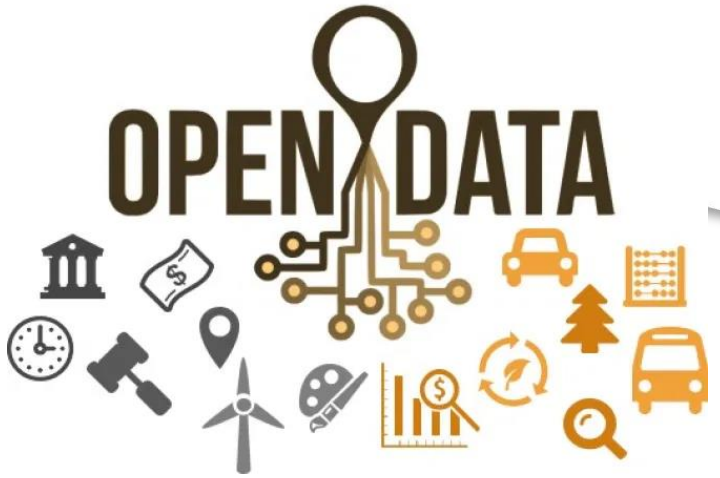

PC
Mini-Note
Internet
Notebook
Mobile
Remote Desktop
Database
Remote Server


Certification
DPO
Minimisation des données
Droit à la portabilité
PRIVACY BY DESIGN
RESPONSABILISATION
RGPD
Notification de violation de données
Registre
Coopération européenne


BIG
DATA
ANALYTICS
STORAGE
INTERNET
MANAGEMENT
SHARED
SOCIAL
RELATIONAL
DISK
PETABYTES
SYSTEMS
INCLUDE
PARALLEL
MASSIVELY
GROW
SEARCH
DATABASES
NETWORKS
HUNDREDS
RECORDS
CAPACITY
BIOLOGICAL
PROCESSING
OMPLEX
CONTINUES
CITATION
SET
USE
ER
ALSO
TOOLS
TARGET
APPLIED
SENSOR
DEFINITION
MOVING
WITHIN
PRACTITIONERS
CAPTURE
BUSINESS
SETS
TERABYTES
TOLERABLE
ABILITY
SIZE
SAN
EXAMPLES
EVERY
DIFFICULTY


OPEN DATA

Tour de table

Tour de table : 3 questions

- Présentation personnelle Qui ? (Je suis …)
- Entreprise/Contexte professionnel
- Attentes (les attentes de chacun par rapport à la formation)

Objectifs

Comprendre les enjeux du Big Data

Découvrir les opportunités du Big Data

Acquérir les connaissances nécessaires permettant de mettre en œuvre les technologies Big Data

Evaluer et sélectionner des outils appropriés dans le cadre d'un plan de mise en œuvre du Big Data


BIG
DATA
ANALYTICS
COMPLEX
SOCIAL
INTERNET
STORAGE
SETS
MANAGEMENT
PETABYTES
SYSTEMS
INCLUDE
HUNDREDS
DATABASES
NETWORKS
PROCESSING
BIOLOGICAL
CAPACITY
SEARCH
PARALLEL
SHARED
DISK
SENSOR
DEFINITION
CAPTURE
BUSINESS
TERABYTES
TOLERABLE
SIZE
ABILITY
MOVING
WITHIN
MASSIVELY
GROW
DIFFICULTY
GARTNER
RELATIONAL
LARGER
CONTINUES
TENS
CITATION
EVERY
EXAMPLES
TOOLS
TARGET
APPLIED
MAY
ALSO

BIG DATA

Sommaire

Présentation du Big Data

Explosion de la quantité des données

Le partage des données

La recherche des données

Le stockage des données

Le traitement des flux de données

Qu'est-ce que Big Data ?

Big Data

- Mégadonnées
- Donneés massives

Question :

- Big Data = Le volume de données?

Réponse : Faux

Qu'est-ce que Big Data ?

Big Data désigne un ensemble très volumineux de données qu'aucun outil classique de gestion de base de données ou de gestion de l'information ne peut manipuler convenablement.

Exemples :

- ✓ environ 2,5 trillions d'octets de données chaque jour.
- ✓ ce sont des informations provenant de sources diverses :
  - messages
  - vidéos
  - informations climatiques
  - signaux GPS
  - transactions
  - etc.

Qu'est-ce que Big Data ?

Il n'existe pas de définition précise du Big Data. Chacun la définit selon ses points d'intérêt ou son domaine d'activités :

- ✓ Traitement des données volumineuses dépasse la capacité des logiciels de base
- ✓ Nouveaux outils qui nous aident à trouver des données pertinentes
- ✓ La convergence de l'informatique d'entreprise et de consommation.
- ✓ Le passage (pour les entreprises) du traitement des données internes à la fouille des données externes.
- ✓ Le passage (pour les individus) de la consommation de données à la création de données.

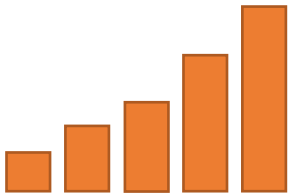
Qu'est-ce que Big Data ?

- La taille n'est pas l'unique caractéristique.
- Flux de données continu (big data stream).
- Plus personne ne veut jeter les données.
- La variété des sources de données indépendantes nécessite l'intégration des données.
- Hétérogénéité des données.
- Souvent associé aux données non structurées, mais non exclusivement.

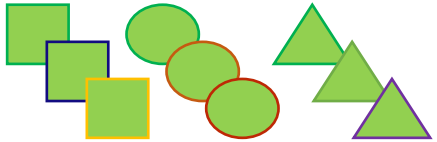
Qu'est-ce que Big Data ?

La règle des **3V**

Volume



Variété



Vélocité

Qu'est-ce que Big Data ?

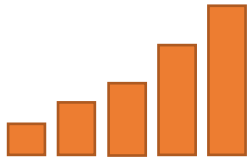
Plus **2 V**

Véracité

Valeur

Qu'est-ce que Big Data ?

Volume



Le volume fait référence à la grande quantité de données

- ✓ générées
- ✓ les e-mails
- ✓ messages Twitter
- ✓ photos
- ✓ clips vidéo
- ✓ données de capteurs

Sur Facebook nous
Envoyons 10 milliards de messages par jour
Cliquons sur le bouton "J'aime" 4,5 milliards de fois
Téléchargeons 350 millions de nouvelles images chaque jour

Données générées depuis le début jusqu'au l'an 2000

=

Données que nous générons aujourd'hui chaque minute!

Qu'est-ce que Big Data ?

Vélocité

La vélocité fait référence à la vitesse
- ✓ de la générations des nouvelles données
- ✓ à laquelle les données se déplacent

Une petite pensée aux
Productions des messages de médias sociaux en quelques minutes. La rapidité avec laquelle les transactions par carte de crédit sont vérifiées pour des activités frauduleuses.
Aux millisecondes pour analyser les réseaux sociaux afin de détecter les signaux qui déclenchent des décisions d'achat ou de vente.
Etc…

La technologie Big Data nous permet aujourd'hui d'analyser les données pendant qu'elles sont générées sans les stocker dans des bases de données.

Qu'est-ce que Big Data ?

Variété

La variété fait référence aux différents types de données utiliser.

Avant :
Stockage des données structurées dans les bases de données relationnelles.

Aujourd'hui :
80% des données ne sont pas structurées
Difficile de les stocker dans des bases de données relationnelles

- ✓ photos
- ✓ vidéo
- ✓ audio
- ✓ messages réseaux sociaux

Avec la technologie Big Data nous pouvons désormais exploiter différents types de données, et les rassembler avec des données structurées et traditionnelles.

Qu'est-ce que Big Data ?

Véracité

La véracité fait référence au désordre ou à la fiabilité des données.

La qualité et la précision sont moins contrôlables

- ✓ messages Twitter avec des hashtags
- ✓ les abréviations
- ✓ les fautes de frappe
- ✓ Etc…

La technologie Big Data et Analytics nous permet désormais de travailler avec ces types de données.

Qu'est-ce que Big Data ?

Valeur

La Valeur fait référence à notre capacité à transformer nos données en valeur.

Tous les V qu'on a vu jusqu'ici

- ✓ Volume
- ✓ Variété
- ✓ Vélocité
- ✓ Véracité

doivent être transformés en Valeur !

La Valeur est le V du Big data qui compte le plus.

Pourquoi Big Data ?

Difficulté de traiter la masse de données produite chaque jour

Les informations sont produites en temps réel

Problème d'optimisation de base de données

Le traitement des données non structurées

Perspectives offertes par le Big Data

Les revenus du marché mondial des Big Data pour les logiciels et les services devraient passer de 42 milliards de dollars en 2018 à 103 milliards de dollars en 2027, atteignant un taux de croissance annuel (Compound Annual Growth Rate (CAGR)) de 10,48% selon Wikibon.

Perspectives offertes par le Big Data

L'ensemble du marché mondial des logiciels devrait générer des revenus de 628 milliards de dollars, dont 302 milliards de dollars provenant des applications.

Perspectives offertes par le Big Data

Selon une étude d'Accenture, 79% des dirigeants d'entreprises s'accordent à dire que les entreprises qui n'adhèrent pas au Big Data perdront leur position concurrentielle et pourraient être menacées d'extinction. Plus encore, 83%, ont poursuivi des projets Big Data pour se positionner face à la concurrence.

Selon PwC, 59% des dirigeants affirment que les Big Data de leur entreprise seraient améliorées grâce à l'utilisation de l'intelligence artificielle.

Perspectives offertes par le Big Data

84% des entreprises ont lancé des analyses avancées et des initiatives Big Data pour améliorer leur précision et accélérer leur prise de décision.

Les initiatives Big Data axées sur ce domaine ont également le taux de réussite le plus élevé (69%) selon la dernière enquête NewVantage Venture Partners.

Plus du tiers des entreprises, soit 36%, affirment que cette zone est leur priorité absolue en matière d'analyse avancée et d'investissement dans le Big Data.

Freins de la mise en place du Big data

Coût,

Manque de compétences,

Manque de visibilité sur les opportunités,

Difficile de quantifier le ROI des investissements Big Data,

La collecte des données doit surpasser les canaux traditionnels,

Les données sont non structurées (nécessité de compétences pour les traiter).

