
# ALTIC Big Data Stack

Tugdual SARAZIN
tugdual.sarazin@altic.org
Guillaume WEILL
guillaume.weill@altic.org

http://altic.org

# Altic

Born in 2004

Integrator company specialized in Business Intelligence

Open-Source oriented philosophy

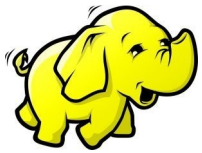
Provide training courses, systems integragtion...

Altic SARL | 95 avenue Victor Hugo | 93360 Neuilly-Plaisance
+33 (0)155 28 89 40 | www.altic.org | contact@altic.org

EXPERT BI, DATA INTÉGRATION
ET PORTAIL OPEN SOURCE

# Who we are

**Tugdual SARAZIN** : PhD student on BigDataMining



Spark

Scala

**Guillaume WEILL** : BI & BigData consultant


talend
*open data solutions

+ableau
SOFTWARE

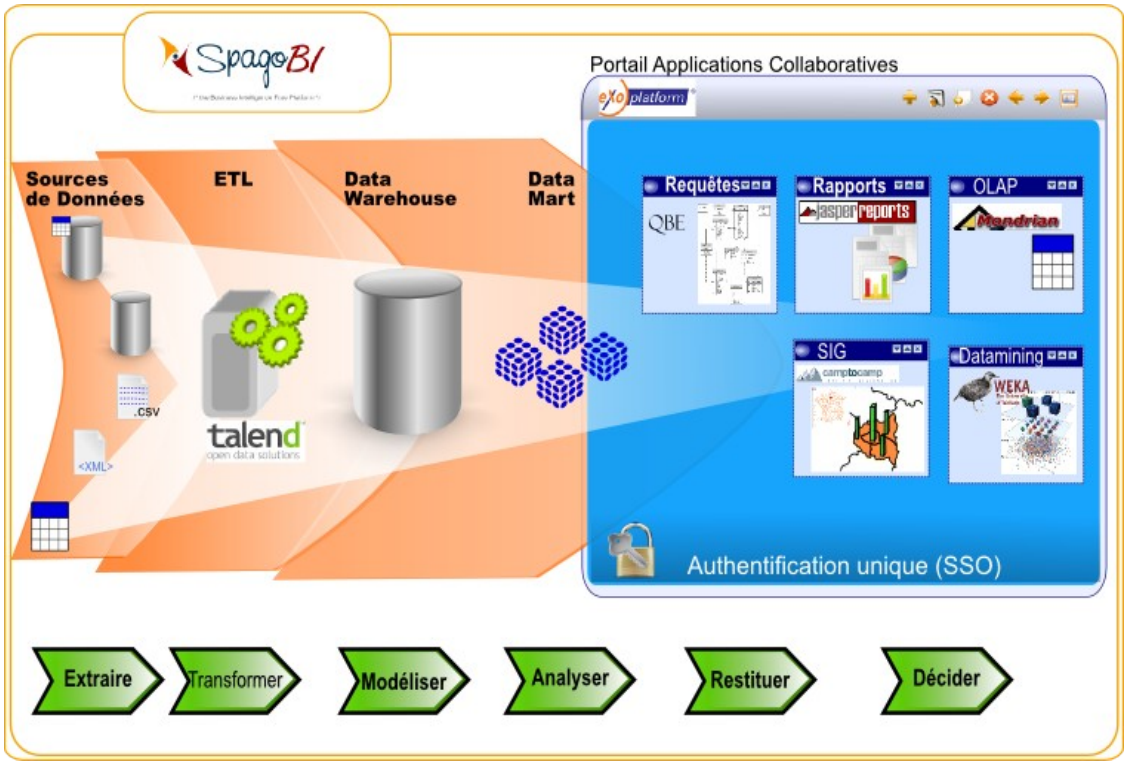
SpagoBI

Altic SARL | 95 avenue Victor Hugo | 93360 Neuilly-Plaisance
+33 (0)155 28 89 40 | www.altic.org | contact@altic.org

EXPERT BI, DATA INTÉGRATION
ET PORTAIL OPEN SOURCE

# Our historical tools

- ETL : Talend
- Reporting : JasperReports, Bi
- OLAP : Mondrian, Palo
- BI platform : SpagoBI




Altic SARL | 95 avenue Victor Hugo | 93360 Neuilly-Plaisance
+33 (0)155 28 89 40 | www.altic.org | contact@altic.org

