06/03/2014

Ecole d’ingénierie Informatique : B3 INIT

**M DEBAS THOMAS**

**M DESZCZ SEBASTIEN**

**M JAGOURY MATTHIEU**

**M MAURICE REMY**

**M MEURILLON GREGOIRE**

**MELLE STALTER MARIANNE**

CLOUD & BIG DATA

L’AUTOMATISATION des DATA CENTERS

SOMMAIRE

[Le Cloud Computing 2](#_Toc381884562)

[I – Le Cloud : généralités 2](#_Toc381884563)

[II – Le Cloud, plus précisément 2](#_Toc381884564)

[II.1 – Types et branches du Cloud 2](#_Toc381884565)

[II.2 – Fonctionnalités 2](#_Toc381884566)

[II.3 – Avantages 3](#_Toc381884567)

[II.4 – Risques 3](#_Toc381884568)

[III – Etude de marché 5](#_Toc381884569)

[III.1 – Les cibles 5](#_Toc381884570)

[III.2 – Les concurrents : Data Center Automation 5](#_Toc381884571)

[III.2.1 - Qu’est-ce qu’un Data Center ? 5](#_Toc381884572)

[III.3 – Les offres proposées 6](#_Toc381884573)

[IV – Les outils d’orchestration 7](#_Toc381884574)

[IV.1 - L’outil d’orchestration, c’est quoi ? 7](#_Toc381884575)

[IV.2 – Des exemples d’outils 7](#_Toc381884576)

[Le BIG DATA 9](#_Toc381884577)

[V – Le BIG DATA, généralités 9](#_Toc381884578)

[VI – Le BIG DATA, plus précisément 9](#_Toc381884579)

[VI.1 – A quoi sert-il ? 9](#_Toc381884580)

[VI.2 – Comment ça marche ? 9](#_Toc381884581)

[IX – Avantages 10](#_Toc381884582)

[X – Inconvénients 10](#_Toc381884583)

[XI – Limite du BIG DATA 10](#_Toc381884584)

[XI – Etude de marché 10](#_Toc381884585)

[XI.1 – Les cibles 10](#_Toc381884586)

[XI.2 – Les fournisseurs 12](#_Toc381884587)

[XI.3 – Supports 12](#_Toc381884588)

[XI.4 – Les offres proposées, la « BIG Solution » 12](#_Toc381884589)

[XI.5 – Exploitation des données 13](#_Toc381884590)

[XII – Les outils d'orchestration 13](#_Toc381884591)

[XIII – Intérêts perçus du BIG DATA 14](#_Toc381884592)

# Le Cloud Computing

## I – Le Cloud : généralités

Le Cloud Computing, littéralement informatique dans les nuages, désigne l'utilisation de serveurs distants (en général accessibles par Internet) pour traiter ou stocker l'information. L'accès se fait le plus souvent à l'aide d'un navigateur Web. Enregistrer des fichiers via Internet sur un serveur en est un exemple. Le logiciel lui-même peut être déporté sur l'ordinateur distant.

Le Cloud se base sur la virtualisation : Celle-ci permet aux entreprises d’utiliser des serveurs virtuels, disséminés quelque part dans le Cloud, plutôt que des serveurs physiques. L’avantage premier réside dans le fait que l’entreprise utilise les serveurs selon ses besoins, et fait une économie non négligeable dans le coût de ses serveurs.

Le fait de pouvoir mettre plusieurs systèmes d’exploitation sur un seul et même serveur est de même un avantage, car l’entreprise n’est plus obligée d’entretenir elle-même un serveur dédié, on s’en occupe à sa place.

## II – Le Cloud, plus précisément

### II.1 – Types et branches du Cloud

Le Cloud Computing se décompose en 3 parties :

* IaaS (Infrastructure as Service) : Permet d’externaliser les serveurs, principalement pour la gestion de sauvegarde : l’entreprise garde une copie de ses données, par exemple, dans le Cloud. Concrètement, seule l’infrastructure matérielle est dématérialisée.
* PaaS (Platform as Service) : L’infrastructure est, d’une part, dématérialisée. D’autre part, certaines applications le sont aussi : bases de données, environnement de développement…
* SaaS (Software as Service) : Couche finale du Cloud. Permet à chaque employé d’une entreprise d’accéder à toutes les applications de son entreprise, hébergée dans un environnement sécurisé. Il n’a qu’à entrer son identifiant et mot de passe pour y accéder. Le temps passé sur l’application sera alors facturée à l’entreprise.

