Sierra Systems

DATA DISCOVERY VS ENTERPRISE BUSINESS INTELLIGENCE

INTRODUCTION

Over the past several years a new category of Business Intelligence (BI) tools has appeared in the BI marketplace under the heading of data discovery (also known as visual analytics). Data discovery tools offer highly interactive and graphical user interfaces built on in-memory architectures to address business users need for ease-of-use. It allows business users to analyze unstructured data sets, get answers quickly, and not have to rely on IT to build new business subject areas in the enterprise BI solution.

The traditional Enterprise BI solutions tend to be highly governed, based on a centralized data warehouse with corporate reports and dashboards made available across the organization. While analytical capabilities are included, such as online analytical processing and ad hoc query, they are typically not fully embraced by all business users due steep learning curves, time constraints and limited self-serve capabilities.

TOOLSETS

The following provides an overview of the two toolset categories.

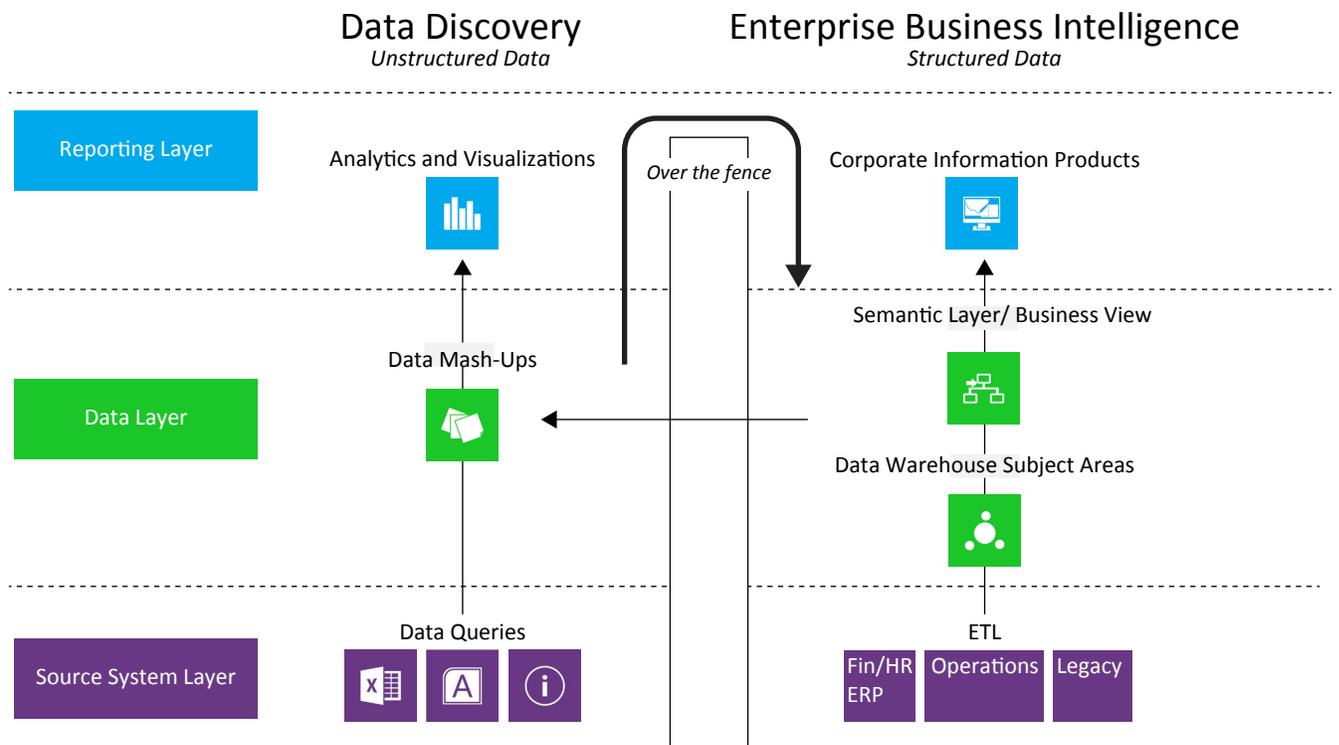|	Enterprise Business Intelligence	Data Discovery
Key buyers	IT	Organizational departments
Main sellers	Large vendors (SAP, IBM, Oracle)	Small, fast growing independent solution providers
Use Case	Reporting, business analysis, monitoring	Rapid one off analysis, what-if analysis, prototyping, data mashups
Query	Navigation of known questions / pre-defined business processes	Rapidly explore and visualize the unknown
Data Structure	Structured Data - predefined data models (Kimball subject areas)	Unstructured Data - raw data as it exists in the real world
Approach	Top down, IT modeled, query existing repositories	Bottom up, user mapped (data mashups), into dedicated repository
User interface	List/grid reports, key performance indicators (KPI), dashboards	Visualization, in-memory queries (fast)
Developers	IT consultants and or in-house BI staff	Departmental power users
Data Access	Typically via an ETL process (business rules applied, data validation)	Typically on the fly data mashup
Resources	IT centric/dependent - often constrained by available technical resources	Departmental business power users

Complimentary Technologies

The rapid growth of data discovery tools should be viewed as a compliment to the success of BI in the corporate mainstream over the past 20 years. BI has grown to be so successful that IT can no longer keep up with business demands for additional datasets and reporting - given the average industry norm is 60 to 90 days to create a business subject area.