Gains de la mise en place du Big data

Anticipation des enjeux stratégiques liés à une meilleure utilisation des données internes et externes,

Diversité des données collectées et des canaux de collecte,

Adoption de nouvelles technologies d'exploitation de la donnée.

Les acteurs du Big Data

Les acteurs historiques de solutions IT :

HP

Oracle

IBM

SAP

Figurent parmi les principaux acteurs du Big Data.

Les acteurs du Big Data

IBM :

- ✓ BigInsights Enterprise

Microsoft :

- ✓ privilégié l'utilisation du framework Hadoop
- ✓ Windows Azure et Window Server

Amazon :

- ✓ Amazon Web Services
- ✓ Elastic MapReduce (EMR)

Les acteurs du Big Data

Parmi les spécialistes des solutions Big Data

- ✓ CapGemini
- ✓ Accenture
- ✓ EMC
- ✓ MapR
- ✓ Teradata
- ✓ Sopra Steria
- ✓ Atos

Les acteurs du Big Data

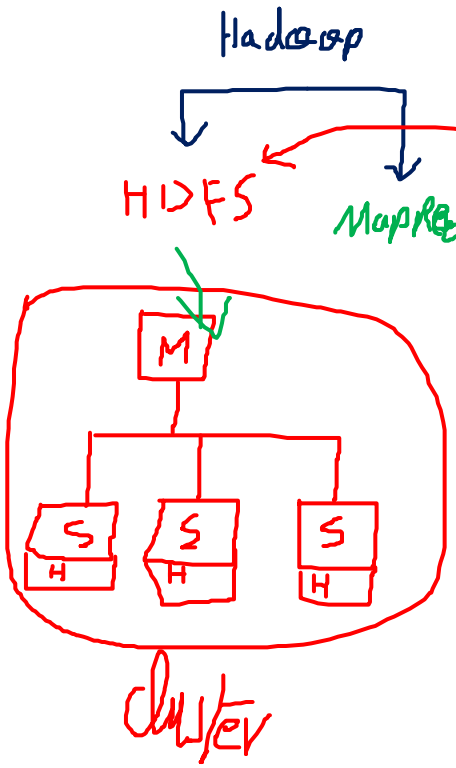
On retrouve également les spécialistes de l'analytique :

- ✓ Qliktech
- ✓ SAS
- ✓ Micro-strategy

ou encore les fournisseurs spécialisés tels que :

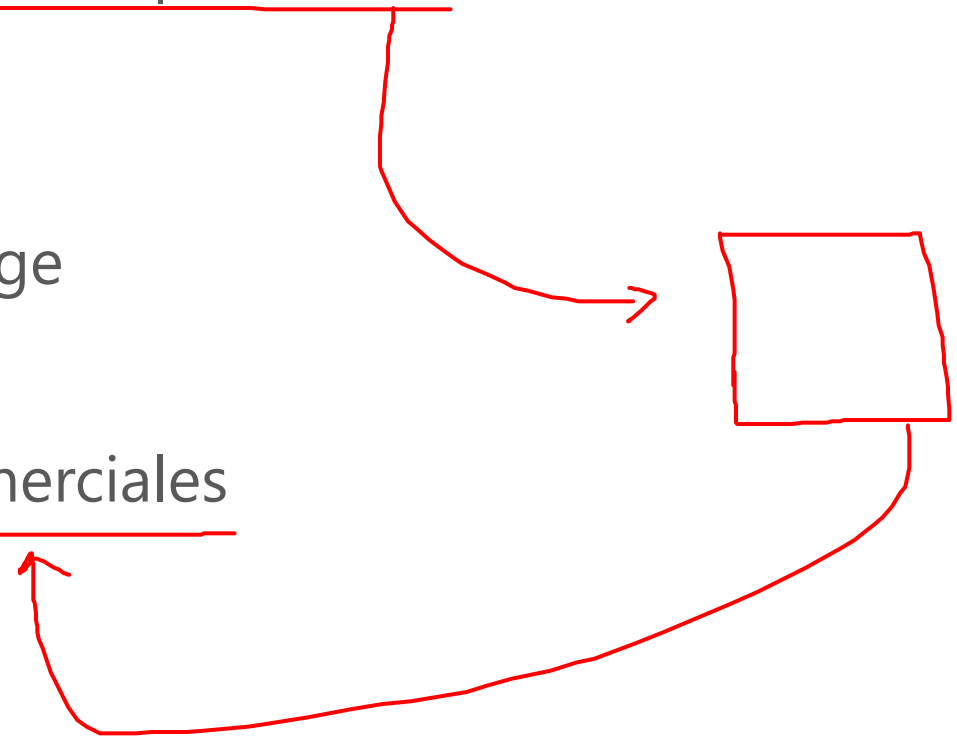
- ✓ Zettaset
- ✓ Datameer

Distributions majeures



Hadoop est le Framework logiciel open source au cœur de la révolution Big Data.

- ✓ sortie en 2011
- ✓ solutions pour le stockage
- ✓ l'analyse de données
- ✓ Avec distributions commerciales



Distributions majeures

Un facteur important à prendre en compte dans le choix d'une distribution.

Hadoop sur site ou sur le cloud ?!!

Distributions majeures : Cloudera

Cloudera a été le premier fournisseur à proposer Hadoop en tant que package et continue d'être un leader dans l'industrie.

Cloudera CDH avec des composants open source

Premier à proposer SQL-for-Hadoop avec son moteur de recherche Impala.

La sécurité et les interfaces pour l'intégration avec des applications tierces

Distributions majeures : Cloudera

cloudera manager | Clusters | Hosts | Diagnostics | Audits | Charts | Backup | Administration | Search (Hotkey: /) | Support | admin

Home

30 minutes preceding November 3, 2015, 1:44 PM PST

Status | All Health Issues | Configuration 16 | All Recent Commands | Add Cluster

Cluster 1 (CDH 5.5.0, Packages)

- Hosts 4
- FLUME-1
- HBASE-1
- HDFS-1 1
- HIVE-1
- HUE-1 1
- IMPALA-1
- KAFKA-1
- KS_INDEXER-1
- KUDU-1
- MAPREDUCE-1
- OOZIE-1
- SOLR-1
- SPARK_ON_YA...
- SQOOP-1
- SQOOP_CLIENT-1
- YARN-1
- ZOOKEEPER-1 1

Charts — 30m 1h 2h 6h 12h 1d 7d 30d

Cluster CPU

100% / 50% / 0% (percent) — 01:15 — 01:30

Cluster 1 2.4% Cluster 2 1.6%

Cluster Disk IO

977K/s / 488K/s / 0 (bytes / second) — 01:15 — 01:30

Cluster 1, Total Disk... 0 Cluster 1, Tot... 297K/s
Cluster 2, Tot... 68.3b/s Cluster 2, Total ... 99K/s

Cluster Network IO

39.1K/s / 19.5K/s / 0 (bytes / second) — 01:15 — 01:30

Cluster 1, Tot... 25.3K/s Cluster 1, Tot... 23.6K/s
Cluster 2, Tot... 19.2K/s Cluster 2, Tot... 23.3K/s

HDFS IO

20b/s / 0 (bytes / second) — 01:15 — 01:30

HDFS-1, Total B... 1b/s HDFS-1, Total ... 2.8b/s
HDFS-2, Total B... 1b/s HDFS-2, Total ... 2.8b/s

