Components of a Business Intelligence software solution

Daniel BĂLĂCEANU
TotalSoft
www.totalsoft.ro

Business intelligence (BI) is a business management term which refers to applications and technologies which are used to gather, provide access to, and analyze data and information about company operations. Business intelligence systems can help companies have a more comprehensive knowledge of the factors affecting their business, such as metrics on sales, production, internal operations, and they can help companies to make better business decisions.[3]
The ability to extract and present information in a meaningful way is vital for any business management application. Business Intelligence enables companies to make better decisions faster than ever before by providing the right information to the right people at the right time. Employees increasingly find that they suffer from information overload and need solutions that provide the analysis to effectively make decisions. Whether they are working on the strategic, the tactical, or the operational level, business intelligence applications provides tools to make informed decisions a more natural part of all employees everyday work experience.
Keywords*: business intelligence, KPI (Key Performance Indicator), software architecture, data warehouse, pivot table*

Introduction

The reporting and analytics solutions that form a BI solution could be delivered as part of the core of business applications, while others can be additional products that extend the business applications. Key goals of these solutions are to provide:

- An integrated platform and applications
- A secure and personalized user experience
- A collaborative environment
- A total solution that is cost effective and comprehensive

Figure 1 shows all the components of a business intelligence architecture including the users. This article describes all of the components and their relationships.

A solid BI architecture: the multidimensional data warehouse

The multi-dimensional data warehouse is the core of the business intelligence environment. Basically it is a large database containing all the data needed for performance management. The modeling techniques used to build up this database are crucial for the functioning of the BI solution.
Typical characteristics of the data warehouse are that:

- it contains time invariant data, in other words a report ran 2 months ago can be reproduced today (if launched with the same parameters)
- it contains integrated data, where integrated means that the same business definitions are use throughout the data warehouse
- it contains atomic data, and not aggregated data as often assumed, because users may (and often do) need the lowest level of detail

To get a good understanding of what a multi-dimensional data warehouse is, it is important to understand the multi-dimensional modeling techniques that assure the above characteristics.
A data warehouse is the main repository of the organization's historical data, its corporate memory [1]. For example, an organization would use the information that's stored in its data warehouse to find out what day of the week they sold the most widgets in May 1992, or how employee sick leave the week before the winter break differed between California and New York from 2001-2005. In other words, the data warehouse contains the raw material for management's decision support system. The

critical factor leading to the use of a data warehouse is that a data analyst can perform complex queries and analysis (such as data mining) on the information without slowing down the operational systems.

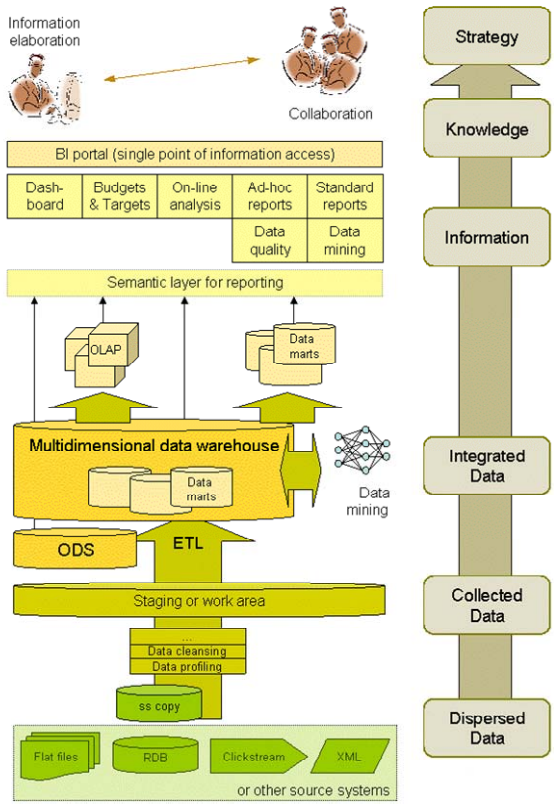

Fig.1. Components of a BI solution

While operational systems are optimized for simplicity and speed of modification (online transaction processing, or OLTP) through heavy use of database normalization and an entity-relationship model, the data warehouse is optimized for reporting and analysis (on line analytical processing, or OLAP). Frequently data in data warehouses is heavily denormalised, summarized and/or stored in a dimension-based model but this is not always required to achieve acceptable query response times.

