Rethinking Data Discovery

The new research and experimentation paradigm for analytics and discovery

Nitesh Jain
General Manager and Vertical Head – CPG (Europe & Emerging Markets)

Table of Contents

Abstract

Most enterprises across industries are realizing the growing importance of harvesting their data and drawing meaningful analysis for driving business growth and building competitive advantage. Many of them have already invested in building enterprise data warehouses to generate analytical insights, but it limits them to only structured data within the enterprise.

In today's digital and Big Data era, data can be structured or unstructured and most of it resides outside the enterprise. It has, thus, become crucial to have a platform that can be used to integrate all kind of information, perform a lot of discovery-based analysis (to separate noise from real signals), perform analytics on top of it and finally consume it through an intuitive visualization interface. This avenue is now being explored in the form of data discovery platform in many enterprises.

This paper discusses data discovery as the new Research and Experimentation platform for an analytics-driven discovery. It also throws light on why enterprises should go ahead with the shift to build a competitive edge for their businesses.

Introduction

Every day, huge volume of data is being generated in various forms and from different sources. For instance, the retail giant, Wal-Mart, handles 1 million customer transactions every day, thereby adding 2.5 petabytes of information to the database[1]. The telescope installed at New Mexico, under the project Sloan Digital Sky Survey, collected more than 140 terabytes of information in a decade. Its successor, the Large Synoptic Survey Telescope scheduled to come online in 2016, in Chile, will gather that much of data every week[2].

Such enormous volume of digital information is the storehouse of meaningful insights such as valuable business trends, shifting consumer behavior, onset of an epidemic, changing weather patterns and rising crime rate, to mention a few. When managed well, this data can provide business houses and governments an opportunity to unlock new business avenues and solutions for better governance.

For quite some time, only large data-centric enterprises such as Facebook and Google or e-commerce firms were using Big Data to analyze trends and derive insights for developing new products and services. They engaged specialized quantitative analysts called data scientists for this. Soon, these analysts began to realize the potential of analytics for every other industry and business. Enterprises, too, in other industries, began to recognize the latent potential of analytics.

