



Improving Access to Data for Successful Business Intelligence

Part 2: Supporting Multiple Analytical Workloads in a
Changing Analytical Landscape

By Mike Ferguson
Intelligent Business Strategies
March 2014

Prepared for:



Table of Contents

Introduction	3
The Changing Analytical Landscape.....	4
From data warehouse to analytical ecosystem.....	6
The Growth of Self-Service BI Tools	7
What Is Self-Service BI?	7
The Importance of Multi-Source Data Access In A Self-Service BI Environment.....	7
Conclusion	9

INTRODUCTION

Business is demanding business insight from data acquired from more data sources

This series of three white papers on “Improving Access to Data for Successful Business Intelligence” looks at the challenges companies are facing in delivering successful business intelligence (BI) during an era where the data landscape is growing in complexity and there is increasing business demand for more data from more sources to deepen business insight. In addition, there is an expectation from business that BI can be delivered more quickly in an agile way, regardless of an increasingly distributed data environment.

Business demand to analyse data from new data sources is resulting in a need to support multiple analytical workloads that go beyond the capabilities of a data warehouse

In Part 1 in this series, we looked at how data is on the march to becoming more distributed across more systems. We discussed how multiple instances of transaction processing systems exist in many organisations and how some of these are being deployed on the cloud. We also explored the growth in analytical systems and the emergence of big data platforms and NoSQL databases. The paper examined why the number of data sources is steadily climbing, with much more data coming into the enterprise from external and unfamiliar data sources, such as social media sites and external bodies.

There is also a growth in the deployment of self-service BI tools for visual discovery

In Part 2 in the series, we will look at the impact the growing numbers of data sources has had on analytical systems and the changing analytical landscape that has emerged from the need to support multiple analytical workloads that require platforms beyond that of a data warehouse. We will also look at the growth in adoption of self-service BI tools in use in business areas, both of which are driving up demand for maintaining good data access in an increasingly complex analytical environment.

In the final Part 3 we will look at how one vendor, Progress Software, is addressing these challenges in order help organisations be successful with their business intelligence initiatives.

TODAY'S CHANGING ANALYTICAL LANDSCAPE

New data sources are required to deepen customer insight

The need for deeper insight into customers and business operations has meant that companies now want to analyse new data sources both inside and outside the enterprise over and above the transaction processing systems already providing data into data warehouses and data marts.

Popular big data sources now in demand to get deeper insight into customers include:

- Web log clickstream data to understand what customers and prospects are doing online
- Customer interactions recorded in internal or cloud-based CRM systems such as in-bound emails
- Social media profile and interaction data (e.g. Twitter, Facebook)
- Review web sites
- Blogs

Clickstream, email and social media data are all in demand

Note that some of these data sources are unstructured and can include text from social media sites, emails, review web sites and others.

Sensor networks are being deployed to get data to improve insight into operations

With respect to operations, sensor data is now in demand as companies start to deploy sensor networks to instrument their operations. Sensors can measure everything from temperature, light, vibration, GPS location, pressure and more. This kind of data makes the following possible:

- Supply/distribution chain optimization
- Asset management and field service optimization
- Manufacturing production line optimization
- Location-based advertising (mobile phones)
- Grid health monitoring (e.g. electricity, water, cell network)
- Smart metering

There are many vertical industry use cases where analysis of sensor data can help optimise operations

Sensors are also being deployed inside products and equipment, opening up a whole set of new data about the performance and location of a product or asset.

Both structured and multi-structured data are needed to help deepen business insight

The issue with big data is that it is increasing in complexity in terms of data volume, data variety and data velocity (i.e., the rate in which data is being generated). Today, both structured and multi-structured data are needed for analysis. In the case of most structured data from OLTP systems, personal data stores (e.g. Excel, Access), these data are well understood and a schema is often defined. Multi-structured data on the other hand is often un-modelled and not well understood. Such data may also need new analytics to be able to analyse it.

The need to analyse these new kinds of data has taken us beyond the traditional data warehouse into a world where multiple analytical workloads now exist that warrant the use of new platforms beyond that of the data warehouse. Popular new big data analytical workloads include:

1. Analysis of data in motion
2. Complex analysis of structured data
3. Exploratory analysis of un-modelled multi-structured data
4. Graph analysis (e.g. social networks, fraud)
5. Accelerating extract, transform, and load (ETL) and analytical processing of un-modelled data to enrich data in a data warehouse or analytical appliance
6. The storage and re-processing of archived data

Several new analytical workloads have emerged to analyse big data

New big data analytical platforms are also needed to support these workloads

To support these new analytical workloads companies are starting to extend existing data warehouse architectures to accommodate new analytical platforms like Hadoop, NoSQL Graph DBMSs, streaming analytics systems and analytical relational DBMSs, in addition to traditional data warehouses. These are shown in Figure 1.

