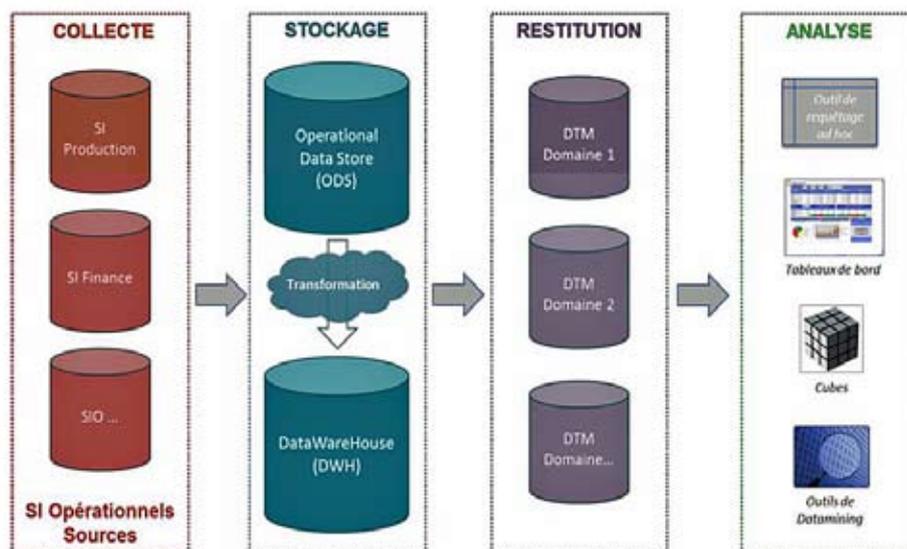
La Business Intelligence est une discipline à la frontière du management et de l'informatique, dont le développement a été tiré au cours des 20 dernières années, d'une part par les progrès technologiques de stockage et d'accès de plus en plus rapide et de plus en plus facile à de grands volumes de données centralisées ou centralisables, et d'autre part par la demande croissante des entreprises et de leurs managers d'accéder à des données de qualité, compréhensibles et partagées, permettant de faciliter le pilotage et la prise de décision.

A l'image des États qui se sont dotés de lourds outils pour piloter les comptes de la nation et administrer leurs populations, les entreprises ont également ressenti le besoin de se doter de systèmes informatiques complémentaires à leurs systèmes classiques de comptabilité, pour piloter leur activité commerciale, maîtriser leur gestion financière et améliorer leur efficacité opérationnelle.

Pour l'essentiel, il s'est agi pour elles de mettre en place une vision rétrospective la plus fidèle possible de leur situation sur ces différents axes d'analyse en se dotant de moyens industrialisés leur permettant de pratiquer une statistique souvent descriptive et parfois explicative, et plus rarement, prédictive ou prescriptive.

Pour cela, les acteurs de la Business Intelligence ont développé des outils visant à donner à leurs clients les moyens d'atteindre ces objectifs de restitution et de pilotage.

Le marché est aujourd'hui mature et a conduit la plupart des entreprises à mettre en place des systèmes décisionnels que l'on peut schématiser de la façon suivante :



Ces outils s'articulent autour de quatre couches principales :

- La première est une couche de collecte et de transformation des données (couche technique),
- La deuxième est une couche de stockage des données (couche technique),
- La troisième est une couche de restitution de l'information (couche métier),
- La dernière est une couche d'analyse de l'information (couche métier).

LES LIMITES DE LA BI CLASSIQUE

La première de ces limites, et probablement la plus sévère, est liée au temps de mise à disposition des données aux utilisateurs finaux. L'évolution des pratiques commerciales et l'accélération de la prise de décision face aux contraintes économiques actuelles se heurtent en effet à un certain manque de souplesse des systèmes décisionnels. Les autres limites correspondent à des nouveaux usages non encore incorporés dans les solutions classiques : ainsi, par exemple, dans une économie de plus en plus interconnectée, il apparaît de plus en plus vital de pouvoir croiser les données internes (clients, produits, ventes, production...) avec les données externes (réseaux sociaux, données macro-économiques...). Or les systèmes décisionnels sont historiquement conçus pour centraliser les données internes à l'entreprise et de fait, ils sont encore souvent faiblement outillés pour répondre à ce type de problématiques, pourtant de plus en plus pressantes. De même, les solutions actuelles ne sont pas adaptées à un usage mobile. Alors que l'utilisation des appareils mobiles dans le cadre professionnel se généralise, rares sont encore les produits qui permettent de consulter des rapports ou de procéder à des analyses sur une tablette, par exemple dans le cadre d'une négociation commerciale chez un client. Enfin, l'aspect visuel et les options d'interaction des outils actuels sont peu avancés en comparaison de ce que permettent Excel (dans sa dernière version) ou les éditeurs spécialisés. Bien évidemment, les outils évoluent et commencent à proposer des représentations dépassant le simple tableau ou diagramme en barres. De nouvelles techniques voient le jour afin de représenter les données de façon synthétique et visuelle et de mettre en valeur le message que l'on souhaite faire passer, mais peu de produits les intègrent encore à ce jour.

DE GRANDES ÉVOLUTIONS EN COURS

Sous l'effet de plusieurs mouvements de fond (démocratisation de l'outil informatique, Cloud Computing, usages mobiles et temps réel, volume grossissant et de plus en plus déstructuré des données), la brique de restitution est actuellement en pleine transformation. On assiste progressivement au basculement d'un outil verrouillé et relativement peu dynamique vers une plateforme collaborative, personnalisée et interactive. Et dans les faits, l'essentiel des transformations en cours s'articule autour de deux mouvements :

DataDiscovery : placer l'utilisateur au cœur du dispositif pour mieux répondre à ses attentes

La société Gartner a dévoilé le Magic Quadrant, version octobre 2013, pour les éditeurs de solutions de Business Intelligence et de plateformes analytiques. Pour la première fois, dans la catégorie des leaders aux côtés d'Oracle, Microsoft, IBM ou SAP, on trouve des éditeurs plus modestes tels que Tableau Software, Tibco Spotfire ou QlikTech, qui ont précisément investi sur les fonctionnalités répondant aux attentes exprimées précédemment.

En plaçant l'utilisateur au centre de l'outil, elles font de celui-ci un véritable acteur plus qu'un simple consommateur et les caractéristiques différenciantes de ces outils découlent de fait de cette stratégie :

- L'architecture est plus souple et plus simple,
- L'utilisateur a directement accès à la donnée et peut en modifier la structure,
- Les tableaux de bord sont dynamiques et interactifs,
- Un mode collaboratif est mis en place,
- L'utilisation devient possible sur n'importe quel périphérique ...

DataViz : mettre en scène l'information pour améliorer l'impact visuel et faciliter la communication

Un terme est actuellement très présent dès que l'on parle de restitution de données : la DataViz (raccourci de Data Visualisation). Regroupant l'ensemble des techniques permettant de représenter de façon visuelle ou graphique une information, il ne s'agit pas à proprement parler d'un outil d'exploration des données mais plutôt d'une forme nouvelle de visualisation de l'information.

Le principe sous-jacent à cette tendance est de dépasser les représentations graphiques très convenues (courbes, histogrammes, diagrammes circulaires, nuages de points...) pour se diriger vers des supports permettant de faire passer un message clair et ayant un impact visuel plus fort.

La DataViz est fortement liée à l'émergence du Big Data. En effet, les volumes augmentant de façon rapide et déstructurée, il n'est plus possible de représenter rapidement et intuitivement les informations sous forme simple.

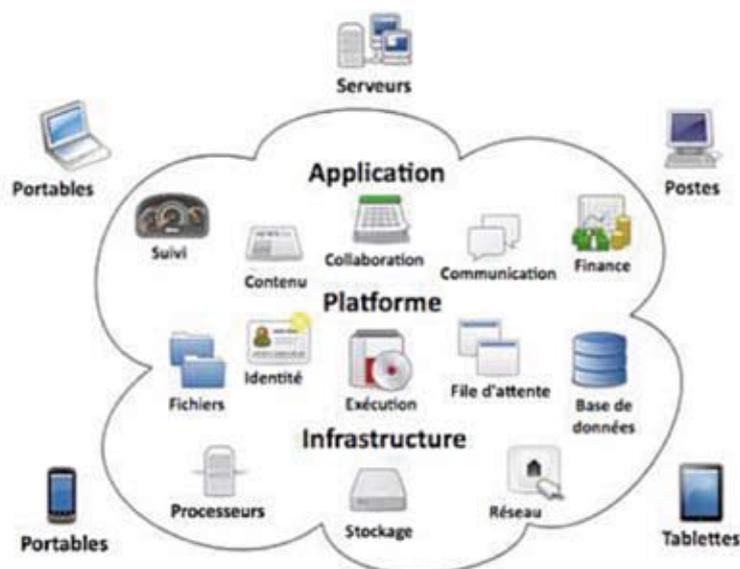
La DataDiscovery et la DataViz permettent à l'utilisateur d'avoir une première vision de ces données et des usages qu'il souhaite en faire, et donc l'aident à exprimer ces besoins, notamment en informatique décisionnelle.

Les besoins d'évolutions de la Business Intelligence sont de pouvoir intégrer, stocker, exploiter et partager tous types de données structurées et non structurées, quels que soient le format et la vitesse à laquelle elles sont produites, puis de disposer d'une plateforme collaborative et surtout évolutive, capable de s'adapter à la demande, tout en maîtrisant les coûts.

► CLOUD ET BIG DATA SONT UNE OPPORTUNITÉ POUR LA BUSINESS INTELLIGENCE

Le Cloud Computing offre l'élasticité, une plateforme de partage et de collaboration pour les systèmes décisionnels en gardant le principe de la facturation à l'usage.