Running MapReduce Jobs

(jobs) — 01:15 — 01:30

MAPREDUCE-1 0 MAPREDUCE-2 0

Completed Impala Queries

(queries / second) — 01:15 — 01:30

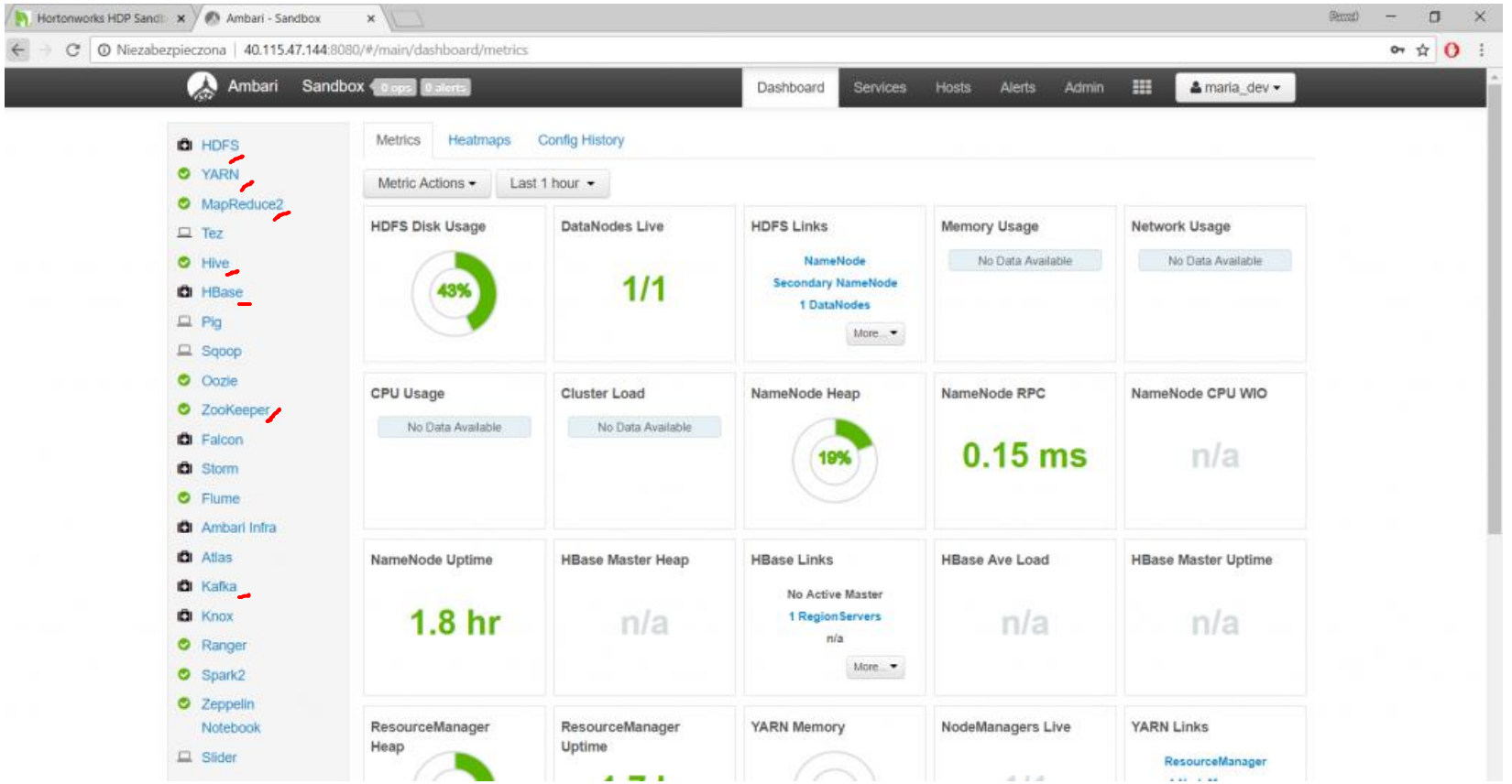
IMPALA-1 0 IMPALA-2 0

Distributions majeures : Hortonworks

Hortonworks est entièrement open source

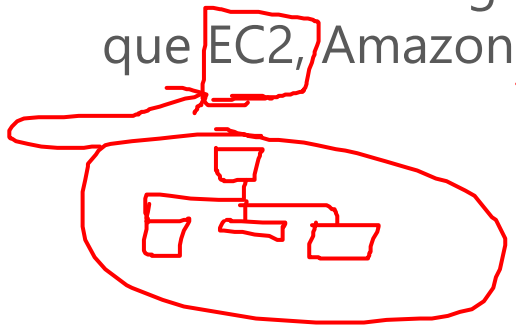
Amazon et IBM proposent désormais Hortonworks en tant qu'options sur leurs propres plateformes

Distributions majeures : Hortonworks



Distributions majeures : Amazon EMR

- Amazon Elastic Map Reduce
- Amazon propose une plate-forme cloud-only Hadoop-as-a-service via ses Web Services.
- Utilisation de
  - ✓ Framework de traitement de flux Apache Flink (temps réel)
  - ✓ Kafka …
- Il se connecte également aux autres infrastructures de cloud d'Amazon telles que EC2, Amazon S3 et DynamoDB et AWS IoT.



Distributions majeures : Amazon EMR

AWS Management Console

https://console.aws.amazon.com/elasticmapreduce/home

aws.amazon.com | AWS | Products | Developers | Community | Support | Account | Welcome, Benjamin Langmea

Amazon EC2 | Amazon Elastic MapReduce | Amazon CloudFront

5. Job flow debugging

Your Elastic MapReduce Job Flows

Region: US East | Create New Job Flow | Terminate | Debug | Show/Hide | Refresh | Help

Viewing: All | 1 to 3 of 3 Job Flows

Name	State	Creation Date	End Date	Elapsed Time	Normalized Instar
Myrna	RUNNING	2010-05-20 21:10 EDT	-	0 hours 42 minutes	40
Myrna	COMPLETED	2010-05-18 10:24 EDT	2010-05-18 11:18 EDT	0 hours 52 minutes	40
Myrna	FAILED	2010-05-18 09:43 EDT	2010-05-18 09:43 EDT	0 hours 0 minutes	0

Steps | Bootstrap Actions

1. Job status

3. Time and instance-hours used

State	Start Date	End Date	Jar	Main Class	Args
COMPLETED	2010-05-20 21:13 EDT	2010-05-20 21:13 EDT	s3://us-east-1.elasticmapreduce/libs/script-runner/script-runner.jar	-	s3://us-east-1.elasticmapreduce/libs/
COMPLETED	2010-05-20 21:13 EDT	2010-05-20 21:23 EDT	/home/hadoop/contrib/streaming/hadoop-0.18-streaming.jar	-	-input s3n://ec2-jeff-tmp/example/ye hdfs:///myrna/intermediate/32190/p -compress=gzip -stop=0 -maxperfile= push=HDFS:///myrna/intermediate/3 org.apache.hadoop.mapred.lib.NLineI emr/0.1.0/Get.pm#Get.pm -cacheFil cacheFile s3n://myrna-emr/0.1.0/To emr/0.1.0/AWS.pm#AWS.pm -cache emr/0.1.0/samtools64#samtools -job

2. Step status

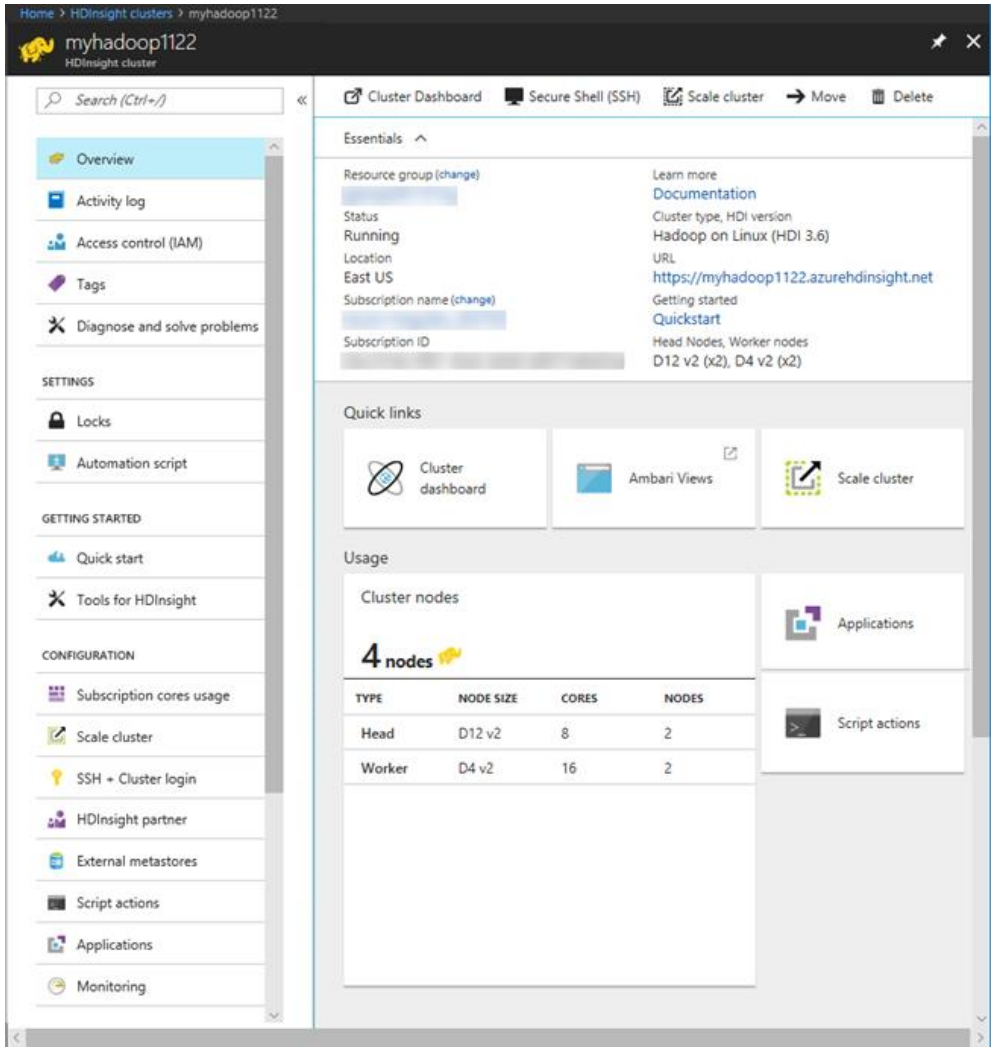
4. Step command line

© 2008 - 2009, Amazon Web Services LLC or its affiliates. All right reserved. | Feedback | Support | Privacy Policy | Terms o

Distributions majeures : Microsoft

- Azure HDInsight
  - ✓ service en cloud
  - ✓ propose des installations gérées de plusieurs distributions Hadoop (Hortonworks, Cloudera et MapR)
  - ✓ Il les intègre à sa propre plate-forme Azure Data Lake pour offrir une solution complète de stockage et d'analyse basée sur le cloud.
  - ✓ Fournit les services cloud Spark,Hive, Kafka et Storm

Distributions majeures : Microsoft



Distributions majeures : MapR

Comme Hortonworks et Cloudera, MapR est un fournisseur axé sur la plateforme, plutôt qu'un fournisseur de services gérés, comme Amazon ou Microsoft.

