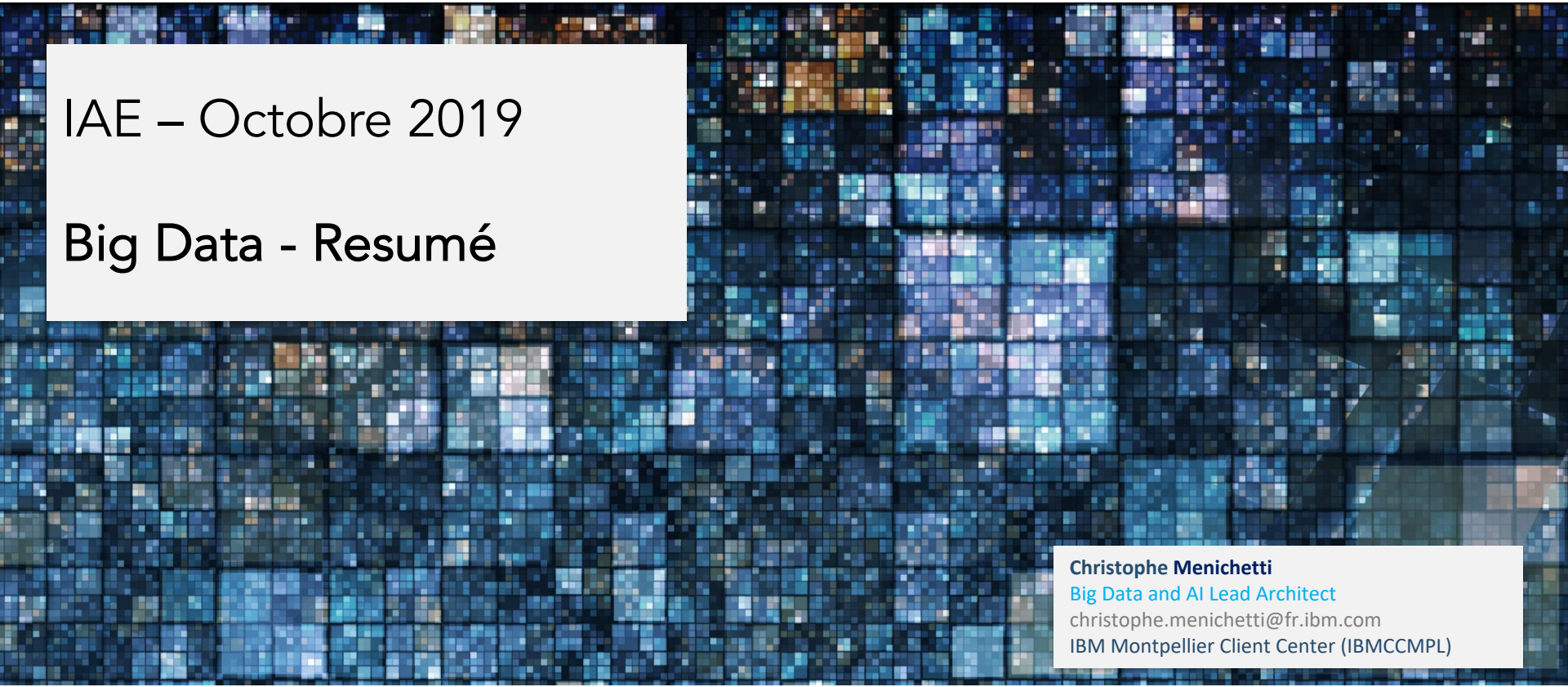


IAE – Octobre 2019

Big Data - Resumé

Christophe Menichetti
Big Data and AI Lead Architect
christophe.menichetti@fr.ibm.com
IBM Montpellier Client Center (IBMCCMPL)

AGENDA

1. De la BI au Big Data : Introduction
   - *Introduction à l'analyse de la donnees*
   - *Architecture et Solution Data Warehouse*
   - *Les nouvelles problematiques de la data (3V)*
   - *les objets Connectes (IoT)*
2. Usages et Technologies Big Data
   - *Nouvelle approche de stockage/traitement donnee - - -*
   - *l'ecosysteme Hadoop*
   - *Focus sur Map Reduce et Spark*
   - *Les differentes offres du marché*
3. Du Big Data vers la Data Science
   - *La donnee : le nouveau petrole*
   - *Le Machine Learning et Intelligence artificielle*
   - *Les differents Metiers du Big Data et roles clés*
   - *Focus sur le Data Scientist*
4. Architectures BI, Big Data et AI
   - *architecture On Premises*
   - *Le Cloud et solutions Data dans le cloud*
   - *Etudes de Cas*
   - **Synthese de tout le cours**

Data : the new oil

Transactions (ERP, CRM, ...)

