SpagoBI and Big Data: 
Next Open Source Information Management suite

Monica Franceschini
SpagoBI Architect - SpagoBI Competency Center - Engineering Group
SpagoBI Suite now

The **business intelligence comprehensive suite**
All BI capabilities, many analytical areas, various engines

**Unique solutions**
Location Intelligence, Real-time BI and Mobile BI

**Suitable for any requirement**
Just-in-time BI, agile and sustainable developments

**100% open source, forever**
Pure open source model, no software lock-in

**A complete range of support services**
Pay-as-you-go support services
What Big Data are

**VALUE**
Extract value from information

**VOLUME**
Large amounts of data

**VARIETY**
Different types of structured and unstructured data

**VELOCITY**
Data that needs to be analyzed quickly
SpagoBI Basic concepts

**Data Source**: Data storage (DWH, Database)

**Data Set**: query over the Data Source to get a result that can be used across SpagoBI documents

**Engines**: produce different documents

**Documents**: different Analysis that can use the same data sets
SpagoBI and Big Data now

Talend for Big Data offers a BI ETL approach

Load, extract and improve data located in these diverse data sources and to govern big data projects centrally

Talend integration enables the NO-Sql Data Management

Behavioural model
Cross services
Data Set and Data Sources evolution

Use of Data Source/Data Sets as connections to new Data Storages

Data Sets as PLUGINS

Analytical DBMS
- Netezza
- Vertica
- InfoBright
- Teradata
- Vectorwise

NoSQL
- MongoDB
- HBase
- Cassandra

Hadoop
- Hadoop
- Hortonworks
- Cloudera

Cross services

Behavioural model
SpagoBI - Big Data Scenario #1

Scenario #1

- **Data Sources**: distributed data storages for Big Data such as Hadoop HDFS, HBase, Hive
- **Data Sets**: with Query Languages depending on the data source, such as HQL, Cloudera Impala, Pig

Not Real Time Analysis
SpagoBI - Big Data Scenario #2

Scenario #2
Complex Event Processing Engine CEP
- **Data Source** populated with Storm/S4/Druid
- **Data Set** with Esper (EPL query language)

Real Time Analysis

BYTES STREAM

Console document
Advanced platform for the monitoring of heterogeneous services and applications and the historical analysis of data
Scenario #3

- **Data Sources**: Storages for RDF-semantic format
- **Data Sets**: Interrogation with SPARQL syntax or others, depending on the data source (HBase, Hadoop...)

Not Real time/ Real Time depending on the storage

Linked Data Analysis
SpagoBI & Big Data – the complete view

BIG DATA

BYTES STREAM

SEMANTIC
Visit SpagoBI website:
www.spagobi.org

Contact us:
spagobi@eng.it

Download SpagoBI from OW2 Forge:
http://forge.ow2.org/
or visit us at Engineering Group's stand