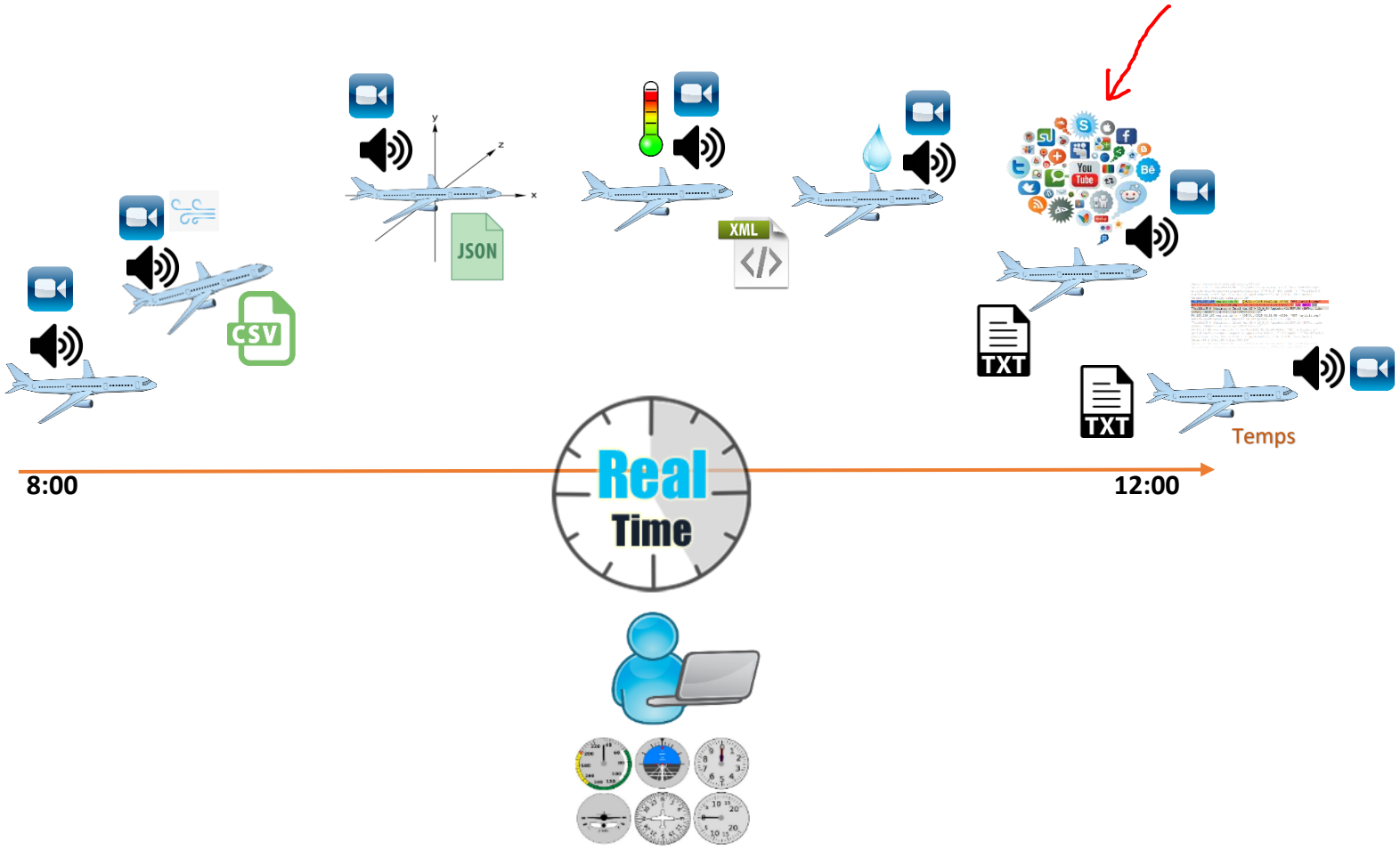
MapR intègre son propre système de base de données - MapR-DB

- ✓ Selon MapR – il est (4 à 7 fois) plus rapide que la base de données Hadoop

En raison de sa puissance et de sa rapidité, MapR est souvent considéré comme un bon choix pour le plus grand des projets Big Data.


BIG
DATA
ANALYTICS
COMPLEX
SOCIAL
INTERNET
MANAGEMENT
PETABYTES
SYSTEMS
STORAGE
PARALLEL
SETS
INCLUDE
SHARED
DISK
SIZE
HUNDREDS
DATABASES
SEARCH

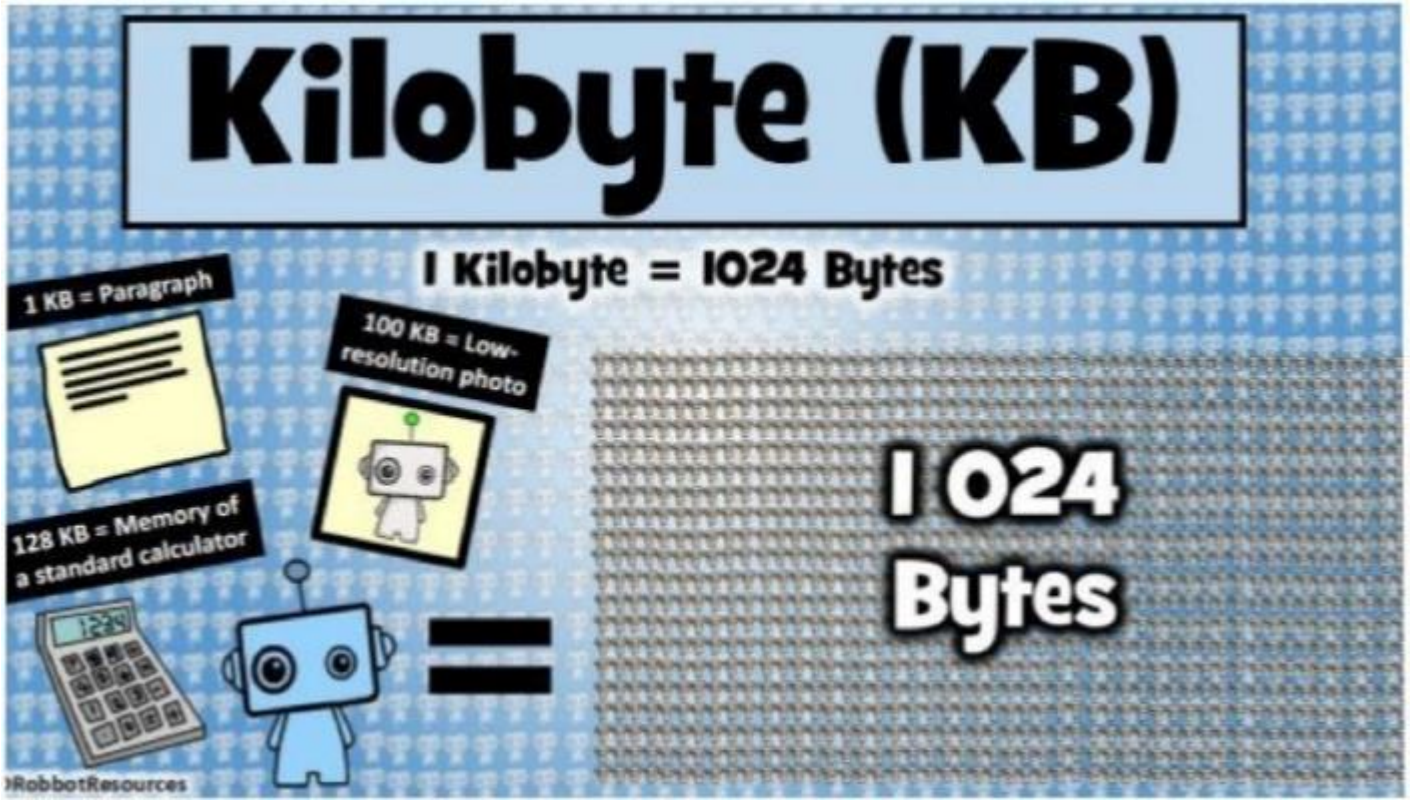
Exemple


CSV
JSON
XML
TXT
TXT
Temps
8:00
Real
Time
12:00


BIG
DATA
ANALYTICS
COMPLEX
SOCIAL
INTERNET
MANAGEMENT
PETABYTES
SYSTEMS
STORAGE
PARALLEL
SEARCH
HUNDREDS
DATABASES
KEYWORDS
SETS
INCLUDE
SIZE
SHARED
DISK
SENSOR
MOVING
SHARED
SAN

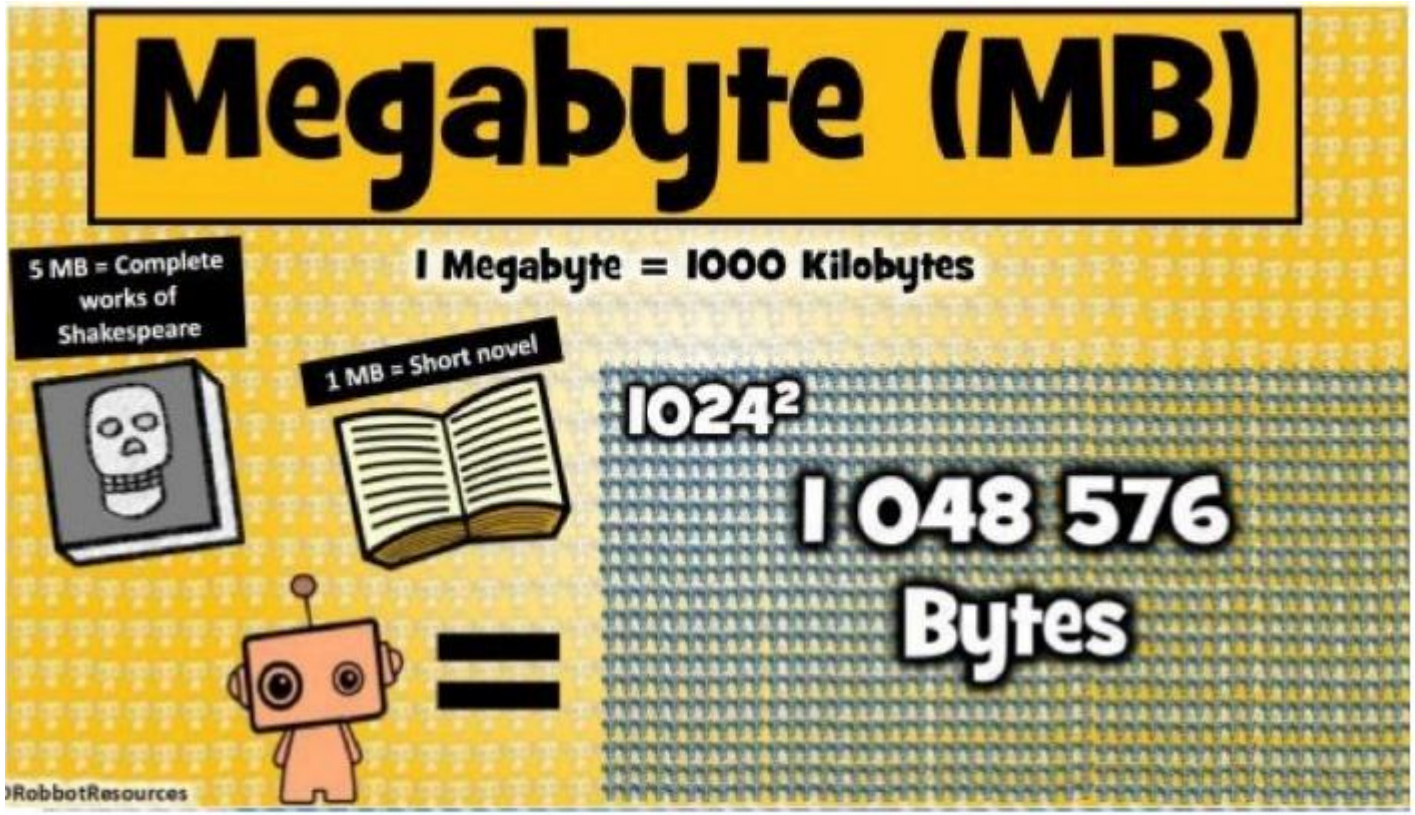
Exemple

Pourquoi utiliser le Big Data ?