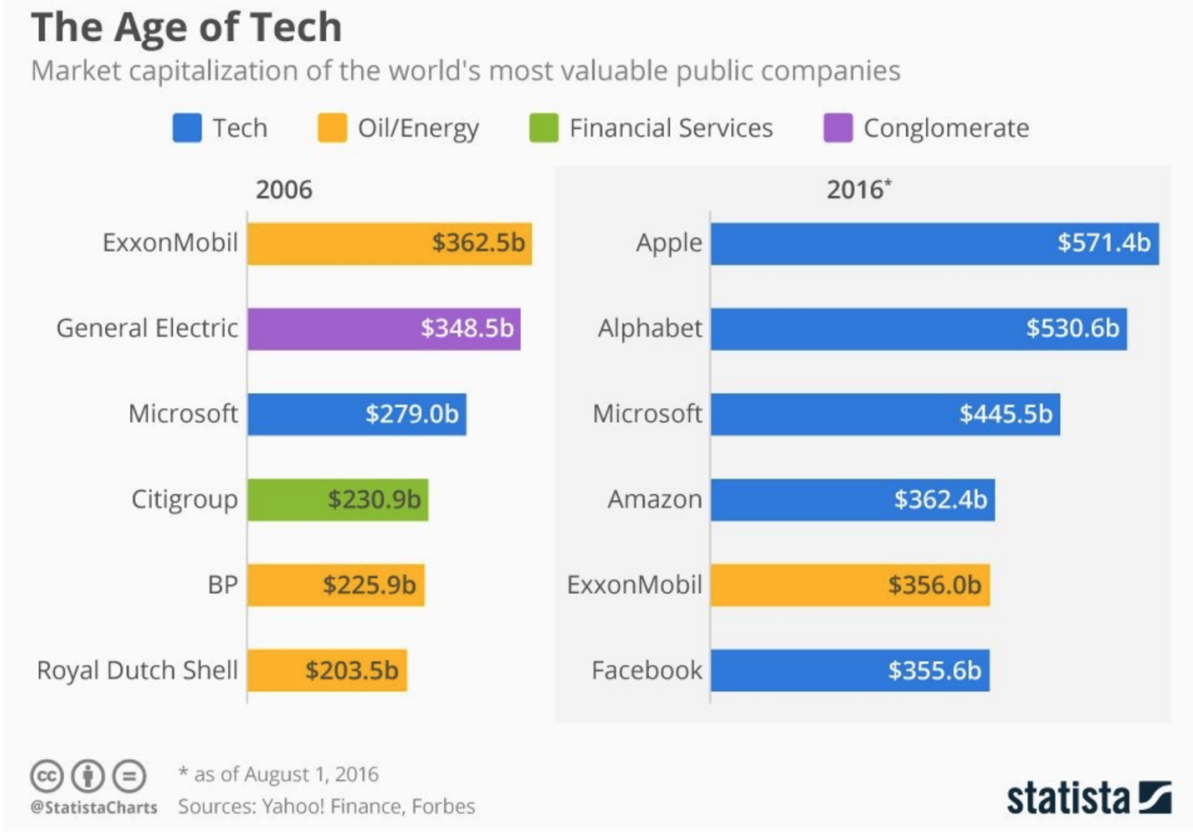
Graph

Image

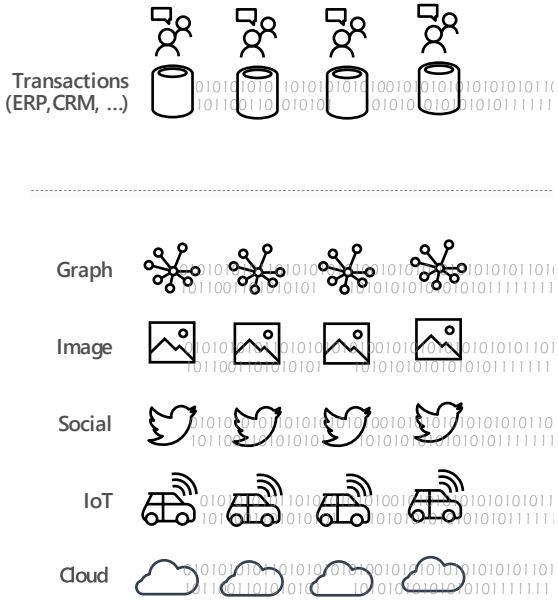
Social

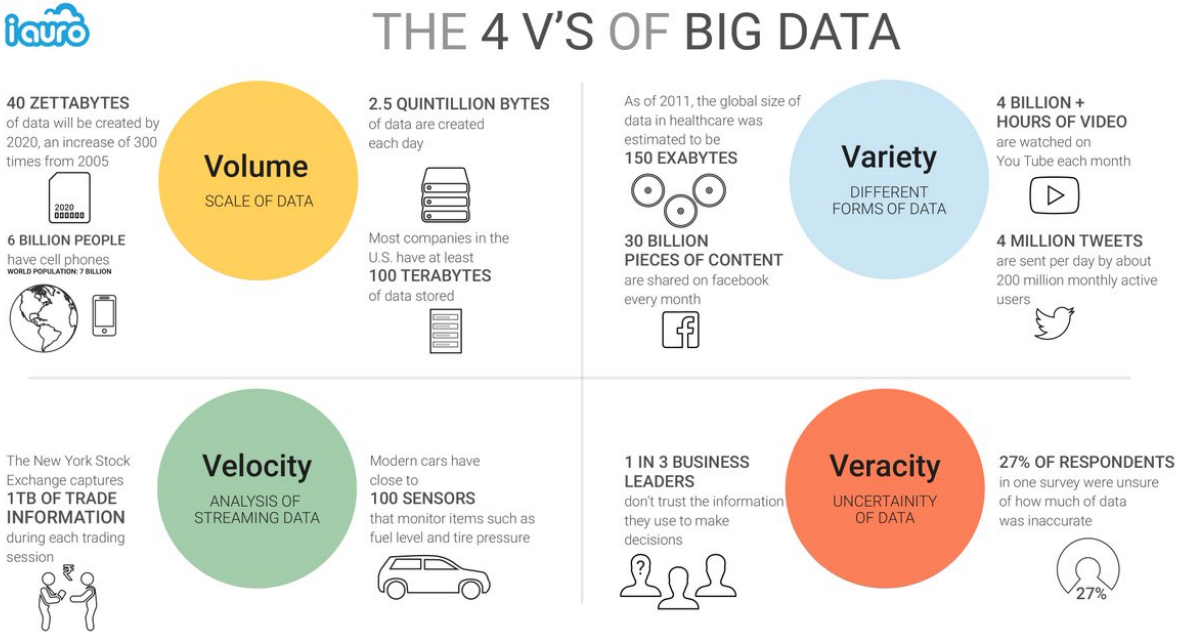
IoT

Cloud



Data : but new challenges





Reference : http://www.ibmbigdatahub.com/infographic/four-vs-big-data

Data : different types

Transactions (ERP, CRM, ...)

STRUCTURED = Static schema, easy to insert into RDBMS and request with Structured Query Language

---

Graph

Image

Social

IoT

Cloud

UNSTRUCTURED = Dynamic schema, time and resources consuming (hard) to insert into Relation DB

Data : how to store it ?

Transactions (ERP, CRM, ...)

Graph

Image

Social

IoT

Cloud

STORE

Data Warehouse

Big data

Design : Transactions | : Consistent | : easy to request

Data Warehouse

STRUCTURED

UNSTRUCTURED

Hadoop/NoSQL (Big Data)

