Open Source BI Platforms: a Functional and Architectural Comparison

Matteo Golfarelli

DEIS – University of Bologna

Agenda:

1. Introduction
2. Conduct of the Comparison
3. Platforms description
4. Discussion

Open Source BI Software

- BI market is historically dominated by commercial vendors
  - Microstrategy
  - Oracle BI suite
  - Microsoft BI
  - …
- Only recently Open Source solutions appeared as single tools first, and as complete platforms later

An OS BI Platform provides a full spectrum of BI capabilities within a unified system that reduces the overhead for the development and management of each application, and lets the user feel like she was using a single BI solution

Commercial platforms are commonly considered superior to OS ones, that, on the other hand, can evolve faster.

Conduct of the comparison

- This work compare three different OS BI platforms
  - JasperSoft
  - Pentaho
  - SpagoBI

  the versions considered are those released by December 31 2008

- The outcome of the analysis is the fusion of our independent analysis and the testing and evaluation activities of three consulting firms specialized in BI projects.

  1. We initially defined an evaluation grid describing in details the aspects to be investigated.
  2. The resulting grid was shared with the consultant firms and further discussed and integrated.
  3. Each consultant firm carried out one or more porting of real projects previously implemented through commercial BI suites.
  4. The compiled grids were finally shared and discussed with the other participants.

Top level of the comparison grid

- **Non-technical:** platform philosophy, type of licensing and availability of enterprise editions.
- **Architectural:** in terms of the global framework, modules and their relationships, programming languages and supported operational systems.
- **Functional:** in terms of functionalities provided natively by the platforms or made available to the users through the integrated BI tools.
- **Meta-data:** in terms of expressiveness, completeness, standardization and level of reusability.
- **Security:** in terms of functionalities provided for authentication and profiling of the users, interfaces to external authentication systems and secure data transmission.
- **Usability:** both from the user viewpoint, in terms of level of transparency in using the different tools, and from the developers' and system administrators' viewpoint in terms of complexity of installation and administration as well as development of applications, quality of manuals and forums.

Non-technical aspects: open source model

- Commercial open source: provides for separate product releases.
  - community edition meets the user's basic needs and it is completely free
  - enterprise edition of the product can be purchased and it usually includes enhanced features as well as support and training services.

  Jasper and Pentaho fit into this model.