Il existe trois types de Cloud : public, privé, et hybride.

* Cloud public : Géré par un fournisseur. Plusieurs personnes distinctes peuvent utiliser les ressources informatiques mises à disposition par le fournisseur.
* Cloud privé : Exploité par une entreprise – la sécurité des données y est assurée. Si le Cloud est hébergé par un prestataire, il ne sera accessible que via des réseaux sécurisés (VPN).
* Cloud hybride : Destiné aux entreprises qui, en cas de forte utilisation du cloud privé, peuvent utiliser le Cloud public temporairement.

### II.2 – Fonctionnalités

* **Un accès en libre-service à la demande** aux capacités de calcul. Ce service sera le plus souvent effectué par le fournisseur de service de façon automatique sans nécessiter d’interaction humaine.
* **Un accès ubiquitaire au réseau** : les capacités sont disponibles sur le réseau et accessibles par des mécanismes standards, qui favorisent l’accès au service par des clients lourds, légers, des smartphones… via des plates-formes hétérogènes.
* **Une mise en commun des ressources** : les ressources de calcul sont mises à disposition des clients sur un modèle multi-locataires, avec une attribution dynamique des  ressources physiques et virtuelles en fonction de la demande. Le client n’a généralement aucun contrôle ou connaissance sur l’emplacement exact des ressources fournies. Toutefois, le client peut imposer de spécifier l’emplacement à un niveau plus haut d’abstraction (par exemple le pays, l’état ou le Data Center).
* **Une « Elasticité rapide »** : les capacités proposées peuvent rapidement augmenter ou diminuer en fonction des besoins.
* **Un service mesuré en permanence** : les systèmes contrôlent et optimisent automatiquement l’utilisation des ressources par rapport à une moyenne estimée de  consommation du service. L’utilisation des ressources peut être gérée, contrôlée et communiquée, fournissant ainsi de la transparence au client et au fournisseur.

### II.3 – Avantages

* Diminution des frais d’acquisition et d’entretien du parc informatique ;
* Faire de la place dans les volumes de stockage internes de l’entreprise ;
* Avoir régulièrement des mises à jour des logiciels et applications que le prestataire ou le fournisseur mettent à disposition ;
* Bénéficier d’un ajustement de la bande passante en fonction des besoins ;
* Payer uniquement les services consommés.
* **Diminution du coût du service informatique** :
  + Paiement sur les connexions,
  + Paiement sur l’espace de stockage utilisé
  + Disparition des frais de maintenance et d’investissement sur un serveur interne.
* Ainsi, les outils fournit par le prestataire seront en constante évolution pour offrir un service de plus en plus performant.

Ensuite, le Cloud permet à ses clients une meilleure gestion de leurs trafics réseaux, puisqu’elle offre une réponse instantanée lors d’une demande vers un serveur distant. Cette flexibilité propose donc aux clients une meilleure gestion de leurs bandes passantes et de l’utilisation des données, en fonction de leurs besoins.

Par ailleurs, le cloud aide à une meilleure compétitivité entre les entreprises. Les petites structures peuvent rivaliser avec les grandes multinationales puisqu’elles ont accès aux mêmes technologies, tant pour le traitement des données que pour leurs sauvegardes.

De plus, toujours pour cette disponibilité, le Cloud procure une sécurité dans le domaine des sauvegardes de données. Les données étant décentralisés vers des entreprises spécialisées, elles seront plus surement sauvegarder en cas de dysfonctionnement du système d’information où de la perte par un employé. Mais, elles seront aussi plus rapidement récupérées, comme le démontre l’étude de « Aberdeen Group ».

Enfin, cette solution va permettre une meilleure mobilité des données. Effectivement, n’importe quelle utilisateurs disposant d’un outil informatique et d’une connexion internet pourra accéder à ses données de travailles. De surcroit, elle solutionne le problème de gestion des données par des utilisateurs à travers le monde. Ainsi, chaque donnée traitée sera accessible par tous les employés à n’importe quelle heure.