Design : Flexible | : Scalable | : cheap for perf

Unstructured data

The university has 5600 students.
John's ID is number 1, he is 18 years old and already holds a B.Sc. degree.
David's ID is number 2, he is 31 years old and holds a Ph.D. degree. Robert's ID is number 3, he is 51 years old and also holds the same degree as David, a Ph.D. degree.

Semi-structured data

```
<University>
  <Student ID="1">
    <Name>John</Name>
    <Age>18</Age>
    <Degree>B.Sc.</Degree>
  </Student>
  <Student ID="2">
    <Name>David</Name>
    <Age>31</Age>
    <Degree>Ph.D. </Degree>
  </Student>
  ....
</University>
```

Structured data

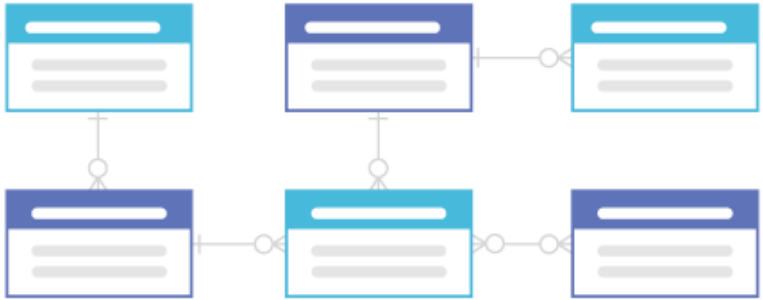
ID	Name	Age	Degree
1	John	18	B.Sc.
2	David	31	Ph.D.
3	Robert	51	Ph.D.
4	Rick	26	M.Sc.
5	Michael	19	B.Sc.