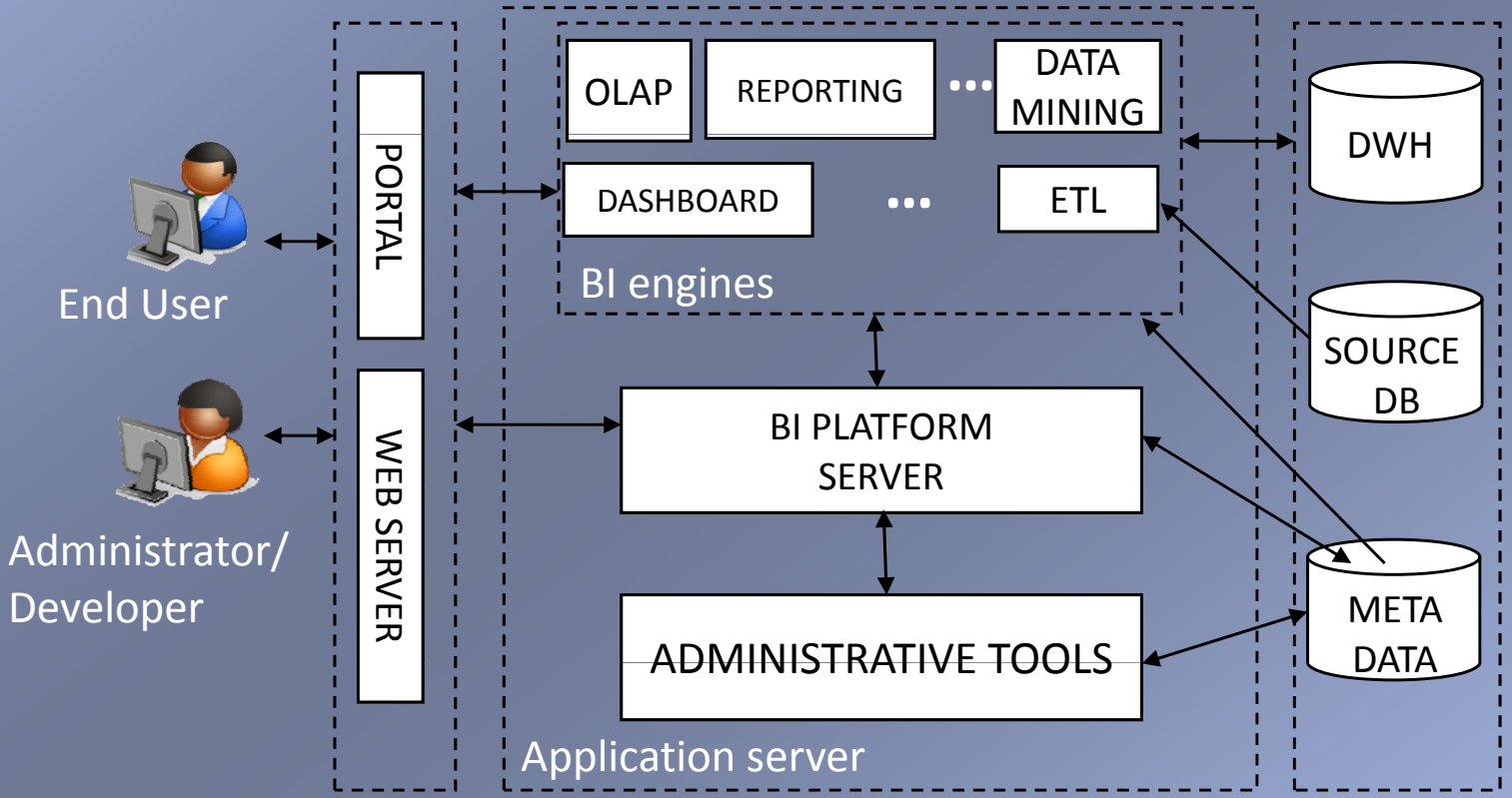
- Free and Open Source Software (FOSS): the product is completely free, no enterprise solution is available, thus all the functionalities are available to the community for free. SpagoBI fits into this model.

Non-technical aspects: plugging

- Integration: a software interface is defined in order to control and to exploit module functionalities directly and transparently through the platform.
  - The intellectual property of the software does not change, and the original developers remain in charge of maintaining and evolving the module.
  - SpagoBI is strictly based on integration
- Acquisition: the intellectual property of the software is acquired and the original project terminated.
  - The buyer will be in charge of maintaining and evolving the module.
  - Pentaho often has recourse to acquisition (e.g. Pentaho ETL comes from the Kettle project)
- Technological partnership: stands in the middle between integration and acquisition.
  - The original project remains alive and it is maintained by the original developers.
  - The partner that incorporates the module influences its evolution and collaborates to its maintenance. The module usually appears with a different name in the new platform.
  - Jasper mainly exploits partnerships (e.g. JasperETL was developed through a partnership with Talend that still maintains Talend Open Studio).