### II.4 – Risques

Le risque majeur du Cloud est lié à la protection des données personnelles. En effet, aux Etats Unis, le FBI a le droit de consulter tous les serveurs du Cloud sans en informer l’utilisateur principal. L’USA Patriot Act est la loi qui permet cette utilisation des données.

Dans l’Union Européenne, cependant, une loi garantie la protection des données. Il est important de noter que si l’on fait appel à une entreprise européenne, le risque de vol de données confidentielles est moindre.

Cependant, il existe de nombreux autres risques liés au Cloud :

* ***Cadre légal***

Aucune connaissance de l’endroit où sont stockés les données, et aucun accès physique à celle-ci. En cas de perte de ces données par nous ou par le fournisseur, il nous est donc impossible de les retrouver sauf en contactant celui-ci.

* ***Connexion internet***

Le Cloud est un service puissant, mais celui-ci nécessite une connexion internet haut débit pour des transferts plus rapides. Dans le cas d’une basse connexion internet, les transferts effectués pourraient durer plusieurs heures. Cela engendre le fait qu’il faudrait gérer une planification des données à transférer et donc perdre du temps à ce genre de tâche.

* ***Problèmes techniques***

Disponible uniquement via internet, si une panne de courant devait arriver, l’entreprise serait privée de ses applications et de ses données sur le Cloud.

Il y a également les interruptions de services inévitables tels que les catastrophes naturelles, les erreurs humaines ou d’autres causes intermédiaires. Le temps d’arrêt est agaçant, mais si celui-ci s’éternise, cela peut avoir des conséquences désastreuses pour l’entreprise. Il faut prévoir un service en continue pour éviter ce genre de catastrophe.

* ***Coût***

Les entreprises ne font souvent pas attention aux frais de transferts, qui peuvent s’avérer être importants, selon l’utilisation que l’entreprise fait du Cloud. Ces frais de transferts sont donc à prendre en considération et à surveiller.

* ***Optimisation***

Malgré une connexion internet rapide, certaines applications web peuvent s’avérer être très lentes et plus limitées que des applications fonctionnant sur les propres ordinateurs de l’entreprise.

* ***Sécurité***

La sécurité du stockage : si les données sont conservées dans un seul disque, si celui-ci devait tomber en panne, les données seraient perdues. Si elles sont entre plusieurs unités de stockage, sont-elles réparties à plusieurs endroits ? La sauvegarde des données est un critère important à prendre en compte pour éviter aucune perte et prendre le moins de place possible.

La sécurité et la confidentialité des données : si le fournisseur assure des tests portant sur sa sécurité et si ses tests sont réguliers. Il faut donc surveiller les tests effectués par le fournisseur pour savoir si sa sécurité est fiable et optimale.

La sécurité des locaux : accessibilité pour tous ? Si n’importe qui accède à des données qui devraient justement être limité à certaines personnes de l’entreprise, quelles conséquences cela pourrait-il avoir ?

* ***Piratage***

Certaines applications sont très soumises aux attaques. Le piratage d’un compte d’entreprise pourrait avoir des conséquences dangereuses pour la réputation de l’entreprise tandis que l’utilisation imprudente des applications par un salarié pourrait offrir aux cybercriminels l’opportunité d’entrer dans le réseau et de soustraire des informations confidentielles.

* ***Pérennité***

Il est nécessaire de se demander si l’hébergeur cloud va durer dans le temps. Cet élément est important à prendre en compte car un changement d’hébergeur peut prendre du temps, et peut nécessiter un recodage des applications.

* ***Productivité***

Il est nécessaire que les employés de l’entreprise sachent se servir du cloud. En effet, même si l’entreprise fait des économies sur le stockage et le traitement d’informations, si les employés passent plus de temps pour leurs taches à cause du cloud, l’entreprise risque d’y perdre plus que d’y gagner.

## III – Etude de marché

### III.1 – Les cibles

Le Cloud est destiné aussi bien aux entreprises, qu’aux artisans ou qu’aux particuliers.

En somme, tous les acteurs de l’économie sont ciblés, dans la mesure où le service proposé est entièrement adapté au client.

