

OW2 BI Initiative

The reasons to go open source for your BI
The **Xaltia** success story



- ➔ OW2 BI Initiative

- ➔ An integrated scenario: the Xaltia project
 - ➔ Who is Xaltia
 - ➔ The project functional requirements

- ➔ The VAS BI Project
 - ➔ Technical features: huge volumes, performances
 - ➔ Chosen platforms: SpagoBI and Talend
 - ➔ The project
 - ➔ Lessons learnt

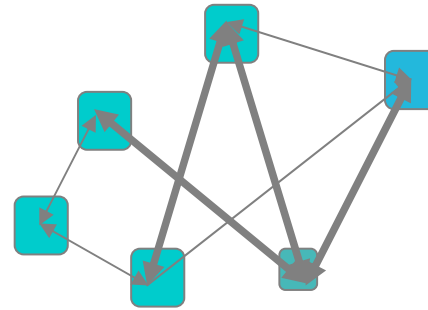
The BI Initiative

A joint effort of OW2 and not-OW2 members set-up to:

- ➔ Improve the coordination effort in the OS BI Area
- ➔ Increase utilization of OS BI solutions at enterprise level
- ➔ Strengthen connections between OS companies and users and the research communities
- ➔ Attract more attention from the EU Funded Research activities to foster “Innovative BI solutions and practices”

VISION & GOALS

OS Business Intelligence qualification
BI investigation and research
Use cases and best practices



STAKEHOLDERS

Software Vendors
Systems Integrators
Academia
Enterprises
SMEs
Public Entities
Consultants

Open Source BI Business Ecosystem

TRACKS

Data quality and Security

Business Process Control

more (in progress)

Scenarios

Project Synergies

Innovation

Open Source BI Vision

OW2 CODE BASE



Gartner

Research

Publication Date: 16 April 2008

ID Number: G00156326

Who's Who in Open-Source Business Intelligence

Andreas Bitterer

Many of the commercial business intelligence (BI) vendors have a long history and large marketing budgets, resulting in high visibility and mind share. Their lesser known open-source counterparts, such as Actuate BIRT, JasperSoft, Pentaho, or Spago have, however, started to gain traction in the market, beyond simple report writers for small shops and even larger enterprises are becoming aware of available open-source BI options.

Key Findings

- Open-source BI is here to stay.

An integrated scenario: the **VAS BI** project

- ➔ XALTIA S.p.A. (www.xaltia.it) is operating as one of the most important mobile networks operators in designing, developing and managing Value Added Services for consumer and business users, locally and internationally
- ➔ Innovation, Technology and new projects drive the Xaltia strategy, allowing to rapidly react to the market changes, turning all challenges in the fixed and mobile communication domain into recognized success for its customers

- ➔ Xaltia design and develop:
 - ⇒ Specific platforms to develop VAS services on mobile channels
 - ⇒ Frameworks and distributed systems for mobile browsing and messaging
 - ⇒ Advanced platforms for multimedia and interactive services on circuit and packet networks
 - ⇒ Solutions for video streaming, multimedia downloads, games and IVR solutions

- ➔ Several outsourcing VAS services are provided through the Xaltia infrastructure: vertical solutions to manage complex and innovative applications in the mobile context

- ➔ Delivery of multimedia contents on innovative technologies

- ➔ Interactive services based on a multichannel logic



➔ Engineering (Xaltia)

Via dei Vasari, 10

05018 Orvieto (TR)

+39 0763 3131



ORVIETO

⇒ Value Added Services (VAS)

- ⇒ VAS are services provided by mobile operators on several channels (SMS, MMS, WAP)
- ⇒ They allow to receive different multimedia contents as:
 - Information (news, horoscopes, traffic news, ...)
 - Entertainment (music, games, ...)

