Applications in Finance for BIG DATA

Thomas P. Oberst **March 18, 2015**

Advanced Strategic Technology
27 Snow Street, Sherborn, MA 01770
Phone: (508) 650-3561 | Email: thomasoberst@advancedstrategictechnology.com

Corporate Finance

Corporate Finance - Mergers and Acquisitions

Financial Accounting, Auditing, Compliance, Reporting

Federal Reserve - Currency

Federal Reserve - Interbank Lending

Capital Markets - Government Borrowing

Capital Markets - Commercial Paper

Capital Markets - Venture Capital

Capital Markets - Crowd Funding

Capital markets - Currency

Brokerage & Trading on Financial Assets

Banking - Consumer

Banking - Residential

Banking - Commercial

Banking - Investment

Banking - International

Mutual Funds Management

Insurance - Fiduciary & Business Activity & Assets

Insurance - Consumer Asset

Insurance - Life

Real Estate

Money Management

Big Data = Big Data + Big Data Analytics + Machine Learning + Predictive Modeling

Big Data = Volume, Variety, Velocity, Veracity, (Viability and Value)

IBM has coined a worthy V – "veracity" – that addresses the inherent trustworthiness of data. The uncertainty about the consistency or completeness of data and other ambiguities can become major obstacles. As a result, basic principles as data quality, data cleansing, master data management, and data governance remain critical disciplines when working with Big Data.

Our first task is to assess the viability of that data because, with so many varieties of data and variables to consider in building an effective predictive model, we want to quickly and cost-effectively test and confirm a particular variable's relevance before investing in the creation of a fully featured model. [1]

Correlation does not mean causation.[2]

Call for Papers

2015 IEEE International Conference on Big Data (IEEE BigData 2015)

In recent years, "Big Data" has become a new ubiquitous term. Big Data is transforming science, engineering, medicine, healthcare, finance, business, and ultimately society itself. The IEEE Big Data has established itself as the top tier research conference in Big Data. The first conference IEEE Big Data 2013 (http://cci.drexel.edu/bigdata/bigdata2013/ , regular paper acceptance rate: 17.0%) was held in Santa Clara , CA from Oct 6-9, 2013 with more than 400 registered participants from 40 countries. The IEEE Big Data 2014 (http://cci.drexel.edu/bigdata/bigdata2014/index.htm, regular paper acceptance rate: 18.50%) was held in Washington DC, Oct 27-30, 2014 with more than 600 registered participants from 45 countries. The 2015 IEEE International Conference on Big Data (IEEE BigData 2015) will continue the success of the previous IEEE BigData conferences. It will provide a leading forum for disseminating the latest research in Big Data Research, Development, and Applications. We solicit high-quality original research papers (including significant work-in-progress) in any aspect of Big Data with emphasis on 5Vs (Volume, Velocity, Variety, Value and Veracity) relevant to variety of data (scientific and engineering, social, sensor/IoT/IoE, and multimedia-audio, video, image, etc) that contribute to the Big Data challenges.

[1] THE MISSING V'S IN BIG DATA: VIABILITY AND VALUE http://www.wired.com/2013/05/the-missing-vs-in-big-data-viability-and-value/

[2] THE MISSING V'S IN BIG DATA: VIABILITY AND VALUE http://www.wired.com/2013/05/the-missing-vs-in-big-data-viability-and-value/

Does Value also pertain to the source of the data having the value set that you desire? [34]

Does Value pertain to the variables you want to select?

