

BI at 1000 Feet

© Mishtu Banerjee, Harmeny Systems Ltd. And Scientificals Design Inc.

May 2006

V 0.9



4203-315 Southampton
Drive, SW,
Calgary, AB
T2W 2T6

403-472-7531
info@harmeny.com
www.harmeny.com

Copyright Notice



For more information go to: <http://creativecommons.org/licenses/by-nc-nd/2.5/>



4203-315 Southampton
Drive, SW,
Calgary, AB
T2W 2T6

403-472-7531
info@harmony.com
www.harmony.com

Introduction

“Get to 1000 Feet and Stay There” -- Guy Kawasaki, “The Art of the Start”

This briefing is a high level overview of Business Intelligence(BI) for CIO's and CTO's. We will cover:

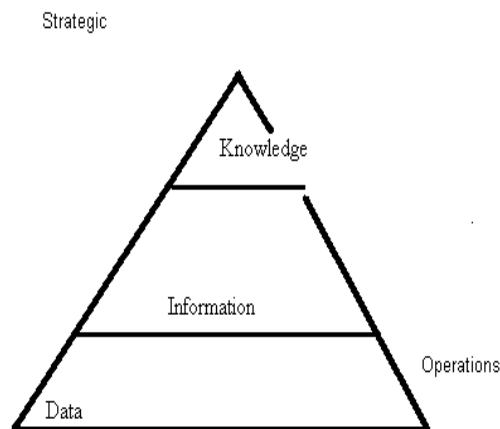
- Strategic Versus Tactical: BI provides the information path to resolve both views in a company.
- A Brief History of BI: BI is the confluence of several mature IT technologies, and some new ideas.
- BI Core Skills: Simply put, Data Organization, Data Analysis, Data Presentation.
- BI Tools: Fall into several categories: Reporting/Charting, Modeling/Integration, Analysis/Mining
- BI Business Value: Reducing friction so information flows more smoothly into supporting decisions.
- BI Limitations: BI fails without process, and without training.

This briefing, at 1000 feet, will not compare vendor products – there are a large range of resources on the Internet with such comparisons. Instead the focus will be on the core capabilities and skill sets required to implement a successful BI program.

Strategic and Tactical Information.

Strategy is the 10,000 foot view of a company, while tactics is the 100 foot view, or occasionally the 1 mm view. Companies gather masses of information internally, and externally. However, many a manager about to make a critical decision has had the queasy feeling: “The information is here, somewhere in the system but where, and how do I get access to it?” At the level of daily operations and tactics, most companies create masses of data as part of due diligence, process quality control, marketing and customer relationship management. But how do we get our heads around so much information? Only through a process of summarizing, simplifying, and filtering.

At the end of the day, the mantra of BI is this: The right information, to the right person, in the right form, at the right time, so it can be used for a decision. The path by which this is achieved is three fold: Data Organization, Data Analysis, Data Presentation.



The process of leveraging corporate information is often presented in the context of a pyramid where at the base is a mass of data, in the middle are information summaries, and at the top is knowledge in the form of trends,



relationships, models. In this view, the goal of BI is to facilitate the flow of data into knowledge. BI is the latest flavour of a decades long effort to reach such an ideal of a knowledge organization. A bit of potted history will place BI in context.

A Brief History of BI

Business Intelligence as a term was coined around 1989 by Howard Dresner of the Gartner group. It was used as a catch-all term to describe decision support systems that leveraged corporate information. As such BI can be considered as an extension of the term "Decision Support Systems" (itself quite loosely defined), and a successor to the failed "Expert Systems" movement of the 80's which tried to develop rule-based systems using symbolic logic. In BI the focus switches to defining and leveraging the large stores of data accumulating in a corporations "electronic vaults" in various forms.

BI is the confluence of several critical technologies. Many of those technologies are mature and have been stable well understood components of IT for decades. Other technologies are newer. Furthermore there is a great deal of terminological confusion in BI because it brings together several distinct technologies, each with their own jargon. Additionally, it brings together practitioners of those technologies, each of which view the field from their own area of strength.