Data discovery tools should be regarded as complimentary to enterprise BI platforms. Typically data discovery lives alongside the enterprise BI solution, albeit often as a departmental solution. Data discovery tools can read from structured or un-structured data sets, existing business subjects areas, operational data sets, XLS, web data sets, and other sources. The intent is to expand the use of analytics to a broad range of information consumers and non-traditional BI users.

Once business users have analyzed a new data mashup, performed queries, produced various information products and feel comfortable with what they have generated, the output can be 'tossed over the fence' for IT to institutionalize it as part of the overall enterprise solution (ie: create a business subject area). Within the enterprise solution business rules, formal data modeling and ETL structures are applied. It is here that the corporate 'shared truth' is achieved.

Organizational Information Management

Data Discovery
Unstructured Data

Enterprise Business Intelligence
Structured Data

Reporting Layer	Analytics and Visualizations	Over the fence	Corporate Information Products

Semantic Layer/ Business View

Data Layer	Data Mash-Ups		Data Warehouse Subject Areas

Source System Layer	Data Queries		ETL

Fin/HR ERP   Operations   Legacy

In reality this fence is the proverbial "chinese wall" in that there is not really a fence between data discovery and enterprise BI. The two can co-exist and it is a matter of managing your organization's information as a whole.

The role of the BI professional is to work with the business users to guide and mentor them in the use of the data discovery toolset. This ensures business users understand how to effectively acquire and merge both unstructured and structured data, and can advise on data mashups, common dimensions, etc. with an eye to the day when the data discovery results may be "tossed over the fence" to the enterprise BI solution. It is about establishing sound practices that match the methods to the needs.

Rapid Market Growth

Data discovery tools have experienced phenomenal growth over the past several years. Organizations are rapidly embracing the idea of empowering business users with an ability to navigate and visualize data in a "surf and save" mode as an alternative to a report- based architecture. Driven largely by business users, data discovery products such as Tibco™ Spotfire®, QlikView and Tableau® have grown to become leading product sets in the BI market place in only a few short years.

The BI market is transforming from enterprise BI systems used primarily for measurement and reporting to those that also support data discovery including analytics and visualizations. Data discovery vendors are increasingly embedding both traditional reporting and dashboards into their products. Enterprise BI vendors are adding interactive analysis, advanced analytics and location intelligence into their products. The challenge is to find one vendor that does both well.

Leading Data Discovery Vendors

Based on our experience, we have found the three leading data discovery vendors are Tibco Spotfire, QlikView by QlikTech, and Tableau Software. They are all based on in memory data store and direct query access. These vendors offer relatively limited enterprise BI capabilities.

- TIBCO Spotfire, provides an easy-to-use platform for business user data discovery, interactive visualization, ad hoc analysis, and predictive modelling. On average its deployments still tend to be focused on departments however they are adding enterprise BI features with each new release.

- Qlik View (by QlikTech) enables business users to explore and find connections, patterns and outliers in data without having to model those relationships in advance. It has highly complex analytical and great visualization exploration capabilities. It has a reputation for being a bit more difficult to learn.

- Tableau Software provides purpose-built, business-oriented data mashup capabilities with direct data connectors. Tableau's products often fill an unmet need in large organizations that already have an enterprise BI standard, which makes them frequently deployed as a complementary toolset.

Leading Enterprise BI Vendors

The two leading enterprise BI vendors are IBM® Cognos® and Microsoft®. Both of these two vendors have reacted to the emerging data discovery market but only one is currently providing both enterprise BI and data discovery capabilities.

IBM Cognos has long been rated and widely accepted as the leading enterprise BI platform in the market. Their data discovery toolset is Cognos Insight which has some integration with Cognos 10. Cognos Insight is not to be confused with Cognos Business Insight which is their dashboard building tool. Cognos Insight imports data into its own data store and has the ability to publish and distribute via TM1. While it can publish to the Cognos Portal, it does not access data via the common FM layer which means it defines and uses its own business rules, calculations, dimensions, etc. Essentially this is "TM1 Lite", IBM's reaction to the emergence to data exploration tools. Cognos Insight lacks much of the polish, integration and visualization capabilities of the Tableau, QlikView, and Spotfire. Cognos is planning to release cloud based "smart data discovery" based on its Watson project soon which will dramatically change its data discovery story.