More formally, Bill Inmon (one of the earliest and most influential practitioners) defined a data warehouse as follows [2]:

• **Subject-oriented**, meaning that the data in the database is organized so that all the data elements relating to the same real-world event or object are linked together;

• **Time-variant**, meaning that the changes to the data in the database are tracked and recorded so that reports can be produced showing changes over time;

• **Non-volatile**, meaning that data in the database is never over-written or deleted, once committed, the data is static, read-only, but retained for future reporting;

• **Integrated**, meaning that the database contains data from most or all of an organization's operational applications, and that this data is made consistent.

In multi-dimensional modeling often the words star-schema and data mart are used as synonyms. In the architecture pictured here (**Fig.1.** Components of a BI solution), it is used the same naming convention. Therefore the data warehouse is a collection of data marts connected by conformed dimensions.

A data mart (DM) is a specialized version of a data warehouse (DW) [5]. Like data warehouses, data marts contain a snapshot of operational data that helps business people to strategize based on analyses of past trends and experiences. The key difference is that the creation of a data mart is predicated on a specific, predefined need for a certain grouping and configuration of select data. A data mart configuration emphasizes easy access to relevant information.

Sometimes data marts may be deployed physically on a different platform or database than the data warehouse, but in my concept of data warehousing this is done solely for technical reasons as e.g. performance or security.

There can be multiple data marts inside a single corporation; each one relevant to one or more business units for which it was designed. DMs may or may not be dependent or related to other data marts in a single corporation. If the data marts are designed using conformed facts and dimensions, then they will be related. In some deployments, each department or business unit is considered the "owner" of their data mart which includes all the "hardware, software and data."[6] This enables each department to use, manipulate and develop their data any

way they see fit; without altering information inside other data marts or the data warehouse. In other deployments where conformed dimensions are used, this business unit ownership will not hold true for shared dimensions like customer, product, etc.

The source systems

Actually source systems aren't really part of the business intelligence environment. But they feed the BI solution and so they are at the basis of your whole architecture.

Important in the set-up of BI environment is that you consider all the types of data you may need to include in your analysis. Don't stop at considering only information included in databases, but consider also external data feeds (from partners) e.g. through XML; or capturing customer behavior on your website (clickstream analysis); connecting to ERP systems on application server level (rather than trying to read a 10.000 table SAP data model :-( ) or uploading Excel files (almost always needed); and many more candidate sources.

When you start working with the data from these systems you'll soon get to know that "you can't live with them and you can't live without them". I often use this phrase to describe that the data provided to you by source systems is crucial for your analysis, but often that data is so bad that it is impossible to work with it. Most companies run into the data quality problem sooner or later in their implementation, and it is a key element at the start of any BI project.

Another phenomenon that occurs frequently is the off-line copy of a source system. Sometimes a 24x7 online application is so business critical that even simple operational reporting on this application jeopardizes the availability of the application. In this case a mirror of the application database is set up to run the reports on. Obviously if such a situation is in place, the data warehouse should extract from the mirror rather than the online database.

ETL: Extract, transform and load

In order to build up a multi-dimensional data warehouse, data from all the sources previously described needs to be extracted and brought to your BI environment. Given the variety of source systems, possible connectivity is a key word here.

After extraction, the data needs to be transformed. Transformation can mean a lot of things, but it includes all activities to make the data fit the multi-dimensional model that makes up the data warehouse. Given the significant difference between ER models and multi-dimensional models, the transformations may become quite complex. If we add to this the extra work done to clean up and 'harmonize' the data coming from different systems, then one might understand why some authors describe ETL work as 70% of the IT side of a BI project. Usually a separate database or specific zone in the data warehouse is reserved as storage space for intermediate results of the required transformations. This area is called often staging area or work area. After all the transformation work, the prepared data can be loaded into the multi-dimensional model. Although this step is not as complex as the transformation part, a lot of attention needs to be given to this step, since data loaded to the data warehouse may be 'released' or 'public' to end users.