1 KB = paragraphe
100KB = photo basse résolution
128 KB = mémoire d'une calculatrice standard

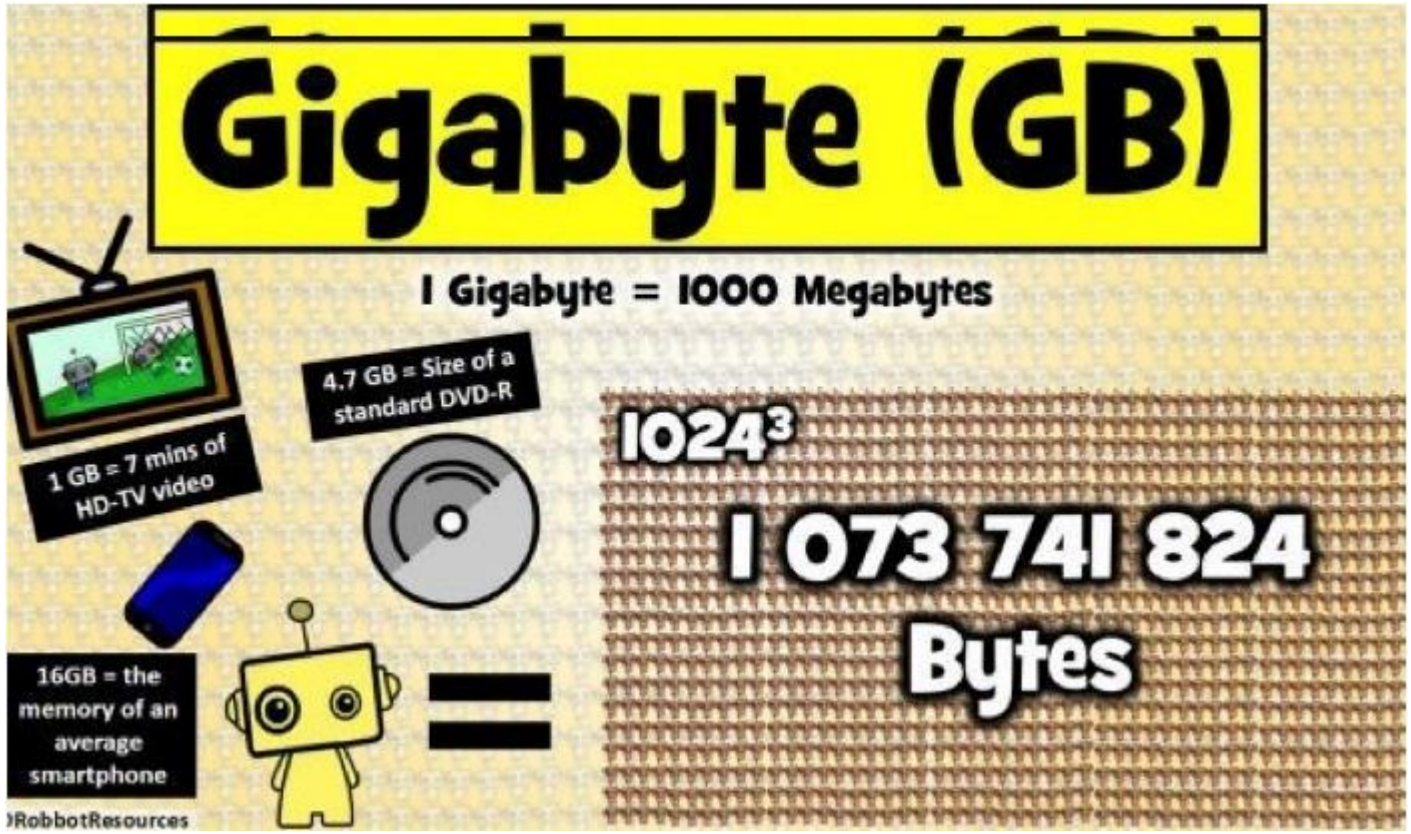
Pourquoi utiliser le Big Data ?



1MB = roman court

5 MB = œuvres complètes de Shakespeare

Pourquoi utiliser le Big Data ?



1GB = 7 minutes d'une vidéo HD

4.7 GB = Taille standard d'un DVD-R

16 GB = mémoire d'un smartphone moyen

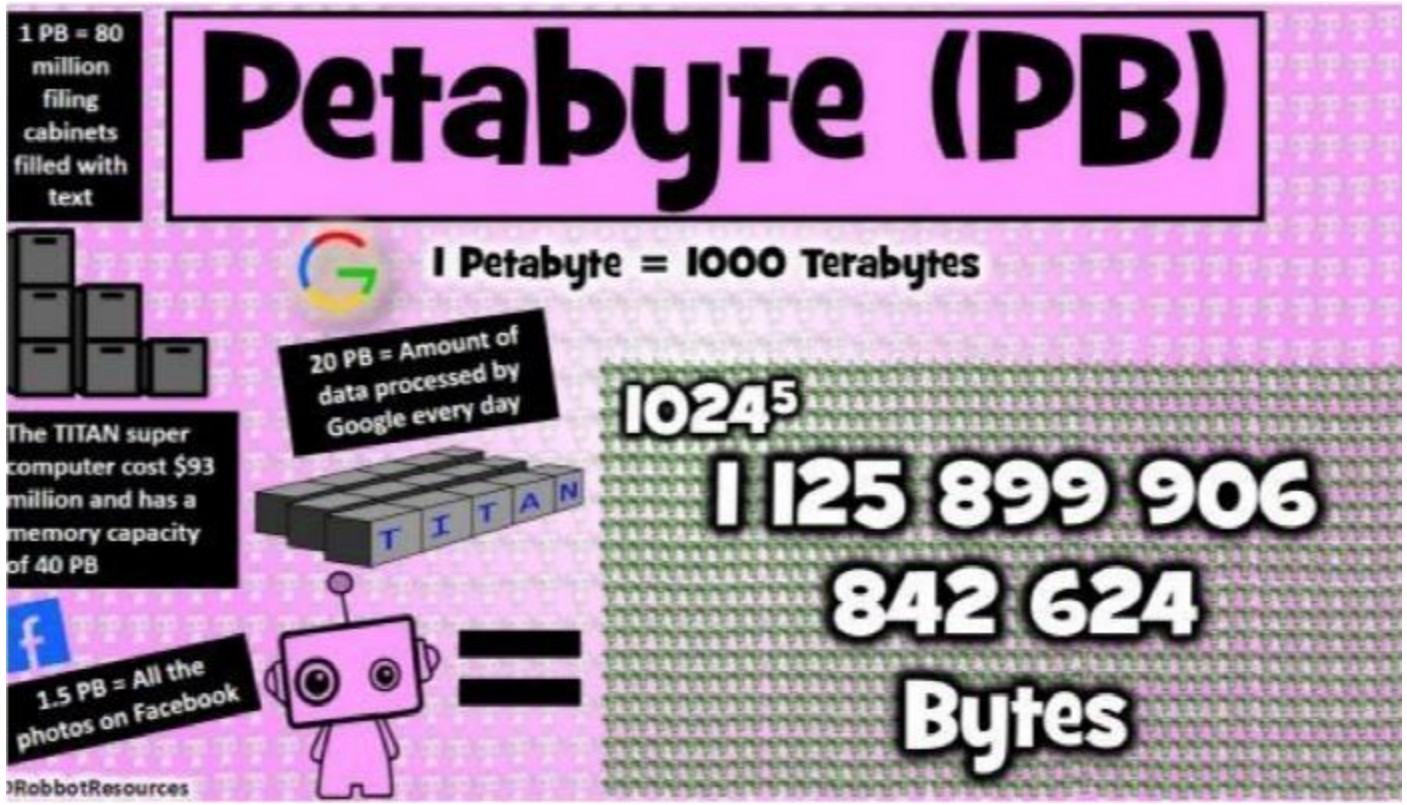
Pourquoi utiliser le Big Data ?

1TB = Disque dur d'un ordinateur portable moderne

10TB = assez pour stocker tout ce que vous regardez pendant un an

1TB = 50 000 arbres en papiers imprimés

Pourquoi utiliser le Big Data ?