Three general kinds of technologies (with multiple origins) are coming together, from different directions:

- Data Organization – typified by relational databases, and relational data modeling.
- Data Analysis – typified by statistics programs supporting business statistics, and OLAP programs with their dimensional Modeling paradigm.
- Data Reporting – the convergence of tabular and graphical reporting technologies introduced independently from the statistics, OLAP, and relational database fields.

Let us first review the mature technologies that are part of BI:

- Relational Databases and Relational Modeling
- Online Analytical Processing (OLAP) and Dimensional Modeling
- Statistics and Sampling
- Business Graphics and Statistical Charting
- Business and Statistical Reporting (tabular)

Additionally there are some newer immature technologies that are associated with BI:

- Data Mining
- Data Model Integration
- Web Services

Finally there are some technologies that seem to be long-going but never quite mature which contribute to the design and data access components of BI

- Computer Assisted System Engineering (CASE) tools
- Extraction-Transformation-Loading (ETL).



Lets place these technologies in a brief time-line:

- 1970s : **First Commercial Relational Databases and Statistics Software**. Relational databases provide a generic access process to information. OLAP exists as a set of custom mainframe oriented programs, and in a language most often used by actuaries, APL. Analysis is the purview of statistical programs.
- 1980's. **First Expert Systems and Decision Support Systems**. CASE tools become more prevalent, including those for Relational Databases. There's a lot of talk of "4GLs", 4th Generation Languages, that will be programmed in wait for it something like English. Relational Databases move onto smaller computers suitable for smaller businesses. Expert Systems come from academia to business, and prove themselves not suitable for prime time. Expert Systems and DSS get a bit of a bad name. Both Relational Databases and Statistics programs leverage increased power on small computers to emphasize reporting. Statistics programs particularly go wild on charting. Business graphics programs are widely adopted.
- 1990's. **Focus on Data Consolidation via Data Warehouses**. Databases introduce greater support for statistical processing. Statistics package introduce greater support for relational data access. CASE/4GL and reporting tool providers who play a middle-man role in system integration begin to specialize in a new field called, "BI" Much of the initial focus is creation of Data Warehouses and Data Marts, often using dimensional design techniques. Certain tension between Dimensional and Relational design paradigms. Data mining tools become available on PC's and begin their escape from academia.
- 2000s. **BI oriented companies begin to take off**. Major database players introduce BI solutions, including OLAP. Major statistics players introduce BI solutions, including OLAP. Increasingly the emphasis for BI moves beyond structured data in databases, and the large amount of unstructured data available within and outside the corporation via XML based technologies. Small Data Mining, and AI oriented R&D firms are bought up by the larger relational database, statistics, and BI vendors.
- 2010s???? **AI Returns as Search**. Expert systems come back into style based on the soft AI paradigm of the 90's and data mining/machine learning techniques, rather than the hard AI paradigm of the 80's. Search becomes the interface for BI.

What Are the Core Skills of BI?

BI's core skills are fairly simple to state:

- **Data Organization**. This is a design heavy endeavor; and there are two approaches - relational design which seeks highly normalized structures, and dimensional design which represents information as a "hypercube", "dimensions" (axes of said hypercube) and "facts" (cells in the hypercube). While the jargon in these two areas varies; they are less different than they appear to be. Given the prevalence of relational databases, SQL becomes the defacto "low level" language for expressing the organization of information and querying it. OLAP and dimensional design are also facilitated by matrix algebra, as a formal process for Modeling data cubes.
- **Data Analysis**. This area often also incorporates the manipulations to move data into the correct form for analysis: "Extraction, Transformation, Load (ETL). Most business oriented analyses are focused on relatively simple statistical summaries. Sampling theory is required to understand sources of bias in corporate data and correct for it. The largest trend in this area is the growth of data mining techniques. These techniques emphasize machine learning procedures where large datasets are processed and



converted into rules, classifications, models. There is a certain tension between Data Mining automated approaches, and approaches stressing data visualization and human interpretation. Data mining represents the greater availability and commercial application of several technologies that originated in the research communities as “soft AI”. The defacto “low level” language for expressing classic statistical data transformations is matrix algebra.

