



Use of Big Data Technologies in Capital Markets

- Ruchi Verma, Sathyan R Mani

Abstract

Data is growing at a tremendous rate with an increase in digital universe from 281 Exabyte's (year 2007) to 1,800 Exabyte's (year 2011). However the increase in data is, in itself, a minor problem, but the increase in percentage of unstructured data in the overall data volume is what is concerning all, including Wall Street.

Capital market firms have been transforming organically to handle the big data over years. This article discusses the key transformations that capital market firms are undergoing to handle big data, drivers for use of big data technology in capital markets and relevant uses cases.

The article further outlines the future big data technologies to handle big data complexity and responsiveness – a sense & response system, a federated database and in memory analytics.

Data is growing at a tremendous rate with an overall increase in digital universe with petabytes of data flowing from messaging, emails, open source surveys, blogs, Twitters, Facebook and so on. According to studies conducted by top market research firms, the big data market is expected to grow from US\$3.2 billion in 2010 to US\$16.9 billion in 2015. The next big challenge for capital markets will be the velocity of data (value of extremely large and diverse data sets), considering the frequency of data generation and delivery in financial services industries like investment banking (highest stored data per firm), and the need to be still analysed in near real-time. Responding to these trends, capital market firms are setting aside a bigger pie of IT spend on big data technologies.

According to a report by a major analyst firm, using big data analytics in the field of customer intelligence in the investment banking sector in UK has reaped benefits of £554 million and it is expected to go up to £5,275 million by the end of 2017. Investment banks in UK are spending 4% of their gross operating surplus on big data R&D.

Capital market firms are also seeking competitive advantages by exploring the plethora of information from all possible sources and by transforming organically to handle the big data. The key technological transformations taking place in capital market to handle Big Data are –

Scale-out storage infrastructure that grows with data

The massive amount of stored and generated data means the scope for scale-up is reduced and the focus is more on “scaling out” which means use of shared pools of storage devices that deliver more capacity with same operational cost. Scale-out storage is the key to managing Big Data.

Convergence of data sources

A unified analytics platform that helps to collaborate between the structured and un-structured data. The increase in the unstructured data means more data to analyse. The applications that are used to analyse the Wall Street data (capital markets data) must be capable of collaboration and correlation. Big Data is just bits (data that makes no sense) until its value is unlocked.

Innovative applications

NASDAQ and NYSE Euronext stock exchanges have implemented various data-warehouse applications like IBM Netezza and EMC Greenplum to store big data. These tools/applications store the unstructured data got from various sources, handle the data and then process the data to make more sense out of the unstructured data.

Re-modelling of data models to make better, faster and less risky decisions

The traditional RDBMS (Relational Database Modelling Systems) cannot handle big data and hence there is a shift towards specialized, non-relational databases like Hadoop. Microsoft SQL, which was the standard for querying data out of RDBMS systems, is now being replaced by Map Reduce, which is the distributed querying and data processing engine used to extract data from big datasets hosted on clusters in any typical Hadoop implementation. The systems which have both, the structured and unstructured data are currently operated in mutually exclusive mode without interoperability. Vendors are trying to bridge the connectivity gap between RDBMS and the Hadoop systems

Use of cloud storage

The usage of shared pools via cloud storage is increasing the need for virtual infrastructure products.

Capital Market - Drivers for Big Data Technologies

Data management has been a big challenge for capital markets firms for decades. The financial industry has collectively spent billions of dollars over the years trying to create appropriate datasets. Data sharing across institutions continues to be a challenge, as each business unit prefers calculating using its own set of data, making enterprise-wide data analysis tremendously difficult. The sources of unstructured data from which data needs to be collected, analysed and stored are increasing. For instance, traders who are interested in particular companies or industries news, and advanced trading operations have developed tools to analyse news in real time and hence help them make efficient trading decisions. Now, with news coming in via additional channels such as audio, video or even through social media (Twitter), some firms are trying to come up with new ways to analyse all of this data. Firms are also analysing data from different sources such as documents, data from surveys, "click" data from websites, web search traffic data and more. As terabytes grow into petabytes, traditional relational database structures are having a difficult time in storing and handling the data.

Technologies play an important role in storage, analytics and processing of Big Data. In-memory tools can likely be deployed in four main areas of the business: front office, credit valuation adjustment, market risk management and client reporting. The current economic environment necessitates the business drivers for any investments.

Following are the key business drivers for using Big Data Technologies in Capital Market

Changing regulatory landscape

Financial services at large are seeing lots of regulatory make over in last few years. Capital Market firms are facing this new regulation landscape and are gearing up for meeting the upcoming regulations (Basel II, EMIR, Dodd Frank, MiFID etc.)