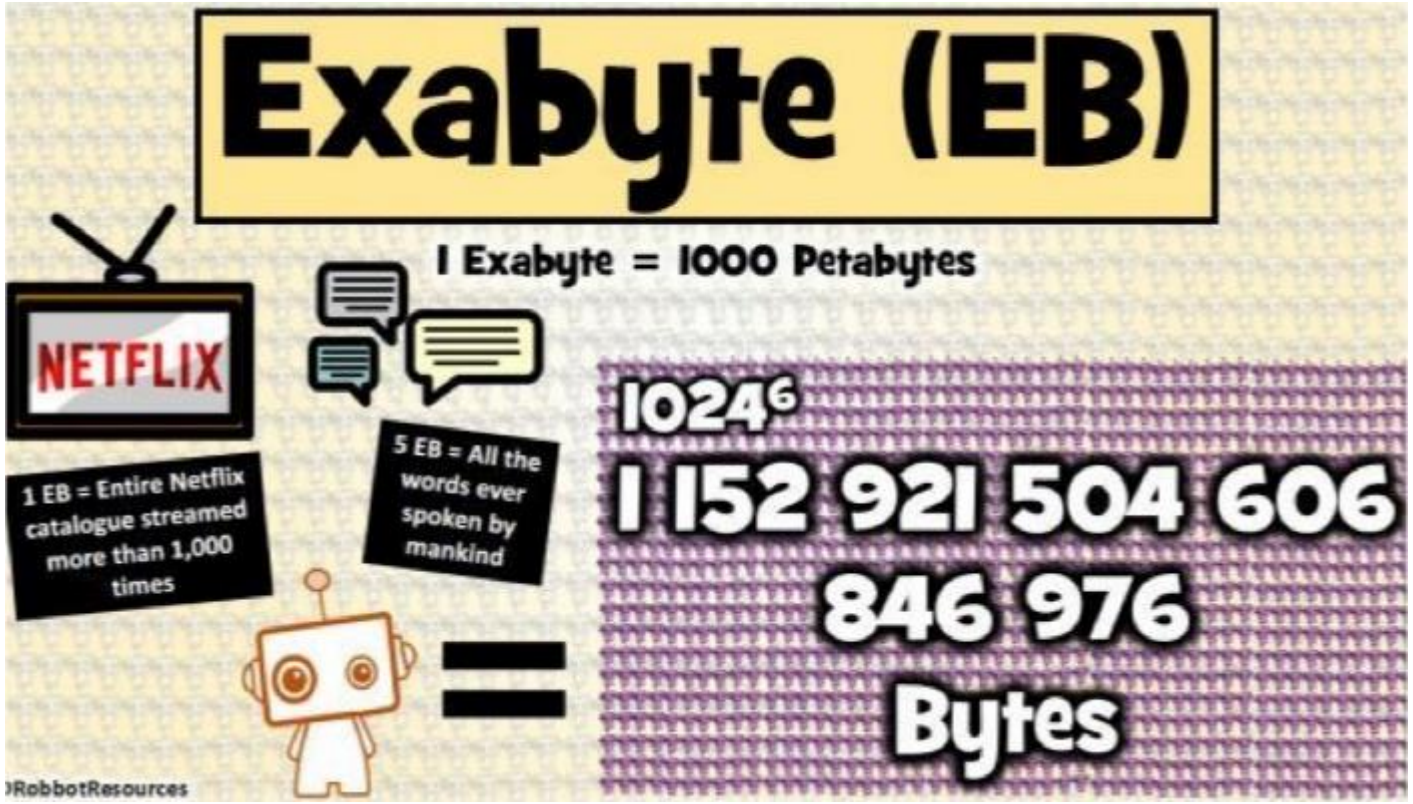
1PB = 80 millions de classeurs remplis de texte.

Le super supercalculateur Titan a coûté 93 millions de dollars et a une capacité de mémoire de 40 PB.

1.5 PB = Toutes les photos sur Facebook

20 PB = La Quantité de données traitées par Google chaque jour.

Pourquoi utiliser le Big Data ?



1EB = l'ensemble du catalogue Netflix diffusé plus de 1000 fois.

5EB = Tous les mots jamais prononcés par l'humanité.

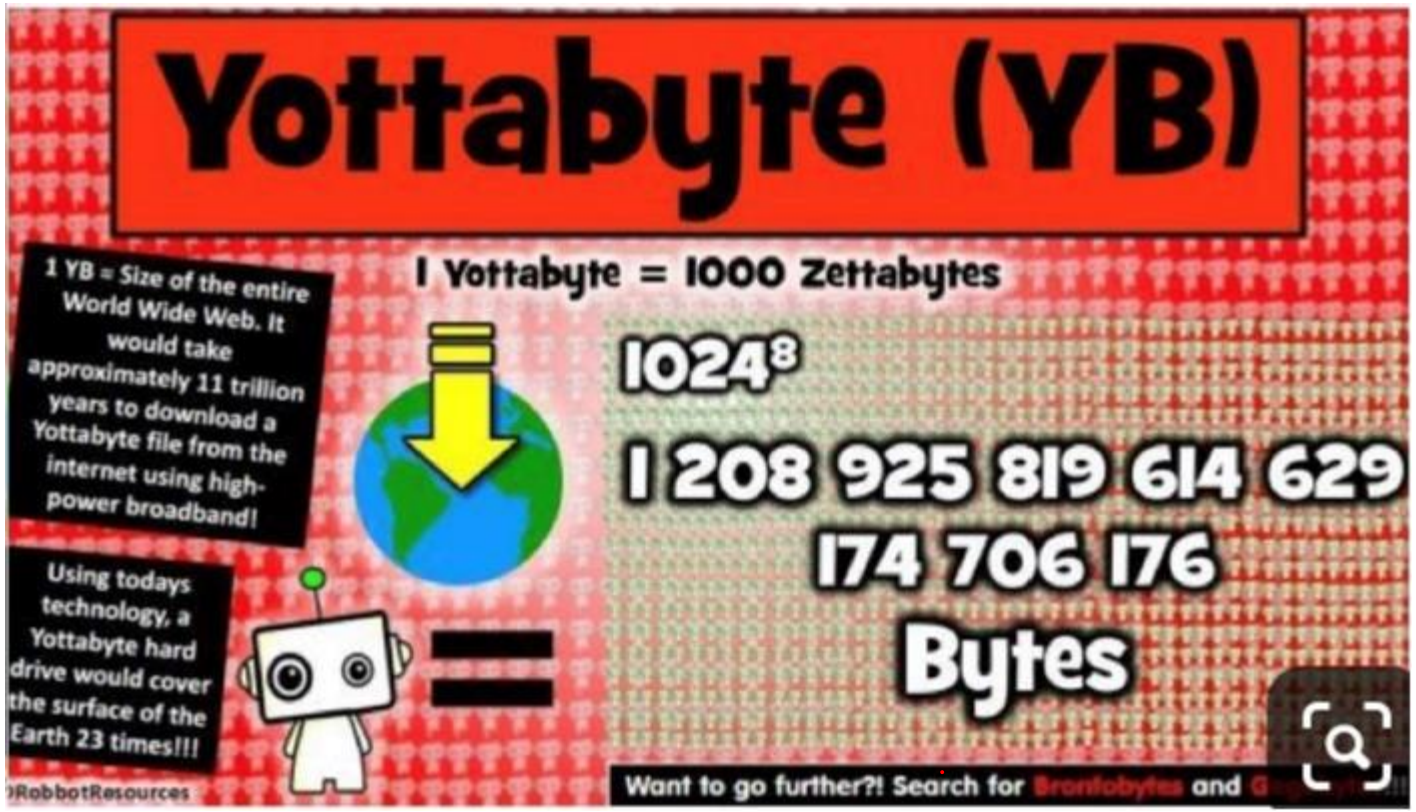
Pourquoi utiliser le Big Data ?

1ZB = Chaque jour, en moyenne, environ 500 000 000 tweets sont tweetés sur Twiter.
Si ces données sont stockées, il faudrait environ 100 ans pour égaler un zettaoctet de données.

1ZB = 250 billion DVD

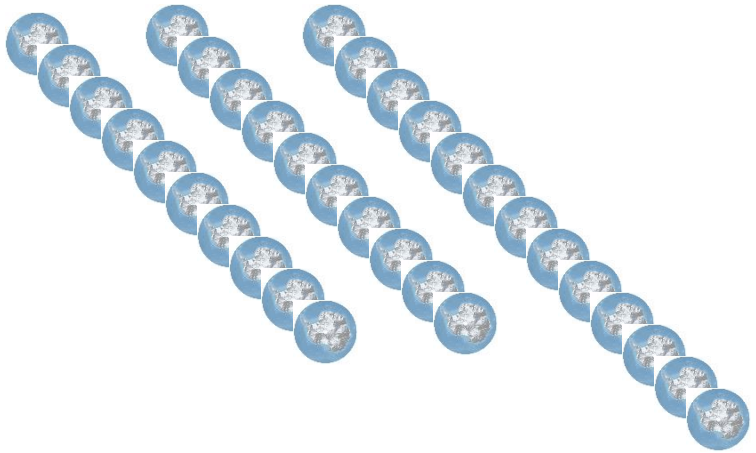
En utilisant la technologie d'aujourd'hui, un disque dur Zettabyte aurait la taille de l'Antarctique.

Pourquoi utiliser le Big Data ?



1YB = Taille du World Wide Web. il faudrait environ 11000 milliards d'années pour télécharger un fichier Yottabyte à partir d'Internet à l'aide d'une bande à haute puissance.

En utilisant la technologie d'aujourd'hui, un disque dur de Yottabytes couvrirait la surface de la terre 23 fois !!!



L’écosystème du Big Data

.

Systèmes de fichiers

Quantcast File System (QFS)

Bases NoSQL

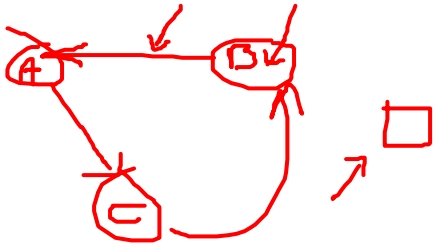
mongoDB.