The Need for Domain Specific Knowledge

Distributed Data Storage/File System

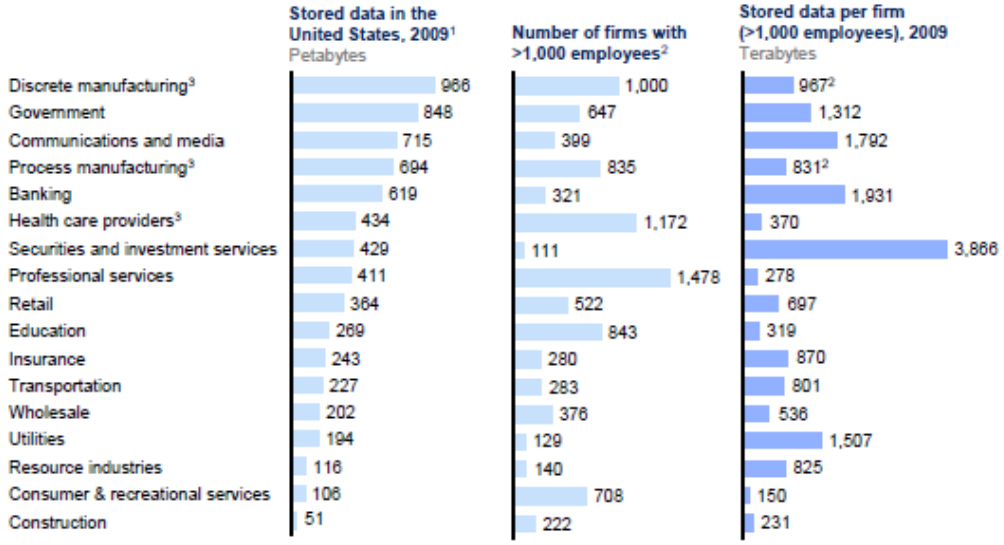
Structured Data - Fixed Fields - Relational - Spread Sheets

Semi-structured - XML

Unstructured - Free formed text, audio, video, imaging

Exhibit 7

Companies in all sectors have at least 100 terabytes of stored data in the United States; many have more than 1 petabyte



1 Storage data by sector derived from IDC.
2 Firm data split into sectors, when needed, using employment
3 The particularly large number of firms in manufacturing and health care provider sectors make the available storage per company much smaller.

SOURCE: IDC; US Bureau of Labor Statistics; McKinsey Global Institute analysis

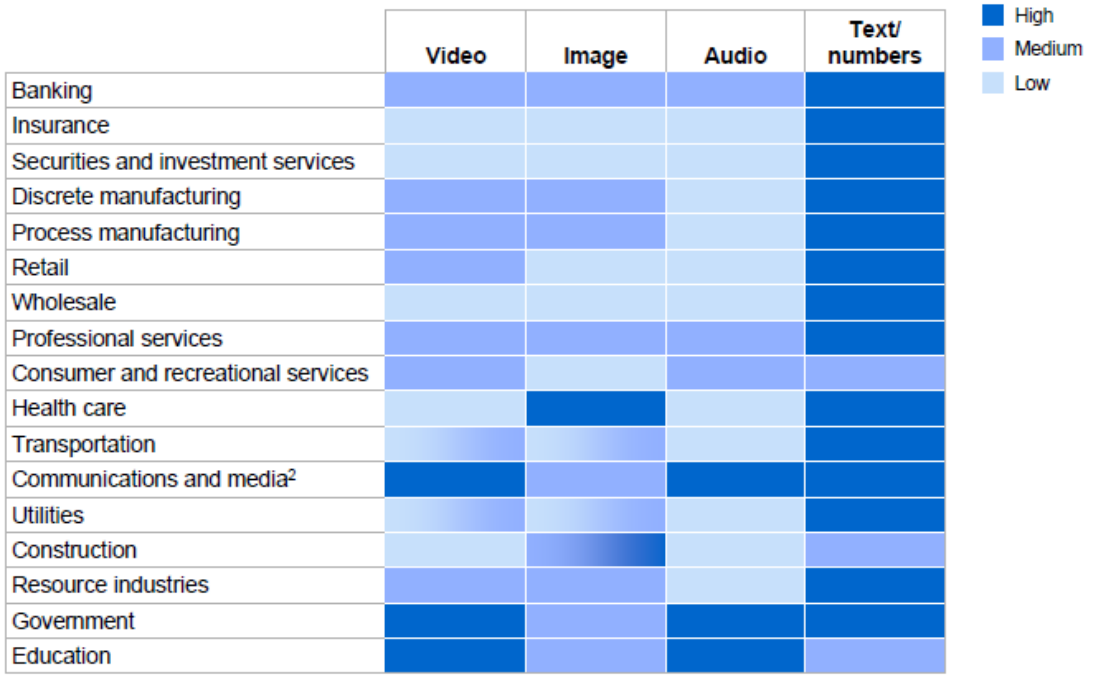
5

[3] **A Survey of Network Traffic Monitoring and Analysis Tools** http://www.cs.wustl.edu/~jain/cse567-06/ftp/net_traffic_monitors3/index.html#Section3.0