Changing trading strategies

Capital markets have evolved from simple strategies, like 1980s-paired models, to the intricate gaming strategies of today. Trading strategies have started including unstructured data (an example is Twitter-based trading strategies). In 2012, analysts expect to see an increased growth in the use of unstructured data in trading strategies and alpha discovery (Returns generation from various trades).

Increasing data volume

Market data volumes is growing really big. For years, capital markets firms have used advanced technology to store historical market data for their quants. However with the growing data sources (social media, Internet, mobility), the need for big data storage will become more profound

Increasing complexities of processes

There is an increase in data due to the increase in the complexities of the underlying business process

Stringent timeliness

In this tough economic environment, it is very important to find value in the data and be as fast as possible. Real-time execution is been seen to take the physical extremes by collocated high-frequency trading.

Need for handling speed, agility and control

The speed required for retrieving the right data, manipulating it and using the analysis to gain control of the market dynamics is one of the biggest business drivers.

Visibility of data

Since front and back offices are physically more distributed but more closely integrated from a business perspective, there is a need to adapt to changes which allows everyone to see what they need to see.

Need for integration of siloes systems

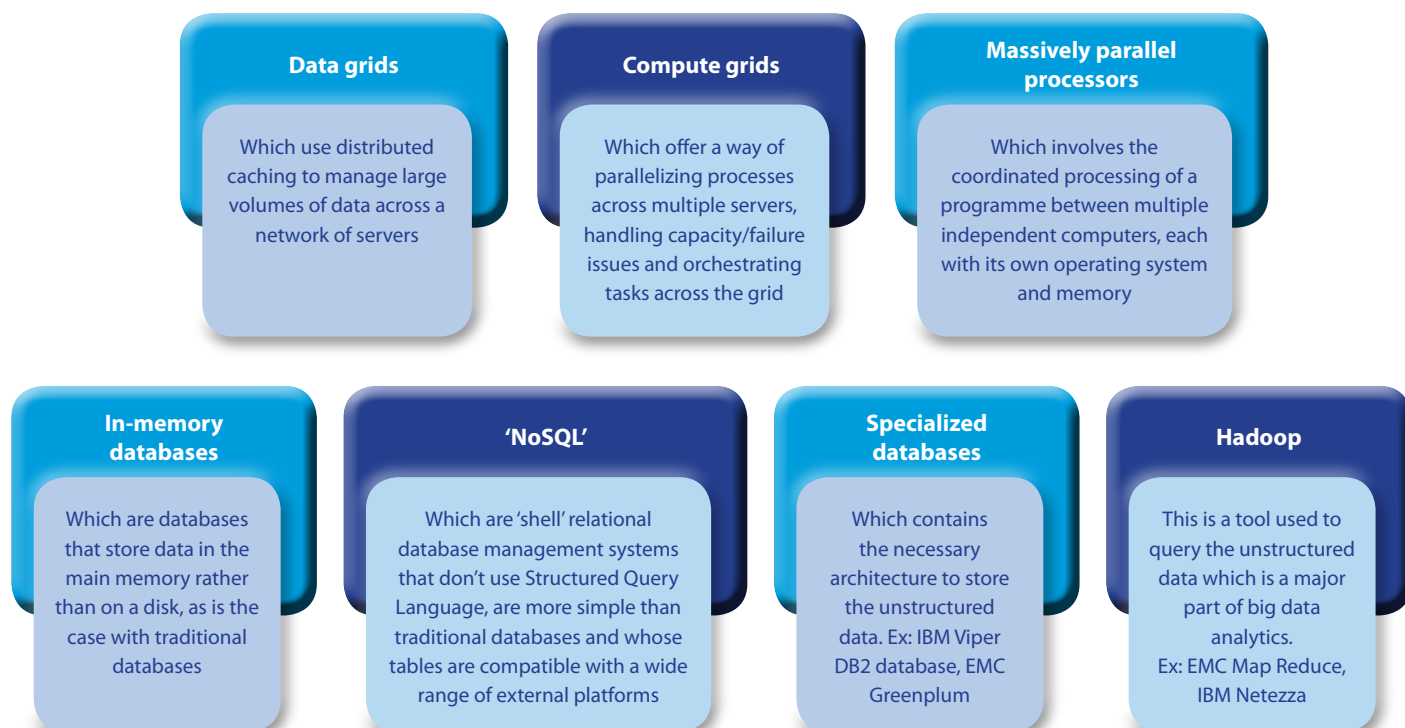
In order to understand the risks an investment bank is accumulating, it is important to connect the data from various disparate systems. The investment bank will not have an accurate view of potential risks unless they have a system in place which connects the risks and the analytical tools. This in turn means that they have to have a central platform for risk calculation.