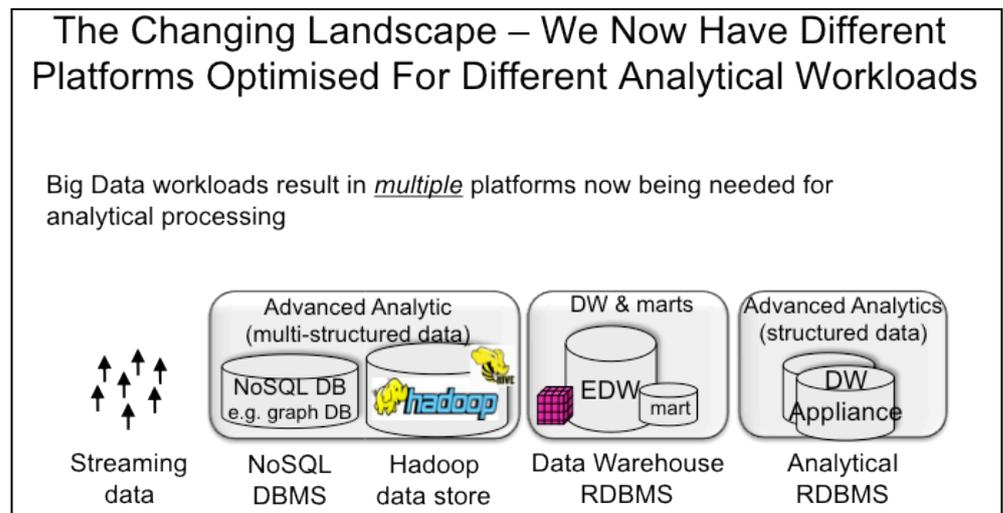


Figure 1

FROM DATA WAREHOUSE TO ANALYTICAL ECOSYSTEM

What this means is that we are extending beyond the traditional data warehouse into a new, more complex analytical ecosystem as shown in Figure 2.

The demand for access to new data sources together with new analytical workloads has resulted in the need for an enterprise analytical ecosystem

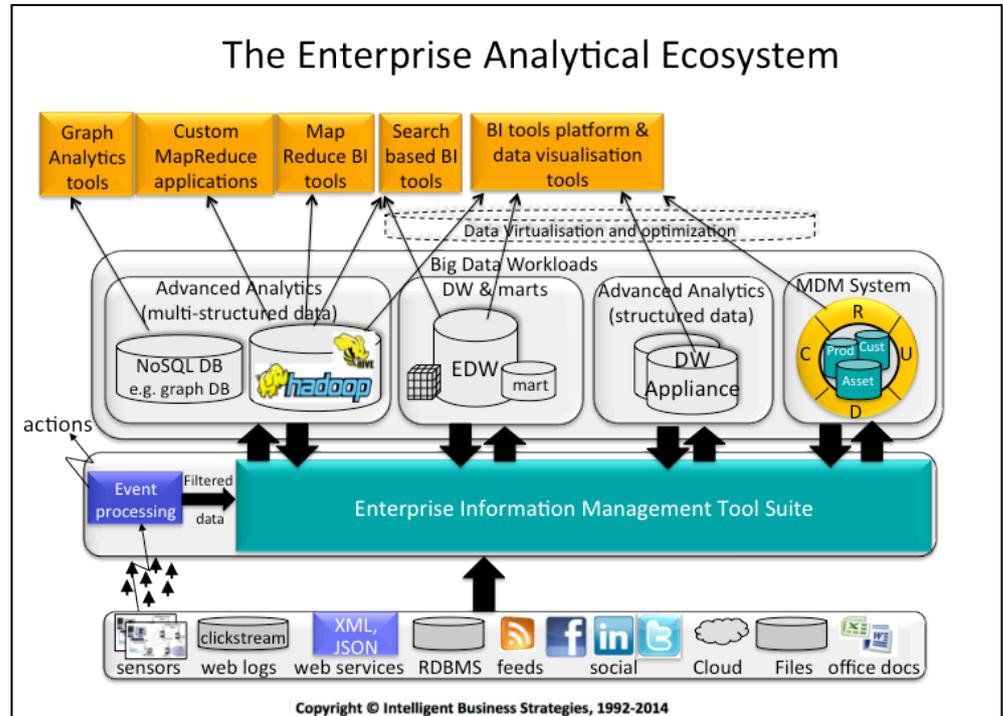


Figure 2: Big data workloads result in multiple platforms now being needed.