[4] **List of network protocols (OSI model)**
http://en.wikipedia.org/wiki/List_of_network_protocols_%28OSI_model%29

Exhibit 8

The type of data generated and stored varies by sector[1]

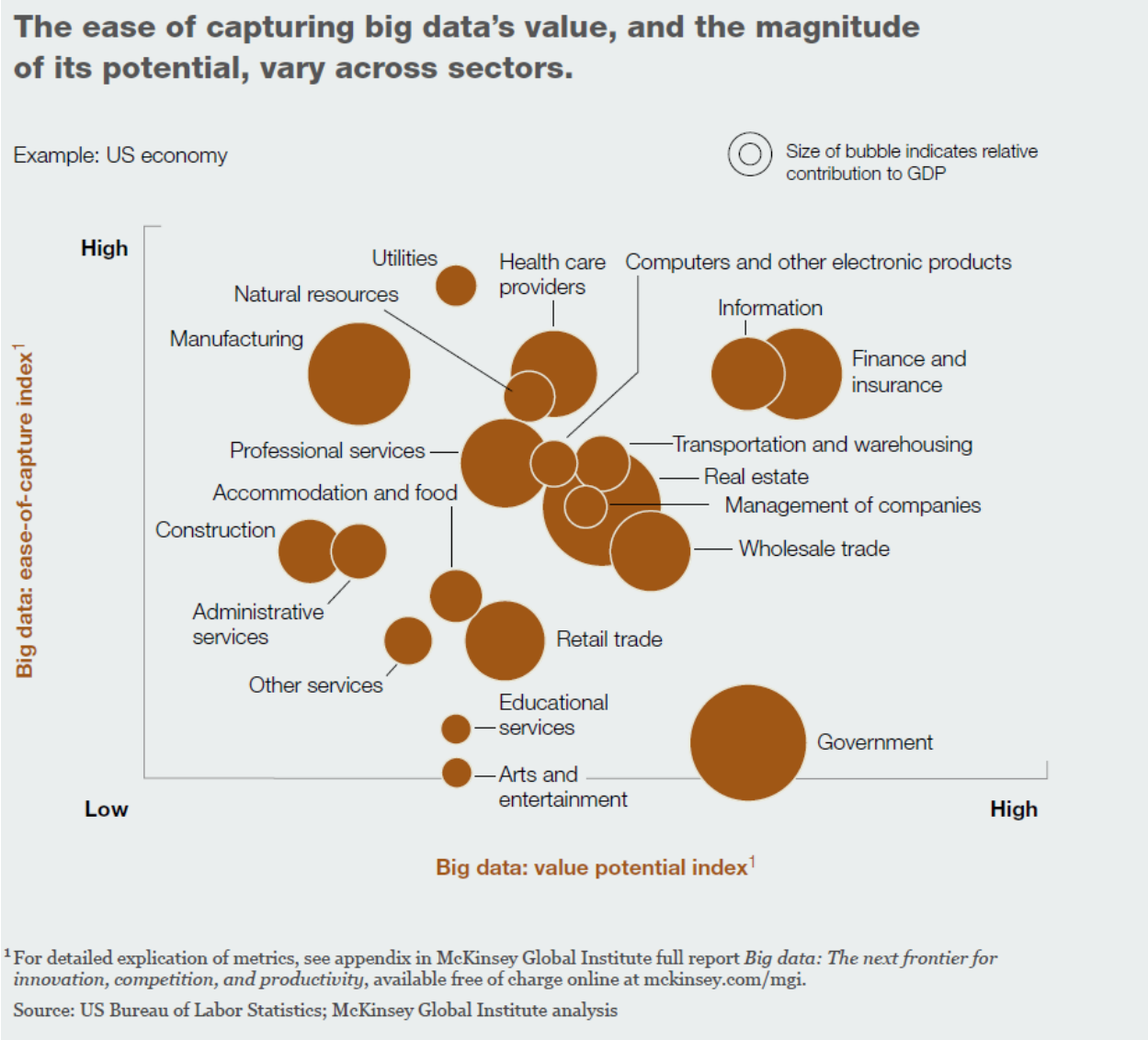


	Video	Image	Audio	Text/ numbers
Banking				
Insurance				
Securities and investment services				
Discrete manufacturing				
Process manufacturing				
Retail				
Wholesale				
Professional services				
Consumer and recreational services				
Health care				
Transportation				
Communications and media[2]				
Utilities				
Construction				
Resource industries				
Government				
Education				

1 We compiled this heat map using units of data (in files or minutes of video) rather than bytes.
2 Video and audio are high in some subsectors.
SOURCE: McKinsey Global Institute analysis

Satellite imagery

[5] **Big data: The next frontier for innovation, competition, and productivity**
McKinsey Global Institute May 2011



Insurance

Consider how this has affected underwriting in personal auto insurance. Instead of relying only on internal data sources such as loss histories, which was the norm, auto insurers started to incorporate behavior-based credit scores from credit bureaus into their analysis when they became aware of empirical evidence that people who pay their bills on time are also safer drivers. While the use of credit scores in private-auto-insurance underwriting has been a contentious issue for the industry with consumer groups, the addition of behavioral and third-party sources was a significant leap forward from the claims histories, demographics, and physical data that insurers analyzed in the past[6]

[6] **Unleashing the value of advanced analytics in insurance**

Workshop on High Performance Computing for Industry

http://www.rpi.edu/hpcw/program.html

http://mediasite.itops.rpi.edu/Mediasite5/Play/637cf8369ab74e7199e88bc169400ecc1d?catalog=d11e2ef6-6f93-4a09-a858-93c9510909c8

Dr. John E. Kelly III

