RESEARCH NOTE

DATA DISCOVERY AND BUSINESS INTELLGENCE –UNCOVERING THE BEST FIT

THE BOTTOM LINE

Companies have traditionally adopted business intelligence solutions to support business decision making on a consistent daily basis, bringing data from disparate systems into a common data structure or warehouse for reporting, analysis and creating analytic applications. Emerging data discovery tools have focused instead on providing data-savvy users with free-form, more tactical one-off analysis on single data sets. Both are valuable in maximizing the value of data to an organization because they take into account the varying half-lives of data. Moving forward, companies will maximize value from BI and data discovery by driving rapid analysis of new data and bringing those models into the broader BI reporting environment for greater breadth and repeatability

When considering adoption, organizations should take into account the "single source of truth" available in BI solutions, the skill sets required for each solution and BI solutions that integrate discovery as well traditional reporting and dashboard capabilities.

THE SITUATION

The concept and market definition of what is business intelligence (BI) software has been around for many years. There are many established vendors in the market. BI includes the applications, infrastructure, and tools that enable access to and analysis of information to improve and optimize business decisions and performance.

BI solutions perform functions such as reporting, dashboarding, analytics, data exploration, benchmarking, predictive analytics, and prescriptive analytics from a variety of data sources, and can be consumed in interactive, scheduled and static fashions. In most implementations, there is universal definition of the data for everyone, or a single version of the truth, so the users of the solution within an organization would have the same business definition of the data. In industry parlance, this is known as a semantic (or logical) layer, and lends business context to raw data.

Over the past forty years, companies such as IBM, MicroStrategy, Oracle, and SAP have built up market presence, expertise, and vertically diverse industry solutions. These traditional BI vendors have grown by internal development, or by acquiring, and integrating additional components.

Emerging data discovery vendors such as Tableau Software, QlikTech, and Tibco Spotfire allow users to quickly analyze data based on a variety of visualization and geo location-based methods. Data discovery tools are often searched based or visualization driven and enable data savvy users to develop and refine views, and analyses of quantitative, structured and unstructured data. Generally, a common business view (semantic layer) of the data is not defined in data discovery implementations. Users commonly have greater flexibility in their choice of data and data sources in their investigative analysis.

Both traditional and new BI vendors have started including data discovery and visualization capabilities in their solutions. While the terms BI and data discovery have been used interchangeably, they have very different usage patterns, and deliver different benefits to organizations. In looking at the usage and benefits of data discovery and BI solutions, Nucleus found a number of factors organizations should consider when evaluating their overall analytics strategy.

TIME TO DEPLOY

When considering the time to deploy, there is a significant difference between the implementation of a business intelligence solution, and the installation of a data discovery tool. Generally, business intelligence approaches include a data warehouse, and as the majority of end users are not data scientists, they frequently employ a single source of truth (semantic layer) for their analysis. End users may want to bring in external data, but are not looking to build relationship diagrams, nor are they necessarily skilled on primary keys, column characteristics or other metadata rules.

The time to deploy a business intelligence solution is often measured in weeks and months, depending on the complexity of the business intelligence reporting and analysis requirements. The benefit of this approach of course, is that the extended BI community is leveraging a universal view of the data frequently spanning multiple applications and data sources. The tools themselves can be deployed very quickly to a wider audience, especially in the case of cloud based implementations, where the time to deploy can be measured in hours. Data discovery tools, as they do not require a predefined data model or semantic layer, can be deployed much faster, sometimes measured in days.

THE HALF-LIFE OF DATA

The half-life of data is important to consider because not all data has the same value or relevance over time. Data that has a shorter life is better served by discovery tools where that quick analysis is important, whereas for data that has a longer life, business intelligence tools are a better fit.

The suitability of data discovery tools for analyzing new data versus bringing them into the traditional business intelligence environment is in large part determined, at least initially, by the expected half-life of that data (Nucleus Research *m36 - Measuring the half life of data,* May 2012). The value of data diminishes based on the cadence of decisions. Decision tempos are tactical (driving process changes in near real time), operational (driving changes that take days or weeks to implement), or strategic (driving changes that become part of a quarterly or longer planning and implementation process). Data discovery tools allow users to analyze, and gain insight from data that has a half-life and will not be part of the enterprise data environment.

