

# Big Data : comment réussir leur intégration dans le SI ?

Plus qu'un effet de mode, la révolution des Big Data vient bousculer les modèles et les technologies informatiques connues jusqu'à lors, tout en révélant de nouveaux enjeux auxquels devra répondre le SI de demain.

Parmi les domaines d'études de l'histoire du monde, nous avons la paléontologie. C'est la discipline scientifique qui étudie les restes fossiles des êtres vivants ayant vécu par le passé.

Dans le domaine de l'informatique, il n'existe pas encore de discipline qui permet d'étudier l'histoire et l'évolution des technologies. Mais une chose est sûre, les mutations permanentes de l'ère numérique se font à un rythme effréné et laissent derrière elles un certain nombre de fossiles, pour ne pas dire de cadavres.

**Que peut-on penser alors des téléphones portables du début des années 2000 en comparaison des Smartphones d'aujourd'hui ? Ce sont des dinosaures.**

Dans les SI (systèmes d'informations) des entreprises, un phénomène équivalent se produit où les (nouvelles) technologies entrantes poussent vers la sortie les plus anciennes. Tel est le cas avec le [Big Data](#) très en vogue actuellement. Plus qu'un effet de mode, cette révolution déjà en marche vient bousculer les modèles et les technologies informatiques connues jusqu'à lors, tout en révélant de nouveaux enjeux.

**En effet, le SI de demain devra répondre à deux grands défis :**

- La gestion de grands volumes de données,
- La création d'applications qui délivrent une analyse de ces données « just in time ».

En résumé tout, et tout de suite.

**Les personnes qui intégreront les entreprises de demain seront formatées à l'ère de l'Internet où l'information est disponible instantanément.**

Comment comprendre alors que pour obtenir une information il faille passer par des recueils de besoins, des spécifications, des développements, des recettages et une mise en production avec des délais parfois très longs (plusieurs semaines voire plusieurs mois) ? Inconcevable aujourd'hui.

**Pour répondre à cette problématique**, les entreprises devront intégrer une philosophie de gestion des données différente. Celles-ci devront être stockées dans d'immenses silos de réception afin d'en gérer les énormes volumes et les différents formats. Ces silos comparables à des puits de données devront également être en mesure d'intégrer des données externes (open data, réseaux sociaux, etc.), gonflant d'avantage la note en terme de volume. Les informations devront ensuite être traitées et analysées afin de répondre à la variété des demandes clients (analyse marketing, ciblage client, analyse commerciale, contrôle de gestion, etc.).

**Les futurs outils et technologies informatiques devront s'adapter à cette volumétrie et à ces structures de données différentes. Cela paraît maintenant inévitable.**

Cependant, des solutions existent et sont déjà employées par certains grands acteurs, notamment dans le secteur de l'Internet.

## Quelles sont ces technologies ?

L'une d'entre elle, jouissant déjà d'une très belle notoriété, est issue des travaux de Google et de Yahoo ! : Hadoop. Cette technologie permet de stocker et de traiter des grands volumes de données structurées ou non. Elle s'appuie à la fois sur un système de fichiers distribués (de un serveur jusqu'à plusieurs milliers) et sur un système d'analyse pourvu d'un algorithme très puissant, Map/Reduce.

## Un écosystème très riche gravite également autour de ces briques de base :

- Flume : un outil de chargement temps par flux.
- Hbase : une base de données temps réel pour stocker et requêter de forts volumes de données.
- Hive : un data warehouse qui s'appuie sur le système de fichier Hadoop.
- Mahout : un système de datamining puissant.

Toutes ces technologies contribueront demain à révolutionner l'approche du traitement de la donnée dans les entreprises. Et en présence d'une technologie open source, l'argument massue jouant en faveur de cette nouvelle tendance sera bien évidemment le prix.

**La plus grosse difficulté sera d'intégrer ces technologies dans les SI des entreprises.**

**En effet des spécialistes de la data devront nécessairement officier, ces types de compétences restant encore rares à l'heure actuelle.**

Aujourd'hui des entreprises répondent déjà à ces besoins en formant directement dans les écoles d'ingénieurs et les laboratoires informatiques les futurs spécialistes de cette nouvelle révolution.

Les entreprises au cœur de cette mouvance et les *data scientists* comme on les appelle, ne connaîtront certainement pas la crise aux regards d'une demande qui risque d'exploser dans les 5 ans à venir.

La révolution du SI est en marche.