Les ressources dans le Cloud peuvent être représentées de la manière suivante :



Le Nuage (Wikipedia)

Les caractéristiques essentielles du Cloud Computing sont la disponibilité mondiale en self-service, l'élasticité, l'ouverture, la mutualisation et le paiement à l'usage :

- **Ressources en self-service** : adaptation automatique à la demande. La capacité de stockage et la puissance de calcul sont adaptées automatiquement au besoin d'un consommateur. Ceci contraste avec la technique classique des hébergeurs où le consommateur doit faire une demande écrite à son fournisseur en vue d'obtenir une augmentation de la capacité – demande dont la prise en compte nécessite évidemment un certain temps. En Cloud Computing, la demande est automatique et la réponse est immédiate.
- **Ouverture** : les services de Cloud Computing sont mis à disposition sur l'Internet et utilisent des techniques standardisées qui permettent de s'en servir aussi bien avec un ordinateur qu'avec un téléphone ou une tablette.
- **Mutualisation** : elle permet de combiner des ressources hétérogènes (matériel, logiciel, trafic réseau) en vue de servir plusieurs consommateurs à qui les ressources sont automatiquement attribuées. La mutualisation améliore l'évolutivité et l'élasticité et permet d'adapter automatiquement les ressources aux variations de la demande.
- **Paiement à l'usage** : la quantité de service consommée dans le cloud est mesurée à des fins de contrôle, d'adaptation des moyens techniques et de facturation.

Les environnements « Big Data » apportent la possibilité de stocker et de manipuler tous types de données de manière distribuée, indispensable pour les nouvelles architectures décisionnelles.

Le nouveau Framework technologique du Big Data a été conçu pour contourner les limites de performance intrinsèques aux calculs en réseau (les échanges de données entre machines), tout en bénéficiant des avantages du stockage réparti (le Cloud public ou/et privé). Ce challenge ayant été relevé, il apparaît alors que les principaux apports de ces technologies nouvelles s'inscrivent dans trois dimensions principales, que l'on peut résumer de la façon suivante :

- **Axe Opérationnel** : la possibilité de placer des fonctions analytiques complexes au cœur d'applications opérationnelles temps-réel (notion de Data Centric Application) permettant d'imaginer de nouveaux produits et services.

- **Axe Analytique** : la possibilité d'opérer des analyses statistiques complexes et des calculs mathématiques en des temps réduits sur de plus grands volumes de données et sur des données de nature différente (non structurées) .
- **Axe Financier** : la promesse de réduction de coûts liée à la possible mutualisation des infrastructures (data centers) et à la scalabilité desdites infrastructures.

Il serait illusoire de dresser une architecture Big Data type, tant celle-ci doit dépendre des usages que l'on souhaite faire de la technologie et de la nature des données manipulées. Cela étant, il me semble utile d'esquisser une sorte de typologie des différentes formes architecturales possibles, ainsi que des principaux systèmes de gestion de bases de données nativement compatibles :

Trois grandes classes d'architecture se dégagent assez clairement, chacune avec leur propre socle de référence :

- Les architectures Big Data en traitement batch (Hadoop + Map Reduce),
- Les architectures Big Data proche du temps-réel (Hadoop + Message Object Middleware - Kafka),
- Les architectures Big Data temps-réel (Storm, Hadoop YARN 2.0, IBM Streams).

Quant au système de stockage des données, coexistent cinq grands groupes de bases de données Big Data dont les caractéristiques peuvent se recouper :

- Les bases de données orientées lignes et massivement parallèles selon un modèle MPP (Teradata, IBM Netezza, Microsoft SQL Server PDW, SAP HANA...),
- Les bases de données orientées colonnes (HP Vertica, Cassandra, HBase, SAP Sybase IQ...),
- Les bases de données orientées documents (MongoDB, CouchDB, Redis...),
- Les bases de données orientées graphes (Neo4J...),
- Les bases de données In Memory (CouchBase/Membase...),

Sachant que la plupart des bases de données fournissent la possibilité de traiter les données en mémoire.

► LE LAC DE DONNÉES : UNE NOUVELLE ARCHITECTURE D'ORGANISATION DES DONNÉES

Le Lac de données est une plateforme qui permet de collecter, de stocker tous types de données, quelles que soient leurs tailles, leurs formats et leurs complexités, et d'en extraire rapidement la valeur en la partageant avec les applications et les utilisateurs.

Aujourd'hui, l'entreprise a besoin de nouvelles architectures décisionnelles permettant d'intégrer l'exhaustivité de données internes et externes, structurées et non structurées, et capable de garantir une meilleure scalabilité.

Bien que les entreprises commencent à mettre leurs outils Business Intelligence dans le Cloud, elles sont cependant plus réticentes à faire la même chose pour leurs données. Les raisons sont principalement la sécurité des données et le respect de la vie privée. Ceci est en train de changer pour au moins les trois raisons suivantes :

- Le Cloud apporte l'extensibilité quasi infinie (l'élasticité, scalabilité),
- Les coûts dans le Cloud sont de plus en plus intéressants,
- Le Cloud est de plus en plus sûr.

A l'heure du digital, les entreprises produisent de plus en plus de données, en même temps qu'elles en reçoivent (depuis le Web, les réseaux sociaux, les objets connectés...). Ces données internes et externes s'avèrent précieuses et portent de la valeur à l'entreprise pour son développement, sa performance, à condition qu'elle soit capable de les exploiter. Les principales difficultés sont d'ordre technique et résident dans la collecte et le stockage de ces données massives, variées et volatiles.

L'apparition des systèmes de stockage distribués comme Hadoop a ouvert de nouvelles perspectives. Lorsque le besoin est de stocker de gros volumes de données à structure variable, mais surtout dont on ne sait pas à l'avance comment elles vont être utilisées et analysées, apparaît le concept de Lac de données (Data Lake). Les technologies Big Data sont aujourd'hui une solution à ces différentes problématiques techniques et ont montré leur capacité à collecter et stocker tous types de données venant de différentes sources et à plusieurs vitesses, et à les traiter d'une manière performante pour en tirer de la valeur (Le Savoir).

Le Big Data et le Cloud apportent la technologie et le modèle économique nécessaires aux nouvelles architectures décisionnelles dont l'entreprise a aujourd'hui besoin. Le Big Data intègre l'exhaustivité des données internes et externes structurées et non structurées. Le Cloud offre une plateforme multi-tenante garantissant une scalabilité à des coûts maîtrisés.

Avec la valeur de la donnée et son potentiel après exploitation, la plateforme Lac de données permet d'adresser des besoins émergents de différents secteurs, comme :

- Social & Web analytics,
- Flux de données temps réel,
- Advanced analytics.

Le Lac de données permet également d'apporter des améliorations techniques et de répondre à des nouveaux défis techniques, par exemple :

- Le besoin de supprimer les silos et d'intégrer les données non structurées,
- Améliorer les temps de mise à disposition des outils et des données,
- Réduire les coûts,
- Maîtriser les données,
- Offrir de nouvelles possibilités d'exploitation des données externes à l'entreprise.

Un Lac de données est une manière radicalement différente dans le stockage, le partage et le traitement des données de l'entreprise. Elle doit cependant adapter ses processus et sa culture. Voici quelques défis que les entreprises doivent surmonter :

- **Les politiques de partage de données :** le Lac de données stocke et partage les données. Au sein d'une même entreprise, les propriétaires de ces dernières peuvent résister au partage de leurs données pour des raisons de sécurité ou de confidentialité. L'entreprise doit mettre à jour sa politique de sécurité en fonction de ces contraintes et de cette nouvelle architecture. Ne pas oublier que la donnée, c'est le pouvoir.
- **La complexité de la migration des données existantes :** les systèmes hérités ont subi plusieurs opérations de maintenance corrective pour régler des anomalies ou des erreurs de conception. Des solutions de contournement ont souvent été mises en place et ont assez souvent impacté les données. Par conséquent, les données brutes peuvent fournir une valeur limitée en dehors de leur contexte. Or le Lac de données fonctionne de manière optimale lorsqu'il est utilisé avec des données non altérées par les systèmes sources.
- **Le cycle de vie des métadonnées :** les Lacs de données nécessitent des méthodes avancées de gestion des métadonnées, des fichiers de données et de suivi pour chaque opération de transformation.
- **La question de l'intégration :** elle est nécessaire pour transformer les données en informations exploitables. C'est un défi important. Bien que l'intégration de données ait lieu dans la couche Hadoop, la contextualisation des métadonnées (offrant une vue sur les données sélectionnées) est réalisée au moment de la création du schéma. Une couche d'intégration sécurisée est donc nécessaire pour relier le Lac aux diverses sources des données.

En résumé, le concept de Lac de données est à recommander pour les gros volumes de données dont on ne connaît pas a priori les structures. Il est donc complémentaire des entrepôts des données qui restent la structure la mieux adaptée à l'analyse répétitive et comparative des données structurées et historisées de l'entreprise.

LES PRINCIPALES CARACTÉRISTIQUES

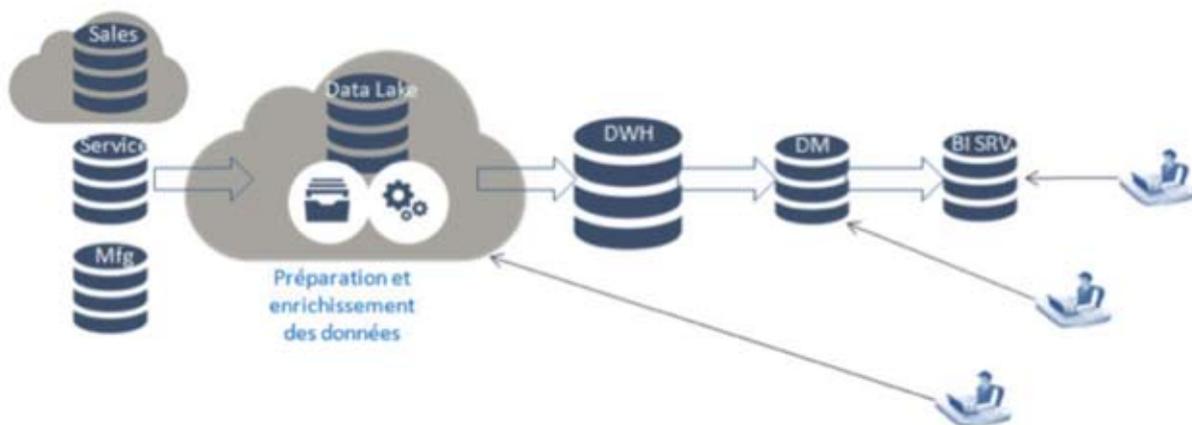
Le Lac de données doit tout d'abord, assurer la sécurité d'accès dans un contexte de multi-tenants ; ensuite, permettre la scalabilité inhérente à la croissance des données ; et bien sûr, garantir la disponibilité d'accès aux données.