- Data Presentation. On the statistical side, Edward Tufte clarified and essentially defined the area of “Information Design” in a trio of books: “The Visual Display of Quantitative Information”, “Envisioning Information,” and “Visual Explanations”. A great deal of research on information presentation was also done in the context of “Exploratory Data Analysis” in the 80's. These bodies of work influenced much of the range of reporting and data visualization formats encountered in modern BI presentation software.

Obviously, this is a broad range of skills to look for in an individual. Those coming to BI from the database side tend to be strongest at relational (and sometimes dimensional) Modeling skills. Those coming to BI from the Data Analysis side are strongest at information processing tasks; and often at data presentation.

As Wikipedia notes: “People working in business intelligence have developed tools that ease the work, especially when the intelligence task involves gathering and analyzing large quantities of unstructured data. Each vendor typically defines Business Intelligence their own way, and markets tools to do BI the way they see it.”

The consequence of this trend is for BI core skills to be confounded with particular vendor tool sets. Once the core skills of data organization, analysis, and presentation are emphasized it easier to evaluate the capabilities and skills needed for BI.

Let us quickly summarize the core skills for several BI oriented tasks:

- BI Application Development: Strength must be in the particular area of the application (Data Organization, Analysis, or Presentation) and an appropriate programming language and development methodology. Verifying questions concern identifying the key area of strength, the subsidiary areas of strength, and the breadth across areas. In particular, for application development, it is important (particularly for an Application Development Manager) there be the ability to implement an application development methodology.
- BI Architect: Strength in design is the strongest criteria. Given history, strength is usually in either Relational or Dimensional design - but there should be familiarity with the other field. Verifying questions concern verifying design skills, and sufficient business process experience to be able to architect a system nested within a given corporations process. Since the BI Architect often plays the role of intermediary between end users and the development team, strong soft skills in communication are necessary.
- BI Analyst. Strength can either be in design or in analysis. But there should be an emphasis on balancing out whichever is the weaker skillset. Verifying questions concern identifying whether strength is on the design or analysis side; and establishing that there is a basic familiarity with relational techniques and SQL, and basic statistical processing and data presentation.
- BI Tool Driver. Strength is usually in a particular tool, and intimate knowledge of it. There should be an emphasis on learning new tools, and developing a broader understanding across tools. Verifying questions here become tool specific.

Where do BI specialists come from. Given the convergence of several technologies, BI specialists come from a range of backgrounds (as is clear from the above section), with as much disparity as a BI application developer coming



from a programming background stressing numerical methods, a BI architect coming from either a relational or OLAP design background, a BI analyst coming from a statistical background, a BI tool driver being at one extreme a programmer and at another extreme an administrative clerk trained up on, and with deep knowledge of a particular tool. The origins vary – check your assumptions at the door, and confirm the core skill sets: Data Organization, Data Analysis, Data Presentation.

Finally, let us mention what might be called the “0th” skill for BI – since it sets the context for the leveraging of all other skills: Business Process and Business Domain Knowledge. BI is not a set of uniform recipes that can simply be applied routinely. The context for BI is the culture, process, and specific domain of a business. Since BI specialists like other IT migrants tend to move across industries, skill number 0 is simply: to learn the business process, culture, and domain models of a new business. The goal of BI is to leverage the intuition of knowledge workers by presenting them with timely accurate information for decisions – which means you must be familiar with the thought models specific to your organization. Success in BI relies more heavily on “soft skills” such as communication, leadership and teamwork, than other areas of IT.

BI Tools

The BI tools available (from various vendors, and in open source incarnations) can be evaluated along two dimensions of complexity: Data Organization and Data Analysis. At one extreme are data integration tools, which are highly oriented towards data organization. At the other extreme are Data Mining tools which are highly oriented towards Analysis. In the middle are OLAP tools which integrate organization and analysis.