EXPERT BI, DATA INTÉGRATION
ET PORTAIL OPEN SOURCE

# Our first Big Data project at Altic

- eFraudBox project (2010 – 2013)
  - Goal : predict frauds on Internet
  - Context :
    - Customer : GIE carte bancaire
    - European Research and Development project
    - Lot of industrial and academic partners
  - Data :
    - Type : Banking transactions
    - Volume : One GB per day

# How did we start our first BigData project ?

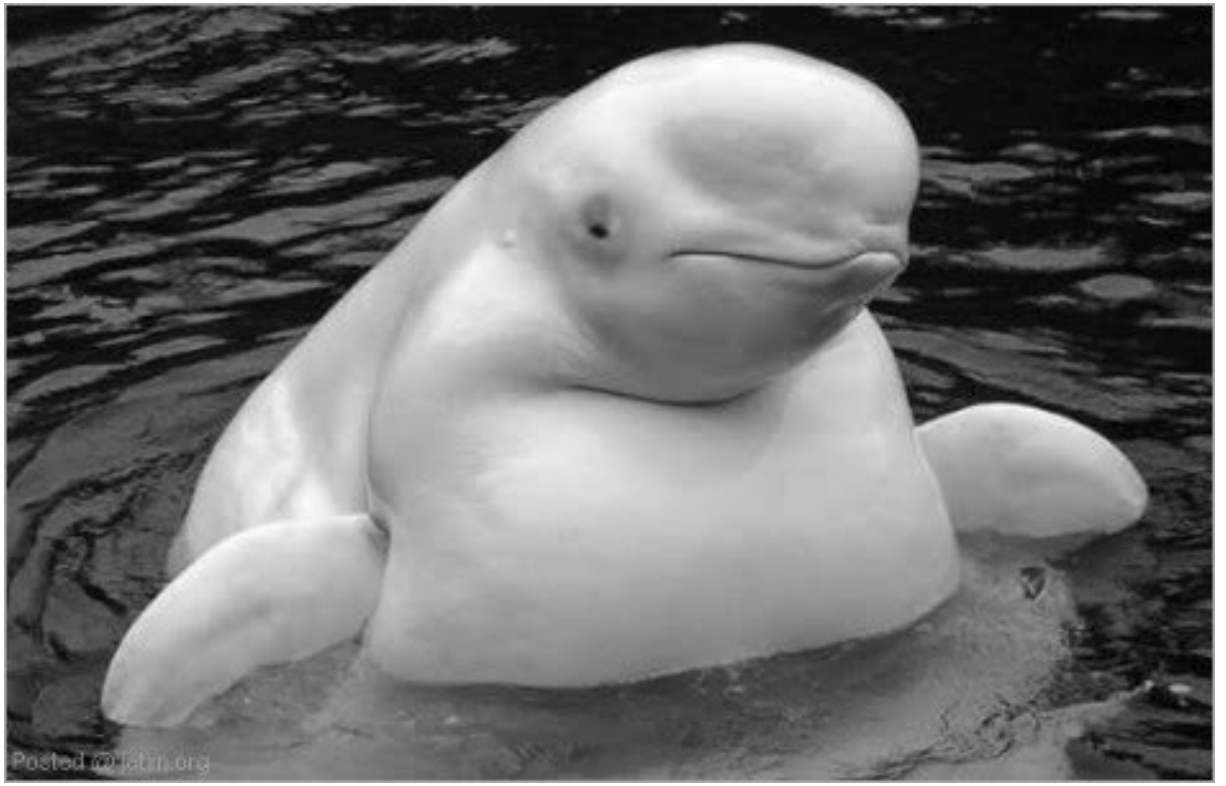
Altic SARL | 95 avenue Victor Hugo | 93360 Neuilly-Plaisance
+33 (0)155 28 89 40 | www.altic.org | contact@altic.org

EXPERT BI, DATA INTÉGRATION
ET PORTAIL OPEN SOURCE

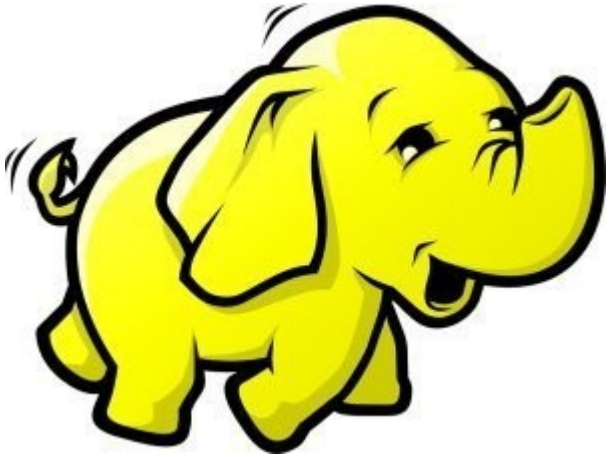
# We want to store and query data


?