Senior Vice President and Director, IBM Research

http://www.rpi.edu/hpcw/bios/JohnKelly.html

Stanley Young

Chief Executive Officer, NYSE Technologies

http://www.rpi.edu/hpcw/bios/StanleyYoung.html

http://mediasite.itops.rpi.edu/Mediasite5/Catalog/Full/d11e2ef66f934a09a85893c9510909c821

Jeremy Kepner

https://scholar.google.com/citations?user=BSrwwfYAAAAJ&hl=en

https://scholar.google.com/scholar?hl=en&as_sdt=1,22&q=%22Jeremy+Kepner%22&scisbd=1

http://arxiv.org/abs/1406.4923

Achieving 100,000,000 database inserts per second using Accumulo and D4M

http://arxiv.org/ftp/arxiv/papers/1406/1406.4923.pdf

Jeremy Kepner, William Arcand, David Bestor, Bill Bergeron, Chansup Byun, Vijay Gadepally, Matthew Hubbell, Peter Michaleas, Julie Mullen, Andrew Prout, Albert Reuther, Antonio Rosa, Charles Yee (MIT)

(Submitted on 19 Jun 2014)

The Apache Accumulo database is an open source relaxed consistency database that is widely used for government applications. Accumulo is designed to deliver high performance on

unstructured data such as graphs of network data. This paper tests the performance of Accumulo using data from the Graph500 benchmark. The Dynamic Distributed Dimensional Data Model (D4M) software is used to implement the benchmark on a 216-node cluster running the MIT SuperCloud software stack. A peak performance of over 100,000,000 database inserts per second was achieved which is 100x larger than the highest previously published value for any other database. The performance scales linearly with the number of ingest clients, number of database servers, and data size. The performance was achieved by adapting several supercomputing techniques to this application: distributed arrays, domain decomposition, adaptive load balancing, and single-program-multiple-data programming.

Article

Adapting to digital consumer decision journeys in banking

A host of emerging technologies are poised to personalize consumer experiences radically. Here's how banks can prepare.