Different big data analytical workloads are suited to different big data platforms

As a general guideline, the six new analytical workloads outlined above would likely run on the following platforms.

Analytical Workload	Big Data Platform
1. Analysis of data in motion	Streaming analytics
2. Complex analysis of structured data	Analytical appliance
3. Exploratory analysis of un-modelled multi-structured data	Hadoop
4. Graph analysis e.g. social networks, fraud	NoSQL Graph DBMS
5. Accelerating ETL and analytical processing of un-modelled data to enrich data in a data warehouse or analytical appliance	Hadoop
6. The storage and re-processing of archived data	Hadoop

Many new data sources now need to be accessed to support traditional data warehouse and big data analytical workloads

This new analytical ecosystem consists of multiple analytical platforms and not just a data warehouse. Note the sheer number and variety of data sources shown at the bottom of Figure 2 that business now needs to access. There is a need for both structured and multi-structured data that has to be captured, cleansed, integrated and provisioned across the entire analytical ecosystem. Succeeding in this more complex analytical environment is now even more dependent on the ability to access data from a much larger set of data sources both inside and outside the enterprise.

THE GROWTH OF SELF-SERVICE BI TOOLS

In addition to the emergence of a multi-platform analytical ecosystem, one area of business intelligence has experienced unrivalled growth over the last few years. This area is self-service BI.

WHAT IS SELF-SERVICE BI?

Self-service BI is growing rapidly

Self-service BI can be defined as the creation of a BI environment whereby business users can create and access BI reports, queries, and analytics without the need for IT.

Self-service data discovery and visualisation tools have been widely adopted

A new type of BI has emerged to support self-service BI and that is data discovery and visualisation tools. These include tools such as Tableau, QlikView and Tibco Spotfire. Business analysts use self-service data discovery and visualisation tools to explore data, detect patterns and produce insight without the need for IT. In particular, they use these types of tools to:

Self-service BI tools allow business analysts to access and analyse data in multiple data sources

- Analyse personal data (e.g. Excel data) with more visual tools
- Quickly mash, explore and investigate personal and corporate data to determine if there is any value in it
- Explore analyse and visualise big data
- Produce insight quickly that they can share with others

They are highly interactive, support automatic charting and visualisations that make complex data easy to understand

There are many reasons why self-service data discovery and visualisation tools are popular with business units and business analysts. They include the fact that they speed time to value and allow exploration of personal and corporate data. They are highly interactive and improve business agility. Business can therefore clear the IT backlog. In addition, smarter visualizations (for example, graph matrices, bullet graphs, tree maps, heat maps) mean that information that is hard to understand in a tabular form can be easily visualised. Business analysts using these tools can produce dashboards to present information and guide people to take the correct action. They can also quickly share insights via user-assembled dashboards by publishing them to the web and to mobile devices.

THE IMPORTANCE OF MULTI-SOURCE DATA ACCESS IN A SELF-SERVICE BI ENVIRONMENT

One of the greatest strengths of self-service data discovery and visualisation tools is their ability to access multiple data sources and do data blending. Business analysts can therefore answer questions quickly, even though all the data they need is not in a single database. A good example of this is shown in Figure 3.

Having access to multiple data sources is critical to allowing business analysts to produce the insights they need

It means that business does not have to wait for IT to design a database and consolidate all the data required

