# Business Intelligence and Data warehousing


**Business Intelligence**

Business intelligence is a broad category of applications and technologies for gathering, providing access to, and analyzing data for the purpose of helping enterprise users make better business decisions. The term implies having a comprehensive knowledge of all factors that affect a business, such as customers, competitors, business partners, economic environment, and internal operations, therefore enabling optimal decisions to be made.

BI applications include the activities of decision support systems, query, and reporting, online analytical processing (OLAP), statistical analysis, forecasting and data mining. Business intelligence applications can be:

- Mission-critical and integral to an enterprise's operations or occasional to meet a special requirement
- Enterprise-wide or local to one division, department
- Centrally initiated or driven by user demand

This term was used as early as September, 1996, when a Gartner Group report said:

By 2000, Information Democracy will emerge in forward-thinking enterprises, with Business Intelligence information and applications available broadly to employees, consultants, customers, suppliers, and the public. The key to thriving in a competitive marketplace is staying ahead of the competition.

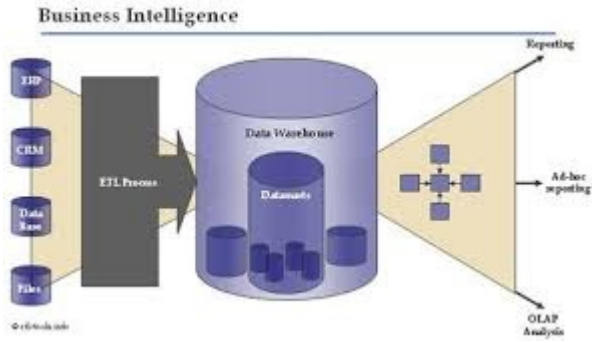
Making sound business decisions based on accurate and current information takes more than intuition. Data analysis, reporting, and query tools can help business users wade through a sea of data to synthesize valuable information from it - today these tools collectively fall into a category called "Business Intelligence."


**Traditional Business Intelligence**

Traditionally business intelligence has been used for many years for strategic and tactical decision making. This type of processing involves intensive analytical processing of historical and summarized data managed in an enterprise data warehouse. Data performance issues caused by centralizing data in an enterprise data warehouse have led to the creation of data marts, which solve performance problems by spreading the BI processing across multiple data stores.

The problem with data marts is that organizations often build them directly from business transaction databases, rather than the enterprise data warehouse. This is because it is often quicker and easier to build a data mart than to incorporate additional data into the enterprise data warehouse and then build the data mart from the data warehouse. Another problem is that many organizations have more than one “enterprise” data warehouse. Multiple disconnected data warehouses and data marts leads to data consistency issues, which data warehousing was supposed to solve in the first place.

One of the more interesting developments to complicate the data warehouse environment in recent years has been the explosion in Big Data. Big Data is the massive volume of data sets, structured and unstructured, whose size and complexity make it difficult to manage and process using traditional tools – like those represented by data warehouses. The concern had been the need to find ways to scale up IT systems so that Big Data could lead to business value as a function of BI.




Today, scaling up the data warehouse to manage Big Data remains a concern, but it’s manageable, and increasingly, businesses are finding ways to team them up to drive better

BI insights.

Philip Russom, director of The Data Warehouse Institute Research for Data Management, notes that Big Data from new social and digital sources ranges from structured to semi-structured to unstructured; most data warehouses are not designed to store and manage the full range. And Big Data is often fed continuously and in real time – something most data warehouses are not equipped to accommodate.

“Most of the business value coming from Big Data is derived from advanced analytics based on the combination of both traditional enterprise data and new data sources,” according to Russom.

“Big Data and the data warehouse can be a powerful team, providing many new analytic applications that enterprises need to stay competitive. Achieving this, however, requires some modifications to existing infrastructure, tools and process to integrate Big Data into the existing data warehouse environment.”