La sauvegarde de données propose aussi bien aux entreprises de sécuriser leurs données, et aux particuliers de stocker leurs photos. Le service est d’autant plus intéressant que les données peuvent être accédées de n’importe quel pc ou support.

### III.2 – Les concurrents : Data Center Automation

### **III.2.1 - Qu’est-ce qu’un Data Center ?**

Un Data Center (centre de traitement des données) est un lieu où se trouvent différents équipements électroniques ([ordinateurs](http://fr.wikipedia.org/wiki/Ordinateur), systèmes de [stockage](http://fr.wikipedia.org/wiki/Stockage_d%27information), équipements de [télécommunications](http://fr.wikipedia.org/wiki/T%C3%A9l%C3%A9communications)…). Cela sert à traiter les informations nécessaires aux activités d'une entreprise. Par exemple, une [banque](http://fr.wikipedia.org/wiki/Banque) peut recourir à ce service, lui confiant les informations relatives à ses clients, tout en traitant les transactions de ceux-ci. En pratique, presque toutes les [entreprises](http://fr.wikipedia.org/wiki/Entreprise) de taille moyenne l’utilisent.

#### III.2.2 - Différence Data Center et Cloud Computing ?

Ces deux services, permettent de stocker des données. Ce qui différencie donc Data center et Cloud Computing est que le Cloud stocke des données sur internet, tandis que le Data Center stock les données sur du matériel présent physiquement dans le local de l’entreprise.

Les services du cloud sont sous-traités par des prestataires de services qui effectuent toutes les mises à jour et maintenances nécessaires. Les data center sont dirigés par le service informatique de l’entreprise.

#### III.2.3 - Quelle est la meilleure solution ?

Le Data Center est idéal pour les entreprises ayant besoin d’un système qui leur donne un contrôle total sur leurs données et leurs équipements. L’entreprise va utiliser la puissance de cette infrastructure pour exécuter des nombreux types d’applications et des charges de travail complexes. Cependant, le Data Center a une capacité limitée. Une fois que celui-ci est mis en place, il n’est pas autorisé de changer la quantité de stockage et la charge de travail.

Quant à lui, le Cloud est une solution extensible aux besoins du client. Contrairement au Data Center, il a la capacité d’être potentiellement illimité. Mais il a également ses défauts. Le cloud ne permet pas autant de contrôle qu’un Data Center géré par l’entreprise. Il est cependant possible d’installer un cloud privé dans le réseau de l’entreprise.

#### III.2.4 - Question prix ?

L’architecture du Data Center est construite totalement par l’entreprise. Elle nécessite des outils, de l’entretien, une administration, du personnel, de la gestion, et surtout du temps et de l’argent. Cette solution peut coûter aux entreprises 7 à 18 millions d’euro (10 à 25 millions de dollars) pour fonctionner correctement.

Contrairement au Data Center, le Cloud est une solution plus rentable pour les petites entreprises. Elle ne nécessite pas de temps et peu de capital. Les forfaits de celui-ci sont abordables, répondent au budget et aux besoins de performance.

Les tarifs du Cloud varient en fonction du fournisseur, de l’offre, et des attentes du client (capacité de stockage, service attendu, …). Le client peut demander un tarif à l’heure ou au mois dans la plupart des cas. Ces tarifs se trouvent en moyenne entre 120 et 500 euros par mois en fonction des critères cités précédemment.

Finalement, pendant que le Data Center met du temps à se mettre en place, les services du Cloud sont disponibles pour une utilisation presque immédiate après un enregistrement.

### III.3 – Les offres proposées

Quasiment toutes les offres de Cloud Computing sont payantes.

Il existe cependant plusieurs moyens de quantifier des paiements :

* L’utilisateur peut choisir un tarif au mois, indépendamment de ce qu’il en fait avec.
* L’utilisateur peut choisir un tarif qui va se baser sur :
  + Le volume de bande passante utilisé ;
  + La durée d’utilisation ;
  + La quantité de données hébergées et transférées.