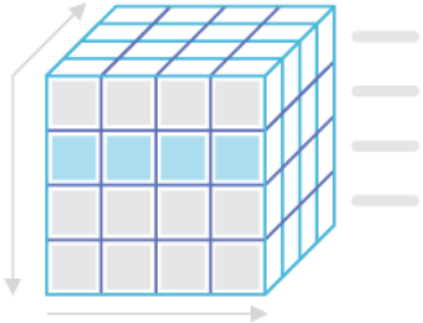
SQL

NO ONE FIT ALL

NoSQL

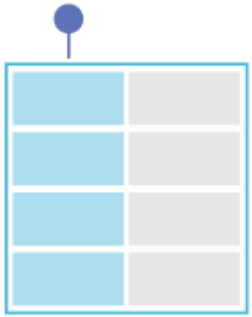
Relational Database Management Systems (RDBMS)



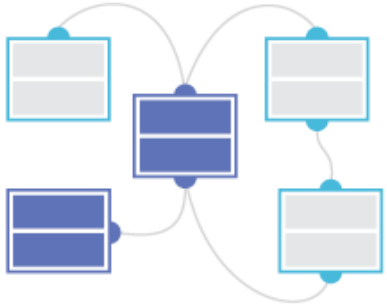


Online Analytical Processing (OLAP) Cube