# But we have too much data !



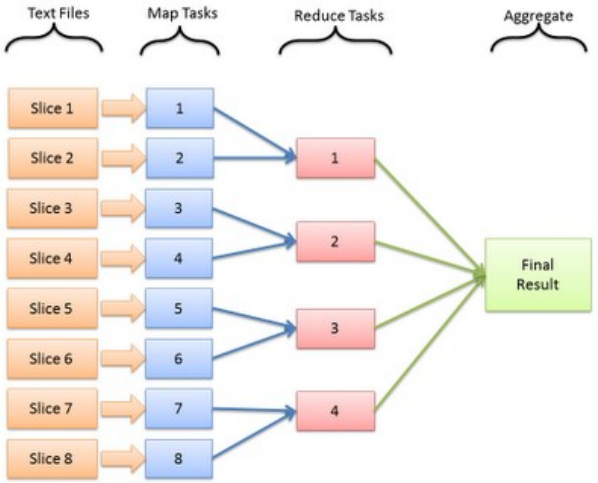
# Let's have a look at Hadoop ?



**Specs :**

- Distributed file system
- MapReduce processing
- OpenSource
- Infinite scale

# How do we query Hadoop ?



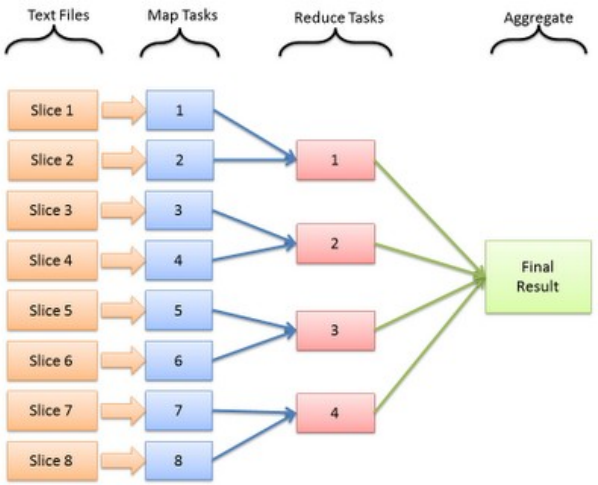


- Java
- Very optimised
- Very customisable

- Pig Latin
- Easy syntax
- Support unstructured data

- SQL like
- Easy development

# How do we query Hadoop ?





- Need to code evertything
- Why not ?
- We already know SQL !

# Ok, we have our storage, but how can we store data ?

- By using ETL for Big Data of course !