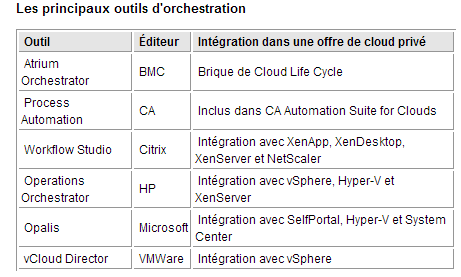
CloudWatt – Cloud public (stockage de données)

## IV – Les outils d’orchestration

### IV.1 - L’outil d’orchestration, c’est quoi ?

* Gère l’enchaînement des tâches d’exploitation, décrites dans des processus.

### IV.2 – Des exemples d’outils



**Selon le cabinet Gartner, le cloud public devrait connaître une croissance de 18,5 % entre 2012 et 2013 pour finalement peser 131 milliards de dollars cette année.**

# Le BIG DATA

## V – Le BIG DATA, généralités

Aussi appelé « Grosses données », le BIG DATA est utilisé pour désigner des ensembles de données qui deviennent tellement volumineux qu’ils en deviennent difficiles à travailler avec des outils classiques de gestion de base de données ou de gestion de l’information.

## VI – Le BIG DATA, plus précisément

Ce sont toutes les données numériques que nous produisons continuellement. Cela représente une telle quantité qu’elles ne peuvent être gérées par les outils de gestion de données traditionnels.

### VI.1 – A quoi sert-il ?

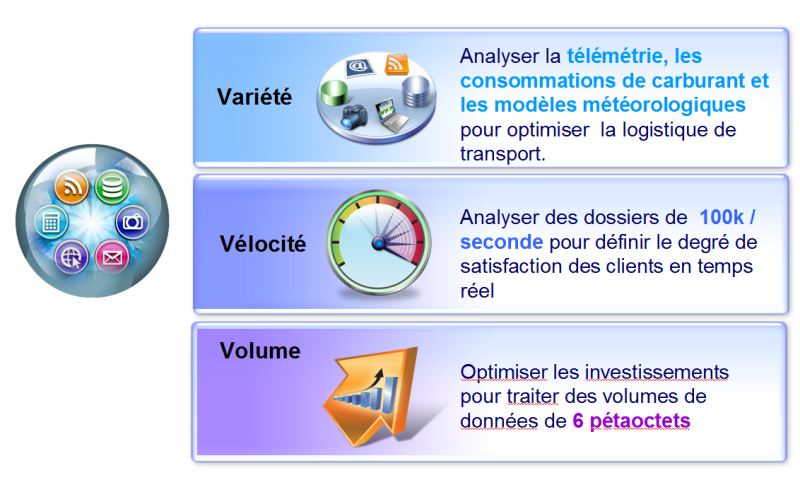
Le BIG DATA à une grande utilité économique. Aujourd’hui il est important de savoir analyser le comportement du consommateur, c’est ce qui permet aux entreprises de connaître nos habitudes de consommation, nos attentes, et ainsi savoir quoi proposer, ou le proposer, quand le proposer et comment le proposer. Ainsi une marque pourra précisément **comprendre le comportement de ses clients et à terme l’anticiper**.C’est pour cela que le BIG DATA représente un réel atout économique dans un monde consumériste.

### VI.2 – Comment ça marche ?

Le BIG DATA se concentre sur analyse des données provenant de sites web, de smartphones, de web analytics, via des algorithmes de calculs puissants et des technologies comme les bases de données très développées, des infrastructures de serveurs et le stockage des données. Il se caractérise souvent sous la forme de **3 valeurs** :

* **Le Volume** : c’est la quantité de données générées, en constante augmentation.
* **La Vitesse** : c’est la vitesse à laquelle les données sont générées et partagées.
* **La Variété** : c’est la diversité des types de données, provenant de différentes sources,

mobiles, médias sociaux, machines.