However, the biggest challenge for these enterprises was assimilating such volumes of data, segregating the irrelevant information from the relevant one and classifying them into meaningful categories. Enterprises had been using traditional Business Intelligence (BI) systems that use historical data to test a hypothesis. This approach is not only lengthy and comes with its own processing time, but also carries the risk of, "what if the question is wrong?" If the initial question asked is wrong, instead of throwing away the entire output, the report is tweaked to align the analysis with the question asked. This is not prudent for making crucial customer and market-relevant decisions at the frontline of a business operation.
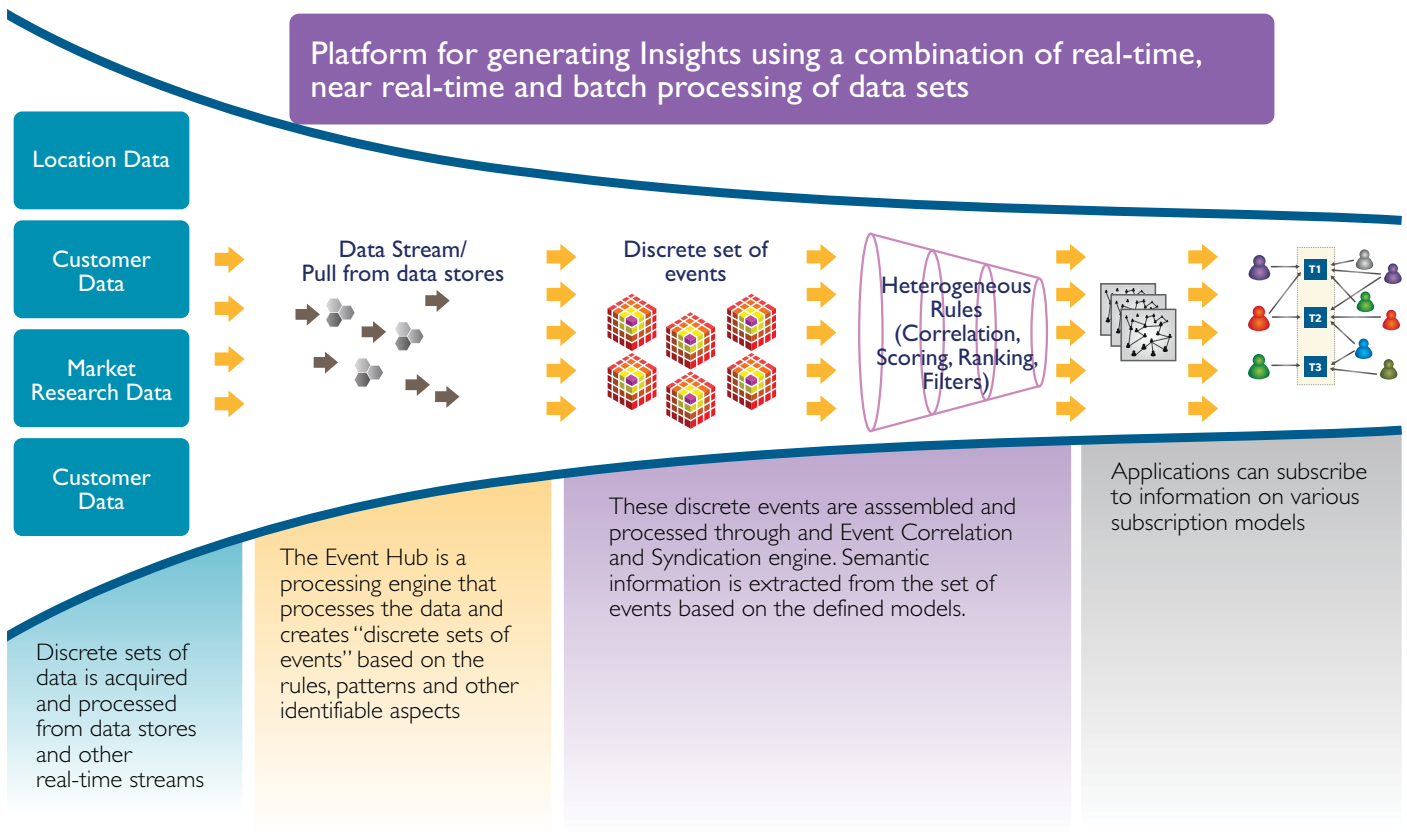
At the same time, the decision to migrate to a new data discovery platform is marred by the lack of proper knowledge and a number of associated concerns related to security, ROI, ease of use, etc. That apart, new data sources keep on adding to the enterprise data input.

The question here is whether the latest data discovery platform will be able to create placeholder signals dynamically for the new kind of data that is coming in or not.

Data Discovery Platform

Today, enterprises need a comprehensive discovery platform that can process huge volumes of data in real time to provide meaningful insights. This discovery or decision-making platform is able to assimilate data, structure, refine, provide exploratory capability, identify and evaluate various patterns, and help make insight-based decisions faster.

It offers layers that allow one to explore data, project trends and draw valuable insights using the known methods of analysis such as customer path analysis, social network analysis, mathematical analysis and more. The results/insights are readily visible on the dashboard. They are visual and easy to understand for the intended audience/decision maker.

A data discovery platform is agile. It allows more types of data (as shown in Figure 1.– location data, market research data, customer data) to be considered for the analysis, more number of variables/cases to be used in models. Also, data with short shelf life can be analyzed fast, thus enabling enterprises to make swift decisions with respect to launching new products and services faster.

A potential data discovery platform can be visualized in a simplistic manner (see Figure 1).

Figure 1: A Data Discovery Framework

Platform for generating Insights using a combination of real-time, near real-time and batch processing of data sets

Location Data

Customer Data

Market Research Data

Customer Data

Data Stream/ Pull from data stores

Discrete set of events

Heterogeneous Rules (Correlation, Scoring, Ranking, Filters)

T1 T2 T3

Discrete sets of data is acquired and processed from data stores and other real-time streams

The Event Hub is a processing engine that processes the data and creates "discrete sets of events" based on the rules, patterns and other identifiable aspects

These discrete events are asssembled and processed through and Event Correlation and Syndication engine. Semantic information is extracted from the set of events based on the defined models.

Applications can subscribe to information on various subscription models

Understanding the Data Discovery Framework

A data discovery platform is business user-friendly: Typically, the data scientists work in the back office assessing and monitoring volumes of data and produce reports, based on a hypothesis, such as, the success or failure of a product launch, customer satisfaction, campaign effectiveness, etc. for the benefit of the key decision makers in an enterprise.