APACHE HBASE

elasticsearch

cassandra

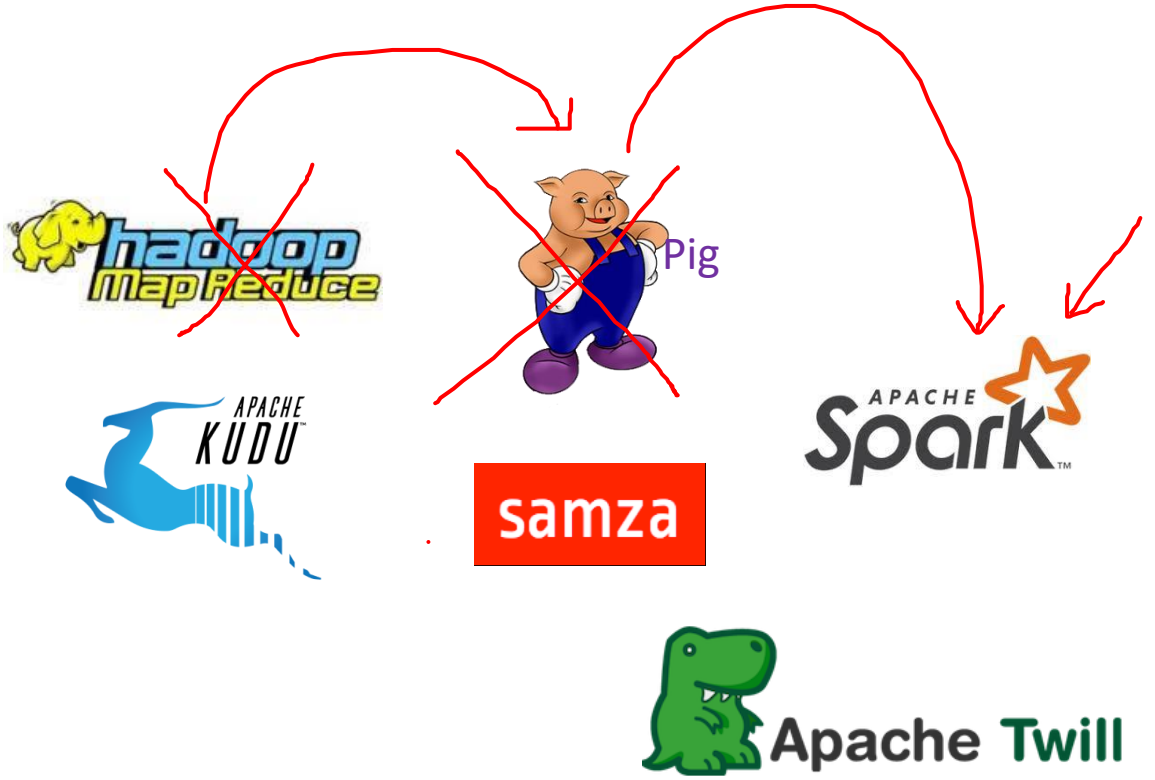
neo4j



Bases NewSQL

Intégration de données

Programmation distribuée



Machine Learning

Spark MLlib
The Machine Learning Library

mahout

WEKA
The University of Waikato

H_2O

julia

Visualisation

IBM COGNOS

QlikView

Spotfire

SAP Business Objects

MicroStrategy

tableau

Pentaho reporting

kibana

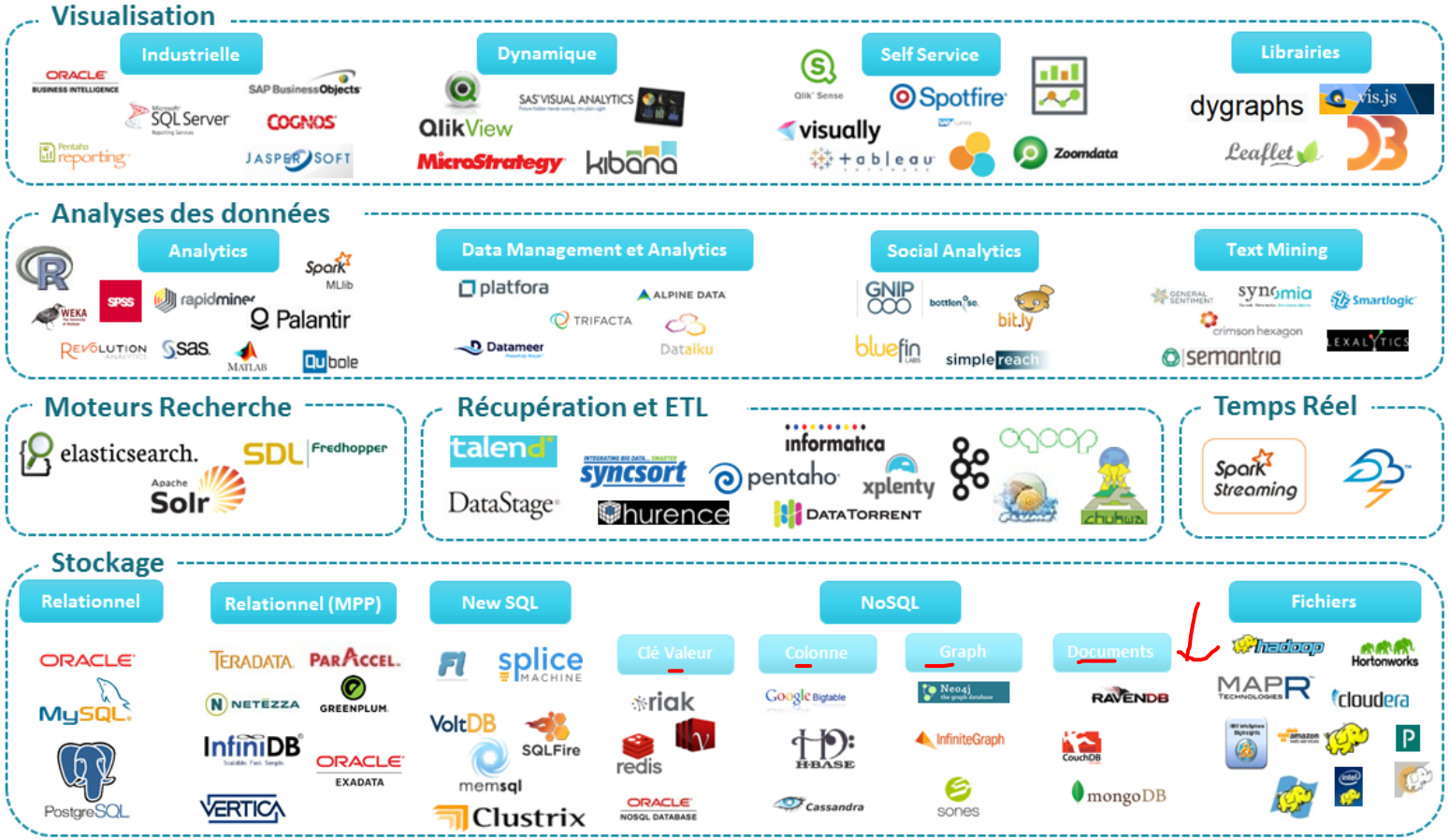
ZOOMDATA

Sécurité

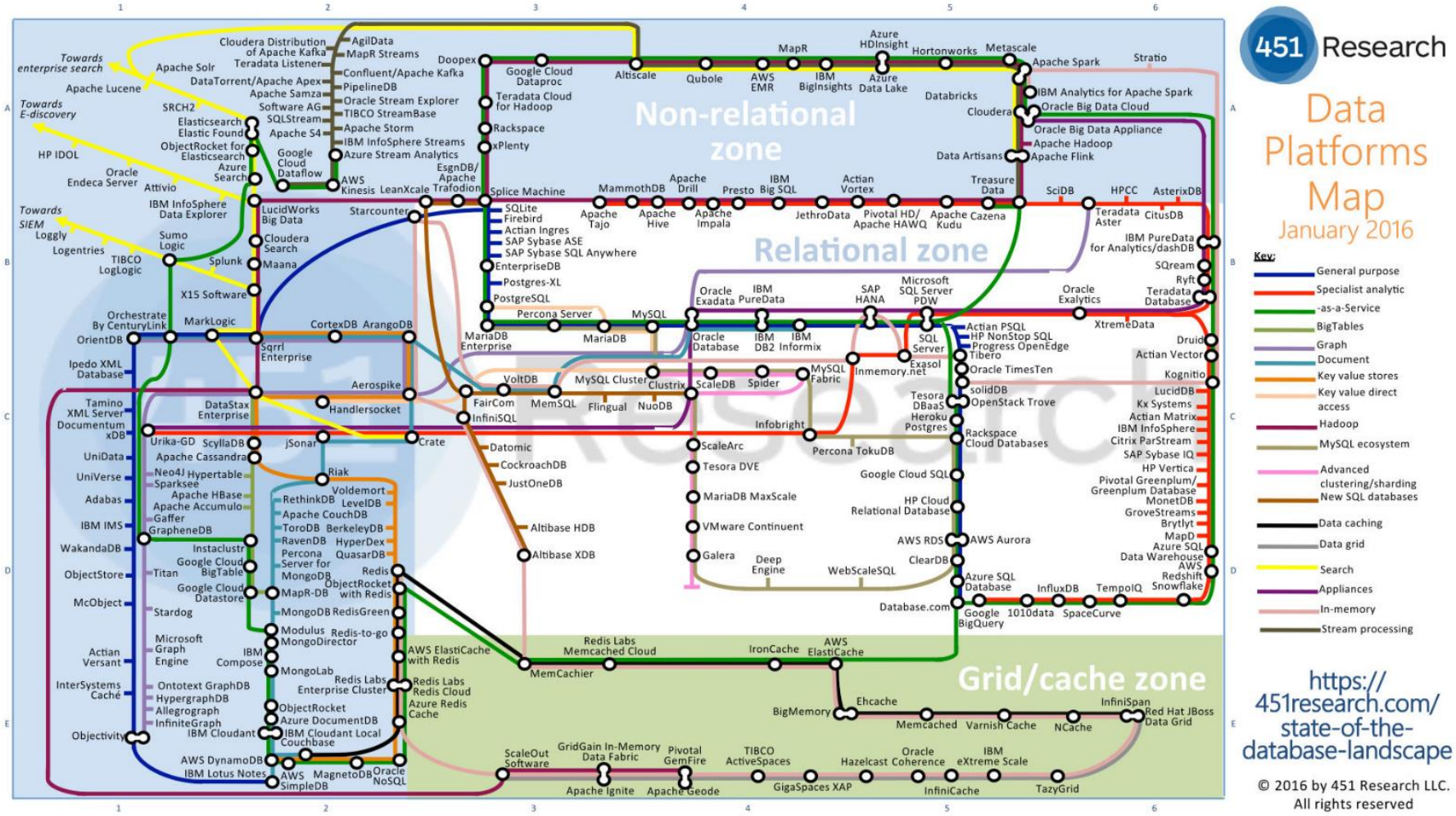
Sentry

Apache Ranger

L'écosystème du Big Data



L'écosystème du Big Data



451 Research

Data Platforms Map

January 2016

Key:

- General purpose
- Specialist analytic
- -as-a-Service
- BigTables
- Graph
- Document
- Key value stores
- Key value direct access
- Hadoop
- MySQL ecosystem
- Advanced clustering/sharding
- New SQL databases
- Data caching
- Data grid
- Search
- Appliances
- In-memory
- Stream processing

https://451research.com/state-of-the-database-landscape

Sommaire

1. Généralités
2. **Les opportunités qu'offre le Big Data pour la DSI**
3. Open Data
4. Cas d'usage du Big Data
5. Le stockage dans le Big Data
6. Les technologies du Big Data
7. Le traitement des données en Big data
8. Compétences autour du Big Data
9. Les étapes d'un projet Big Data

Evolution avec le Big Data ?