February 2015 | byEdwin van Bommel and David Edelman

http://www.mckinsey.com/insights/financial_services/adapting_to_digital_consumer_decision_journeys_in_banking

Discovery. Banks must apply advanced analytics to the large amount of structured and unstructured data at their disposal to gain a 360-degree view of their customers. Their engagement strategies should be based on an empirical analysis of customers' recent behaviors and past experiences with the bank, as well as the signals embedded in customers' mobile or social-media data.

Example: The new normal

Diane is on a business trip. She lands at Chicago's O'Hare Airport and walks through the terminal toward the gate for her connecting flight to Toronto. As she passes a billboard for her bank, she receives a text message offering her a credit card upgrade, one with better travel perks than the one she has. When she opens the message, she is led to a customized web page that provides a benefits-based comparison of her existing card and the new one the bank is recommending. She simply needs to tap an "apply" button to start the purchase process. When she does, another message appears prompting her to take and upload a "selfie" so the bank can authenticate her.

The card is added to her mobile wallet, and Diane is given the option of making it her default payment method for any of the top ten online merchants she deals with. She selects seven of the ten, and within minutes of completing the task, she receives a message from the bank thanking her for her response and offering her an online coupon for a free drink from Chicago Coffee Roasters, a regional chain that happens to have a stand only two gates from where Diane is waiting for her flight.

The Case for Big Data in the Financial Services Industry

WHITEPAPER

Michael Versace Karen Massey

September 2012

It is increasingly vital for firms to harness Big Data into insights that help inform actionable, optimized, and timely decisions; keep risks at anticipated and acceptable levels; and uncover opportunities to stay ahead of the competition. For example:

- The NYSE creates 1 terabyte of market and reference data per day covering the use and exchange of financial instruments. In comparison, Twitter feeds generate 8 terabytes of data per day (or 80MB per second) of social interactions.
- 10,000 payment card transactions execute per second across the globe.
- 210 billion electronic payments were generated worldwide in 2010. This number is expected to double by the end of the decade.
- Between 2009 and 2014, the total number of U.S. online banking households will increase from 54 million to 66 million.
- In 2012, 46% of financial services CIOs are exploring the potential of cloud computing, up from 33% in 2010.
- Market data volumes grew 10x between 2007 and 2011 and are still growing strong.
- Some of the top European insurers report a sixfold increase in the amount of data and analytic reporting required by just the first pillar of Solvency II insurance reform regulation.
- IDC Financial Insights estimates that worldwide spending on core financial crime and fraud management solutions and infrastructure will top $28 billion in 2012, a growth rate of over 8% compared with 2011.

TABLE 1

Big Data Use Cases: 2012 Examples

Case	Achievements
European hedge fund	In-memory analytics to optimize price discovery and investment strategies for large portfolio trades and swaps
Global investment bank	Common operational data stores to speed post-trade settlement, confirmation, and access to common data
Retail banking innovation leader	Using geolocation data to create merchant intelligence and assist in optimizing offers and pricing to retail customers; driving mobile banking growth
Asia/Pacific national bank	Tracking social media into finely tuned market campaigns
Expanding U.S. property insurer	Granular microtargeting of customer segments and individuals with specialized pricing based on historical risk performance and forecast data
Global European institution	Using social analytics to gauge sentiment toward key product and service initiatives
Investment research institution	Using data from private sector satellite companies to understand traffic patterns and parking lots and fill rates of major retailers; fill rates and trends inform investment advice
Community bank	Analyzing transactional and unstructured data (voice) collected to anticipate workloads and staffing needs in call centers and branches

Source: IDC Financial Insights, 2012

Use of Big Data Technologies in Capital Markets

http://www.infosys.com/industries/financial-services/white-papers/Documents/big-data-analytics.pdf