A specific aspect of the ETL work which has gained much attention during the last years is everything related to data quality.

Data profiling is the activity of 'measuring' the level of data quality one might expect to get from a specific source system prior to starting the ETL work. This analysis might be quite complex, so specific tools have been developed for this. However the value of knowing any data deficiencies before engaging in an expensive BI adventure is enormous. Building a technically perfect BI solution with perfectly unusable data in it is an expensive mistake.

Data cleansing refers to specific activity of cleaning up data. Mostly it is about address data. Specific tools are available where one can validate customer addresses against an external database of addresses so all address information in the data warehouse is identical and verified. Similar methods of cleansing are available for product classification codes

(as UNSPSC).
Obviously the usefulness of data quality solutions extends beyond the area of business intelligence. Due to the enormous impact of data quality on the ROI of your BI application however their first use is often in this context.

ODS or Operational Data Store

Often customers ask about the difference between an ODS and a data warehouse. The difference is almost as big as between an operational system and a data warehouse.
An ODS is an off-line copy of 1 or more source systems:

- with a data model that respects the data model of the source system (ER model)
- with an added functionality of storing historical versions of the data
- potentially with some or full integration between the ER models of different applications

In any case the ODS is not built to support strategic decision making, but rather support operational reporting, related to a the functional scope of a specific application.

OLAP cubes

While most data warehouses or data marts today are still deployed on relational databases, also multi-dimensional databases are available, often called OLAP cubes. The multi-dimensional modeling techniques using facts and dimensions are applicable to both relational databases and multi-dimensional databases.
Even though OLAP offers superior performance in analysis of the data, historically the size of an OLAP cube was limited and incremental loading of the cubes sometimes difficult. OLAP cubes have had very high success rates for environments where the BI solution is used for what-if analysis, financial simulations, budgeting and target setting (requiring write-back).

Semantic layer for reporting

The objective of a BI solution is to offer a tool where end users can easily access data for analytical purposes. Unfortunately databases (relational or multi-dimensional) are notoriously user-unfriendly. To make data access easier for end users most vendors of reporting solutions have provided the possibility to create a layer between the database and the reporting tool. In this layer the database fields can be translated into 'objects', each having a clear business definition. These objects can be dragged and dropped onto a report, making report creation a 'piece of cake'.

Reporting

The semantic layer and the reports produced on top of your data warehouse are only the tip of the iceberg for a person sitting in the IT department, but for most of the end users they are the only visible part of your BI solution. IT professionals tend to give less attention to the reporting part of the BI application as the big IT work goes into the construction of the data warehouse. It is in this part of the architecture that we need to make a huge distinction between how IT persons view the BI environment and how end users view the BI environment. For end users the availability of reports, access to the reporting tools, correctness of titles and descriptions on the reports is crucial, and lack of attention on this side may result in utter failure of the BI project, even with a correctly modeled data warehouse, perfect architecture and good quality data.
The richness of reports that can be provided to end users is sometimes outside of the comprehension of typical IT persons, but also outside of the imagination of the business people. IT people often know too little of the business to format a report into the most readable, easiest understandable graph, while business user have to few insight into the technical possibilities of the platform and have been forced to think within the limits of the same graphs and tables they have always worked with. An extreme attention needs to be paid to the collaboration between IT and business to build a good set of reports that maximizes the return on investment of the BI project.
Usually the reports are structured in the following typology:

- **Standard reports**: Fixed reports scheduled

and made available to the users a specific point in time or at user request without any further required input.

• **Parameter reports**: Reports of fixed lay-out, available at user request, requiring input of some parameters when the user launches the report.

• **Ad-hoc analysis**: Report created by the user on the spot, either starting from an existing standard or parameter report or from scratch.