Capital Market – Use cases of Big Data Technologies

Capital market firms are using big data (Unstructured data) primarily in five key areas –



Capital market firms are primarily using the following big data technologies –



Other Use Cases –

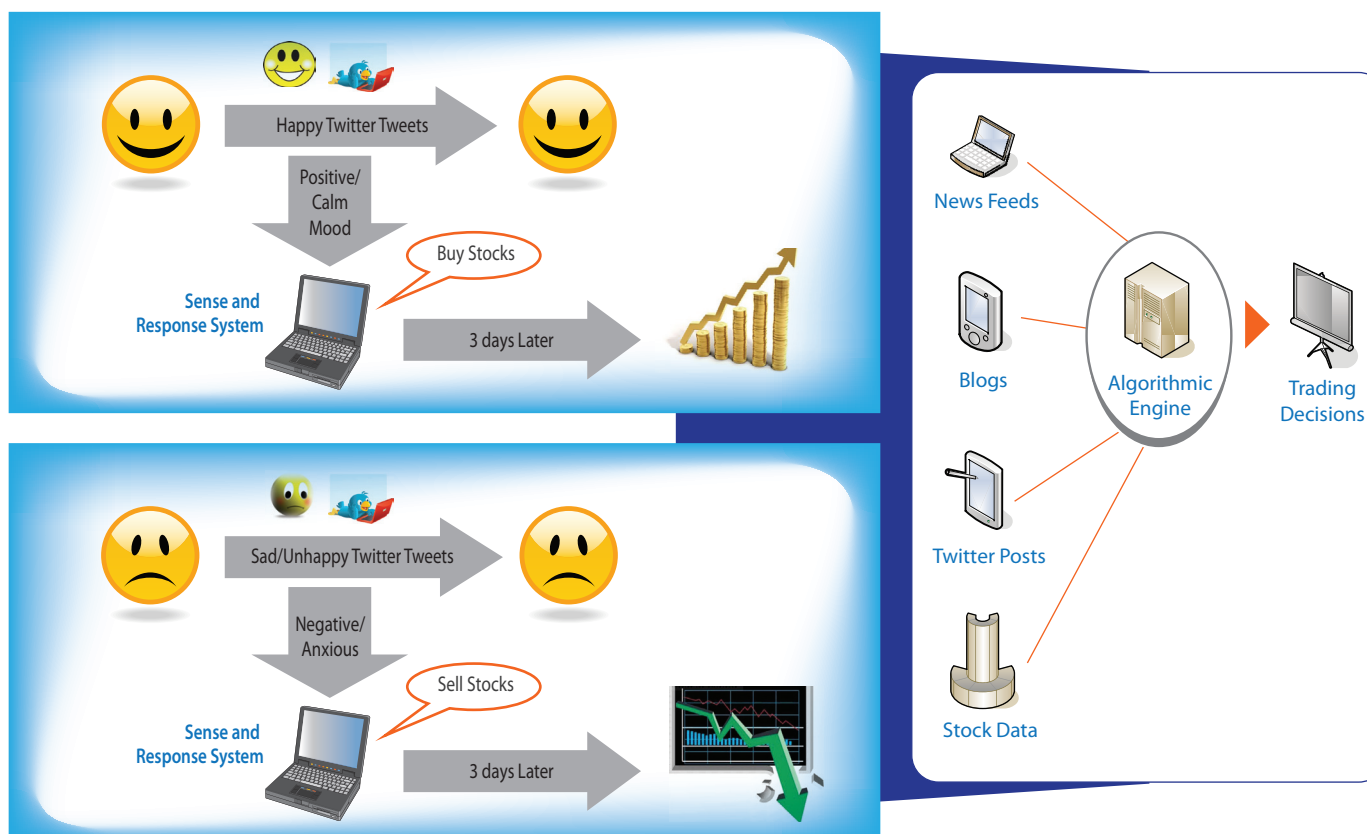
Firms	Big Data Use Cases	Areas
Investment Bank	An investment firm with assets of over US\$1 trillion and operations in approximately 50 countries uses big data technology to deliver Reference Data to the Murex trading platform and other downstream systems.	Reference Data Management
Investment Bank	An investment firm with assets of over US\$1 trillion and operations in approximately 50 countries uses big data to manage risk exposure through real-time communication across bond, futures and credits trading	Risk Analytics
US Investment Bank	An investment bank shifted the risk management and P&L towards a real-time environment. Big Data technologies were leveraged to help the firm to gather all relevant data into one place	Regulation
European Investment Bank	An investment bank used Big Data analytics to track performance monitoring, risk analytics and reporting	Risk Analytics & Regulation
Asian Investment Bank	An investment bank used Big Data technologies to generate on-demand performance metrics for risk measures across multiple global trading businesses.	Trading Analytics
European Investment Manager	An investment manager used Big Data technology to gather relevant details so as to respond as a witness to a litigation action against a prime broker	Compliance
US Investment Manager	Investment manager used Big Data technology to centralize data and applications to apply governance policies and mitigate risk of damages from litigation discovery	Risk Analytics & Regulation
Global Exchange	A major global exchange used Big Data technology to provide global market participants with on-demand access to data and data-mining tools for trading, analytics and risk management in a cloud-based/hosted environment	Trading Analytics & Risk Analytics
US Regulator	A US regulator used Big Data technology to create a searchable library of research, econometric and other information generated by the regulator's activities	Regulation
Buy Side firm	A major buy-side firm uses Big Data technologies for market surveillance, an activity requiring processing of vast quantities of market information	Regulation
Asset Manager	Fiduciary management – a new area of interest in which asset managers outsource management of their portfolios to third-party administrators in order to benefit from economies of scale.	Fiduciary Management – Emerging area
Regulatory compliance and advanced analytics	An investment bank use big-data techniques to handle and manage petabytes of data for regulatory compliance and advanced analytics. The bank used technology from Hadoop, an open source framework that supports data-intensive distributed computing, which allow data to be crunched over a distributed network of computers.	Regulation & Risk Analytics