With data discovery, this is changing. An intuitive interface enables users to explore data without much training or expertise in any analytics tool. The proprietary platform is structured to store and model data gathered from disparate sources (like customer, production, supplier, location, market research data, etc.) while the built-in layers (as shown in Figure 1. – data collection, event hub, event correlation and syndication engine and visual front-end) obviate the need for aggregates, summaries and pre-calculations.

A visual dashboard provides an interactive interface that empowers the frontline users to make important decisions related to a business operation much faster than usual. This greatly improves an enterprise's ability to perform analytics and engage in rapid decision making.

A data discovery platform creates a foundation for data security and protocols: Every enterprise has a set of enterprise security standards that are very rigid in terms of their data usage, data access by employees, types of data that should be accessible and more. However, as the new paradigm of decision-making evolves, components of data security and control, such as data governance, in an enterprise need to evolve as well.

The primary and most important requirement for building a strong foundation for enterprise data security is knowing and understanding the enterprise data and setting up data governance rules. This helps classify data and create data identity, which is the missing link for creating actionable data security and control policies.

Defining data within an enterprise (both internal and external) is crucial. It is important to know what kind of data an enterprise has, where it is stored, how and why is it stored there. This helps application of appropriate policies and controls for data protection.

Data discovery tools and software for visualization, integration, data migration, etc. help enterprises identify and locate sensitive structured and unstructured information and classify them. The entire process is automated, thus preventing anomalies.

BI Vendors Need Rethink

Current vendors who offer traditional BI systems, which comprise structured data warehouses and databases, time-consuming and expensive delivery models, will have to re-think about continuing to offer these products only as these are primarily IT driven and restrict deployment by specialised consultants.

The latest data and analytics tools are expected to provide opportunities to analyze data for the benefit of customers and the market. It is not only about making changes in internal processes, but also about evolving the products and services model in a positive manner.

In short, data discovery tools are affecting business directly at decision-making levels unlike their predecessors that involved processing huge quantum of data, both structured and unstructured, to enable data specialists in making recommendations to business leaders.

Conclusion: Early Adopters

A few enterprises that recognized the importance of customer and market-facing analytics, pioneered the march into adopting data discovery tools. This helped in better understanding the relationships among various types of data as an essential core of a business process.

The two key roles that data discovery tools play are:

- They help determine the new products and services to be developed and launched in the market

- Leverage data and analytics for decision making at speed and scale

Early adopters of discovery tools, while launching products and services, aided by data-driven insights, are also experimenting and learning to make more complex and data-driven decisions at speed and scale. Before implementing a new decision-making process at the production level, enterprises need to run pilots to check the feasibility of the process.

This will help them acquaint themselves with an effective discovery environment that would enable them to make rapid decisions and implement new offerings quickly in the longer run.

About the Author

Nitesh Jain is General Manager and Vertical Head - Consumer Goods (Europe & Emerging Markets) at Wipro.

Nitesh has more than 17 years of experience in spearheading different areas of technology led transformations. He has been leading the practice and business portfolios in the areas of analytics, BI, Trade promotions, Foreign exchange trading solutions and corporate performance management practices at Wipro and other leading organizations. He is actively engaged in the field of consulting, business analytics, performance management, data warehousing and implementing strategic analytical solutions for Retail, CPG, Manufacturing, Healthcare, Medical devices and Banking/Financial Services domains. In his career spanning 17 years, he has helped various leading retailers, Manufacturing Companies and Banks implement strategic analytics driven initiatives.

Nitesh holds a Master's in Business Administration from Pune University and has a Bachelor's degree in Mathematics.

About Wipro Ltd.

Wipro Ltd. (NYSE:WIT) is a leading Information Technology, Consulting and Business Process Services company that delivers solutions to enable its clients do business better. Wipro delivers winning business outcomes through its deep industry experience and a 360 degree view of "Business through Technology" - helping clients create successful and adaptive businesses. A company recognized globally for its comprehensive portfolio of services, a practitioner's approach to delivering innovation, and an organization wide commitment to sustainability, Wipro has a workforce of over 150,000, serving clients in 175+ cities across 6 continents. For more information, please visit www.wipro.com

DO BUSINESS BETTER