- ➔ The goal of the BI project is to develop a system for the analysis of VAS services provided by different mobile operators:
 - ⇒ Managing huge volumes of data
 - ⇒ Managing complex analysis on data
 - ⇒ Easy to use by personnel with no specific technical skill
 - ⇒ Scalable, reliable and secure

- ➔ The solution is a DWH, fed by a series of ETL processes, developed with Talend Open Studio. A portal has been built upon the DWH to present BI analysis by leveraging the SpagoBI platform

The VAS BI project

- ➔ The system is fed by heterogeneous data sources
- ➔ Data need to be harmonized before loading (Talend ETL job do the work)
- ➔ Some data flows need to be loaded at start-up (e.g. VAS services domain tables)
- ➔ The majority of data flows are loaded daily in the DWH and can be stocked from 3 to 14 months

Source data	Size (GB)	frequency	retention
Flow_1	3,00	Una tantum	14 months
Flow_2	0,01	Una tantum	14 months
Flow_3	0,01	Una tantum	14 months
Flow_4	0,01	Una tantum	14 months
Flow_5	0,01	Una tantum	14 months
Flow_6	0,75	daily	3 months
Flow_7	0,05	daily	3 months
Flow_8	0,05	daily	3 months
Flow_9	2,90	daily	3 months
Flow_10	0,23	daily	3 months
Flow_11	1,20	daily	3 months
Flow_12	4,00	daily	3 months
Flow_13	0,00	daily	3 months

- ➔ The DWH system is built on a three level architecture (EDS, ODS and DM) that manages huge volumes of data
- ➔ Each day, near 3M records are loaded in the EDS: the present sizing allow a growth up to 6M record)
- ➔ The same amount of data is loaded in the ODS and multiplied by an X factor to calculate KPIs supporting information analysis. When the ODS is correctly loaded, data can be purged from the EDS
- ➔ The Data Mart level allows the analysis thanks to OLAP cubes

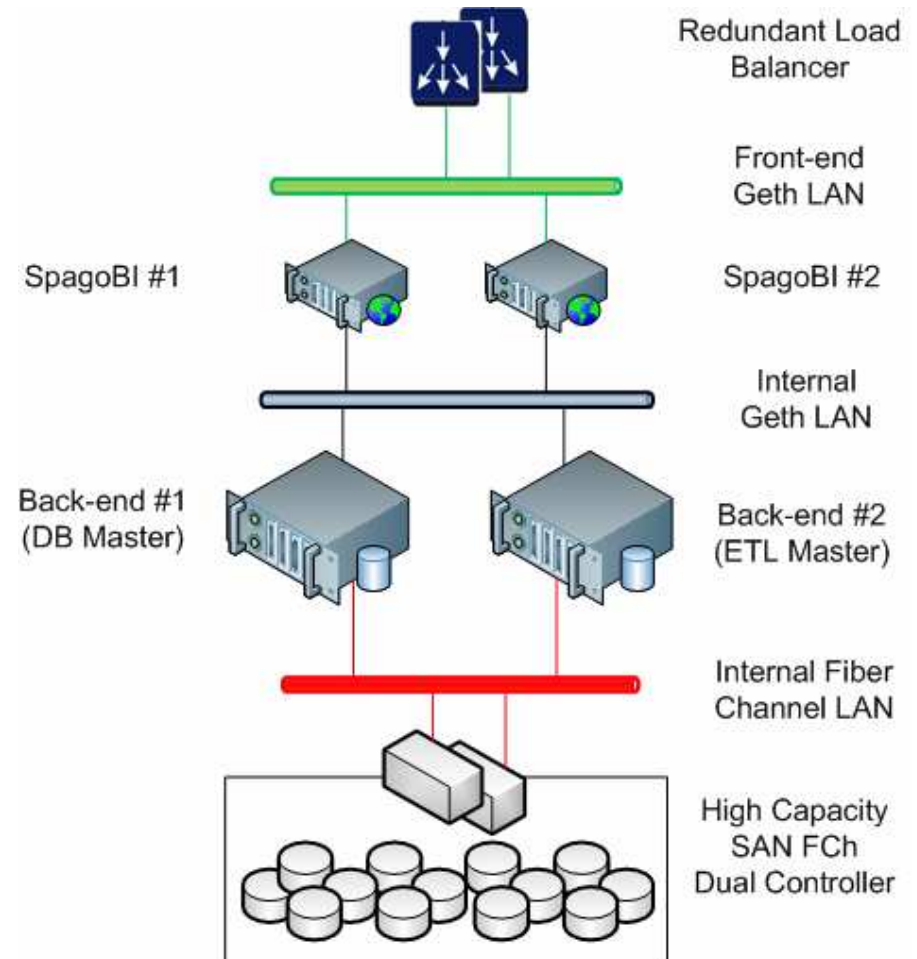
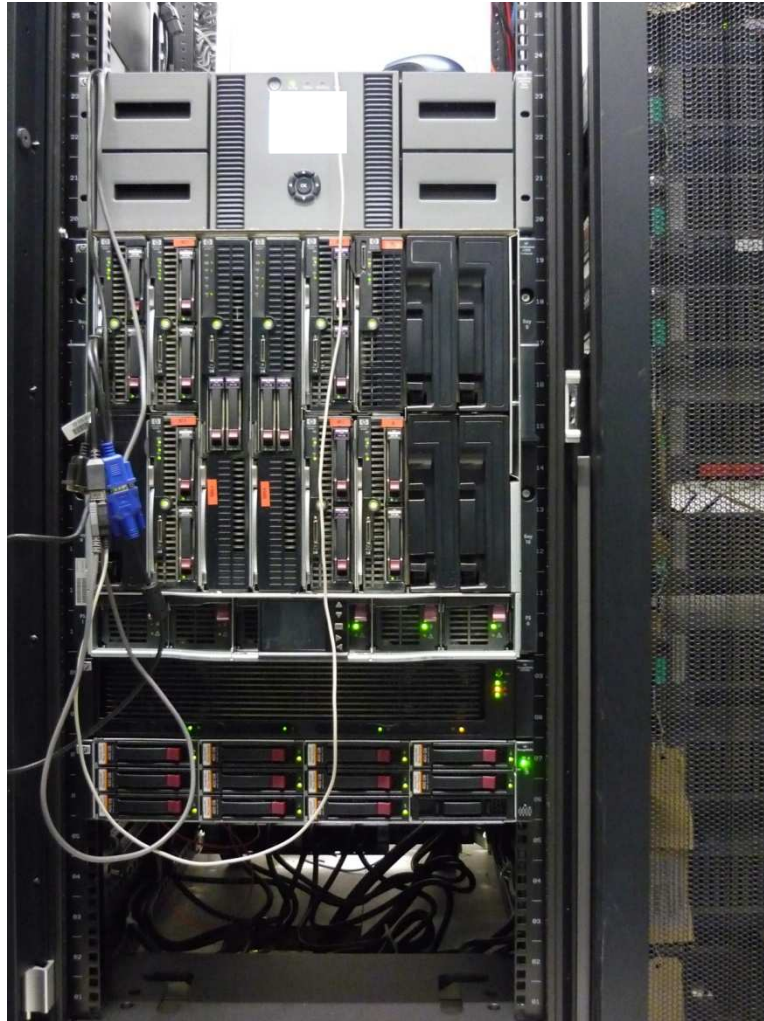
ETL Step	Description	Rows	GB
Feed EDS	ETL from source platform to EDS	Up to 6 Mil	Up to 20
Feed ODS	ETL from EDS To ODS	Up to 6 Mil	Up to 20
Feed DM	ETL da ODS a DM	310.000	0,31