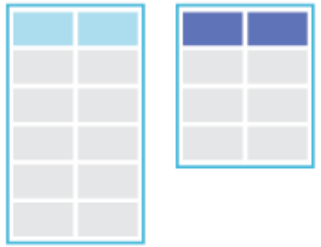
Key-Value



Graph

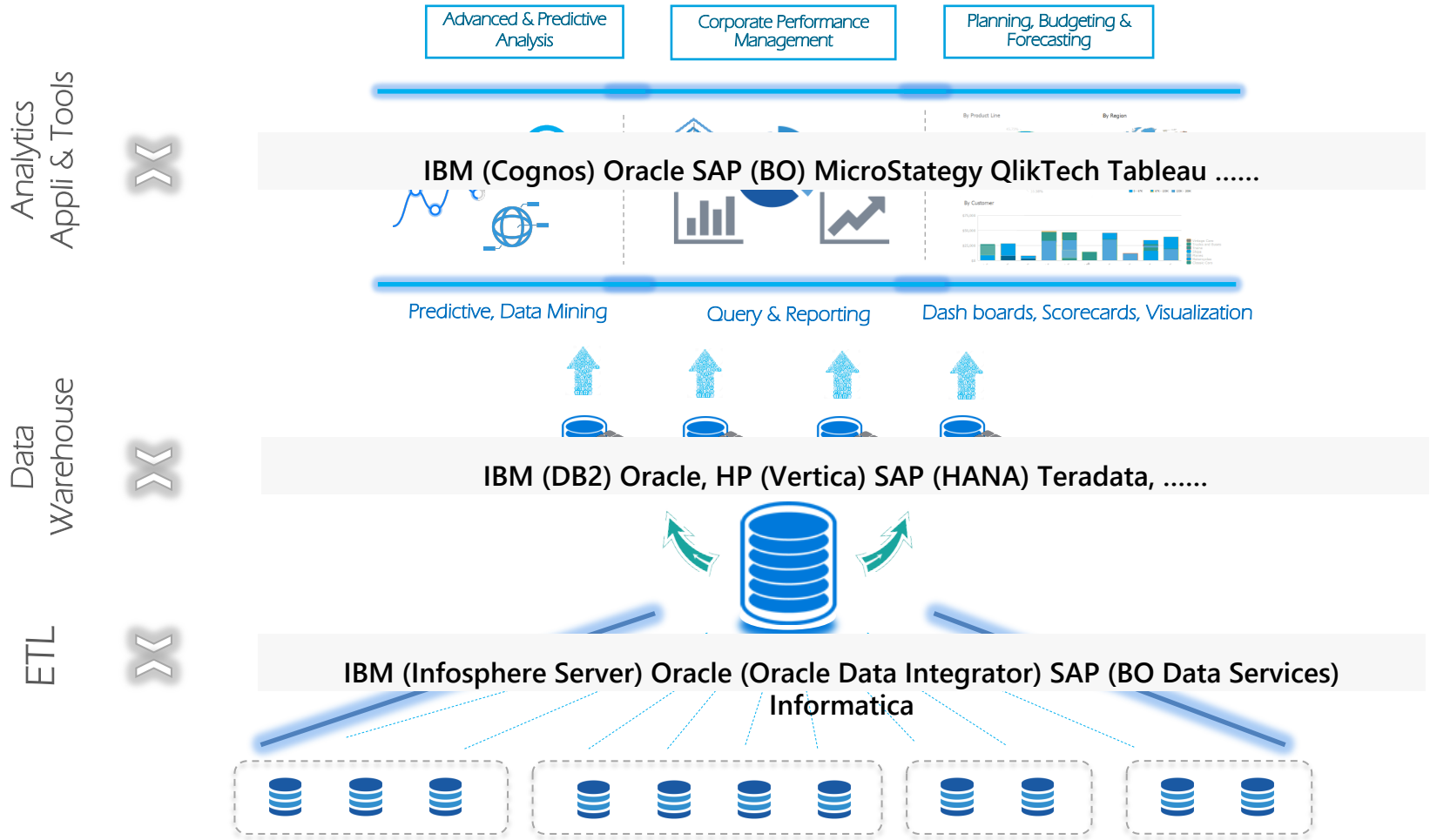


Document

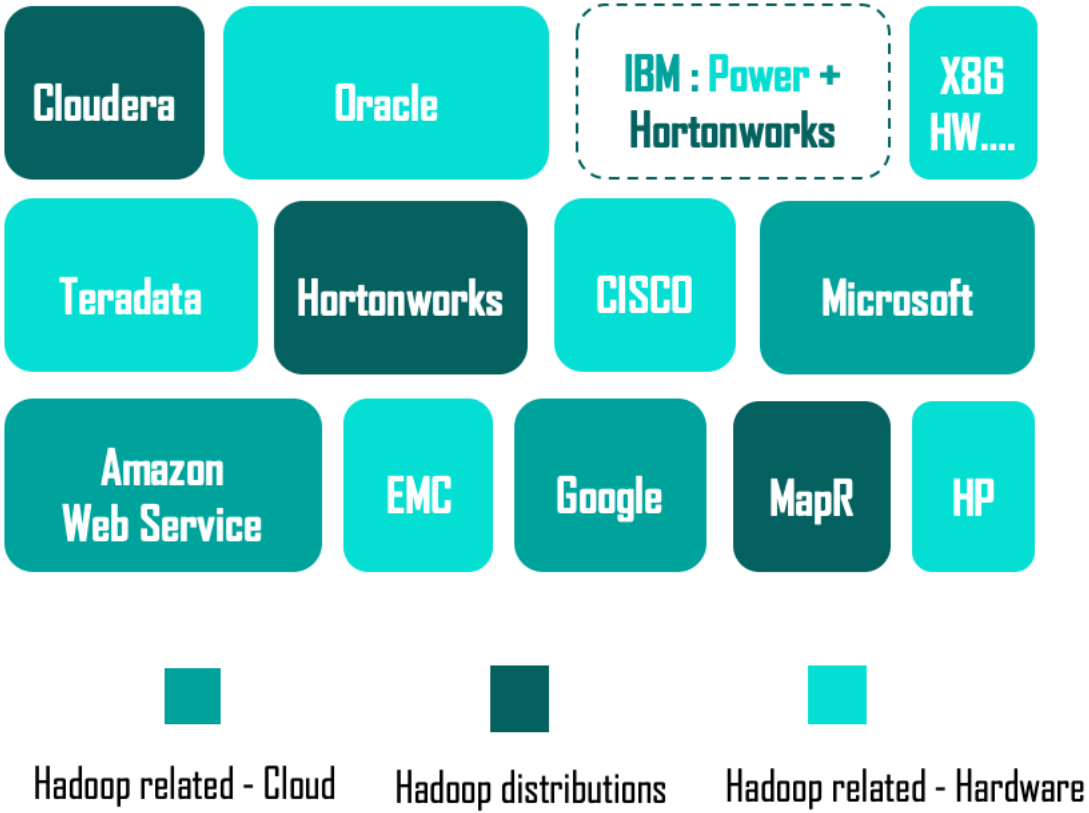


Column store

Les offres du marché BI (non exhaustif)