- ETL – range from custom scripts, to specific data access tools. Many of the tools are low-level converting between various formats, but can integrate higher level capabilities both in organization and analysis of data.
- Data Integration – tools for integrating data across models and frameworks. High on the organization dimension.
- Relational Databases – defacto storage source for much corporate data; and also source of a fair bit of data cleansing/ preliminary analysis.
- Statistics – general tools for statistical data processing; modern tools usually incorporate charting and reporting.
- OLAP – Both a storage solution, and an analysis solution. Often used in small data warehouses, “Data Marts”, while larger systems default to relational databases.
- Reporting/Charting – a range of solutions are available. BI vendors tend to stress fixed reporting/charting templates, integration of reporting with access to relational and OLAP databases; usually incorporates some basic statistics capabilities.
- Data Mining – tools for the automated discovery of rules, classifications, models in large bodies of data. Initial entry into BI of these tools often found them too complicated for non-experts – subsequently these tools/capabilities have often been incorporated into other BI products, “under the hood” rather than broken out as separate products.



BI Value

What is the value proposition of BI?

Simple – BI properly implemented reduces the friction required to make correct decisions. The right information, at the right time, in the right form, is placed in front of the right person who can use it to make a good decision.

In the context of the lack of such information, people use “gut feeling”, “professional experience”, and occasionally “best guess” to make decisions. And that isn't going to change. BI leverages human intuition by organizing, analyzing, and presenting information into a more useful form. As such, it augments human intuition – and makes sure that such intuition is indeed “based on the best facts available”. This allows others to more objectively verify the basis of a decision. Secondly the same systems that provide information to support a system, can be used to further monitor the effectiveness of a decision, in terms of intended and actual consequences.

In this context, the role of BI is that it supports a learning process, by which executives can sharpen their decisions, and by which less experienced employees can benefit from the data and decisions that went before them, and hence leverage the data in the corporation, and the experience in the senior members of the corporation.

If we imagine at one end of the spectrum on corporate information a very technocentric view focused on data storage, organization, analysis and reporting technologies, and at the other end a focus on strategy, business process, and performance, then BI represents the route towards connecting those two perspectives.

IT World View -----> -----> BI -----> -----> Executive World View
(Architecture) (Performance)

BI Limitations

BI limitations are the difference between a lot of the BI hype that is common these days, and the real and measurable gains possible via a BI strategy. Given the mantra, “The right information, to the right person, at the right time, in the right form to make a decision”, what are the barriers to success?

- Strong BI will not overcome weak business process. BI must be nested in a particular corporations culture, and process norms. The BI plan must extend the existing processes, not replace them.
- Cost of retooling information systems. Often there is a trend towards trying to retool existing information systems to a particular BI suite. These costs can be quite high, and must be estimated. Alternatively, look for tools that can adapt to, and augment your existing information systems, and provide a path into the future.
- Integration. Care must be taken in integrating technologies for a BI solution, given the different sources of various BI technologies. One solution to this problem is to deal with vertically integrated BI vendors – the flip side is being locked into a specific vendor.
- Lack of training. It is not enough for specialists to be trained in BI. As BI is an information service supporting decision making – the key training that is needed is outwards from the BI group to those decision makers who could benefit from BI. Such training is not achieved by report formatting workshops



for managers (or their administrative assistants) but via BI staff working with managers and mentoring them in leveraging information; while allowing themselves to be mentored in management perspectives.

- BI hype. BI is not new. It is the convergence of several mature technologies in IT salted with a few new technologies. BI hype can often create unrealistic expectations and buyers remorse.

BI Trends

BI is a moving target. However some trends are clear:

- Vertical Integration of core BI capabilities. That is data organization, analysis, and presentation existing in a single tool, rather than across tools. This is led by the convergence of both relational database and statistics vendors on BI, and both sides integrating each other's context and capabilities into existing and new tools.
- Lowered licensing cost of BI – due to reduction of tool prices, existence of a developing open source community projects for free BI tools.
- Integration into business process – largely through reporting systems that can support specific business processes such as via key performance indicators and dashboarding.
- Component-ization of BI – the opposite of vertically integrated BI are BI components with general API's that allow them to be mixed and matched as needed by an organization.
- Web Services, and Browser as universal interface. BI in this view is a set of web services delivered remotely via a browser. Web services make the possibility of BI leveraging information within and across corporations more realistic. As Drucker has often noted, the most important information for a corporation is not that from inside, but that from outside the corporation.