A cela s'ajoutent des fonctions de recherche et de transformation de données.

Voici quelques exemples d'utilisation de nouvelles architectures Lac de données déployées dans le Cloud.

Le Lac de données est un ELT dans le Cloud

La préparation et l'enrichissement des données sont faits dans le Data Lake, l'entrepôt de données restant local.



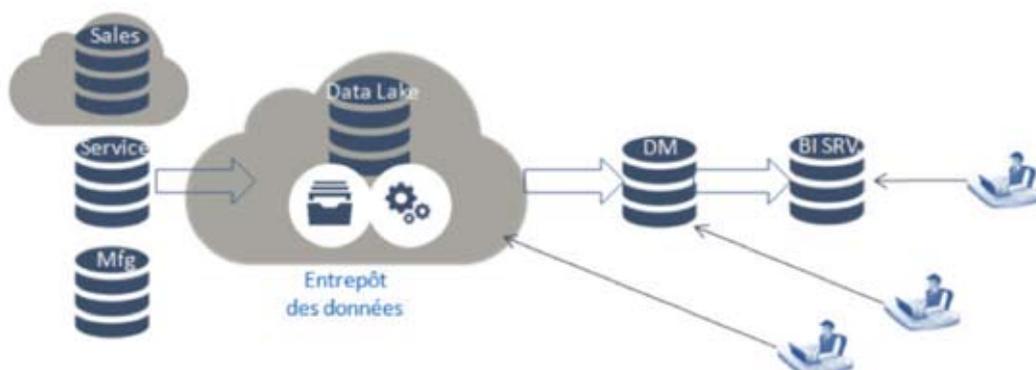
L'avantage d'utiliser un Data Lake pour la préparation et l'enrichissement des données est principalement le fait de pouvoir garder d'une manière exhaustive une version brute des données source en mode schéma à la lecture (schema-on-read).

Cette solution est rapide à déployer et offre à la fois provisionnement dynamique et scalabilité. Ses principaux inconvénients sont la latence entre le cloud et l'entrepôt de données et les aspects de sécurité des données et de confidentialité de la vie privée.

Comme pour les architectures BI classiques, un autre inconvénient de cette configuration est la première migration des données, historiquement dans un environnement classique interne, vers le cloud.

Le Lac de données est un entrepôt performant dans le cloud

Cette solution est une évolution de la précédente, dans un modèle économique plus intéressant qui offre davantage d'opportunités pour de nouvelles applications.

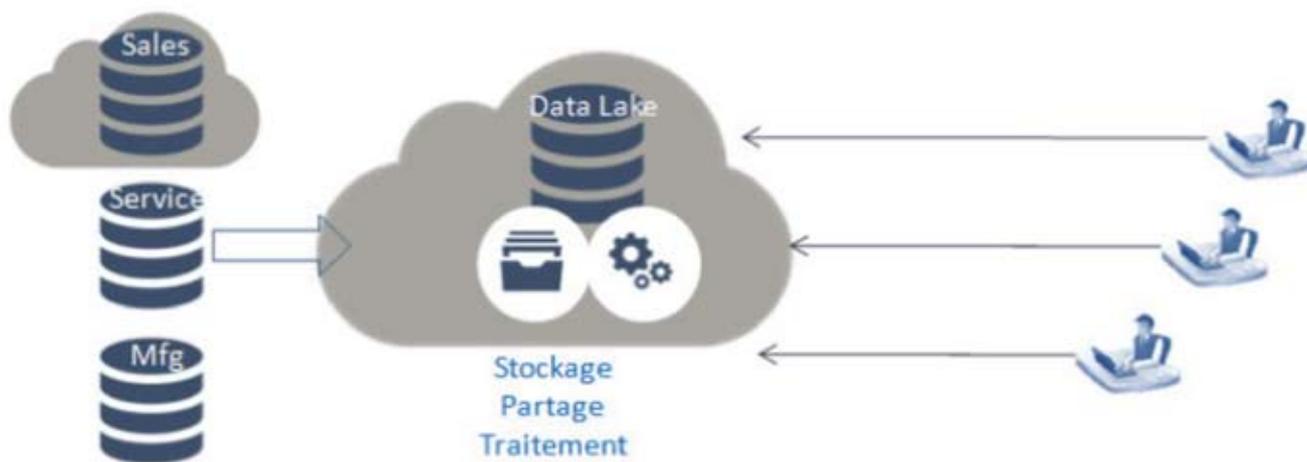


En plus des avantages cités dans la configuration précédente, cette utilisation du Lac de données en tant qu'entrepôt de données apporte plus de performance, mais surtout permet de déporter les traitements vers les données et non pas l'inverse (réduction des coûts et des temps de traitement : voir la partie Architecture BI émergentes – Data Lake).

Néanmoins, il reste à traiter les questions de la première migration des données et de la sécurité.

Le lac de données est un environnement BI et analytique dans le Cloud

Dans cette configuration le traitement des données est intégralement déporté dans le cloud.



Ce cas est le plus intéressant, dans le sens où il permet de tirer un bénéfice maximal des capacités des nouvelles architectures déployées dans le Cloud (public, privé ou hybride).

COMPARAISON D'ARCHITECTURES BI TRADITIONNELLE ET LAC DE DONNÉES

Contrairement à une architecture BI classique, un Lac de données d'entreprise est en mesure de :

- Recevoir et stocker un volume élevé de données structurées, semi-structurées et non structurées en temps réel, à l'aide du matériel de base à faible coût,
- Fournir une plate-forme pour effectuer des analyses en batch et en temps réel,
- Fournir une vue des données adaptée aux besoins des utilisateurs.

Le Lac de données permet de réduire le temps de mise en œuvre des solutions en :

- Éliminant la dépendance entre le modèle et la donnée,
- Réduisant le temps nécessaire pour construire un magasin de données par l'élimination de la couche de métadonnées,
- Fournissant la capacité de voir les mêmes données sous différentes dimensions et découvrir de nouveaux modèles des relations entre les données.

Un Lac de données est une approche simple mais puissante pour résoudre les problèmes de l'entreprise. Il s'adresse à l'évolution constante des besoins des entreprises en permettant le stockage de toutes les données et en fournissant la capacité de découverte des nouvelles pistes d'analyse et d'exploitation, mais il nécessite de changer les processus de gestion et de traitement des données.

Le nouveau processus de traitement de données dans une architecture BI moderne, en cinq étapes :

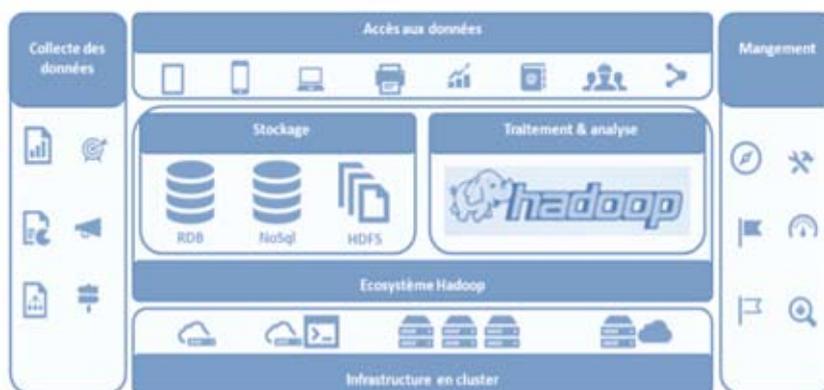


On remarque qu'une étape supplémentaire est apparue dans le processus : c'est la fonction de découverte en amont qui permet de gagner en connaissance et en maturité sur la donnée et son usage.

En plus de la prise en charge des données non structurées, l'architecture Lac de données se différencie des architectures BI classiques par la possibilité de découvrir les données en amont, par la capacité de réaliser les traitements sur les données brutes et par sa flexibilité qui permet de s'adapter aux évolutions des besoins d'analyse.

► LE LAC DE DONNÉES EST UNE ARCHITECTURE TECHNIQUE BASEE SUR DES OUTILS OPENSOURCE

Dans cet article, on ne développera pas l'architecture technique du Data Lake, mais on se contentera de présenter les principaux éléments qui constituent la plateforme. Le schéma ci-dessous met en évidence ces éléments :



L'architecture technique type de Lac de données se repose sur cinq couches principales qui sont les suivantes :

- Une infrastructure en cluster,
- L'écosystème Hadoop,
- Une couche de collecte des données venant de sources diverses,
- Une couche d'accès aux données stockées dans la plateforme,
- Une couche de management de la plate-forme.

► UNE NOUVELLE ORGANISATION ET DES NOUVEAUX ENJEUX HUMAINS

Un des impacts les plus significatifs de l'utilisation d'une architecture Big Data moderne pour la Business Intelligence est le changement ou la transformation organisationnelle nécessaire pour supporter et exploiter la plateforme Lac de données. Les anciens rôles doivent être redéfinis et de nouveaux doivent être créés, ce qui est à la fois une opportunité et un stress pour les collaborateurs et l'entreprise.

LA BUSINESS INTELLIGENCE ET LA DATA SCIENCE DANS UNE PLATEFORME UNIQUE NECESSITE UNE NOUVELLE ORGANISATION

Le Lac de données met en évidence le rapprochement de deux disciplines qui semblaient avant être très éloignées, et qui sont la Business Intelligence et la Data Science. Il est vrai que ces derniers jouent des rôles différents et requièrent également des compétences et des approches différentes, mais l'une ne remplace pas l'autre. Dans une approche Lac de données, les deux disciplines se complètent, l'une tirant avantage des forces et du focus de l'autre.