## IX – Avantages

* Economie de charges dans le processus commerciaux
* Economie de charges informatiques
* Avantages concurrentiels
* Augmentation du chiffre d’affaires grâce à des modèles commerciaux rapportés à BIG DATA

## X – Inconvénients

* **Soucis de sécurité**
  + Attaquants cherchant à repérer ou à créer une faille pour pénétrer le système.
  + Une des plus grande menace : Utilisateurs hors des pare-feu
    - Mobilité
    - Porte ouverte entre les SI et internet
    - Cloud Computing
* **Soucis de protection des données**
  + Problème du tout
    - Considérer toutes les informations comme personnelles
* Défis techniques (intégration, interfaces)
* Risques opérationnels
* Absence de pertinence pour modèle commercial propre
* Volume de données disponibles insuffisant

## XI – Limite du BIG DATA

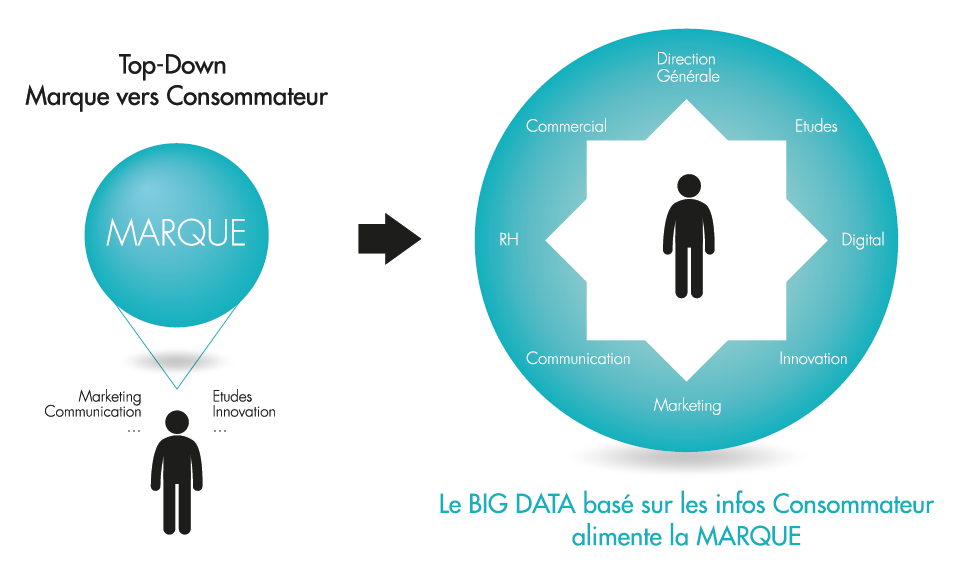
* Epistémologiques
* Méthodologiques
* Ethiques
* Confrontation avec le BIG BROTHER
* Hégémonie mal anticipée
* Outils d’infrastructure et de stockage

## XI – Etude de marché

### XI.1 – Les cibles

L’ensemble des entreprises :

* Tous les départements
  + Direction Générale
  + Commercial
  + RH
  + Etudes
  + Innovation
  + …



Le BIG DATA permet de se défaire de cette angoisse puisqu’elle remet le consommateur au cœur de la stratégie de l’entreprise. Ainsi, le comportement du consommateur, ses centres d’intérêt et ses prises de parole libres générés à grande échelle sont agrégés et analysés afin de créer ce fameux cadre stratégique, qui devient par définition totalement ajusté aux besoins et attentes des clients, et permet à la Marque d’alimenter toutes les directions de l’entreprise de façon innovante et performante.

### XI.2 – Les fournisseurs

Les plus grands fournisseurs de la solution BIG DATA

* + - Oracle
    - SAS
    - IBM
    - Microsoft
    - SAP
    - Qlik View
    - Talend
    - Statsoft

### XI.3 – Supports

Principales technologies du BIG DATA:

* ***Des bases de données NoSQL (MongoDB, Cassandra ou Redis)***
  + Implémentent des systèmes de stockage considérés comme plus performants que le traditionnel SQL pour l’analyse de données en masse (orienté clé/valeur, document, colonne ou graphe).
* ***Des infrastructures de serveurs pour distribuer les traitements***
  + Des dizaines, centaines, voire milliers de nœuds. C’est ce qu’on appelle le traitement massivement parallèle. Le Framework Hadoop est sans doute le plus connu d’entre eux. Il combine le système de fichiers distribué HDFS, la base NoSQL HBase et l’algorithme MapReduce.
* ***Le stockage des données en mémoire (Memtables)***
  + Accélère les temps de traitement des requêtes.