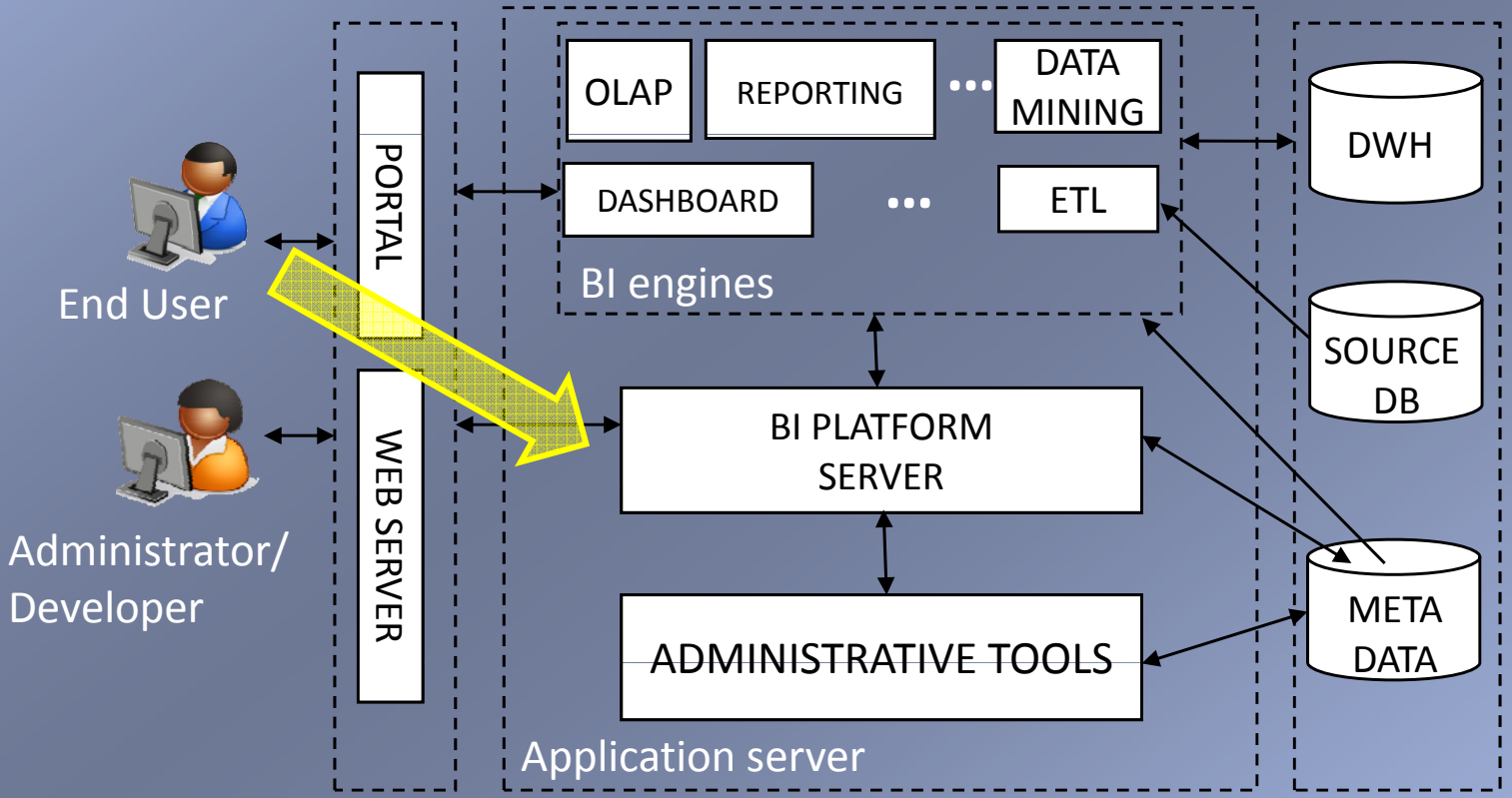
Architectural aspects

- The platforms adopted the same architecture
- OS BI platforms are developed using Java since the modules they rely on are based on this technology.