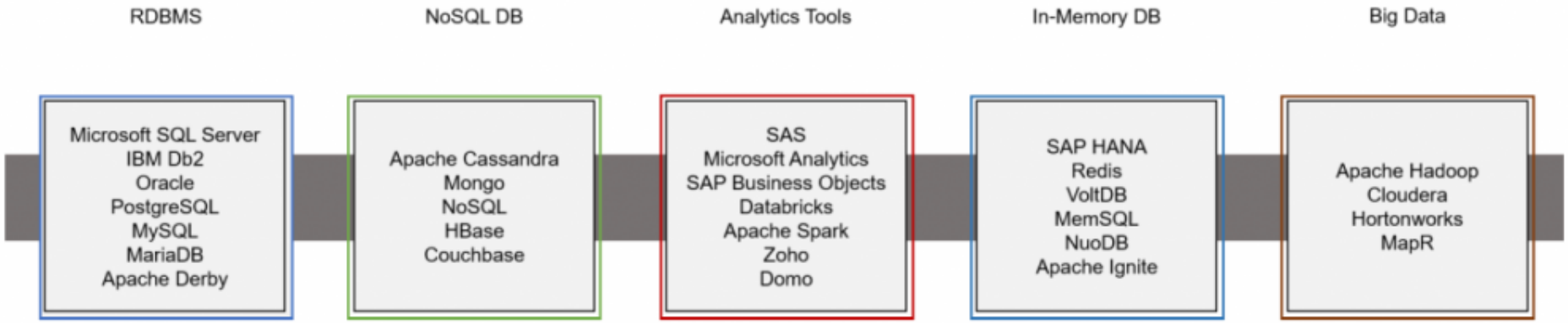
L'offre du marché hadoop (soft + hardware)



Non exhaustive !!!!

