

So how do you want yours served?

The abundance of Business Intelligence tools together with greater regulation and compliance requirements has created new challenges for today's enterprise BI solutions. Chris Bradley considers the role Data Virtualisation and Open Source could play in your BI solution.



Most organisations today have a Business Intelligence (BI) and Data Warehouse (DW) solution of some sort. The maturity of BI & DW has moved on from departmental through operational BI to the position we now see in many corporates to that of Enterprise BI.

The availability of a new generation of tools and BI solutions which easily integrate with ERP systems has undoubtedly provided real benefit in reducing overall time to solution.

However, the information explosion, plethora of tool options and information regulation and compliance presents us with more challenges, including:

- Data migration and take on ETL vs EII (or both?)
- Does Open Source BI have a place for me?
- Historical reporting or Predictive analytics?
- Information Management w.r.t. BI

Time does not allow me to cover all of these in this article so I'm going to highlight the first two.

Data migration and take on ETL vs EII (or both?)

By now most of us will be familiar with the purpose of Extract, Transform and Load tools. Less well known however are the capabilities of the Data Virtualisation or Enterprise Information integration tools such as Composite or MetaMatrix.

Broadly speaking these provide the capability to access data from a massively wide variety of sources without having to move it from the source system. They have extremely rich caching and aggregation capabilities and in my experience have dramatically reduced the time to provide rich access to data. I once heard them described as "views on steroids".

Can EII / Data Virtualisation add value to Data Warehousing?

The use of EII technology in Enterprise Data Warehousing and for data take-on is something that demands serious consideration. There are several ways in which EII can add value to DW solutions; here are just 3 to consider:

a) Prototyping Data Warehouse Development

During DW development, the time taken for schema changes, adding new data sources and providing data federation are often considerable. Using Data Virtualisation to prototype a development environment means you can rapidly build a virtual DW rather than a physical one. Reports, dashboards and so on can be built on the virtual DW. After prototyping the physical DW can be introduced.

b) Enriching the ETL process

Frequently new data sources particularly from ERPs are required in the DW. All too often the ETL lacks data access capabilities to complex sources. Tight processing windows may require access, aggregation & federation activities to be performed prior to the ETL process. The powerful data access capabilities of EII provide rich access and federation capabilities which can present virtual views to the ETL process which continues as though using a simpler data source.

c) Federating Data Warehouses

How many organisations have more than one DW? Is the Information in each completely discrete? I don't think so. Data Virtualisation provides powerful options to federate multiple DW's by creating an integrated view across them. This has particular relevance in providing rapid cross warehouse views following a merger or acquisition.

"The use of EII technology in Enterprise Data Warehousing and for data take-on is something that demands serious consideration"

Data take on considerations ETL or EII?

When providing data into a DW, the use of ETL or EII (or both) needs care. Some of the key considerations include:

Requirement	Implication
Data replicated in DW and Operational System	Update in one or both locations? If data is physically in two locations are there compliance issues (e.g. SoX, HIPPA etc)
Data Governance	Is the data only managed in the originating Operational System?
Currency of the data	How up to date are the data requirements of the DW? Is there a need to see the operational data?
Time to solution	How rapidly is a solution required?
Life expectancy of source system(s)	Are the source systems likely to be retired?
Need for historical / summary / aggregate data	How much historical, aggregated data is required in the DW solution?

Open Source BI

One of the most talked about BI technology trends this year, the theory of “Open Source” is undeniably good. In the commercial world however, remember that “Commercial” Open Source doesn’t mean free. The business model for Commercial open source usually means that support, training, consulting and so on are the supplier income streams.

Thus said, it’s well worth considering these solutions.

For some organisations, management reporting and BI has been provided by spreadsheet based reports and graphs. These have evolved from a few departmental reports to become an intertwined set of Excel reports.

If this example spreadsheet lifecycle sounds familiar, you are not alone. In fact, it is so familiar that many people just try to live with it. Managing data in this way will eventually lead to poor quality data in reports. Depending on the audience of the reports, the implication of poor data quality may be: poor business decisions, loss of credibility, legal compliance issues and possible financial or legal penalties for breaching regulations. If these issues lead to an investigation of data management practises, the spreadsheet daisy chain is going to be hard to defend.

This doesn’t just lead to data quality issues; it also creates a very inefficient chain of data propagation. Everybody in the chain will be dependent upon the previous people. Any issues identified need to be passed back along the chain.

Furthermore, by keeping all the raw data in your spreadsheet, you have far more data stored locally than you need. With the continuous stream of information security failings published in the press, can you defend why your local laptop has a spreadsheet

containing all the low level data, when all you needed to publish were some high level KPIs?

Over time, departments become more and more dependent upon spreadsheets. Before long you have little departmental “Cottage Industries” producing spreadsheet applications, often completely outside the governance of corporate application development strategies. These spreadsheet applications will inevitably need support and enhancement. You may end up with applications which themselves have a total cost of ownership that was never budgeted for.

For these scenarios, open source BI represents a quantum leap and is highly commended.

Not only departmental solutions

Open Source BI is also suitable for larger scale opportunities too, however before taking the plunge; a few questions need to be considered:

- Is the developer interface & capability suitable?
They are often not as refined and capable as their (more expensive) non open source rivals;
- Is the benefit of the open source “community” appropriate in my sector?
The community and development of industry specific solutions is a powerful argument for open source – are there solutions for your industry though?
- Less expensive vs. best?
Fitness for purpose, capability and support must always be considered, not simply the headline price.
- Prototype solution?
These provide a very cost effective way of developing a prototype, and the potential of taking it further;
- Head to head with Microsoft?
At the less expensive end of the BI tools scale, Microsoft SQL server, SSIS, and SSRS solutions are very competitively priced. Open Source tools however are less costly than other well know mainstream rivals.
- Appropriate for end user development?
Not really. However in several organisations the reality is that it’s actually the IT folks who build the Business Objects or Cognos reports too.

The opportunity is thus to see where Open source BI can be beneficial in your overall solution portfolio. It really has now come of age.



Chris Bradley has 30 years of Information Management experience. He is a recognised thought-leader in the Information Management field, is a published author (Data modelling for the Business- A handbook for aligning the Business with IT using high level data models), a regular columnist for the BeyeNETWORK and an International conference speaker. Chris also sits on industry advisory panels. Chris leads IPL’s Business Consulting practice and can be contacted at BusinessConsulting@IPL.com.