Capital Market – Use cases of Big Data Technologies

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

Other Use Cases –

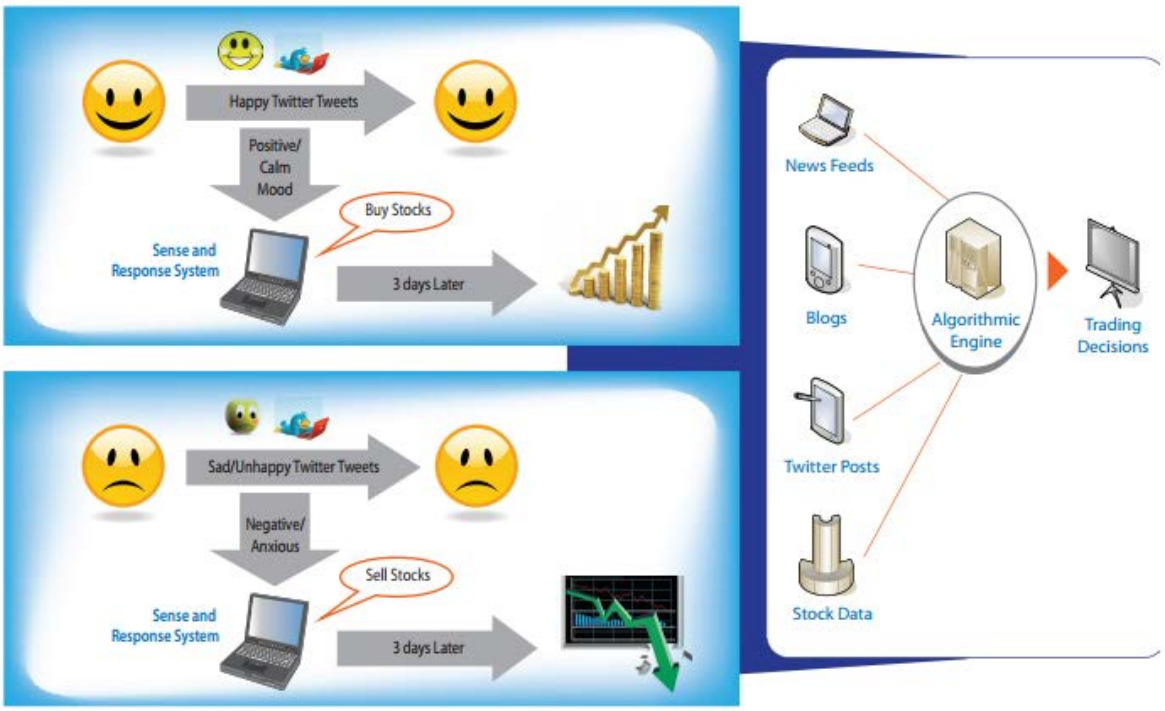
Firms	Big Data Use Cases	Areas
Investment Bank	An investment firm with assets of over US$1 trillion and operations in approximately 50 countries uses big data technology to deliver Reference Data to the Murex trading platform and other downstream systems.	Reference Data Management
Investment Bank	An 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

Must use Federated Approach - fits Big Data - processing where the data is located.

Must use in Memory Analytics - fits with Wall Street (Flash)

High Frequency Trading

Ultra High Frequency Trading



Big Data: Algorithms, Analytics, and Applications

edited by Kuan-Ching Li, Hai Jiang, Laurence T. Yang, Alfredo Cuzzocrea

https://books.google.com/books?id=ylG3BgAAQBAJ&pg=PA329&lpg=PA329&dq=chapter+17+Big+Data+in+Finance&source=bl&ots=PGlxiPkKKW&sig=bI0kUPIFPtoU6Wpgnz6UVjwtR9k&hl=en&sa=X&ei=hZYJVbGQLZOwsAS52YGIDg&ved=0CD8Q6AEwAg#v=onepage&q=chapter%2017%20Big%20Data%20in%20Finance&f=false

CHAPTER 17

Big Data in Finance

Taruna Seth and Vipin Chaudhary

CONTENTS

BACKGROUND

The financial industry has always been driven by data. Today, Big Data is prevalent at various levels of this field, ranging from the financial services sector to capital markets. The availability of Big Data in this domain has opened up new avenues for innovation and has offered immense opportunities for growth and sustainability. At the same time, it has presented several new challenges that must be overcome to gain the maximum value out of it. This chapter considers the impact and applications of Big Data in the financial domain. It examines some of the key advancements and transformations driven by Big Data in this field. The chapter also highlights important Big Data challenges that remain to be addressed in the financial domain.