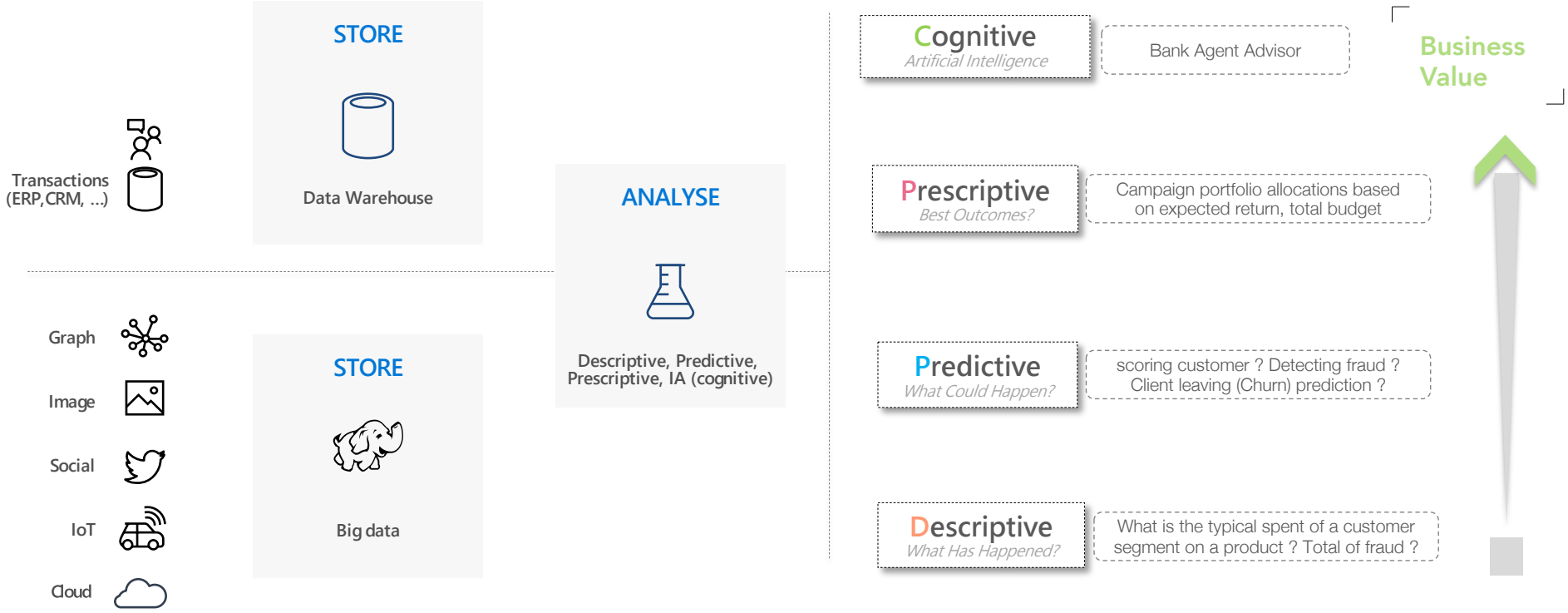
L'offre DATA STORE : vue globale

FIGURE 2 – THE DATA MANAGEMENT LANDSCAPE

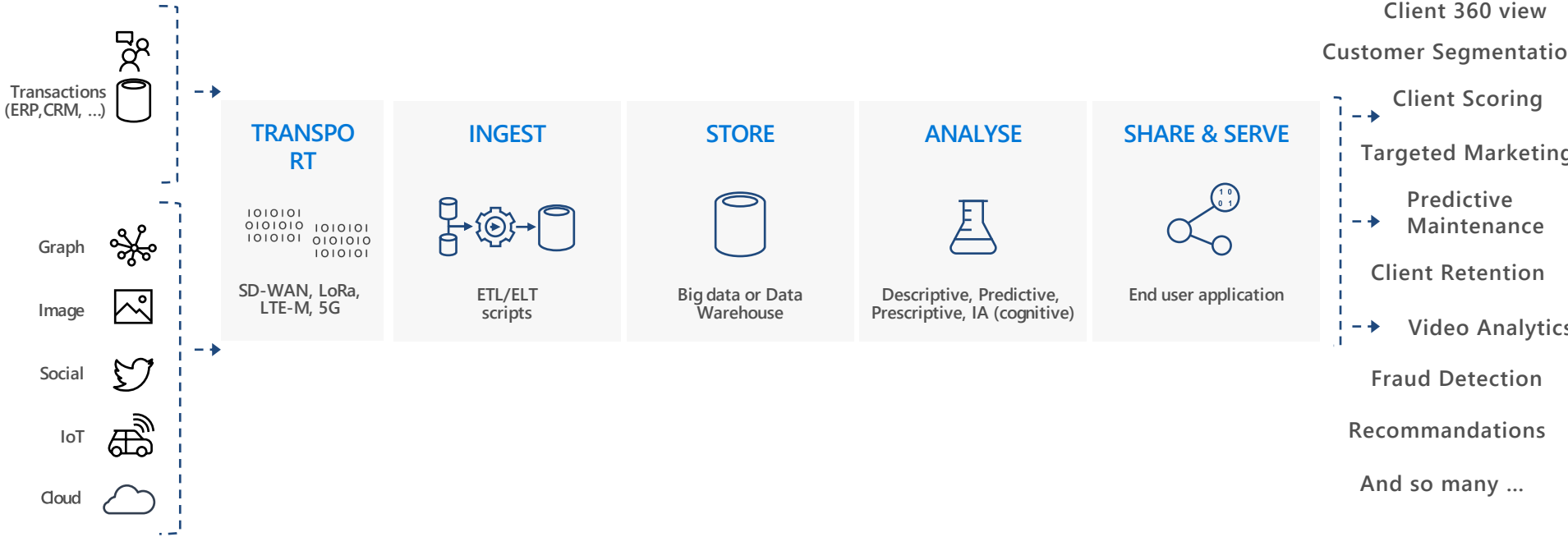


Source: Moor Insights & Strategy

Data : different analysis for different values



The Data's Journey



Le voyage de la donnée : TOUT RESUMER

Structured data = 20%

Transactions (ERP, CRM, ...)

Data is the new oil (competitive adv)

Graph

Image

Social

IoT

Cloud

3V Challenges

How much data store ? Which type of data ? What is your budget ? Your already existing solution ? Your skills ?