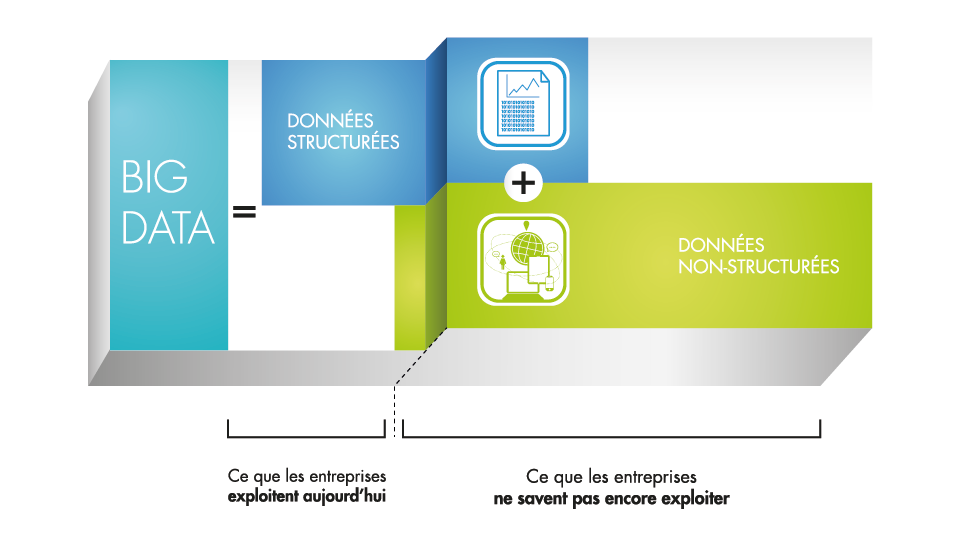
### XI.4 – Les offres proposées, la « BIG Solution »

  Ce tour d’horizon nous a permis de mettre en évidence que les éditeurs conservent leur avantage compétitif dans la Business Intelligence.

* **Oracle** se veut le plus complet en proposant un éventail de solutions allant du stockage à l’analyse des données. Il propose *Oracle Big Data Appliance*, utilisant Hadoop, qui permet  l’acquisition, l’organisation et le chargement des données non structurées dans Oracle data base. La liaison entre Hadoop, Oracle data base et Oracle data integrators est assurée par Oracle Big Data Connectors.
* **IBM** quant à  lui propose IBM InfoSphere BigInsights. Basé sur Hadoop, cette solution permet de stocker, gérer, et analyser un grand volume de données structurées et non structurées. A cela s’ajoute l’optimisation par IBM de Pure data system.
* **Microsoft** reste fidèle à sa stratégie de rendre accessible ses outils aux utilisateurs finaux. Il met à disposition une distribution Hadoop pour Windows Server et Windows Azur. L’objectif étant de faciliter l’accès et l’utilisation de Hadoop aux développeurs. Avec Excel Hive les utilisateurs finaux peuvent analyser les données à partir de Hadoop. La stratégie est claire : faire de JavaScript le langage du Big data, ce qui permettra à terme aux utilisateurs de créer des tâches MapReduce aussi bien en JavaScript, qu’en Hive, ou qu’en Pig latin. Pour se faire Microsoft compte généraliser l’accès à la console JavaScript.
* **SAS**laisse apparaître sa stratégie en matière de BIG DATA dans la phrase suivante: La valeur des Big data sera dans leur usage. SAS conserve son leadership dans le big data analytics (Forrester, 2013)[4]. Il renforce aussi ses solutions analytics en proposant High-Performance analytics dans sa plateforme SAS Business Analytics 9.4. Des solutions In-Memory sont proposées allant de l’analyse statistique au data Mining et le texte Mining.
* **Talend**en diversifiant son offre BIG DATA mérite sa place dans la cour des grands. Il décline trois produits dont Talend Platform for BIG DATA assez large pour répondre à la problématique globale. L’offre Talend utilise : MapReduce, Hadoop, Hbase, Hive, HCatalog, Oozie, Sqoop et Pig.

### XI.5 – Exploitation des données

A l’heure actuelle, les entreprises savent exploiter les données structurées mais ne savent pas exploiter les données non-structurées.



## XII – Les outils d'orchestration

* Amazon développe un nouveau langage : le langage R
  + Environnement de traitements et d’analyse statistique de données reposant sur le langage de programmation.

## XIII – Intérêts perçus du BIG DATA