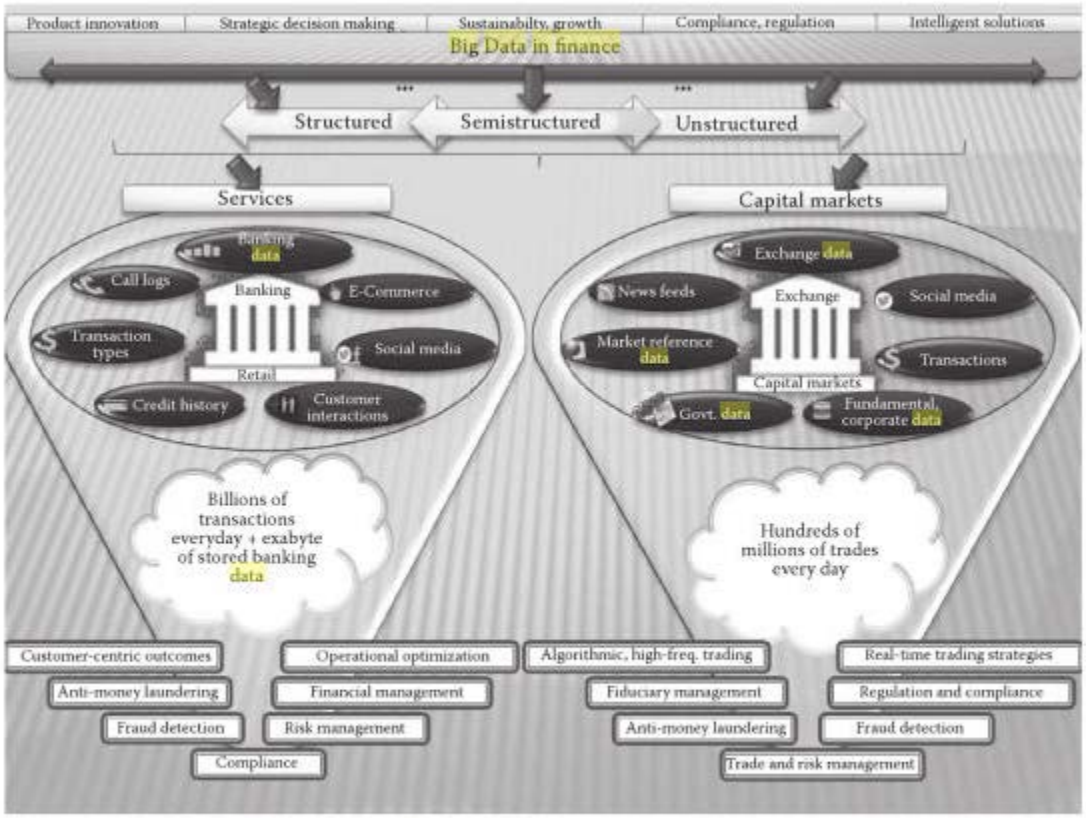


FIGURE 17.1 Big Data applications in key financial domain sectors.