Capital Market – Future Big Data Technologies

Buy side firms, sell side firms, stock exchanges and also regulators are all looking at implementing big data technologies to a greater extent. However as the amount of data is increasing, responding to data demands and handling data complexity will be the two key challenges faced by capital market institutions in handling Big Data.

Evolution of “Sense & Response System” to handle responsiveness

Although capital markets have previously been focused on high velocity market data, it is now moving toward unstructured data due to changing trading dynamics. Unstructured data these days primarily consists of daily stock feeds, Twitter chats and blogs, a reality that would have been considered science fiction even a few years ago. Converting the unstructured information into machine-readable data is the key to gaining an edge in Wall Street trading. This data is then used by algorithmic traders to produce some alpha (returns from the market) from the news. These untapped sources can help traders gain a competitive edge in trading. 20-30 years back, a trader could capture some Alpha (Returns) in a stock even after a week of a company-specific or industry-specific event having occurred. Today, that cushion is gone and stock returns move in milliseconds. A system that examines blogs and Twitter posts is part of a larger, smart platform, known as a “sense and response system”. Wall Street has to look forward to make Sense & Response System a part of its day-to-day processing. Below figure depicts an example of how an Investment Strategy can be determined using big data technology, Sense & Response System (Figure 1).



Use of the output from sense and response system as a parameter to the Algorithmic trading black box not only gives capital market firms a competitive advantage but also helps in timely trading decisions for better alpha.

Evolution of federated database approach for handling complexity of data

Complexity of storage: shared storage and parallel processing are two key aspects to handle complex data especially when data is in structured and unstructured format. The structured data is stored in Oracle Real Application Cluster (RAC) node 1 and the unstructured data in RAC node 2. The RAC is a key to providing shared storage that is used in cloud computing. Taking into consideration the concept of parallel processing, databases like Teradata and Greenplum have multiple buckets for storing data, which helps in accessing the structured and unstructured data in parallel and thereby improving access response times. Using the above proposed mix of shared storage and parallel processing, complex big data can be handled proactively.

Evolution of in memory analytics

The increasing amount of data being made available to investment banks, coupled with the need to enter the capital markets as early as possible and desire to conduct real-time analytics to reduce latency is driving the trend towards in-memory analytics. In-memory software can improve the speed of data analytics by a factor of 14,000. Hence a request that previously took 5 hours can now be accomplished in just one second. The increase in speed is gained by putting data analytics into a CPU, into 'memory'; as opposed to being carried out on disk. Queries sent to on-disk databases take longer because mechanical interaction take place, which is removed when query is, conducted in-memory.

Conclusion

Big Data technology implementation comes with its own cost. In spite of this, capital market firms are seen heavily investing in these technologies not only to stay competitive but also to ensure investors earn optimal ROI from their investments. The 3 V's of data (variety, volume & velocity) play an extremely important role in capital market. Firms, who are still thinking of investing in Big Data technologies, should gear up soon before it becomes too late to remain competitive.

References

- A Team Group Survey – Big Data Solutions in Capital Market
- www.bigdata.com/blog/
- www.emc.com/microsites/bigdata/index.html

About Infosys

Infosys partners with global enterprises to drive their innovation-led growth. That's why Forbes ranked Infosys 19 among the top 100 most innovative companies. As a leading provider of next-generation consulting, technology and outsourcing solutions, Infosys helps clients in more than 30 countries realize their goals. Visit www.infosys.com and see how Infosys (NASDAQ: INFY), with its 150,000+ people, is Building Tomorrow's Enterprise® today.



For more information, contact askus@infosys.com

www.infosys.com

© 2012 Infosys Limited, Bangalore, India. Infosys believes the information in this publication is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of the trademarks and product names of other companies mentioned in this document.