Progression de Hadoop

Une importante réduction des coûts d'analyse des données

Les SGDB NoSQL

La sécurité des données

Le stockage des données sur le cloud

Levier d'amélioration par l'analyse des données

L'information n'est efficace que si elle répond aux besoins de son utilisateur

Pour le Big Data l'entreprise est la partie prenante

Un des principaux enjeux de l'entreprise est la qualité de l'information

Plus les informations sont de qualité et plus les décisions prises sur la base de ces dernières seront pertinentes, créant ainsi de la valeur pour l'entreprise.

Levier d’amélioration par l’analyse des données

L’ analyse des données permet d'aider à la prise de décision dans l'entreprise.

- data mining/analytics

Dans le cas du Big Data, on parle de

- Big Data Analytics

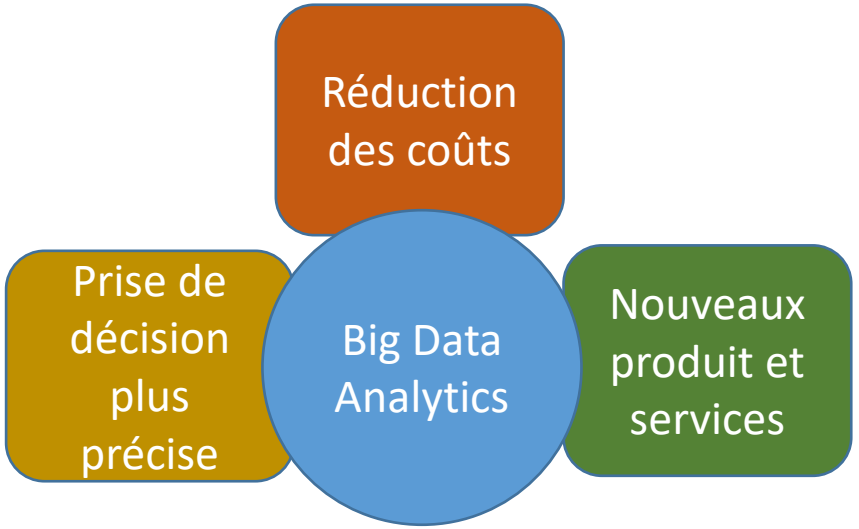
Levier d'amélioration par l'analyse des données

Le Big Data Analytics consiste à examiner l'ensembles de la volumétrie de données afin de découvrir

- ✓ des schémas cachés,
- ✓ des corrélations inconnues,
- ✓ des tendances du marché,
- ✓ des préférences client

Résultat : aider les entreprises à prendre des décisions plus efficace et avec de la valeur.

Levier d'amélioration par l'analyse des données



La transformation des processus métier

S'assurer de la qualité des données est un moyen de limiter les risques

Il est recommandé que les données soient générées et traitées par les processus métier

Il est aussi recommandé d'identifier dans un premier temps les parties prenantes et les enjeux

L'urbanisation de la donnée (suivre l'évolution des données)

Pour urbaniser il faut gouverner ses données

Les impacts du Big Data

Le Big Data peut impacter l'entreprise de différentes manières

Gouvernance : quelles données inclure, et comment définir et assurer la gouvernance du Big Data ?

Planification : comment collecter les données? Comment l'analyser ? Pour quelles finalités ? Comment organiser les résultats ?

Utilisation : quelle infrastructure mettre en œuvre ? Une solution cloud serait la plus adaptée ?

Qualité des données : comment assurer la qualité des données traitée dans le cadre d'un certain nombre de domaines tels que la normalisation, l'harmonisation et la rationalisation ?

Vie privée : comment protéger la vie privée ?

RGPD

Le RGPD est le Règlement Général sur la Protection des Données, une nouvelle réglementation européenne entrée en vigueur le 25 mai 2018.

C'est le règlement no 2016/679, qui constitue le texte de référence en matière de protection des données à caractère personnel. Il renforce et unifie la protection des données pour les individus au sein de l'Union européenne.

Cette nouvelle loi a différents objectifs :

- Renforcer les droits des personnes ;
- Responsabiliser les acteurs traitant des données;
- Crédibiliser la régulation.

Données personnelles

Les données personnelles se répartissent dans trois catégories

- **Données fournies volontairement** : il s'agit des données créées et partagées de manière explicite par des individus (par exemple, les profils sur les réseaux sociaux).
- **Données observées** : il s'agit des données collectées lors du suivi d'actions effectuées par des individus (par exemple, les données de géolocalisation issues de l'utilisation des téléphones mobiles).
- **Données déduites** : il s'agit des données sur les individus résultant de l'analyse des données des deux premières catégories.

RGPD

Le RGPD est la nouvelle réglementation européenne qui va régir la manière dont les entreprises vont gérer les données personnelles qu'elles manipulent.

Les organismes concernés par le RGPD :

- Toutes les organisations réalisant des traitements de données à caractère personnel de citoyens européens sont concernées par le RGPD, quelle que soit leur localisation.
- Le traitement de données personnelles rassemble diverses notions : collecte de données, stockage, analyse… Et le secteur d'activité de l'entreprise ne compte pas, même si ce type de traitement ne constitue qu'une activité secondaire. Les sous-traitants sont également concernés au même titre.
- Les entités publiques comme les hôpitaux et les administrations ou collectivités sont également concernées.

RGPD

Ce qu'il faut retenir :

- Le principal critère d'application du RGPD est en réalité le ciblage des citoyens européens.
- quelle que soit son lieu d'implantation (UE ou hors UE)
- Le RGPD s'appliquera à partir du moment où des données personnelles de citoyens européens seront en jeu

RGPD : La Sanction

RGPD : La Sanction

DATE	TYPE D'ORGANISME	MANQUEMENTS PRINCIPAUX / THÈME	DÉCISION ADOPTÉE
06/01/2021	COMMERCE DE DÉTAIL D'OPTIQUE	Non-respect de l'exercice des droits des personnes Défaut de sécurité des données	Sanction pécuniaire de 250 000 euros et injonction sous astreinte
11/01/2021	SOCIÉTÉ DE DÉVELOPPEMENT DE SOLUTIONS INFORMATIQUES	Défaut de sécurité des données	Sanction pécuniaire de 75 000 euros
12/01/2021	MINISTÈRE	Licéité du traitement Absence d'étude d'impact Défaut d'information des personnes	> Rappel à l'ordre et injonction
03/06/2021	SOCIÉTÉ D'ÉDITION DE LOGICIELS APPLICATIFS	Manquement à l'obligation de traiter des données de manière licite	Sanction pécuniaire de 10 000 euros
14/06/2021	SOCIÉTÉ ÉDITANT UN SITE DE VENTES PRIVÉES DEDIÉ AU BRICOLAGE, AU JARDINAGE ET À L'AMÉNAGEMENT DE LA MAISON	Durées de conservation Défaut d'information des personnes Non-respect des demandes d'effacement des données Défaut de sécurité des données Consentement pour la prospection commerciale	> Sanction pécuniaire de 500 000 euros et injonctions
20/07/2020	ASSURANCE	Durée de conservation Défaut d'information des personnes	> Sanction pécuniaire de 1 750 000 euros
26/07/2021	SOCIÉTÉ SPECIALISÉE DANS LES BIOTECHNOLOGIES AGRICOLES	Défaut d'information des personnes Obligation d'encadrer les relations avec un sous-traitant	> Sanction pécuniaire de 400 000 euros
27/07/2021	PRESSE	Consentement des personnes (cookies)	> Sanction pécuniaire de 50 000 euros

https://www.cnil.fr/fr/les-sanctions-prononcees-par-la-cnil

Sommaire

1. Généralités
2. Les opportunités qu'offre le Big Data pour la DSI
3. **Open Data**
4. Cas d'usage du Big Data
5. Le stockage dans le Big Data
6. Les technologies du Big Data
7. Le traitement des données en Big data
8. Compétences autour du Big Data
9. Les étapes d'un projet Big Data

Donnée (Data)

1640s : première utilisation anglaise du mot "data"

1946 : "données" signifie "informations informatiques transmissibles et stockables".

1954 : première utilisation anglaise du terme "data processing".

Une donnée est une description élémentaire d'une réalité.

- par exemple une observation ou une mesure

Information

Information signifie
- « donner forme à »
- ou « se former une idée de ».

L'information désigne à la fois
- le message à communiquer
- et les symboles utilisés pour l'écrire

Codage de l'information

Exemples : Codage ASCII

Character	ASCII
a	97
b	98
c	99
d	100
e	101
f	102
g	103
h	104
i	105
j	106
k	107
l	108
m	109

Character	ASCII
n	110
o	111
p	112
q	113
r	114
s	115
t	116
u	117
v	118
w	119
x	120
y	121
z	122

Character	ASCII
A	65
B	66
C	67
D	68
E	69
F	70
G	71
H	72
I	73
J	74
K	75
L	76
M	77

Character	ASCII
N	78
O	79
P	80
Q	81
R	82
S	83
T	84
U	85
V	86
W	87
X	88
Y	89
Z	90

Character	ASCII
0	48
1	49
2	50
3	51
4	52
5	53
6	54
7	55
8	56
9	57

Codage de l'information

Exemples :

Code à barres

QR Code

Radio Frequency IDentification

Atelier

Chaque participant doit trouver un sujet Big Data à réaliser par un prestataire afin de créer une valeur.

- **Définir un sujet Big Data**
- **Les données récoltées (!!! RGPD !!!)**
- **La valeur de rentabilité**