Microsoft appears to be emerging as both an enterprise BI and data discovery vendor. Its capabilities are embedded across multiple tools in the MS technology stack including SQL Server, SharePoint and Office. The addition of its Power BI products, on top of the Excel toolset, has the potential to be a game changer. Excel being the one tool, with its familiar and intuitive interface, puts data exploration and visual analytics in the hands of most business users. Power BI enables business users to obtain data using Power Query and perform data mashups and create data models in Power Pivot on demand. These models can then be connected to Power View, Power Map or Performance Point Dashboard and can then be explored in a similar manner to traditional cubes. This functionality is available to individual users or groups of users. While Microsoft lacks much of the polish and visualization capabilities of the leading data discovery vendors it does provide a broad set of functionality in a well-known user interface.

DATA DISCOVERY AND ENTERPRISE BI STRATEGY

Sierra Systems' strategy in regards to this emerging technology is to provide services in both the enterprise BI as well as data discovery arenas. We take this approach as they are both part of the same overall Business Intelligence solution. While one is an enterprise solution and the other a departmental solution, there are common users across both, and a common thirst for information within the organization.

In essence, data discovery represents the fulfillment of the self-serve promise made by the BI industry many years ago. Towards this end Sierra Systems has experience with IBM Cognos in the enterprise BI space, Tibco Spotfire in the data discovery space, and with Microsoft, whose product line effectively spans both enterprise BI and data discovery.

Even with great tools, there is always a very real possibility of creating conflicting or even erroneous information. The tools don't pose a threat to an organization's information quality and integrity - the practices do. When obtaining information in an organization there needs to be a method to determine which side of the house one goes to for management information.

This is a two-step process which includes:

- Matching the method with the need (enterprise BI vs data discovery);
- Identifying the risk level of the chosen method.

Sierra Systems has a process that guides business users in determining which method they should use and then assigning a risk rating index to inform information consumers to the quality of individual information products. This risk index is meant to be presented on every information product produced in the organization to indicate the method used and the associated risk with the results.

It includes a number of factors that are used to determine the Risk Rating Index including:

- Data sources (structured vs. unstructured, internal vs. external);
- Validation processes;
- Number of joins / complexity;
- User experience in data analysis.

By way of example the following describes the Risk Rating Index concept

Class 1	You need to make a financial decision based on an analysis and it needs to be right (your career may depend on it). You want to use a proven financial source system or your enterprise BI data warehouse.
Class 2	You need to present data to the board of directors. It needs to be a good visualization, but your enterprise BI tools aren't that advanced. You might use your data discovery tool against your enterprise data warehouse. There is a risk that you may have joined facts and dimensions incorrectly, but the risk is small.
Class 3	You conduct analysis based on an existing data warehouse subject area fact, but use unstructured data from somewhere else within the organization for a roll-up.
Class 4	Your analysis requires data that is based purely on unstructured data.

From a best practice perspective, Sierra Systems takes the following three step approach to data discovery to guide and mentor the business users in their quest for information.

1. **Data Acquisition and Prep Phase** - We work with business users to identify data sources based on the business issue at hand be they structured, multi-structured or unstructured. We then assist with developing queries to import data into an Operational Data Store (ODS) and applying basic transformations.

2. **Data Analysis Phase** - In this phase, the business users are free to start exploring their data using a variety of analytic and statistic options. They discover relationships, review discoveries and define business rules. Our role is to mentor business users in their analysis activities.

3. **Implementation Phase** - In this phase business users have the option of determining what to do with the results of their analysis and their analytical models.

- Departmental Solution – The results could be treated as a one off piece of analysis or the analysis could be made available for other departmental users on a regular basis.

- Enterprise BI Solution – Once a particular data set / analytic has been successfully used by the department for a period of time, there may be a need to provide the results to a wider audience in the organization as the departmental solution evolves. At this time the departmental solution is "tossed over the fence" and implemented as a formal subject area in the enterprise data warehouse. In this manner business rules are institutionalized and the "shared truth" is achieved across the enterprise.

In summary gaining insight into your business and accessing high quality information in a timely manner is not as simple as choosing a good tool. It involves establishing sound practices that match the methods to the needs, and understanding the associated risks.

For more information please contact us:

contact@sierrasystems.com
1-877-688-1371
sierrasystems.com

The information provided by Sierra Systems contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although Sierra Systems works hard to provide accurate and timely information, we cannot provide a guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act upon such information without appropriate professional advice after a thorough examination of the particular situation.