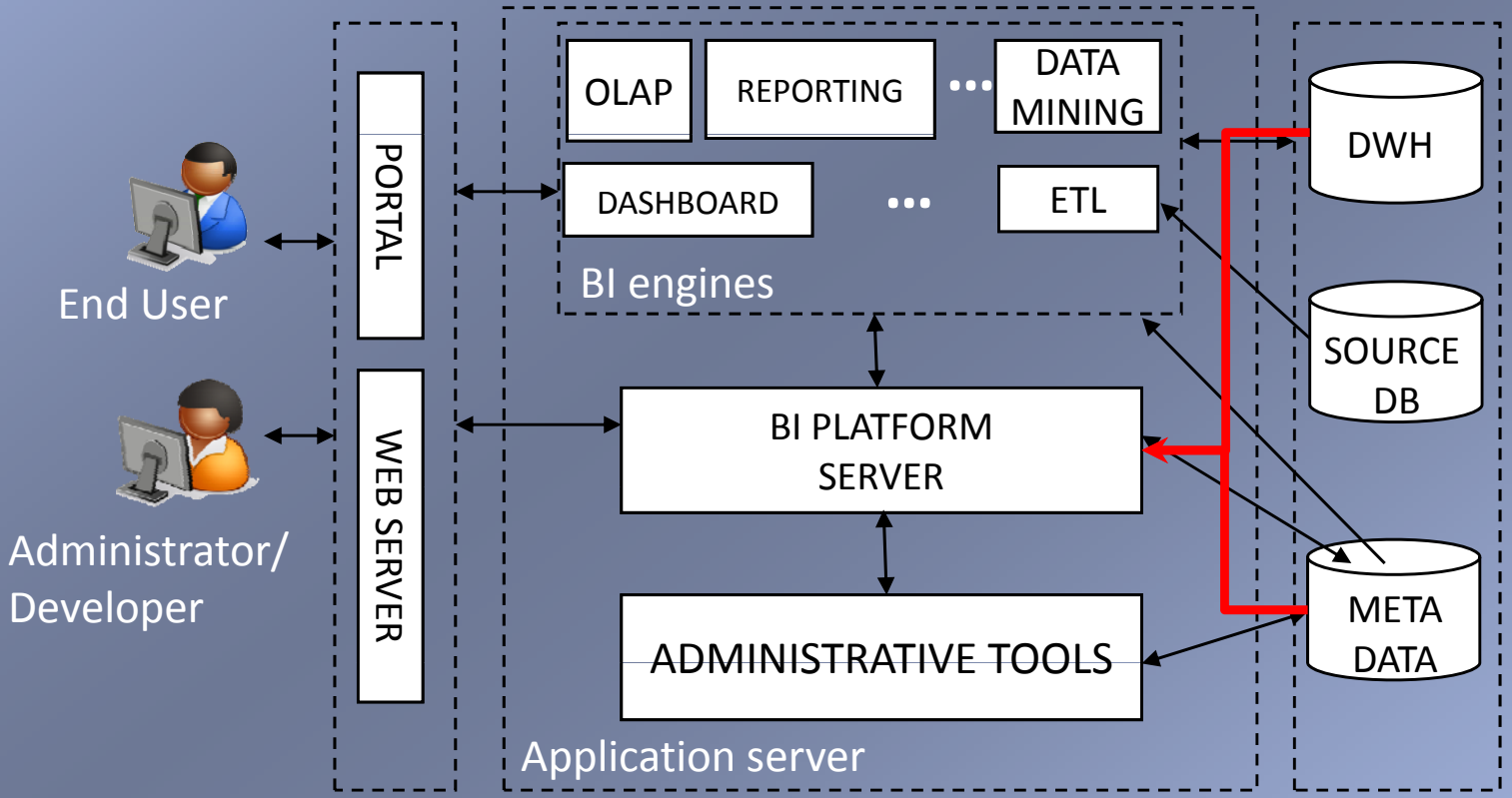
Architectural aspects

- The platforms adopted the same architecture
- OS BI platforms are developed using Java since the modules they rely on are based on this technology.