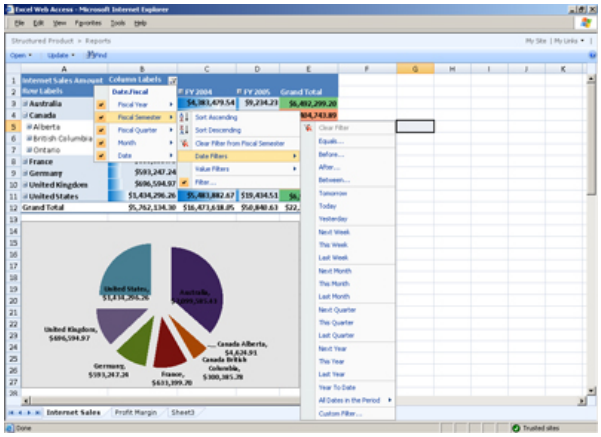

Fig.2. Ad-Hoc reporting with Microsoft Excel 2007 (reference to Microsoft web site, Deliver Business Intelligence through Microsoft Office SharePoint Server 2007)

• **Budgeting & target setting**: These reports show data from the data warehouse and allow input of data with write-back to the data warehouse.

• **Dashboards**: Highly aggregated reports, containing a multitude of KPI's and thus showing an overview of how the company or a department is doing. Mostly dashboard reports combine actual figures with target figures (as well as figures of previous year or month for comparison).

• **Data quality reports**: To follow up the evolution of the quality of the information in your data warehouse usually the data quality steward requires some specific data quality reports. Also data profiling tools and data cleansing tools may provide their own reports that fall into this category.

• **Data mining reports**: As output of data mining exercises, before feeding the results back to the data warehouse usually there are several outputs to be interpreted. These reports, aimed at the analyst working with data mining, are usually part of the data mining tool.

• **IT technical reports**: A series of technical reports to follow up load performance, query performance, number users, data volumes, etc are necessary for the good functioning of the BI solution.

• **Meta data reports**: In order to give end users and system analysts a good insight into what data is available and how it has been transformed to match the company's business definitions, an overview of the metadata is very interesting. Some authors view the IT technical information I described above also as metadata.

It is important to realize for each functional set of reports who are the main users. A good allocation of report ownership, and clear rules for version and release management will help to keep control of the many reports that will be created.

BI portal: single point of information access

The multitude of different reports describe above demonstrates how easy the BI solution may grow into a wilderness of reports. Once the amount of reports begins to grow, the best solution is to create a single point of access to information within the company. Usually this results into some kind of portal solution with pointers to the different reports, clear descriptions of the scope of each report, as well as indication who is the business owner of the report (Fig.3. **BI Portal**).

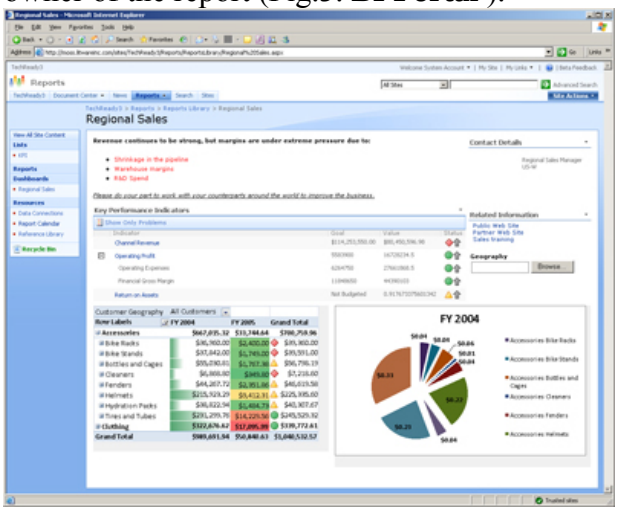

Fig.3. BI Portal (reference to Microsoft web site, Deliver Business Intelligence through

Microsoft Office SharePoint Server 2007)
Another element that the portal application can provide is collaboration. Often figures on reports need interpretation, explanation, side notes and other comments. This textual information needs to be stored somewhere. Given the fact that this information may have multiple forms, a portal solution, which allows document storage (content/document management system) is the right solution to allow this kind of BI collaboration.