ENTERPRISE REPORTING VERSUS TACTICAL ANALYSIS

Nucleus has found that organizations using enterprise reporting have a wider range of users, and very different reporting requirements than those looking for tactical analysis. Traditional business intelligence approaches include a data warehouse, enterprise grade availability and security, and visualization capabilities. Business intelligence has a broad consumption base in most organizations, and is used for identifying trends, performing analysis, and strategic decision making. The majority of end users are not data scientists, and leverage the single source of truth (semantic layer) for their analysis. End users may want to bring in external data, but are not looking to build relationship diagrams, nor are they necessarily skilled on primary keys, column characteristics or other metadata rules. They expect to be and are shielded from the raw data complexity

Data discovery provides many individuals with the ability to creatively explore a shared repository of data, and users can pick and choose data on an ad-hoc basis. Data discovery tends to be focused on more specialized data investigations, used by data scientists or advanced business analysts. These investigations tend to require extremely efficient processing of large volumes of data. In many cases, this data may be from unstructured or external sources, such as social media streams. Analysts are looking for that specific nugget or anomaly in the data, and often use the results of their analysis to feed a wider BI audience.

ONE TO ONE VERSUS ONE TO MANY

Nucleus has found organizations view the business usage of data discovery and business intelligence to divided between individual analysis to be shared between a few, and corporate or multi-departmental analysis shared among the many.

Data discovery tends to be focused on more specialized data investigations, focused on particular sets of data, and used by data scientists or advanced business analysts. The diversity of the sources, formats, and the need for rapid profiling of this high velocity data can present challenges and make this type of data profiling fast and efficient for these analysts. Data discovery best practices require data analysis agility, rapid cycle iterations, and often begin with the end result in mind. Data scientists look for answers to specific

questions, such as which customers are the most profitable or which products are raising the most complaints.

BI provides historical, current, and predictive views of business operations. Common functions of business intelligence solutions are reporting, dashboarding, analytics, data exploration, benchmarking, predictive analytics, and prescriptive analytics. Vendors such as Birst, IBM, MicroStrategy, Oracle, and SAP have developed extensive solutions and tools that provide most, if not all, of these capabilities. Business Intelligence uses technologies, processes, and applications to analyze external, internal, unstructured, and structured data, as well as business processes.

For data analysis, a business intelligence solution handles a variety of analytic capabilities ranging from basic reporting and dashboards to ad-hoc data discovery and predictive analytics. Nucleus has shown how companies increase the value of analytics over time by moving from simple automation of dashboards to enterprise deployments of predictive and consultative analytics (Nucleus Research *m17 – The stages of an analytic enterprise*, March 2012). With this approach companies steadily increase the ROI of their deployments from 188 percent to 1209 percent over time. Ideally, companies starting with basic analytic capabilities should support this evolution while minimizing the software and hardware investment needed to pursue their path to analytic excellence.

DATA SOURCES AND INTEGRATION

The diversity of the sources, formats, and the need for rapid profiling of this high velocity data can present challenges, and make this type of data profiling fast and efficient for the analysts. Data discovery best practices require data analysis agility, rapid cycle iterations, and often begin with the end result in mind. Data scientists look for answers to specific questions, such as which customers are the most profitable or which products are raising the most complaints.

Though the traditional concept of the monolithic data warehouse may be fading as the single source of corporate information, business intelligence still requires the aggregation of key data sources in a consistent manner to support primary business processes and information sources over time. For business intelligence solutions to maintain agility, data management, and data warehouse creation must be automated, when possible, and made easy to use so that data management is not solely in the hands of highly trained data scientists. When analytic input and content are expected, business intelligence solutions have to be secure, support high availability, and be broadly scalable.

CONCLUSION

Data discovery is an important tool to provide ad-hoc and individually driven analysis of specific data sources. Though these solutions are starting to become more collaborative and social, they are not designed to be used as enterprise business intelligence solutions where broad consumption of static reports or dashboards is more common.

When data savvy employees want to rapidly access and analyze new data or do ad-hoc analysis of raw data on an ongoing basis for rapid insights, data discovery solutions are often the best fit. However, Nucleus has found that given the half-life of data, (Nucleus Research *m36 - Measuring the half life of data,* May 2012), in many situations, the initial work done in data discovery provides new structure for analysis that are valuable to be replicated in the broader BI environment. To maximize the value of data and analysis over time while meeting the tactical needs of data analysts, organizations moving forward will reap the most benefit from data discovery tools that easily integrate with their broader BI architecture, and extend initial insights into broader ongoing analysis.