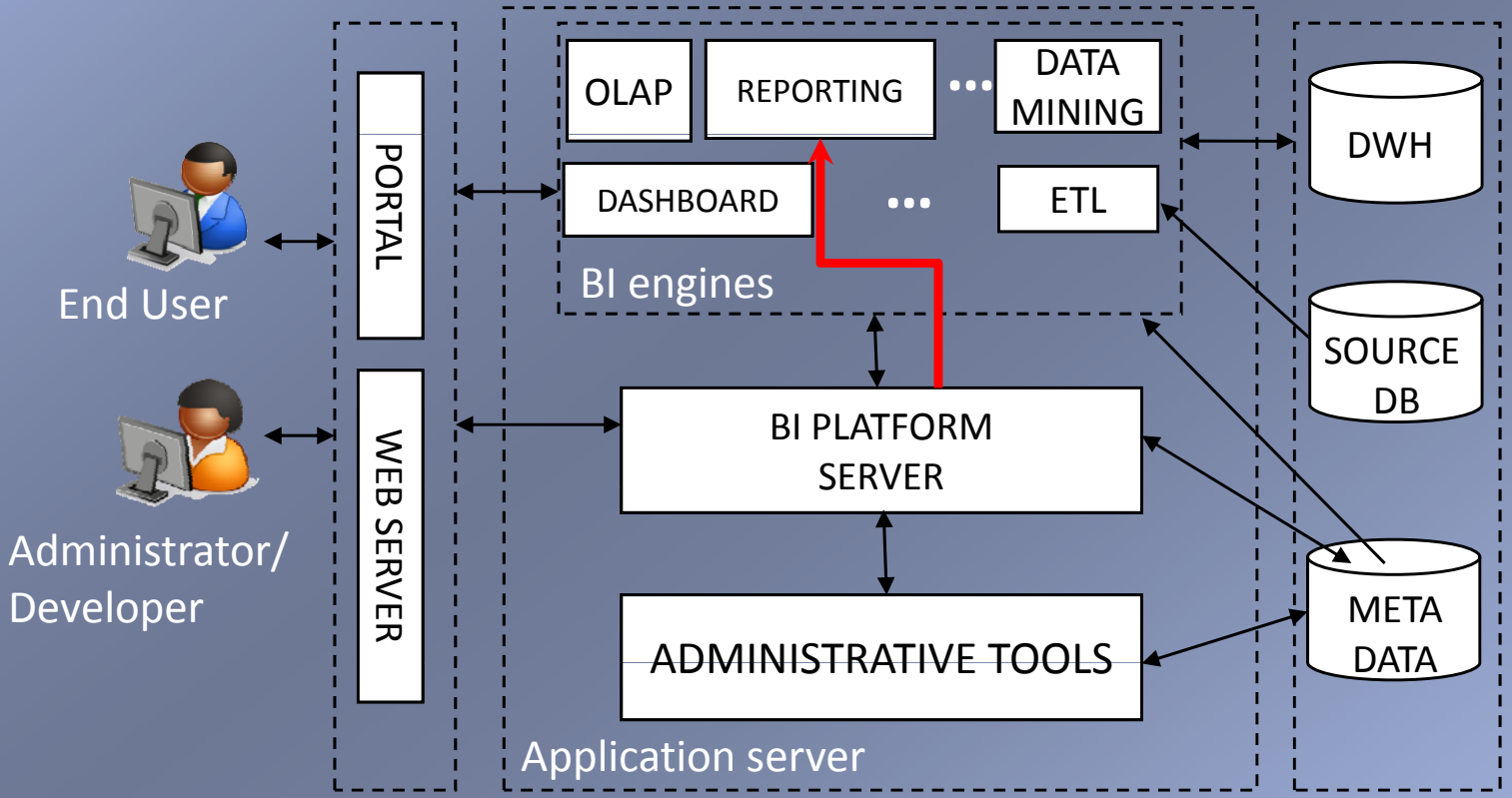
Architectural aspects

- The platforms adopted the same architecture
- OS BI platforms are developed using Java since the modules they rely on are based on this technology.