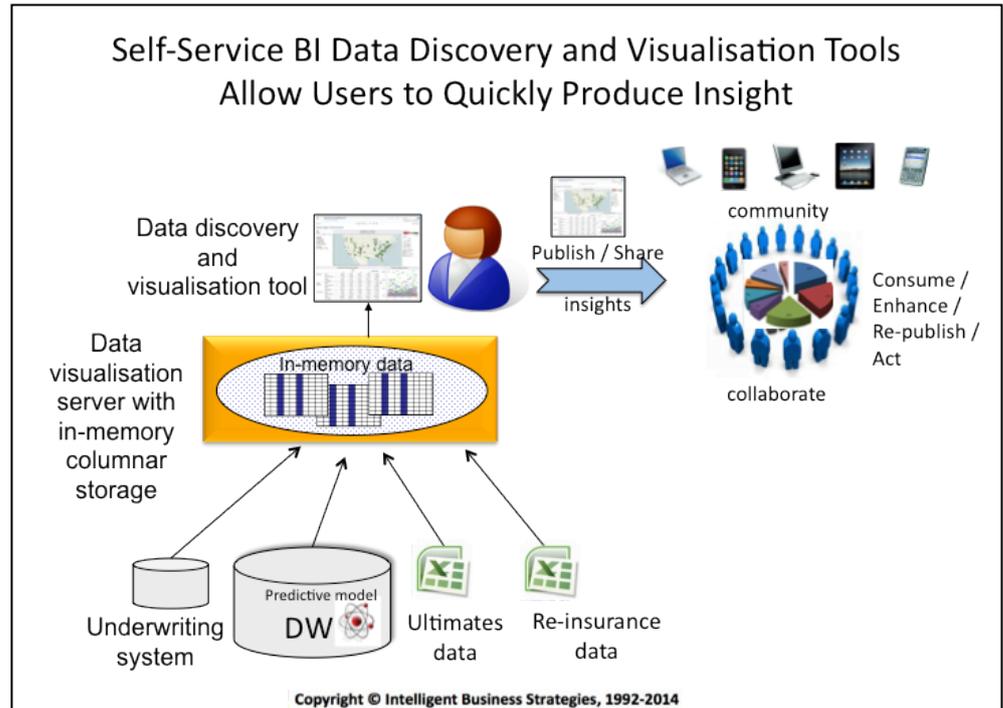


Figure 3: This scenario shows how an insurance non-IT business analyst is able to calculate net premiums and claims even when reinsurance data is not in the data warehouse.

Self-service BI speeds up time to value

Figure 3 shows an insurance example of how a business analyst can make use of a self-service data discovery and visualisation BI tool to access multiple data sources such as an underwriting system, a data warehouse, ultimate loss data and re-insurance data to allow the business analyst to calculate Net Premiums and Claims even when re-insurance data is not in the data warehouse.

CONCLUSION

The business need to understand more about their customers and about business operations is forcing a transition from a data warehouse to an analytical ecosystem

The demand for big data has caused an explosion in the number of new data sources

Self-service BI tool support for multiple data sources allows business analysts to continue to produce insights quickly, even though the number of data sources is increasing

In this paper we have seen that the commercial business drivers outlined in the first paper in this series are driving the need to be able to analyse structured, semi-structured and unstructured data from multiple sources inside and outside the enterprise. These business drivers are fuelling the arrival of big data into the enterprise, which in turn is forcing a transition of the analytical landscape to accommodate new analytical workloads beyond those capable of being supported by a data warehouse.

The demand for big data has caused an explosion in the number of data sources, many of which require external connectivity to cloud, data marketplaces and other hosted data stores.

In addition, the arrival of self-service BI data discovery and visualisation tools is allowing business users to quickly analyse data and produce their own actionable insights that can be easily shared with decision makers to accelerate time to action. Given the increasing distribution of data across more and more systems and the steady increase in the number of data sources, the job of accessing data is getting harder for business analysts. The fact that self-service BI tools support the ability to analyse data from multiple data sources and support lightweight data blending is a key reason why they are so successful.

Part 3 in this series will look at the key requirements for improving data access in an analytical environment and then explore how technology from one vendor, Progress Software, helps companies meet those requirements so that they can be successful with business intelligence.

About Intelligent Business Strategies

Intelligent Business Strategies is a research and consulting company whose goal is to help companies understand and exploit new developments in business intelligence, analytical processing, data management and enterprise business integration. Together, these technologies help an organisation become an *intelligent business*.

Author



Mike Ferguson is Managing Director of Intelligent Business Strategies Limited. As an analyst and consultant he specialises in business intelligence and enterprise business integration. With over 32 years of IT experience, Mike has consulted for dozens of companies on BI/Analytics, big data, data governance, master data management and enterprise architecture. He has spoken at events all over the world and written numerous articles and blogs providing insights on the industry. Formerly he was a principal and co-founder of Codd and Date Europe Limited – the inventors of the Relational Model, a Chief Architect at Teradata on the Teradata DBMS and European Managing Director of Database Associates, an independent analyst organisation. He teaches popular master classes in Big Data Analytics, New Technologies for Business Intelligence and Data Warehousing, Enterprise Data Governance, Master Data Management, and Enterprise Business Integration.



Water Lane, Wilmslow
Cheshire, SK9 5BG
England
Telephone: (+44)1625 520700
Internet URL: www.intelligentbusiness.biz
E-mail: info@intelligentbusiness.biz

Improving Access to Data for Successful Business Intelligence – Part 2

Copyright © 2014 by Intelligent Business Strategies
All rights reserved