INGEST	STORE	ANALYSE	SHARE & SERVE
ETL/ELT scripts	Big data or Data Warehouse	Descriptive, Predictive, Prescriptive, IA (cognitive)	End user application

- BI + Reporting
- Advanced Analytics
- ML/DL/AI

The most important : the data – everything is about DATA (not about intelligence)

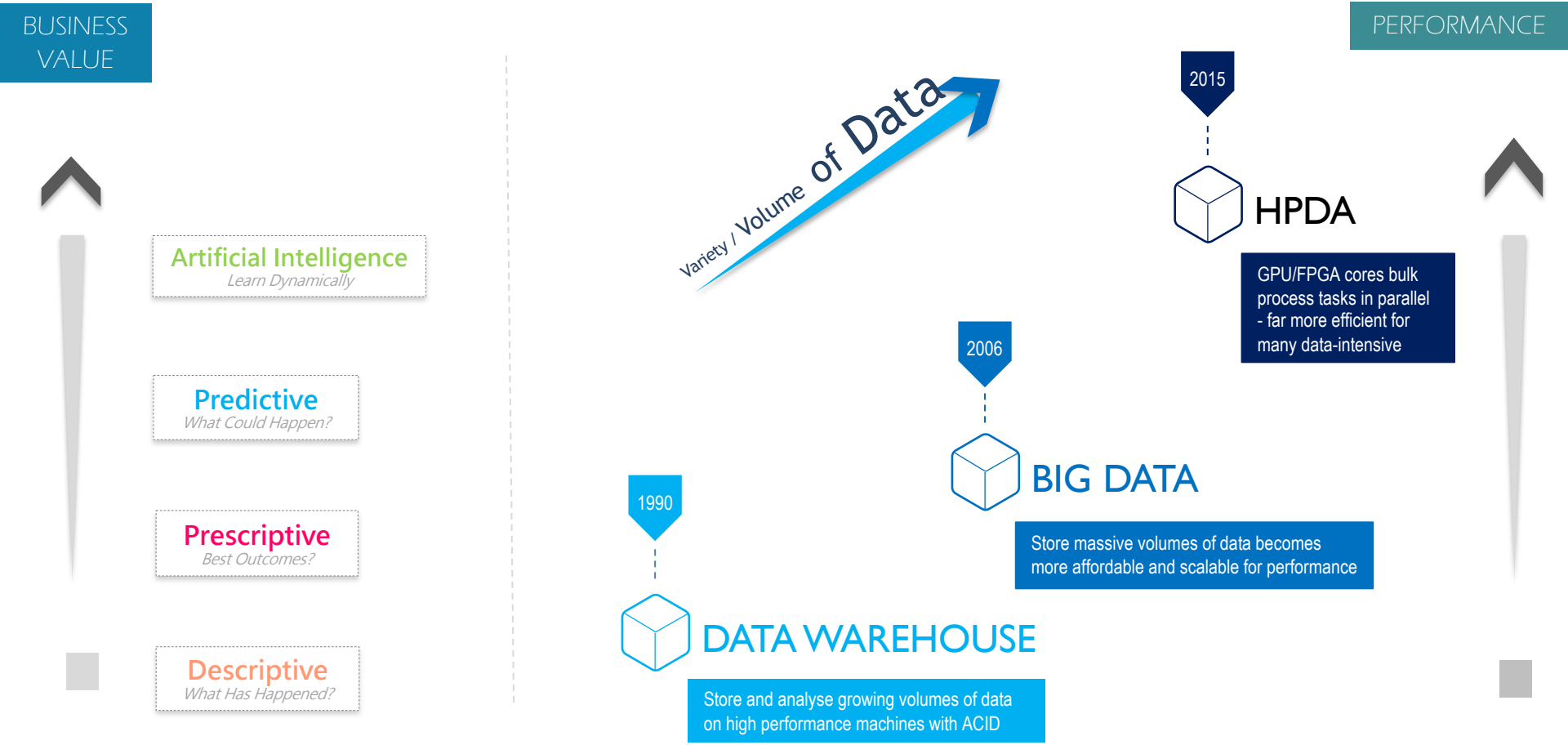
Schema on ingestion (write) or later (on read) ?

Data lake : federate both worlds (SQL and NoSQL- Hadoop and NOSQL DB) into one logical repository

Analysing data brings value to this entreprise, each solution could be deployed and used in stand alone mode but higher value if combined

AI is the current ultimate step in automating human tasks to help and advice humans, augmenting their intelligence to take better and faster decisions for business (or life)

L'analyse de la donnée : enjeux business et IT



Exemple de question

- Qu'est-qu'un entrepôt de données ?
- Quels sont les grands principes de la modélisation multidimensionnelle ?
- A quoi sert un ETL ?
- C'est quoi les 3V ?
- Qu'est-ce qu'une base noSQL ?
- Qu'est-ce que Hadoop ?
- Qu'est-ce que MapReduce ?
- Qu'est-ce qu'une appliance Big Data ?