End users or why did you build your solution in the first place

Of course users aren't part of the BI architecture, however they are in the picture to stress that the roll-out of a BI solution is not comparable to the roll-out of an application. While the functional scope of an application is limited and therefore user training is limited in scope, the possibilities of a data warehouse are limited only by the available data and the creativity of the end users. Unfortunately it is often the creativity of the end user which poses a problem. Most end users are not used to have massive amounts of information at hand. Or in other words, most users are not used to work with a complete BI solution. So while, training users to work with a few pre-defined reports might not take much time, training users to get the most of the data warehouse and to make the actively participate in building the reports part of the BI environment is a tedious and time consuming task.
The link between user education and ROI for BI is an often forgotten part of the project. IT vendors consider the BI project closed when the data warehouse is built, together with a set of requested reports. Unfortunately business people tend to see the project also as closed when IT delivers, while the task of business community actually just starts at that point in time.
Enabling BI in the organization is an often forgotten part of BI projects, requiring change management in how the organization works with KPI's, how managers work with reports and dashboards, definition of new responsibilities, awareness of a 'new way of working'.

Data mining

Data mining is the automated process of discovering previously unknown useful patterns in structured data. The data warehouse is therefore a perfect environment to conduct data mining exercises on. To a certain extent online analytical processing, in which users slice and dice, pivot, sort, filter data to see patterns is a form of human, visual data mining. However the human eye can only see a limited amount of dimensions (mostly three) at the same time and therefore cannot discover more complex relationships. Also discovering relationships between different attributes of dimensions is a time consuming exercise. The field of automated pattern detection has gained a lot of popularity in the last years. Successful implementations of data mining applications are however clearly limited in functional scope and data mining on large scope is not expected before 2010.

Conclusion

According to Gartner [10], CIOs will need to concentrate on information as a leverage point to enhance efficiency, increase effectiveness and support competitiveness. This also corresponds to the continued importance of business intelligence in 2007. As such, CIOs will continue to be responsible for IT — the mechanism. They can further play a greater role in leveraging information — the understanding that drives performance and innovation.

Top 10 Business and Technology Priorities in 2007

Top 10 Business Priorities	Ranking	Top 10 Technology Priorities	Ranking
Business process improvement	1	Business Intelligence applications	1
Controlling enterprise-wide operating costs	2	Enterprise applications (ERP, CRM and others)	2
Attract, retain and grow customer relationships	3	Legacy application modernization	3
Improve effectiveness of enterprise workforce	4	Networking, voice and data communications	4
Revenue growth	5	Servers and storage technologies (virtualization)	5

Improving competitiveness	6	Security technologies	6
Using intelligence in products and services	7	Service-oriented architectures	7
Deploy new business capabilities to meet strategic goals	8	Technical infrastructure management	8
Enter new markets, new products or new services	9	Document management	9
Faster innovation	10	Collaboration technologies	10

Source: Gartner EXP (February 2007)

References

1. Wikipedia, http://en.wikipedia.org/wiki/Data_warehouse
2. W. H. Inmon, Building the Data Warehouse, Wiley, 1005
3. Wikipedia, http://en.wikipedia.org/wiki/Business_Intelligence
4. Ralph Kimball, Margy Ross, The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, Wiley, 2002
5. Wikipedia, http://en.wikipedia.org/wiki/Data_mart
6. Bill Immon, Data Mart Does Not Equal Data Warehouse, DM Direct Newsletter, November 20, 1999 Issue, http://www.dmreview.com/article_sub.cfm?articleId=1675
7. Business Intelligence: Making the data make sense, Dresner, Howard. 2001, Group US IT Expo 2001
8. Larissa T. Moss, Shaku Atre, Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications, Addison-Wesley Professional 2003
9. Joy Mundy, Warren Thornthwait, Ralph Kimball, The Microsoft Data Warehouse Toolkit: With SQL Server 2005 and the Microsoft Business Intelligence Toolset, Wiley 2006
10. Gartner EXP Survey, STANFORD, Conn., February 14, 2007, http://www.gartner.com/it/page.jsp?id=501189