# Ok Hive is filled in with Big Data, but It's a little bit too slow to query...

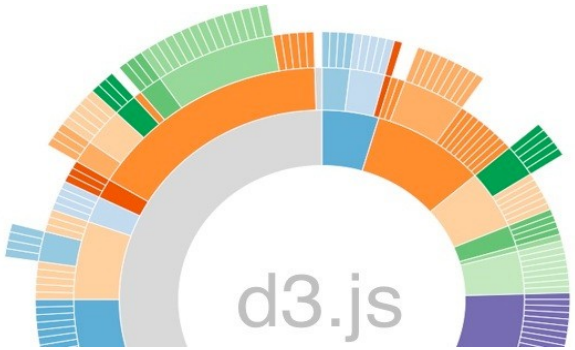
# Aggregate data

- Processing data with Hive and store results in analytical databases

# Ok, now we have our fast queryable datasets, but how can we visualize these ?

SpagoBI — To manage users and visualizations

tableau SOFTWARE — To quickly have a vision of your data




To go deeper in your visualizations

# BigData and Datamining : tMahout


= tMahout

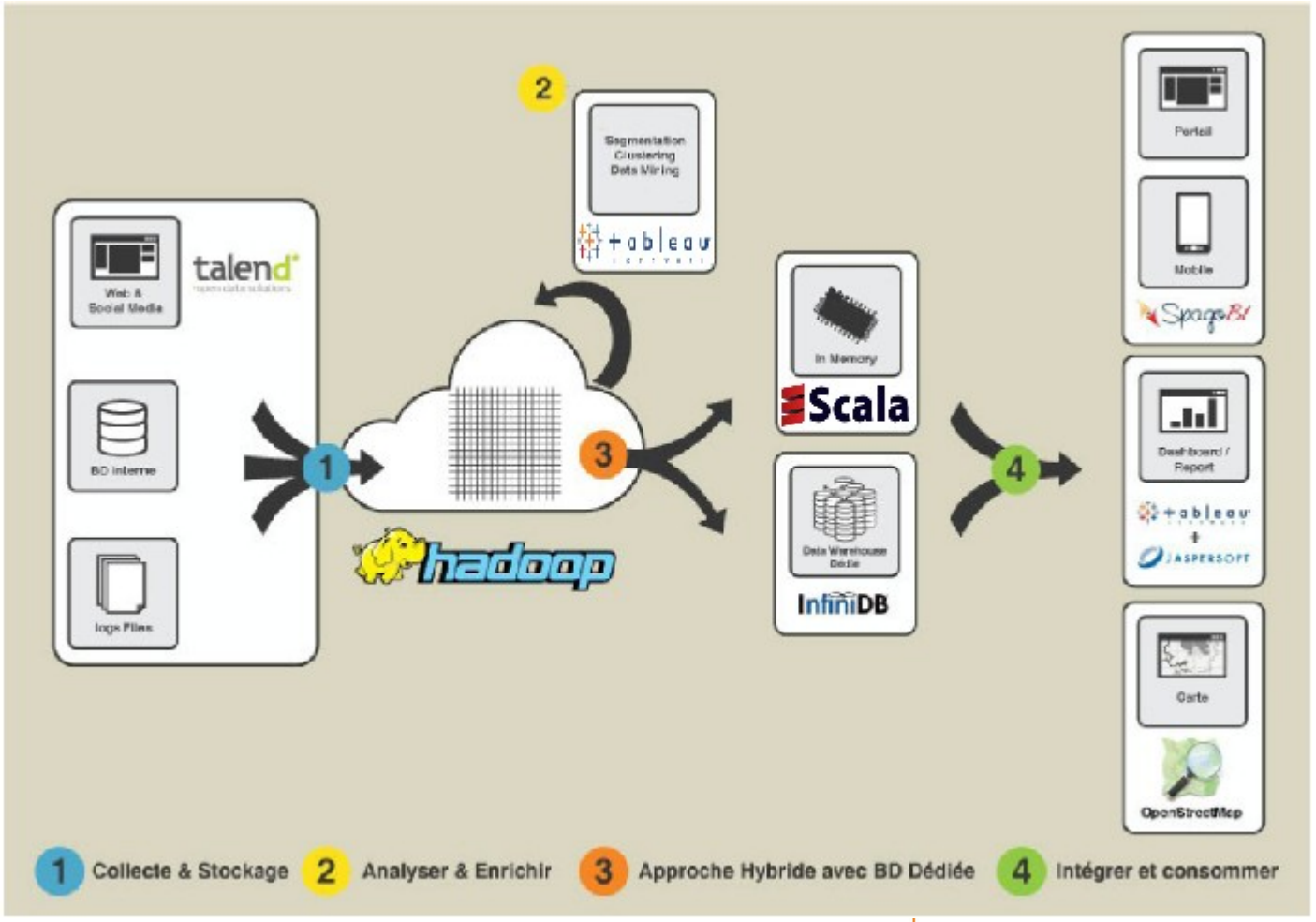
Altic SARL | 95 avenue Victor Hugo | 93360 Neuilly-Plaisance
+33 (0)155 28 89 40 | www.altic.org | contact@altic.org

EXPERT BI, DATA INTÉGRATION
ET PORTAIL OPEN SOURCE

# BigData and Datamining : Spark

- Spark : new InMemory data processing framework
  - Very appropriate for Machine learning
  - MLBase : Machine learning library
  - Spark-clustering : Implementation of SOM algorithm
  - Proof Of Concept : Analysis of mobile telecommunications

# We have now a Big Data stack !




Altic SARL | 95 avenue Victor Hugo | 93360 Neuilly-Plaisance
+33 (0)155 28 89 40 | www.altic.org | contact@altic.org

EXPERT BI, DATA INTÉGRATION
ET PORTAIL OPEN SOURCE

# BI & Big Data for Altic

- Eventually, we still do BI as usual
  - Tools evolve :
    - New storage and processing
    - We do not change our tools, THEY change for us
  - Fundamentals do not really change, only technologies do
    - Hadoop
    - Spark
    - ElasticSearch


# Questions

Thank you !

Tugdual SARAZIN
tugdual.sarazin@altic.org
Guillaume WEILL
guillaume.weill@altic.org

http://altic.org