- ➔ Analytical documents must be available on the portal before 10 a.m. each day
- ➔ Source systems send information not before 5 a.m. each day
- ➔ The system must load all DWH levels and manage the loaded information in a maximum of 3.5 hours
- ➔ An interval of 1.5 hours is mandatory to let technicians evaluate and manage rejected data in such a way to adopt recovery procedures when necessary

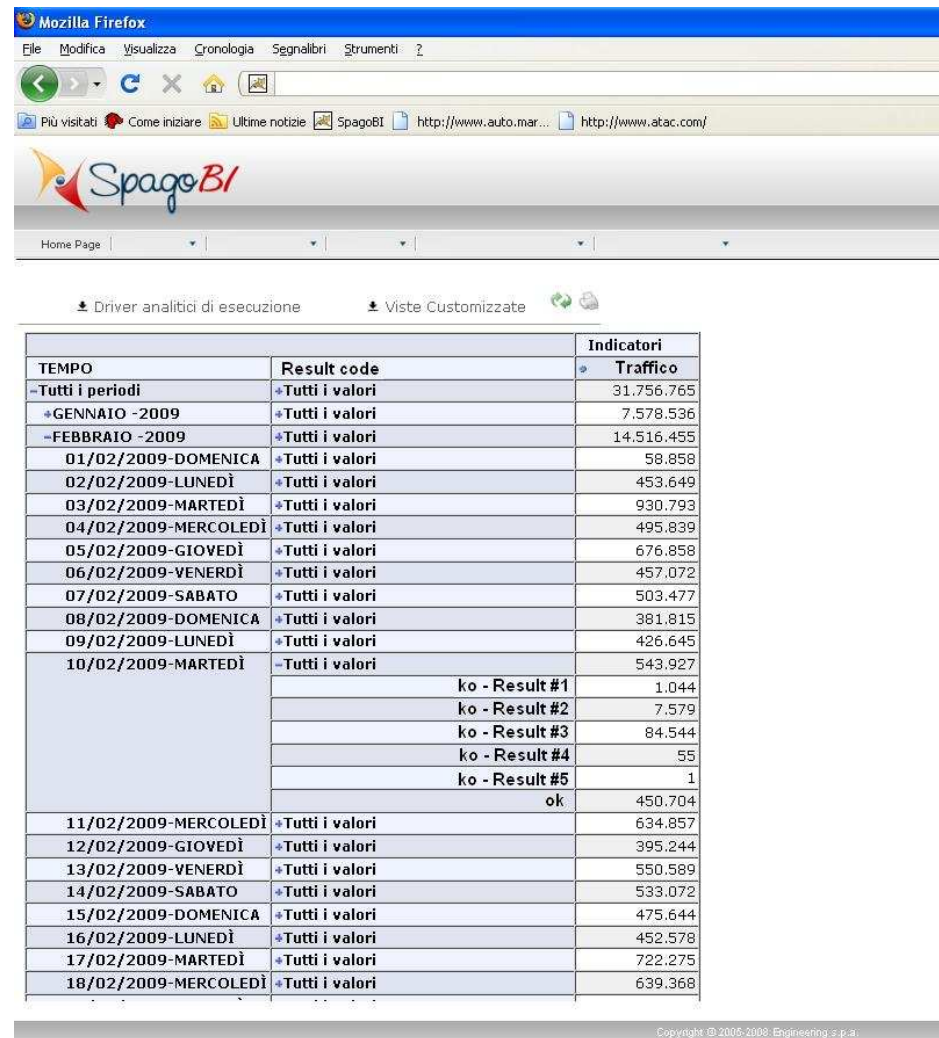
- ➔ Xaltia is already managing open source infrastructure, at the operating system (Linux OS) and database level for their services
- ➔ The market where Xaltia operates is highly challenging and rapidly changing so flexibility at a lower cost is a key success factor
- ➔ Talend Open Studio allows not only to perform data loading but also to organise and document the work. ETL is a complex process not only a technical job
- ➔ SpagoBI facilitate updates and addition of new analysis information

- ➔ In order to guarantee the best performance both in executing ETL jobs and in the DWH interaction, a cluster of two nodes has been implemented by adopting Linux OS features
- ➔ Each node has the following configuration:
 - ⇒ 2 CPU dual core AMD Opteron @ 3GHz
 - ⇒ 16 GB RAM
- ➔ Node A is master in managing the DB, node B is master in managing ETL processes. In case of a fault, each service is loaded on the available node
- ➔ The external storage system to stock data is another critical factor to ensure performance. The infrastructure includes fiber channel duplicated controllers allowing to address 4Tb
- ➔ The infrastructure can scale vertically by adding RAM or by increasing the CPU number

- ➔ The physical infrastructure for the analytical component (SpagoBI) is made up of a two nodes cluster based on the Linux OS, in active-active configuration
- ➔ SpagoBI requests are manage by a Load Balancer pair that guarantees the uniform distribution of the workload between the available nodes and allows to forward requests to a single node when the other is unavailable
- ➔ La nodes configuration for the analytical part consists of:
 - ⇒ A CPU Quad-Core Intel Xeon @ 2833 MHz
 - ⇒ 8GB di RAM
- ➔ The FE infrastructure can scale horizontally by adding SpagoBI nodes in the resource pool managed by the Load Balancer