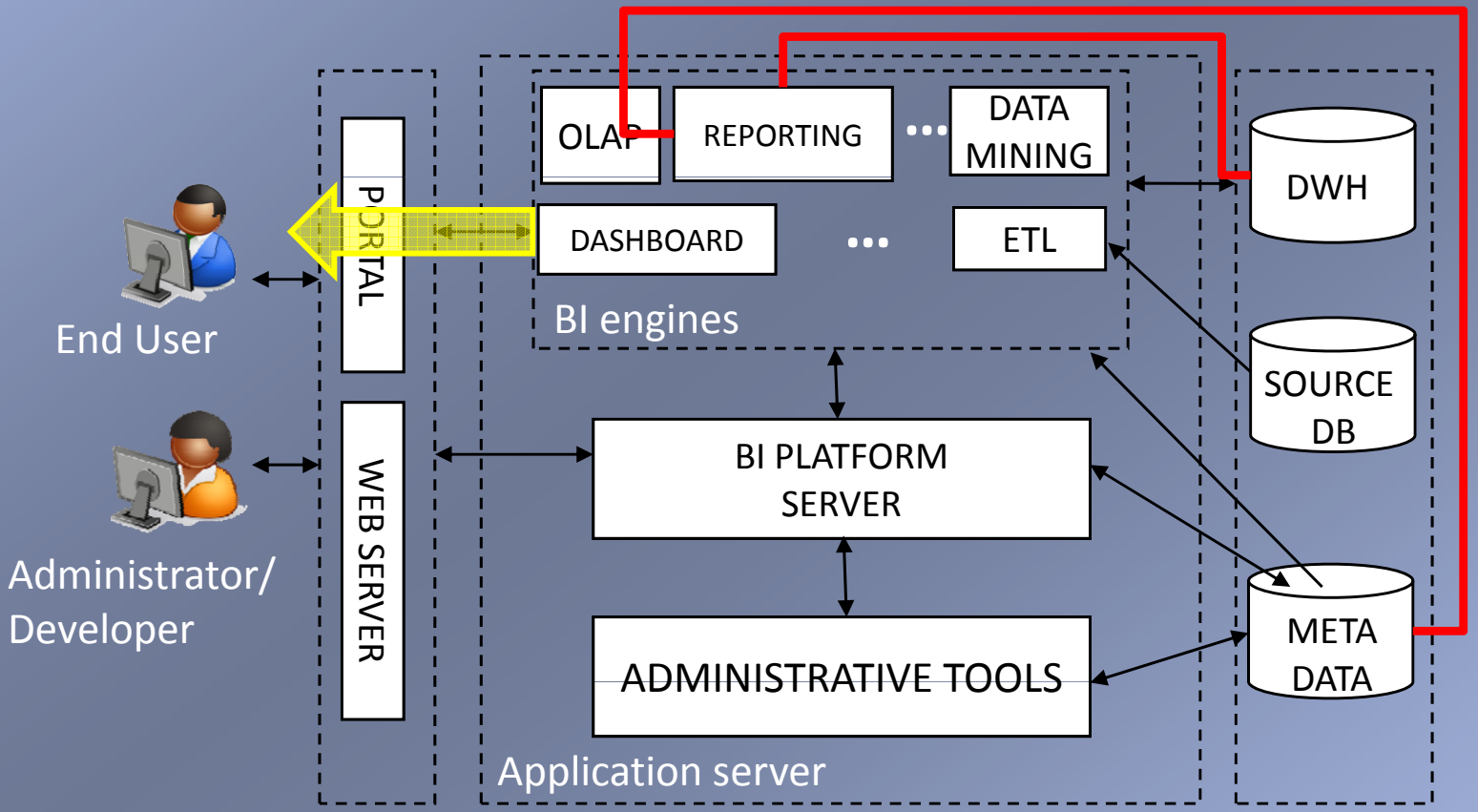
Architectural aspects

- The platforms adopted the same architecture
- OS BI platforms are developed using Java since the modules they rely on are based on this technology.



Architectural aspects

- The platforms adopted the same architecture
- OS BI platforms are developed using Java since the modules they rely on are based on this technology.



Architectural aspects: modules

- Many of the modules are shared
- Some of them are evolutions of a different open source project
- Some modules are standard de facto within BI OS (Mondrian, JPivot , Weka)

Modules	JasperSoft	Pentaho	SpagoBI
Application Server	JBoss	JBoss	JBoss
Authentication and user profiling	Acegi	Acegi	Integrated in eXo Portal
Collaboration	-	-	Dossier
Dashboard	JfreeChart	JfreeChart	Openlaszlo
Data Mining	-	Weka	Weka
DBMS	MySQL, Oracle, SQL Server, PostgreSQL, etc.	MySQL, Oracle, SQL Server, PostgreSQL, etc.	MySQL, Oracle, SQL Server, PostgreSQL, etc.
ETL	JasperETL	Pentaho Data Integration	Talend Open Studio
Geo-referenciation	Google Maps	Google Maps	GEO
Job Scheduler	Quartz	Quartz	Quartz
OLAP	Mondrian&Jpivot	Mondrian&Jpivot	Mondrian&Jpivot
Portal	Liferay	JBoss Portal	ExoPortal, Liferay
Query by Example	-	-	Hibernate
Reporting	JasperReport	Pentaho Report Designer, JasperReport, BIRT	JasperReport, BIRT
Single sign on	Acegi	CAS	CAS
Web Server	Tomcat	Tomcat	Tomcat

Metadata

- Within a BI platform, metadata largely determine the behavior it can exhibit.
- Metadata necessary to specific BI functionalities are usually created outside the platforms by editing an XML file or by exploiting simple graphical tools.
- Although, they model the same information, metadata belonging to different engines are differently coded and cannot be reused. This obviously affects development and maintenance negatively.
- Although all three platforms declare that their metadata are CWM-compliant no interoperability tools have been released yet!

Functional aspects

- SpagoBI community edition overcomes Pentaho and Jasper that make available many of the advanced features only in the enterprise editions.
- All the platforms allow secure data transmission as well as user authentication, while they offer pretty different functionalities for user profiling.