La Business Intelligence se concentre sur la compréhension des processus métiers à un niveau détaillé, dans le but de développer les métriques, les alertes et les rapports qui servent à superviser ces processus. Elle permet donc d'observer les processus de l'entreprise.

La Data Science essaye d'offrir une vision prédictive de l'entreprise et de ses activités afin d'anticiper l'avenir.

Le Lac de données est une plateforme qui répond aux besoins de ces deux acteurs, les analystes BI et les analystes de la Data Science. D'une manière plus générale, le Lac de données est un environnement qui concerne l'ensemble des acteurs intervenant dans la chaîne d'analyse des données.

Le cycle de vie de l'analyse des données :

Afin de tirer bénéfice de l'exploitation du Lac de données, les entreprises doivent développer un processus complet définissant les étapes d'analyse et d'exportation des données, clairement les rôles, les responsabilités et les attentes de tous les acteurs : les utilisateurs métiers, le responsable du Lac de données, l'analyste BI, le Data Scientist. L'analyse du cycle de vie de l'analyse des données permet de mieux comprendre les différents acteurs et les interactions entre eux.

- **L'utilisateur métier** est chargé de définir les processus métiers clés et d'identifier les métriques et les indicateurs clés de performance qui permettent de mesurer ces processus. Dans un second temps, l'utilisateur métier intervient en fin de chaîne pour utiliser les rapports les indicateurs et les différentes métriques produits pour les contrôler et décider de leur efficacité. Ces retours doivent alimenter la prochaine itération du cycle de vie analytique.
- **Le manager de la plateforme** est chargé de définir, développer et gérer la plateforme. Le rôle du manager de la plateforme Lac de données est élargi et concerne aussi l'administration des nouveaux outils Hadoop, les outils de traitement en mémoire et la fédération des données. Le Lac de données prend en charge les données structurées, semi-structurées et non structurées et permet aussi bien l'accès aux données stockées dans la plateforme, issues de plusieurs sources internes et externes à l'entreprise.
- **Le Data Scientist** est chargé de la fouille et la découverte des données internes et externes à l'entreprise, afin d'ouvrir de nouvelles perspectives. Ces spécialistes de la donnée recherchent de nouvelles sources de données capables de fournir des nouveaux axes d'analyse qui peuvent alimenter et améliorer les processus de l'entreprise.
- **L'analyste BI** est chargé d'identifier, gérer et publier les métriques et les indicateurs de performance qui permettent à l'utilisateur métier de contrôler et de mesurer les processus de l'entreprise. Un des nouveaux acteurs qui est mis en évidence dans une approche architecture Big Data/BI est le Data Scientist. Son rôle et ses responsabilités doivent donc être clairement définis.

Les nouveaux rôles dans l'organisation :

La mise en place de la plateforme Lac de données conduit les entreprises à repenser leur manière de gérer, étoffer et protéger leurs actifs « données » issus de gros volumes : perspective analytique, modèles analytiques et bien sûr les données elles-mêmes. Trois aspects sont intéressants :

- **L'expérience utilisateur :** Les grandes entreprises qui ont déjà mis en place des environnements Big Data, ont constaté l'impact direct de la présentation des résultats aux utilisateurs finaux. Une présentation qui n'est pas intuitive et concrète pour les utilisateurs ne peut en aucun cas être acceptée. Elles ont donc compris que la mise en place d'une équipe chargée de l'expérience utilisateur est indispensable.
- **Le rôle de management :** Les données et leurs analyses sont des actifs de l'entreprise qui doit les gérer d'une manière plus organisée. Le responsable des données et le responsable analytique doivent tous les deux gérer la propriété intellectuelle des données et des analyses.
- **Les responsables de la donnée** doivent assurer l'inventaire des données, leurs évaluations économiques, leurs monétisations, les sources de données et la gouvernance des données.
- **Les responsables analytiques** doivent assurer la gestion des actifs analytiques, leurs évaluations, la gestion de la propriété intellectuelle, le dépôt des brevets et la protection de la propriété intellectuelle. Privilégier la créativité (instaurer une culture de l'expérimentation) – la pensée « et si » – est la façon dont le concept d'expérimentation peut libérer la créativité des individus et de l'entreprise dans son ensemble. En effet le Lac de données est l'antidote de la paralysie analytique, car il offre les données, les outils et les méthodologies (dans les outils d'analyse) pour tester des idées, des hypothèses et apprendre de ces tests et aller de l'avant.

LES ACTEURS DU DATA LAKE

Les révolutions Big Data et Cloud Computing ont mis sur le devant de la scène les données massives et leur exploitation par des algorithmes mathématiques, et des solutions et des plateformes informatiques dédiées. Le Data Scientist et l'architecte Cloud/Big Data émergent manifestement comme les acteurs principaux de ces révolutions. Ils ne doivent cependant pas occulter le rôle des autres acteurs :

- L'expert métier,
- L'analyste,
- Le responsable des systèmes d'informations,
- Le fournisseur du service cloud.

Permettre à ces personnes aux talents et aux compétences complémentaires de travailler efficacement ensemble en surmontant les obstacles de leur improbable alliance, conditionne le succès de la mise en place d'une plateforme Lac de données.

On note ici que, parmi les acteurs principaux pour une architecture Lac de données, deux profils sont aujourd'hui très rares à trouver qui sont le Data Scientist et l'architecte Cloud Big Data.

► LES CARACTÉRISTIQUES DE LA RÉUSSITE

Une approche adaptée dès la conception :

Un Lac de données permet en outre à l'utilisateur d'analyser la diversité et le volume des données stockées. Cela nécessite des caractéristiques et des fonctionnalités pour assurer et conserver les données, puis de lancer des analyses, la visualisation et de générer des rapports. Les caractéristiques d'un Lac de données réussi sont :

- L'utilisation de plusieurs outils et produits,
- La spécificité du domaine d'application,
- La gestion automatisée de métadonnées,
- Des workflows d'ingestion configurables,

- L'intégration à l'environnement existant.

La collaboration entre équipe métier et informatique :

Pour réussir les projets Lac de données, il faut que les équipes métier et informatique collaborent afin d'identifier et appliquer une gestion de projet bien adaptée permettant d'atteindre les objectifs clairement définis à l'avance. Dans cette partie, nous présenterons une méthode à utiliser pour guider cette collaboration. Cette méthodologie met en place un processus qui passe par un atelier de créativité et une étude de la valeur attendue jusqu'à la production. Elle englobe les étapes suivantes :

- Identifier la cible métier,
- Déterminer les perspectives,
- Définir la stratégie de données,
- Construire des méthodes analytiques,
- Implémenter le Lac de données,
- Incorporer les perspectives dans les applications.

L'indice de maturité du modèle :

Le Lac de données est un moyen permettant d'exploiter l'analytique et la Business Intelligence sur des gros volumes de données. Cela peut dynamiser les processus de création de valeur. Prendre en compte les attentes de l'entreprise, les processus métiers qui peuvent exploiter les résultats et savoir utiliser les perspectives sont à considérer pour améliorer ou créer des opportunités business.

Afin de pouvoir tester les aspirations de l'entreprise, l'indice de maturité du modèle Big data engendré par le Lac de données peut s'avérer utile. Cet indice est un guide rapide et pratique capable d'aider les managers de l'entreprise à organiser les étapes de maturité, qui sont :

- Superviser l'activité,
- Identifier des perspectives,
- Optimiser l'activité,
- Monétiser les données,
- Transformer l'entreprise.

► CONCLUSION

En conclusion, après les analyses fonctionnelles et techniques effectuées dans le cadre de ma thèse, il apparaît clairement que le Lac de données est une architecture moderne Big Data qui répond pleinement aux attentes, à condition de l'adapter aux besoins et à la culture de l'entreprise et de mettre en place une organisation adéquate. Selon le type de données, leur volumétrie, les besoins métiers ou les architectures existantes des systèmes d'information et des systèmes décisionnels, les outils à mettre en œuvre sont très différents :

- **Enrichir les outils « classiques » existants :** ces derniers ont toujours leur place, quitte à les compléter par des outils Big Data.
- **Utiliser les modèles « solutions clés en main » :** des solutions Business Intelligence, Big Data et Analyse as a Service sont aujourd'hui proposées par plusieurs fournisseurs cloud et peuvent venir enrichir l'existant.
- **Mettre en place sa propre architecture moderne de stockage et traitement des données - son propre Lac de données :** des outils et composants open sources sont aujourd'hui disponibles et ont déjà fait leurs preuves. Les plateformes de Cloud Computing et des technologies du Big Data permettent d'intégrer ces nouveaux outils et écosystème à des coûts raisonnables.

À PROPOS DE L'AUTEUR

Karim Ben Mabrouk est fondateur de KBM Consulting, cabinet de conseil spécialisé dans les nouvelles technologies.

Après avoir passé dix ans dans le conseil en management des systèmes d'information en tant que Manager chez Bluestone Consulting, durant lesquels il a réalisé de nombreuses missions de conception et de mise en place de SI et applications métiers, Karim Ben Mabrouk a préparé le Mastère Spécialisé Expert Cloud Computing et SaaS dispensé par L'ISEP.

Karim Ben Mabrouk possède une forte expérience de la direction de projets et le management des consultants, il maîtrise le management des systèmes d'information et le développement des applications opérationnelles depuis le recueil des besoins jusqu'à la validation finale. Il a développé une très bonne connaissance des métiers de l'énergie, du transport et des télécommunications.



BIG ET LINKED DATA : INTERCONNEXION DE FLUX DE DONNÉES MASSIVES

Par Raja Chiky - Enseignante-chercheuse à l'ISEP

► BIG ET LINKED DATA

La quantité des flux de données collectées par des acteurs hétérogènes est en perpétuelle croissance. Ces flux peuvent provenir de capteurs installés sur de larges parties du territoire, de modèles numériques (comme les prévisions météorologiques), ou de sources de données disponibles sur le Web. Ils sont destinés à nous informer toujours plus précisément sur l'état de divers phénomènes. Cependant, leur difficile intelligibilité et leur conception cloisonnée rendent difficile la réutilisation et l'interopérabilité de ces sources d'information.