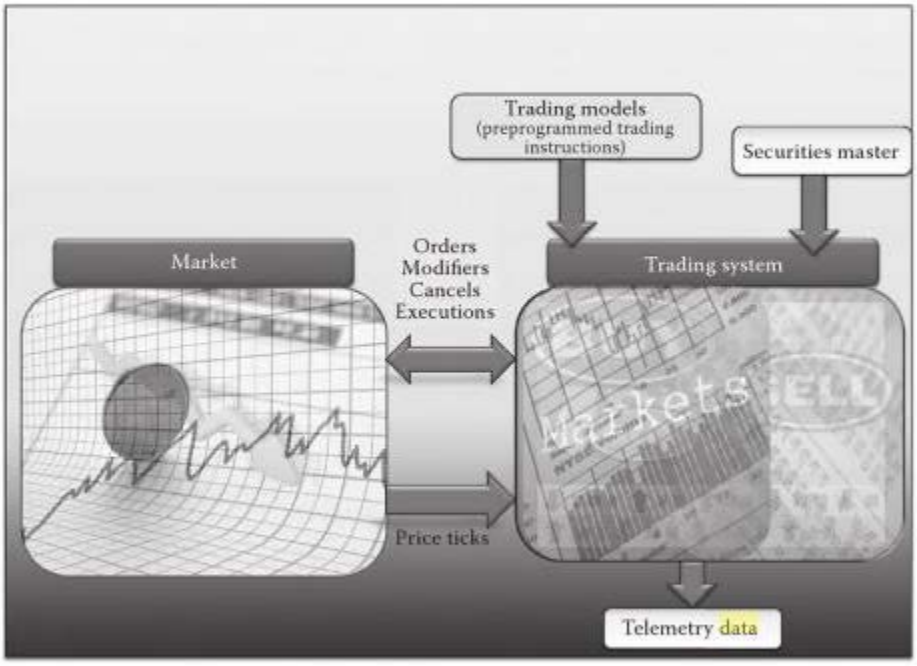


FIGURE 17.2 A typical automated electronic trading system.

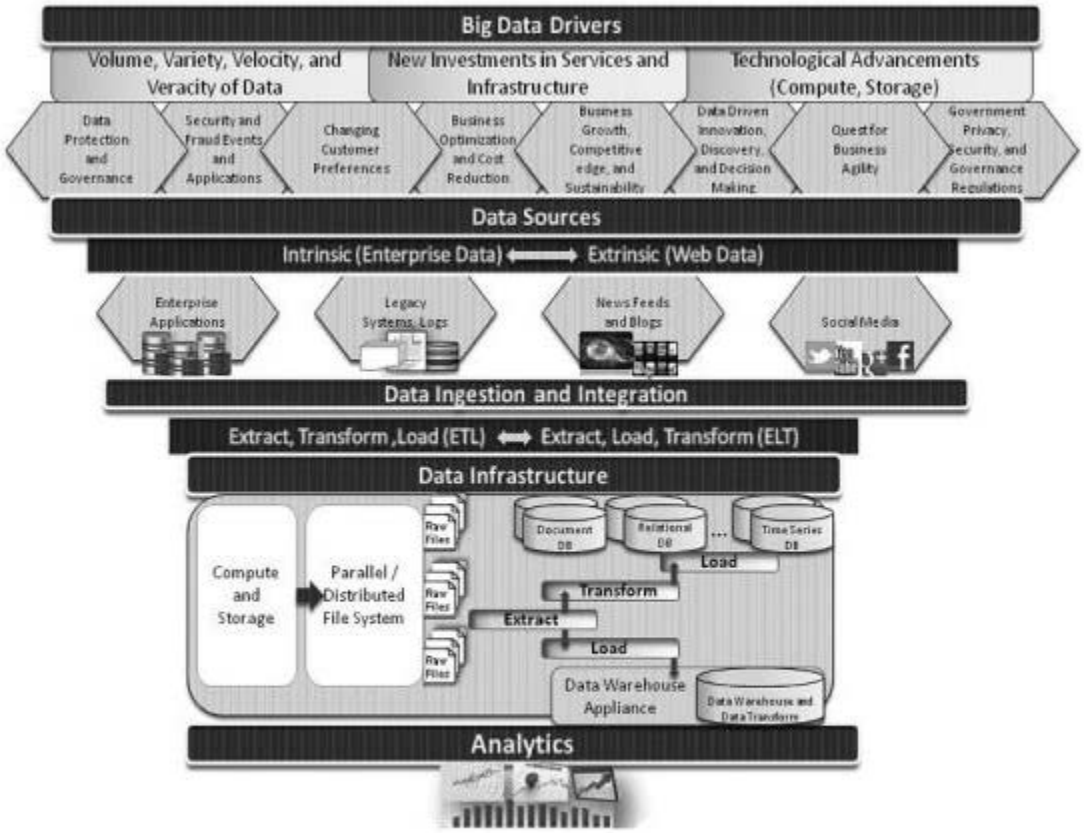


FIGURE 17.4 Big Data ecosystem in finance.