Functionalities	SpagoBI	Pentaho	Pentaho Ent. Ed.	Jasper	Jasper Ent. Ed.
Activities scheduling	√	×	√	×	√
Ad-hoc reporting	×	×	√	×	√
Auditing	√	×	√	√	√
Collaborative BI	√	×	×	×	×
Data Mining	√	√	√	×	×
Dashboard	√	√	√	×	√
Document export	√	√	√	√	√
ETL	√	√	√	√	√
Geo-referenced analysis	√	√	√	×	√
OLAP	√	√	√	√	√
Query by Example	√	×	×	×	×
Report validation workflow	√	×	√	×	×
Reporting	√	√	√	√	√
User profiling	√	×	√	×	√

Functional aspects

- SpagoBI community edition overcomes Pentaho and Jasper that make available many of the advanced features only in the enterprise editions.
- All the platforms allow secure data transmission as well as user authentication, while they offer pretty different functionalities for user profiling.

Functionalities	SpagoBI	Pentaho	Pentaho Ent. Ed.	Jasper	Jasper Ent. Ed.
Activities scheduling	√	×	√	×	√
Ad-hoc reporting	×	×	√	×	√
Auditing	√	×	√	√	√
Collaborative BI	√	×	×	×	×
Data Mining	√	√	√	×	×
Dashboard	√	√	√	×	√
Document export	√	√	√	√	√
ETL	√	√	√	√	√
Geo-referenced analysis	√	√	√	×	√
OLAP	√	√	√	√	√
Query by Example	√	×	×	×	×
Report validation workflow	√	×	√	×	×
Reporting	√	√	√	√	√
User profiling	√	×	√	×	√

Community vs Enterprise editions

- Differences are not only in terms of functionalities available to the users but also in terms of utilities for administrators and developers:
  - Improved administration consoles: the improvement is particularly relevant in Pentaho where the Enterprise console fills the gap with Jasper as concerns usability and functionalities.
  - Wizard based configurations: most configuration activities are based on wizards and do not require a manual access to configuration files or multiple access to menus.
  - Process monitoring: front-end (e.g. query execution) as well as back-end (e.g. ETL) processes can be monitored and analyzed in order to optimize their execution.
  - ETL debugging environment: it is available and determines a strong reduction of the development effort.
- Administrators and developers are further supported through a wider documentation, a knowledge base as well as consultant and training services.
- Such enhancements, together with warranties and certification of the software become more and more relevant when you are developing a mission-critical application or when you are planning to adopt the platform in a large and complex organization.

Usability: users' point of view

- Platforms usability is largely determined by the BI engines composing them.
- We consider the usability of those engines qualitatively satisfactory. Although they do not reach the level of refinement of the commercial suites, their graphical features give the developed applications an appreciable look-and-feel.
- OS BI platforms also succeed in hiding the access to different tools.

Usability: administrators' point of view

- Complexity of the installing and configuring process
  - Installing procedures are in general quite easy. This is particularly true for Pentaho and JasperSoft whose installation procedures completely rely on a wizard
- Administration complexity
  - In SpagoBI and even more in Jasper we appreciated the easiness of the form-based procedure.
- Problem solving and training effort
  - manuals have a good quality and allow most of the problems to be solved.
  - Several practitioners' forums make available a high number of tips.
    - Pentaho more than 20,000 registered users
    - Jasper about 90,000 registered users (but we experienced in many cases longer response time)
    - SpagoBI community is definitely smaller and so the activeness of its forum

Discussion

- Our analysis shows that OS BI platforms determine an added value with respect to single BI tools
  - several functionalities are accessed transparently
  - a set of processes are centralized and simplified
- The main shortcoming of the platforms is the absence of a fully centralized and unified metadata layer.
- The capabilities of the administrative tools could also be improved in the community editions
- SpagoBI functionalities are comparable to the enterprise editions by Jasper and Pentaho.

Discussion

- Although OS BI platforms are still not as sophisticated as commercial ones, they got a sufficient level of reliability and must be considered a valid alternative to commercial suites.
  - This is particularly true in small and medium-sized enterprises where the quantity of data and the workload are not critical points.
  - Several companies are evaluating the use of OS BI in pilot projects where budget constraints are typically very tight.
- The main risks related to an investment in OS technology come from:
  - Unexpected termination of the project
  - The adoption of a more restrictive licensing of the new releases
  - It is impossible to predict if, apart from the initial investment, the companies that are in charge of the platforms will earn enough from services and application developments to stay on the market.