Les technologies du Web sémantique [Aufaure 2013, Aufaure&Chiky2014] peuvent apporter une solution à l'interopérabilité, la réutilisation et l'adaptation de flux de données provenant de sources diverses :

- En sémantisant les flux : en les exposant dans un format sémantique tel que RDF et en les caractérisant à l'aide d'ontologies adaptées ;
- En interconnectant les flux sémantisés ;
- En filtrant les flux en fonction du contexte des applications amenées à les exploiter ;
- En interrogeant efficacement les flux agrégés en fonction de requêtes précises et ciblées.

De plus, quand le volume de données augmente indéfiniment, il devient très coûteux, voire impossible, de stocker la totalité des flux avant leur traitement. Il est donc nécessaire de traiter à la volée les données et de ne stocker que les informations pertinentes ou des résumés de celles-ci. Cet aspect est au cœur des travaux que nous menons [Jain 2013] : traitements sémantiques et qui s'attacheront à l'extension du langage SPARQL (langage de requêtage sur données sémantiques) pour traiter les flux de données sémantisés afin de répondre aux besoins suivants :

- 1. Résistance au flux :** En raison des contraintes de stockage et de performance, les données doivent être traitées à la volée. Les algorithmes de flux en ligne sont restreints à ne faire qu'un seul passage sur les données et tout traitement doit pouvoir se passer d'un stockage impératif des données.
- 2. Traitement des requêtes continues :** Les requêtes traitant les données doivent s'évaluer continuellement afin de prendre en compte les nouvelles valeurs s'ajoutant au flux, et les réponses sont créées aussi en flux. Le modèle de données et la sémantique des requêtes doivent permettre les traitements basés sur le temps ou l'ordre des données. La sémantique du SPARQL est étendue en introduisant la notion de fenêtre (portion infinie d'un flux) et des opérateurs spécifiques au traitement de flux.
- 3. Traitement des données statiques et continues :** Dans de nombreuses applications, il peut être intéressant de comparer des données récentes avec des données historiques ou des ontologies considérées statiques. Pour ce faire, le SPARQL étendu doit pouvoir mêler les données statiques et les données dynamiques. Il doit autoriser le stockage, l'accès et la modification de données statiques et permettre leur combinaison avec des données dynamiques.
- 4. Traitement de la volumétrie :** Le système répond instantanément. Il fournit ainsi des réponses en temps réel en faisant face à un gros volume de données, ce qui nécessite un moteur hautement optimisé. L'incapacité de stocker un flux entier suggère l'utilisation de structures de résumés (approximation).

► L'ÉTAT DE L'ART

Le SPARQL est un langage de requêtes, largement inspiré du langage SQL, qui permet d'interroger des données stockées sous format RDF. Afin de pouvoir réaliser des requêtes sur des flux sémantiques, il a fallu étendre SPARQL, à l'image des extensions faites à SQL dans les systèmes de gestion de flux de données classiques. C-SPARQL [Barbieri 2009] (Continuous SPARQL) est l'une des premières extensions de SPARQL destinée à supporter les requêtes continues.

D'autres projets étendant SPARQL ont vu le jour. Streaming SPARQL étend le SPARQL pour que celui-ci puisse gérer des opérateurs de fenêtrage sans se préoccuper des performances des requêtes. EP-SPARQL [Anicic 2011] est un langage permettant de traiter des événements dans des flux RDF en utilisant Prolog. CQELS [Le-Phuoc 2011], le langage le plus récent, permet d'agir nativement sur les flux RDF et les requêtes continues sans passer par des outils intermédiaires.

Les projets existants dans la littérature ont étendu le SPARQL pour le traitement des requêtes continues avec une prise en compte de l'aspect temporel et des opérateurs de fenêtrage. Toutefois, aucun de ces projets ne s'est préoccupé d'étendre le SPARQL pour effectuer des résumés de flux de données (par échantillonnage par exemple) ou pour implémenter des fonctions d'oubli permettant d'historiser les flux de données avec une granularité variable en fonction du temps et en assurant un espace de stockage limité.

► RÉFÉRENCES

- [Anicic 2011] D. Anicic, P. Fodor, S. Rudolph, and N. Stojanovic. Ep-sparql: a unified language for event processing and stream reasoning. In WWW '11, pages 635–644, 2011.
- [Aufaure2013] M.A. Aufaure, R. Chiky, Massive and Linked Data: Semantic Information Aggregation from streams, social media, Linked Open Data and web documents, Position paper, World Summit on Big Data and Organization Design, 2013
- [Aufaure&Chiky 2014] Marie-Aude Aufaure, Raja Chiky, From Business Intelligence to Semantic data stream management, in MoBid-ER 2014, Springer, 2014
- [Barbieri 2009] Davide Francesco Barbieri, Daniele Braga, Stefano Ceri, Emanuele Della Valle et Michael Grossniklaus, C-SPARQL: SPARQL for Continuous Querying, Poster in WWW 2009, April 20–24, 2009, Madrid, Spain
- [Jain 2013] Naman Jain, Manuel Pozo, Raja Chiky, Zakia Kazi, Sampling Semantic Data Stream: Resolving Overload and Limited Storage Issues, in Daeng, Springer, 2013
- [Le-Phuoc 2011] Le-Phuoc, Danh and Dao-Tran, Minh and Xavier Parreira, Josiane and Hauswirth, Manfred, A Native and Adaptive Approach for Unified Processing of Linked Streams and Linked Data, The Semantic Web – ISWC 2011

À PROPOS DE L'AUTEUR

Raja Chiky est enseignante chercheuse à l'ISEP où elle dirige l'équipe de recherche en Informatique RDI au sein du laboratoire LISITE (Laboratoire Informatique, Signal, Télécommunication et Électronique). Raja Chiky est également responsable des formations en science de la donnée (Big Data, bases de données, fouille de données). Elle est titulaire d'un doctorat en Informatique de Télécom ParisTech obtenu après un Master de recherche en fouille de données et d'un diplôme d'Ingénieur en informatique spécialité Bases de Données. Avant de rejoindre l'ISEP, elle a enseigné les statistiques, les bases de données et la programmation à l'université Paris Dauphine, à l'Université de Paris 12 et à Télécom ParisTech. Elle intervient également en tant qu'experte Big Data dans les formations initiales et continues à l'ISEP, à l'École Centrale de Paris, à l'Institut Télécom et au CNAM.

Sur le plan de la recherche, elle a travaillé en étroite collaboration avec les équipes d'EDF R&D sur des problématiques liées à la gestion de flux de données, aux sondages et à la fouille de données. Mais elle a également participé à des projets de recherche en collaboration avec des universités, des centres de recherche et des entreprises. Elle a aussi pris part dans des conférences, des séminaires et des journées scientifiques aussi bien nationales qu'internationales.

Enfin, tout au long de son activité, Raja Chiky a publié plusieurs articles et revues dans des disciplines aussi variées que les statistiques, la fouille de données, les entrepôts de données, la gestion de flux de données, le Cloud Computing et les systèmes de recommandation.



LE PARADIGME DU BIG DATA ET SON APPLICATION À LA CYBERSÉCURITÉ

Par Gérard Peliks - Président de l'atelier sécurité de Forum ATENA

Si les données occupent un volume tel qu'elles ne peuvent plus être traitées avec les outils classiques, si ces données se présentent dans un format trop peu structuré pour être gérées dans une base relationnelle, comme par exemple les textes bruts, les sons, les vidéos, si les données bougent très vite, c'est là que les technologies du Big Data deviennent indispensables.

► LA DONNÉE

Définissons d'abord ce qu'on entend par « une donnée », au sens du Big Data. Une donnée est une information sans structure particulière, mais porteuse d'un sens. Par exemple « 10 000 euros » n'est pas une donnée, par contre « cette voiture coûte 10 000 euros » en est une. Autour de cette donnée gravitent beaucoup d'informations, non contenues dans les mots qui décrivent directement la donnée. Cette voiture consomme 5 litres aux 100 km, cette voiture appartient à telle personne, cette voiture a été achetée à telle date... sont des métadonnées associées à la donnée, très utiles dans son traitement et donc qu'il ne faut pas dissocier de la donnée qu'ils précisent.

Voyons maintenant ce que change le Big Data.

► LE QUATRIÈME PARADIGME : LA CORRELATION DES DONNÉES ET MÉTADONNÉES

Dans son livre « Le quatrième paradigme », Tony Hey décrit quatre innovations qui ont fait avancer la science. Il y eut d'abord les sciences de l'observation, l'empirisme, qui par l'observation des phénomènes nous permettent, en comprenant leurs causes, de prévoir leurs effets. Sont arrivées plus tard les sciences « dures », mathématiques, physique théorique, par lesquelles il est devenu possible de modéliser ces phénomènes. Le principe de causalité acquiert ainsi une base théorique. Troisième paradigme, tirant parti des deux précédents, la simulation sur ordinateur. A partir d'un échantillon de données en entrée soumis à des algorithmes, la simulation permet de vérifier une hypothèse ou de découvrir les conséquences d'un comportement.

Arrive aujourd'hui le quatrième paradigme : le Big Data qui apporte une nouveauté fondamentale. Pour simuler un phénomène, on ne part plus d'un échantillon en entrée, forcément petit car limité par la disponibilité des données, les tailles mémoire et les puissances de traitement. On part de l'ensemble des données disponibles, de tous formats. On les soumet à des algorithmes de traitement, sans limitation ni du nombre de données, ni de la taille mémoire ou disque, ni de la puissance de calcul.

Les algorithmes corréleront les données entre elles et en sortent des résultats parfois surprenants, souvent imprévus. Leurs conclusions sont basées sur les données et leurs métadonnées et sur les relations reliant les diverses données et métadonnées entre elles, provenant de mondes parfois très hétérogènes. Ces conclusions nous aident à comprendre ou à prévoir des phénomènes parfois inexpliqués jusqu'alors.