## Trending Topics in Business Intelligence and Data warehousing

I liked the post by Ethan Jewett “Toward an Analysis of Data warehouse and Business Intelligence Challenges”. Ethan has worked as a Chief Consultant in Business Intelligence and Data Warehousing since 2005, primarily in the SAP & BusinessObjects world.

### AGGREGATING SILO-ED DATA SOURCES

Silos silos silos. Anyone trying to do data analysis has run into this problem: the data exists, but we can't get at it. The technical aspects of this challenge are many (bandwidth, interfaces, and ETL), but it's worth noting that they are usually dwarfed by the cultural and organizational obstacles (default against sharing, departmental rivalries), many of which are in place for good reason (security and permissions concerns, privacy laws).

### REPRESENTING DATA IN A MEANINGFUL WAY

Historically this feels like one of the least-addressed challenges, but we are finally seeing some serious attention paid to this problem. Challenges in representation of data range from visualization (and the related topic of responsible visualization - as visualization is too often untruthful), to analytical views and tools, through search and guided data exploration.

As we stand, the data in Data warehouses and business intelligence datamarts is too often opaque and misunderstood by most users. Even the most impressive and advanced

visualizations and analysis tools are still highly guided constructs that are often only applicable to predetermined datasets. We have come a long way (finally) over the last decade, but we have a long way yet to go.

**REPRESENTING REPORTING STRUCTURES**

Reporting structures are now fairly well understood, but representing them efficiently in our Data warehouses or BI tools remains a challenge. Some examples of such structures: reporting hierarchies, time-dependency, calculated measures, and derived or dependent characteristics. Challenges revolve around rollup and calculation performance, reorganization due to reporting structure changes, and accessibility to potential users.

**PERFORMANCE**

Traditionally this is the "big one" and it is still very much an unsolved problem. Bound by the Cap tradeoff, we are more or less forced to give up either consistency, availability, or partition-tolerance in order to improve performance under constant resources. Two approaches prevail: architectures that give up one or more of the three in exchange for performance, and architectures that attempt to better optimize for the problem-space in order to improve performance while maintaining all three CAP axes. Both are perfectly legitimate approaches, but it will be important to recognize which architectural approach is being pursued in any given product or technology. As a wise person once said, "there is no such thing as a free lunch".

Further complicating matters, there are multiple performance aspects of Data warehouse and business intelligence applications, and we need to be clear which ones we attempt to optimize for. These aspects include query performance (keeping in mind the random access vs. batch/bulk/scan access difference), data loading (ETL) and reorganization, and (in some systems) writeback or changing of data.

**SECURITY**

Security models pose more of a management problem than a technical problem for Data warehouse and BI applications. Nonetheless, I think they're worth mentioning as a core challenge to keep in mind, just in case someone comes up with a way to make reasoning about security in analytical-processing-oriented datasets less painful.

**DATA LOADING**

Last but certainly not least, data loading is a perennial headache in Data warehouse and BI systems. The three basic types of data loading (batch, real-time/streaming, and write-back/input) all to some extent conflict with each other. Add to that the complexity of managing a profusion of delta mechanisms (many of which exist for good reason, others of which exist because of careless design) and different interface formats and we've got ourselves a real party. Standardization of interfaces and design practices are the key touchstones of conquering this challenge, but as with many of these challenges, this is more of a human problem than a technical problem.

## CONCLUSION - TECHNICAL VS. DESIGN CHALLENGES

If we take one thing away from this enumeration of the challenges of the Data warehouse and business intelligence spaces, I hope it is the fact that most of these challenges are more human in nature than they are technical. They tend to derive from the difficulty in making tradeoff decisions, standardizing interfaces and architectures, identifying and focusing on the problem space, and understanding how people may actually use these systems to greatest effect. Because of this, these challenges are often at least as susceptible to design solutions as they are to pure technical solutions. There is a tendency in the industry to focus on technical answers to these challenges over design answers, perhaps because technical solutions are often more impressive and in some sense physical. I think that's unfortunate.

Best Regards,
Raj Kosaraju, CIO Maxil Technology