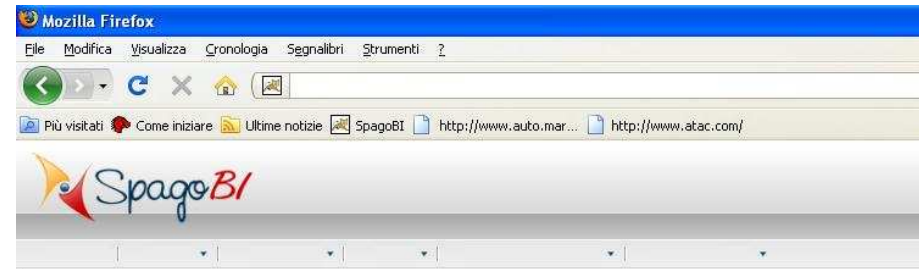


- This analytical document represents traffic data of VAS services on the SMS channel
- The analysis dimension are “Time” and “Result”
- By navigating the OLAP cube we can easily drill to more detailed data, till the single daily events
- The operator can add analysis dimensions as: service type, service name, channel
- This behavior is common to all analytical documents in the OLAP representation



TEMPO	Result code	Indicatori
-Tutti i periodi	+Tutti i valori	31.756.765
+GENNAIO -2009	+Tutti i valori	7.578.536
-FEBBRAIO -2009	+Tutti i valori	14.516.455
01/02/2009-DOMENICA	+Tutti i valori	58.858
02/02/2009-LUNEDÌ	+Tutti i valori	453.649
03/02/2009-MARTEDÌ	+Tutti i valori	930.793
04/02/2009-MERCOLEDÌ	+Tutti i valori	495.839
05/02/2009-GIOVEDÌ	+Tutti i valori	676.858
06/02/2009-VENERDÌ	+Tutti i valori	457.072
07/02/2009-SABATO	+Tutti i valori	503.477
08/02/2009-DOMENICA	+Tutti i valori	381.815
09/02/2009-LUNEDÌ	+Tutti i valori	426.645
10/02/2009-MARTEDÌ	-Tutti i valori	543.927
	ko - Result #1	1.044
	ko - Result #2	7.579
	ko - Result #3	84.544
	ko - Result #4	55
	ko - Result #5	1
	ok	450.704
11/02/2009-MERCOLEDÌ	+Tutti i valori	634.857
12/02/2009-GIOVEDÌ	+Tutti i valori	395.244
13/02/2009-VENERDÌ	+Tutti i valori	550.589
14/02/2009-SABATO	+Tutti i valori	533.072
15/02/2009-DOMENICA	+Tutti i valori	475.644
16/02/2009-LUNEDÌ	+Tutti i valori	452.578
17/02/2009-MARTEDÌ	+Tutti i valori	722.275
18/02/2009-MERCOLEDÌ	+Tutti i valori	639.368

- ➔ This analytical document is a tabular report
- ➔ In the upper part we see the frame to input parameters in order to execute the query (only fields market with a * are mandatory)
- ➔ In the lower part we find the query output format
- ➔ The report header resumes the query parameters and then we have the results
- ➔ This representation has been chosen in all cases when analysis dimensions can be statically configured and query parameters can change frequently



Driver analitici di esecuzione

Viste Customizzate

Driver analitici di esecuzione:

Esegui Salva Cancella

*Data di inizio: 02/02/2009

*Numero Giorni: 30

Parametro #1: [dropdown]

Parametro #2: [dropdown]

Parametro #3: [dropdown]

Parametro #4: [dropdown]

Parametro #5: [dropdown]

Data Inizio 02/02/2009 Parametro #2 Parametro #5
 Numero giorni 30 Parametro #3
 Parametro #1 Parametro #4

Indicatore	Numero	Respinti
Totale Traffico	17.828.871	2.962.131
Messaggi Tipo #1	3.564.793	357.887
Messaggi Tipo #2	656.720	412.517
Messaggi Tipo #3	1.838.645	0
Messaggi Tipo #4	7.491.979	636.936

Causale Errore #1	251736
Causale Errore #2	2576501
Causale Errore #3	152
Causale Errore #4	426
Causale Errore #5	131969
Causale Errore #6	1304
Causale Errore #7	43

- ➔ The Italian antitrust authority in communication (AGCOM) is imposing new rules about premium numbers for VAS services to be used by operators and Content Providers
- ➔ The scenario is still evolving and open to discussion between the different actors: probably the solution will be defining new protocols
- ➔ The reporting functionalities will be enriched to manage new data flows and new analytical documents will be developed in order to manage the new application logic
- ➔ The platform will scale vertically and new data sources will be harmonized

Questions ?

**Thank you for your
attention**