Au principe de causalité emmené par les trois premiers paradigmes se superpose le principe de corrélation emmené par ce quatrième paradigme : le Big Data

► DU PRINCIPE DE CAUSALITE AU PRINCIPE DE CORRELATION

Précisons ce que ce quatrième paradigme apporte. On ne s'attache plus au « pourquoi » telle cause entraîne tel effet (principe de causalité) ; on se concentre sur le « quoi », ce qu'apportent les données entrées, corrélées par les calculs qu'elles subissent, et les conclusions que visualisent les ordinateurs qui les traitent (principe de corrélation). Connaître le « quoi » permet de prendre des décisions rapides, connaître le « pourquoi » devient accessoire et peut être remis à plus tard. Par exemple, si la corrélation d'immenses quantités de logs montre que le système d'Information de tel secteur d'activité est plus attaqué que ceux des autres secteurs, il conviendra d'investir en priorité dans la sécurité du secteur le plus attaqué. Comprendre pourquoi ce secteur est plus exposé que les autres n'est pas ce qui est le plus important et la question peut avoir sa réponse dans un deuxième temps. Les données ont parlé et... « *In Data Veritas* ».

Le Big Data offre la possibilité de corréler des données et leurs métadonnées avec de très nombreuses autres données et métadonnées associées, à travers des algorithmes de traitements mathématiques et statistiques. Le but est de faire parler l'ensemble de ces données qui peuvent occuper un grand volume de stockage, données de tous formats, puis de visualiser les résultats de manière à faire émerger une tendance, une constatation, peut-être une prédiction, en tout cas une information utile à celui qui en a besoin pour prendre une bonne décision. Même si ces résultats ne sont pas ceux qu'on attendait, ce n'est pas un réel problème. Les résultats sont là et on les exploite, quitte à comprendre plus tard la cause qui a conduit à l'effet.

► LES DONNÉES D'ENTREPRISE, UNE MATIÈRE PREMIÈRE INESTIMABLE

Le Big Data s'est imposé récemment avec le Big Bang numérique lié à l'explosion des réseaux sociaux. Il est porté par la vague des smartphones et des tablettes, se poursuit avec l'Open Data et va passer à la vitesse supérieure grâce à l'Internet des Objets. En 2011, le volume de données générées s'est élevé à 1 800 milliards de giga octets. Dans huit ans il devrait être cinquante fois supérieur.

Ces données brutes et complexes représentent une matière première naturelle de grande valeur. Les organisations, qui en ont pris conscience, découvrent qu'elles n'exploitent pourtant aujourd'hui qu'une infime partie de leurs données, alors que celles-ci représentent une mine d'or qui va conditionner leur compétitivité, et peut-être leur longévité. Cet avantage offre aux organisations une connaissance plus fine de leurs clients, ou une prédiction de ce que pourraient être les tendances du marché... pour qui sait exploiter ces données bien entendu. Une donnée peut conserver une vie propre et être indépendante des traitements qu'on peut lui faire subir. Le croisement d'une donnée avec d'autres données en augmente la valeur.

Déjà, les recueils et l'exploitation des données ont des retombées dans bien des domaines et en particulier dans ceux de la médecine, de l'énergie, des transports (adaptation des horaires, fluidification du trafic routier...) et, c'est le sujet de cet article, **la cybersécurité**.

Les données seules n'apportent généralement rien d'exploitable. Les millions, ou plus, de lignes contenues dans les journaux log d'une organisation ne sont pas utilisables en l'état, bien qu'ils traduisent des informations indispensables comme des attaques, des non-conformités ou des vulnérabilités dans le système d'information. Ces données sont déjà plus intéressantes si on les lie avec les métadonnées qui les caractérisent et qu'on les corrèle avec d'autres données, et ainsi de suite, de manière incrémentale. Le résultat obtenu est parfois inattendu, voire incroyable, mais les données ont parlé, et les données ont généralement raison quand elles sont de qualité et quand elles sont traitées par des algorithmes adaptés.

Pour protéger l'Information et le système qui la gère, le Big Data occupe déjà une place qui pourra s'avérer incontournable dans un futur pas très lointain. Les technologies du Big Data permettent de traiter de manière exhaustive tous les événements de sécurité, même les signaux les plus faibles, même les signaux les plus lents, qui seraient sinon passés inaperçus. L'ensemble des signaux corrélés peut mettre en évidence les symptômes d'une cyberattaque furtive.

► LE BIG DATA POUR CONTRER LES APT

Le Big Data apporte une aide précieuse pour diminuer les risques que font peser sur une organisation les attaques venues du cyberspace. Parmi les attaques les plus subtiles et les moins repérables par des contre-mesures classiques, on peut citer les APT (Advanced Persistent Threats).

Une attaque APT commence par une prise de connaissance fine du système d'information d'une organisation, de ceux qui l'utilisent et de ceux qui le gèrent. L'attaquant s'enquiert des failles du système d'information ciblé, interroge les employés de la cible par téléphone, prend connaissance de ce qu'ils racontent dans les réseaux sociaux... Après cette phase, la cible a livré à l'attaquant l'endroit mal protégé, routeur, serveur ou employé par où l'attaque a le plus de chance de réussir.

Personne dans l'organisation ciblée par l'attaquant ne se méfie.

L'attaque commence, l'employé visé reçoit, par exemple dans sa messagerie, un fichier qui semble anodin mais qui contient une amorce de programme malfaisant, indétectable par les antimalwares de l'organisation ciblée.

Le temps passant, l'amorce charge, à l'insu de l'employé, à partir d'un serveur géré par l'attaquant, dit de Commande et Contrôle, l'intégralité du programme malfaisant, petits bouts par petits bouts jusqu'à reconstituer le programme malfaisant au complet, souvent un logiciel espion qui peut de plus se mettre à jour, autres signaux faibles. Il s'agit ici de signaux lents, en général indétectables par les contre-mesures mises en œuvre par la cible.

Le logiciel espion, établi dans le système de la cible, va monter en privilèges et chercher les informations sensibles. Il va les concentrer, par exemple dans une image qu'il va envoyer par morceaux vers son serveur. Sauf si la cible a mis en œuvre une contre-mesure qui sait détecter cette menace en identifiant un comportement anormal, ce qui n'est généralement pas le cas, personne ne s'aperçoit de l'attaque en cours.

La phase d'exfiltration qui suit consiste à envoyer automatiquement vers l'attaquant, les informations ainsi collectées. L'image est divisée en morceaux qui sont chiffrés et envoyés petit à petit vers l'attaquant. Là il s'agit de signaux faibles, en général rien ne permet de les détecter parmi les autres informations qui sortent légalement de l'organisation.

Le pillage des données peut durer plus d'une année avant d'être détecté, et souvent le logiciel malveillant s'autodétruit avant même d'être détecté. L'entreprise a été pillée de son information sensible mais ne s'en est pas aperçue, ou a découvert l'attaque trop tard.

Pourtant, toute action sur un système d'information laisse une trace, qui se retrouve dans des journaux, les fameux fichiers log. Des technologies regroupées dans les méthodes du SIEM (Security Informations and Events Management) savent analyser les traces laissées par les attaques, les vulnérabilités et les non conformités. Le problème est que les traces laissées par les signaux lents, les signaux faibles et les informations comportementales des utilisateurs ne sont souvent pas exploitables. Un autre problème est que les logs sont en trop grande quantité et en trop gros volume pour être traités efficacement.

► SIEM, SOC ET BIG DATA

Les APT, attaques qui sévissent de plus en plus fréquemment et par lesquelles des logiciels espions entrent dans les systèmes d'information des organisations pour piller leurs secrets, ne peuvent le plus souvent, être décelées que par l'analyse de signaux faibles ou lents, sur une période de plusieurs mois. Ces signaux peuvent ne pas sembler porter une valeur quelconque au moment où ils sont remontés, mais, corrélés à d'autres signaux lents ou faibles, ils peuvent se concrétiser en signaux forts et révéler une attaque qui pourrait sans cela ne jamais être décelée. Il convient donc de conserver toutes les données et métadonnées recueillies et de les traiter dans leur ensemble. Bien sûr cela peut représenter un volume de stockage colossal, sur une période de temps suffisante, et requérir une vitesse de traitement et une puissance de calcul importantes, que seul le calcul intensif en parallèle sur de nombreux calculateurs peut donner. Les données peuvent d'autre part se présenter sous de multiples formats, fichiers log, bien sûr, mais aussi conversations téléphoniques, vidéos provenant de caméras de surveillance, etc. C'est là que les technologies du Big Data interviennent en corrélant des événements et des comportements.

La technologie qui permet de remonter et traiter les alertes s'appelle le SIEM, Security Information and Event Management. Dans l'idéal, cela résulte d'une démarche « descendante », demandée par la direction et orientée métier. Mais se protéger de certaines menaces, en particulier celles des APT, en se basant seulement sur des variations anormales de flux sortant ou sur des adresses de destination jugées « anormales » ne suffit pas en général. C'est là que l'approche « ascendante », dans laquelle toute corrélation significative, remontant du terrain est prise en considération et peut être traitée en temps réel ou en différé, rejoint l'approche « descendante » pour une plus grande efficacité de traitement et une plus grande sécurité des utilisateurs

Les corrélations de ces événements remontées vers un SOC (Security Operations Center) – incidents, non-conformités à la politique de sécurité de l'organisation, vulnérabilités non couvertes – provenant de sources multiples, atteignent, le plus souvent, un volume trop important pour être traité en temps réel. Pourtant les traitements de ces événements stockés doivent parfois se faire très rapidement en allant piocher dans d'énormes quantités d'autres événements. Les technologies du Big Data, comme par exemple celles du Framework Hadoop et du calcul parallèle MapReduce, prennent tout leur sens.

Sans verser dans l'utopie, avec le Big Data, nous sommes devant une technologie de rupture. Dans la sécurité aussi, nous allons passer de l'ère de l'informaticien à celle du statisticien et du designer. Il convient d'anticiper cette révolution pour prendre de l'avance sur une technologie qui s'avère incontournable.

À PROPOS DE L'AUTEUR

Gérard Peliks préside l'atelier sécurité de l'association Forum ATENA, qui organise des colloques sur diverses facettes de la sécurité de l'Information. Il est membre du conseil d'administration de l'Association des Réservistes du Chiffre et de la Sécurité de l'Information (ARCSI) et lieutenant-colonel dans la réserve citoyenne de cyberdéfense (RCC) dans la gendarmerie nationale. Gérard Peliks est chargé de cours sur la cybersécurité, dans le cadre de Mastères, dans plusieurs écoles d'ingénieurs. Dans le cadre de l'ISEP, il a été dans le comité de pilotage du mastère spécialisé « Intelligence des risques » et intervient dans des conférences d'ISEP Alumni sur des thèmes comme le Cloud, le Big Data et la Cyberdéfense.

ISSY-LES-MOULINEAUX, UN AN APRÈS

Par Emmanuelle Vatel - Chargée de communication à l'ISEP



Le bâtiment de l'ISEP à Issy-les-Moulineaux - vue de côté

En septembre 2014, l'ISEP a fait le choix d'installer son second campus à Issy-les-Moulineaux. Ce bâtiment, appelé Lorette, offre un cadre de travail exceptionnel et va permettre à l'ISEP d'augmenter significativement ses promotions, répondant ainsi aux besoins des entreprises.

Ce campus supplémentaire (qui s'ajoute au bâtiment de la rue Notre-Dame-des-Champs dans le 6^{ème} à Paris) s'inscrit dans la politique de développement menée par l'école depuis plusieurs années. Les objectifs sont multiples : il s'agit d'augmenter la capacité d'accueil de l'école mais aussi de garantir aux étudiants des choix de personnalisation de parcours encore plus nombreux, grâce notamment au développement de nouveaux partenariats à l'international (plus de 110 universités partenaires à ce jour), des cursus toujours plus variés et une large palette de formations en phase avec l'actualité (Informatique et Libertés, Cloud Computing, Big Data, Business Intelligence, Objets Connectés, etc.).

Résolument tourné vers l'innovation, le bâtiment abrite des salles informatiques modernes, les laboratoires de recherche de l'ISEP, un grand amphithéâtre, des salles de cours et de conférences ainsi que des espaces dédiés à l'apprentissage par projet.

Ce grand bâtiment néoclassique impressionne tant par sa beauté classique extérieure (il date de 1930) que par sa modernité intérieure. Cette installation a d'ailleurs été célébrée lors de l'inauguration du bâtiment en mai dernier, en présence d'André Santini, ancien ministre et député-maire de la ville, ravi d'accueillir la première école d'ingénieurs dans l'histoire d'Issy, puisqu'il s'efforce depuis des années d'en faire une ville tournée vers le numérique, reconnue comme la plus connectée de France.



Le bâtiment de l'ISEP à Issy-les-Moulineaux - vue de face



LA RECHERCHE À L'ISEP

Par Raja Chiky - Enseignante-chercheuse à l'ISEP

La recherche à l'ISEP est organisée dans le cadre d'un laboratoire unique **LISITE** (Laboratoire d'Informatique, Signal et Image, Électronique et Télécommunication) placé sous l'autorité du Directeur de la Recherche et qui est constitué de trois équipes spécialisées mais pouvant interagir dans le cadre de projets multidisciplinaires.

Le LISITE, dispose de trois équipes de recherche : Micro-Nanoélectronique et Radio Communication (**MINARC**), Signal Image et Télécommunications (**SITe**), Recherche et Développement en Informatique (**RDI**).

Pour répondre au souci de fédérer les activités de recherche au sein des équipes, les axes de recherche ont été définis pour répondre à la demande du milieu socio-économique, tout en fédérant, autant que possible, les compétences des enseignants chercheurs. Deux axes principaux sont définis par équipe :

► ÉQUIPE MINARC

Le CMOS Ultime et au-delà

Mots-clés : Silicon On Insulator (SOI), Fully Depleted SOI (FDSOI), Double Gate FDSOI (DG-FDSOI), Static Read Access memory (SRAM), Non Volatile Memory (NVM), Tunneling Field effect Transistor (TFET), Low Power, Low voltage, Process Variability (PV)

Les circuits intégrés RF pour la communication et le biomedical

Mots-clés : Medium Access (MAC), Body Area network (BAN), Medical Implant Communication service (MICS), Radio Frequency Identification (RFID), Ultra Wide Band (UWB), Biomédical, implant

► ÉQUIPE SITe

Gestion des ressources radios dans les systèmes sans fil multi-utilisateurs

Mots-clés : systèmes OFDMA, accès MIMO, algorithmes d'ordonnancement, qualité de service, gestion des interférences, green ICT, répartition des ressources radio

Méthodes avancées d'interprétation et de codage des images

Mots-clés : vision artificielle, multimédia, fusion d'informations, codage des images et des flux vidéo, transformées adaptatives.

► ÉQUIPE RDI

Fouille de données et optimisation dans les systèmes large échelle

Mots-clés : Big Data, Systèmes de recommandation, Web Sémantique, Cloud-Computing, Machine-to-Machine (M2M), Algorithmes d'optimisation, Répartition de charge, Optimisation des coûts.



LA FORMATION CONTINUE À L'ISEP

Par Denis Beautier
 Directeur des formations continues à l'ISEP
 Enseignant à l'ISEP

Saviez-vous que la Formation-Continue@ISEP.fr ?

« **Saviez-vous que la Formation Continue à l'ISEP pour vous ?** »

Quand Michèle Germain m'a demandé d'écrire un article sur la Formation Continue de l'ISEP à destination des anciens élèves, sur ce support numérique, je suis parti de deux idées :

- ▶ D'une part, **les anciens élèves ne savent pas toujours tout ce qu'ils peuvent faire avec la Formation Continue de l'ISEP**, et donc, il peut être très intéressant pour eux que je le leur présente, d'où mes questions : **Savez-vous ? précédées d'une ▶**
- ⇒ D'autre part, comme il s'agit d'un support numérique, je vais **leur donner tous les liens** leur permettant d'accéder au plus vite à toutes les informations qui leurs seront utiles, d'où **mes invitations à cliquer sur les liens précédées d'une ⇒, pour leur permettre de le savoir !**

J'espère que vous trouverez ici une mine d'idées **d'actions** à partir de questions simples que vous n'imaginiez même pas avant de les avoir lues jusqu'au bout !

Agissez, cliquez, et je vous informerai. À bientôt !

▶ POURQUOI ? POUR QUI ?

Vous êtes ancien élève de l'ISEP. Cela signifie que vous avez passé quelques années sur les bancs de votre vénérable **école**. Mais cela signifie aussi que vous avez vécu quelques années de **vos parcours professionnels**.

Vous vous êtes enrichi de ce que vous avez appris lors de **vos formation initiale**, mais aussi lors de **vos expériences professionnelles**.

Mais vous savez que cela n'est pas suffisant et qu'il est parfois nécessaire de **booster** vos connaissances et vos compétences pour être plus à même de **répondre aux besoins** de vos clients. Sinon, ce sont vos concurrents qui y répondent. **Et vous perdrez le marché.**

Cela peut même avoir un impact majeur sur votre parcours, pour éviter de devenir « *has been* » ! Sans aller jusque-là, vous êtes conscient du soin que **vous devez porter à votre capacité à répondre aux besoins du marché**. Vous êtes aussi conscient de votre capacité à **vous positionner** sur tel ou tel marché pour prendre plaisir à vivre votre parcours professionnel. Vous savez aussi qu'il est important **d'être le plus heureux possible** dans votre travail, car il s'agit avant tout de **vos vie !**

Il peut s'agir d'un besoin ponctuel. Mais il peut aussi s'agir de votre employabilité dans la direction que vous souhaitez donner à votre parcours. Elle doit toujours rester optimum. Et le marché n'est jamais stable. Vous savez qu'il est faut s'efforcer de voir loin. Mais ce désir est souvent difficile à mettre en œuvre « *un peu comme un marin au long cours, entouré par les brumes des incertitudes* ».

Vous pouvez opter d'égrainer un chapelet de **formations courtes** qui permettront d'ajuster au coup par coup votre positionnement. Ou alors, vous pouvez choisir un accompagnement sur la durée pour infléchir sensiblement cette courbe à l'aide d'une **formation longue**, souvent diplômante, au moins certifiante.

Vous l'avez bien lu : ceci s'applique certes à **vous-même**.

Mais savez-vous aussi que cela s'applique à **vos collaborateurs**, ceux-là même que vous managez ?

Et encore plus loin : savez-vous que cela s'applique à l'ensemble de **vos collègues** de la structure pour laquelle vous travaillez ?

Et voyons grand : savez-vous que cela s'applique aussi à **vos clients** ?

► SAVEZ-VOUS ?

- **Saviez-vous que l'ISEP**, votre école, peut vous proposer des solutions pour répondre à vos besoins personnels, ceux de vos collaborateurs, de vos collègues, de vos clients ?
- **Saviez-vous que l'ISEP**, votre école, peut mettre en place des formations courtes sur mesure, de quelques jours, pour répondre à un besoin précis et ponctuel ?
- **Saviez-vous que l'ISEP**, votre école, propose des formations longues diplômantes, d'un niveau post-master (bac+6), comme des Mastères Spécialisés®, voire des MBA, comme elle l'a proposé il y a quelques années ?
- **Savez-vous que l'ISEP**, votre école, peut mettre en place ces solutions, dans vos locaux, partout en France, voire plus loin si besoin ? Peut-être la nostalgie vous invitera à revenir sur les bancs de votre école à Paris, ou à découvrir, à cette occasion, son nouveau Campus d'Issy-les-Moulineaux ?
- **Savez-vous que l'ISEP**, votre école, peut chercher à décliner ses solutions en langue anglaise, voire d'autres langues, si besoin ?
- **Savez-vous que l'ISEP**, votre école, peut proposer des formations dans tous les domaines techniques des technologies de l'information : informatique, électronique, télécommunications, réseaux, systèmes embarqués, Big Data, cloud-computing ?

Par exemple, l'ISEP peut former vos programmeurs et chefs de projet informatique aux techniques de codage sécurisé en C, C++ pour l'obtention et le maintien de vos certifications sécurité pour répondre aux exigences du marché dans ce domaine.

- **Savez-vous que l'ISEP**, votre école, peut orienter son contenu dans le secteur dans lequel vous travaillez : sécurité, militaire, aéronautique, santé, médical, automobile, télécom, public, privé, banque, assurances, etc. ?

Par exemple, l'ISEP peut décliner ses formations en « data privacy » au secteur des assurances, voire de la recherche clinique, ce qui va constituer deux formations distinctes, car deux spécificités sectorielles bien différentes.

- **Savez-vous que l'ISEP**, votre école, propose des formations dans tous les domaines transversaux, comme : le management des équipes, la gestion de projet, le marketing, le droit, l'intelligence économique, la gestion des risques, l'innovation, etc. ?

Par exemple, l'ISEP peut former vos collaborateurs à la gestion de projet par les méthodes AGIL.

- **Savez-vous que l'ISEP**, votre école, adapte sa pédagogie pour qu'elle soit le plus proche de vos besoins et fonction des profils et contraintes des stagiaires ?

Par exemple, l'ISEP peut monter une solution d'une journée de formation tous les 15 jours, en 2 ou 3 fois, avec suivi et accompagnement des stagiaires sur leurs projets, ce qui favorise la mise en œuvre du savoir transmis.

Si vous identifiez au moins trois personnes, idéalement cinq, et encore mieux plus, qui auraient un tel besoin, alors, monter une formation intra-entreprise est la meilleure solution. Sinon, il vous faut trouver une solution en inter-entreprise.

► NOS CATALOGUES

Aujourd'hui, vous avez compris que **le catalogue « intra-entreprise »** de l'ISEP touche tous les domaines couverts par l'ISEP dans les technologies du numérique, des sciences transversales et leurs secteurs d'application. Ce très large catalogue ne peut donc pas exister de manière exhaustive. Il vous suffit d'exprimer votre demande pour recevoir une réponse.

► **Identifiez-vous des besoins de formation pour vous, pour vos collaborateurs, pour vos collègues ou pour vos clients, dans un certain domaine, etc. ?**

⇒ **Si oui, pourriez-vous me signaler ce besoin en cliquant ici ?**

Par contre, **le catalogue « inter-entreprise »** de l'ISEP couvre pour l'instant deux expertises :

- L'une dans [le Cloud Computing](#)
- L'autre dans [la Sécurité des Données Personnelles](#), « data privacy » ou « informatique & libertés »

► **Identifiez-vous votre CIL ?**

- *Qu'est-ce qu'un CIL ? C'est un Correspondant Informatique et Liberté !*
- *Il peut se trouver à la Direction Juridique, à la Direction de la Conformité, à la DSI...*

⇒ **Si oui, envoyez-lui un mail** en me mettant en copie et en l'informant de ce lien

<http://www.informatique-et-libertes-formation.fr>

Voici le catalogue des formations courtes qui durent de 1 à 3 jours :

- Dans le [Cloud Computing](#)
 - [le Tour du Cloud – vision à 360°, pour savoir de quoi il s'agit !](#)
 - [comment appliquer une sécurité technique dans le Cloud ?](#)
 - [comment appliquer une sécurité juridique dans le Cloud ?](#)
 - [DataCenters : Que faut-il vérifier avant de confier votre informatique aux nuages ?](#)
 - [comment relever les défis du Big Data ?](#)
 - [comment opérer un projet de migration vers le Cloud ?](#)
 - [comment acheter du numérique dans le Cloud ?](#)
- Dans la [sécurité des données personnelles](#)
 - [kit de survie métier Informatique & Libertés](#)
 - [kit de survie technique Informatique & Libertés pour juristes](#)
 - [kit de survie comment gérer un contentieux avec la CNIL ?](#)
 - [apprendre à piloter un projet d'Anonymisation](#)
 - [kit RH Informatique & Libertés](#)
 - [kit Cyber-Sécurité, Cyber-Surveillance, Géolocalisation et Vidéo-protection Informatique & Libertés](#)
 - [kit Données de Santé Informatique & Libertés](#)

Les formations longues s'étalent sur neuf mois, à temps partiel, à raison d'une journée maximum par semaine, voire une demi-journée par semaine, pour ceux qui placent leurs RTT durant les jours de formation, le **Compte Personnel de Formation (CPF)** pouvant couvrir la demi-journée restante. Le rythme d'alternance permet à des salariés en CDI à temps plein de continuer leur activité professionnelle tout en se formant. Elles sont diplômantes, de niveau bac+6. Ce sont **des Mastères Spécialisés® (MS)** ou des **Bilan d'Aptitude Délivré par les Grandes Écoles® (BADGE)** reconnus par la **Conférence des Grandes Écoles (CGE)**. L'un d'eux est même devenu un diplôme d'état inscrit au **Répertoire National de la Certification Professionnelle (RNCP)** depuis cette année.

Le Financement de ces formations est possible par les OPCA de vos entreprises, car elles sont inscrites à **l'Inventaire**. De plus, le MS inscrit au RNCP est aussi éligible dans le cadre du **CPF**. Vous trouverez ci-après les numéros et les liens pour vous aider à les faire financer dans le cadre de **la Réforme de la Formations Professionnelle...**

Voici le catalogue des formations longues :

- Dans le Cloud Computing :
 - MS **Expert Cloud Computing** (diplôme d'état **CPF n°145653**, inventaire **n°861** et **n°889**)
 - BADGE **Chef de Projet Cloud** (inventaire **n°889**)
 - BADGE **Concepteur de Solution Cloud Computing** (inventaire **n°861**)
- Dans la sécurité des données personnelles :
 - MS en **Management et Protection des Données Personnel** (inventaire **n°966** et **n°972**)
 - BADGE **Praticien légal à la protection des données personnelles** (inventaire **n°966**)
 - BADGE **Praticien technique à la protection des données personnelles** (inventaire **n°972**)

► CONTRATS DE PROFESSIONNALISATION ET THÈSES PROFESSIONNELLES

Savez-vous que vous pouvez aussi enrichir vos équipes en faisant appel à des élèves qui suivent ces formations, sous la forme de stage, de **contrat de professionnalisation** (*solution que je recommande, car financé par les OPCA !*), de **CDD**, de **prestation de service**, voire de **CDI** – pour vos collaborateurs ?

► Pensez-vous avoir besoin de telles ressources ?

- Pour vous aider aux développements de vos solutions **Cloud**
- Pour vous aider dans vos problématiques **Sécurité des Données Personnelles**

⇒ **Si oui**, cliquez sur Formation-Continue@ISEP.fr

Savez-vous que vous pouvez soumettre *des sujets de thèses professionnelles* pour avancer gratuitement sur les verrous que vous souhaitez faire sauter ? Il s'agit de définir un verrou professionnel technique ou de marché. Notre élève, qui se sent investi par ce problème, effectue un état de l'art, propose et défend des pistes de solutions, puis expose des perspectives pour continuer ces travaux une fois qu'il les a soutenus.

► Pensez-vous avoir de tels sujets qui vous préoccupent à travailler ?

⇒ **Si oui**, cliquez sur Formation-Continue@ISEP.fr

Savez-vous que vous pouvez assister librement aux soutenances des thèses professionnelles chaque année ? C'est aussi un excellent moyen de vous tenir informé sur certains sujets qui vous préoccupent.

► Seriez-vous intéressé pour recevoir la liste des thèses qui vont être soutenues ?

⇒ **Si oui**, cliquez sur Formation-Continue@ISEP.fr

Savez-vous que vous pouvez demander nos meilleures thèses professionnelles ?

⇒ **Si oui**, cliquez sur :

- Les thèses **Cloud**
- Les thèses **Sécurité des Données Personnelles**

Savez-vous que vous pouvez aussi être membre d'un jury de soutenance des projets transversaux ou de thèse professionnelle (Cloud, Informatique & Libertés) ? Il s'agit de jouer le rôle d'un membre d'un CODIR ou d'un COMEX. *C'est à la fois pertinent, exigeant et ludique.*

► Seriez-vous intéressé par ce jeux de rôle / « serious game » ?

⇒ **Si oui**, cliquez sur Formation-Continue@ISEP.fr

Savez-vous enfin que la Formation Continue de l'ISEP organise régulièrement *des rencontres professionnelles, AfterWorks, Petits déjeuners, Salons de Professionnels ?*

► Souhaitez-vous être informé des news de la Formation Continue de l'ISEP ?

⇒ **Si oui**, cliquez sur Formation-Continue@ISEP.fr

► CONTACT

Le point de contact pour la Formation Continue de l'ISEP est Denis Beautier, Denis.Beautier@ISEP.fr, 01.49.54.52.20, Directeur des Formations Continues de l'ISEP et Professeur d'Informatique à l'ISEP.

À PROPOS DE L'AUTEUR

Denis Beautier est titulaire d'un DEA Informatique, Réseaux et Systèmes Embarqués en 1989. Il a commencé son expérience dans la formation comme Volontaire Formateur en Informatique dans le cadre de son Service National, puis comme ingénieur dans l'armement chez Thalès durant 5 ans, avant de retourner à ses amours premiers, pour former les collègues qu'il aurait aimé rencontrer, avec un souci de l'excellence technique, managériale, mais aussi humaine. Il a été consultant formateur durant 2 ans, puis professeur d'informatique en école d'ingénieur durant 5 ans, avant de rejoindre l'ISEP en 2001 pour enseigner, prendre la responsabilité des ISTEP et faire évoluer la Formation Continue où elle en est aujourd'hui, notamment la création des Mastères Spécialisés®. Il est aujourd'hui Professeur d'Informatique à l'ISEP où il anime les électifs **Puissance3**, « **Informatique & Libertés** » (dont il est administrateur de l'Association Française des Correspondants à la Protection des Données - **AFCDP**), il est Tuteur d'apprentis du **CFA**, et Directeur des **Formations Continues** de l'ISEP.