

Big Data Analytics: Technologies, Opportunities, and Challenges

Murali K. Pusala

Jan 11, 2018

Big Data Analytics

- **Description:** Essentials of Big Data analytics. Topics include: challenges and opportunities posed by Big Data in a variety of domains, predictive analytics or other advanced methods to extract value from data, innovative statistical techniques to glean insights from data, frameworks for parallelizing data pre-processing and data analytics, such as, Hadoop and Spark, and distributed algorithms to accelerate knowledge discovery.
 - 3 Credit(s). 0 Hour(s) Lab. 3 Hour(s) Lecture
 - Prereq: Math 362 and CMPS 460(G) or permission of instructor required.

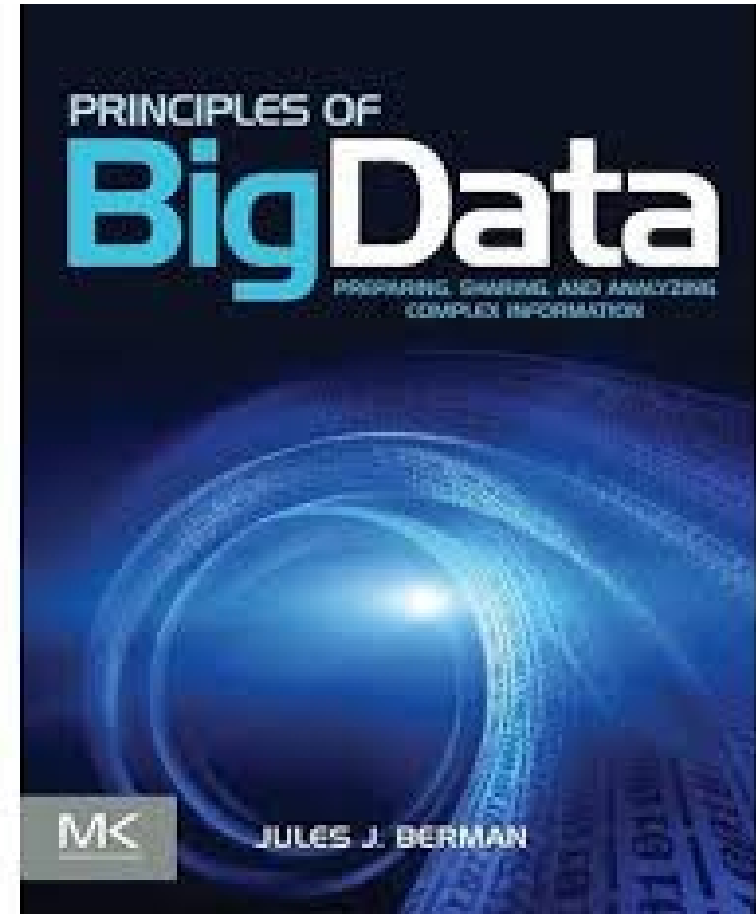
CMPS 490 / CSCE 598

Big Data Analytics

Title Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information

Publisher Morgan Kaufmann Publishers Inc.
San Francisco, CA, USA ©2013

ISBN 9780124047242



Who is Generating Big Data?



- Facebook : 1.39 billion monthly users
- YouTube: 300hrs video uploaded/min
- Instagram:70 Million photos/day



- 4.9 billion devices (2015)



- Walmart: 1 million transactions each hour
- Walmart holds 2.5 petabytes

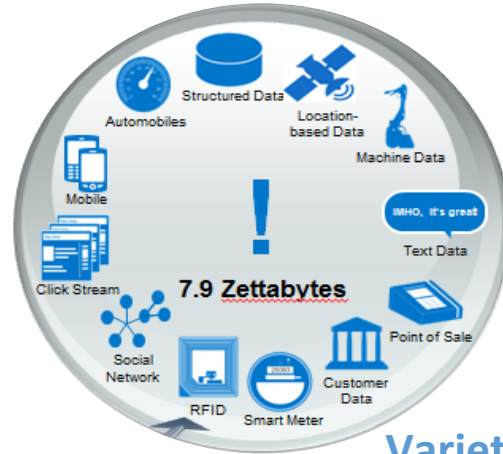


- ASKA Pathfinder: 36 antennas stream 250GB/sec/antenna
- LHC at CERN:15 petabytes/year

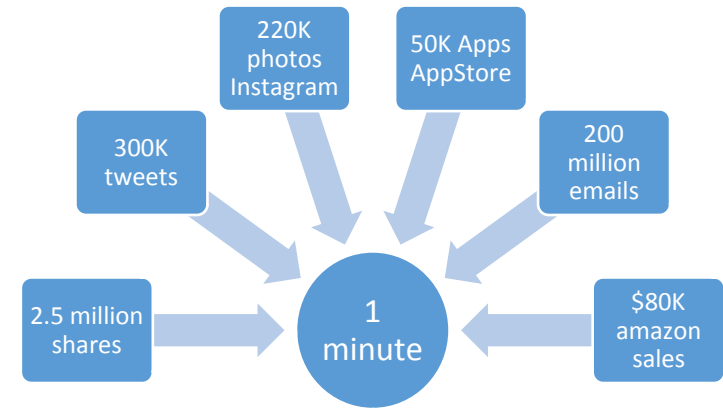
Big Data Overview



Volume



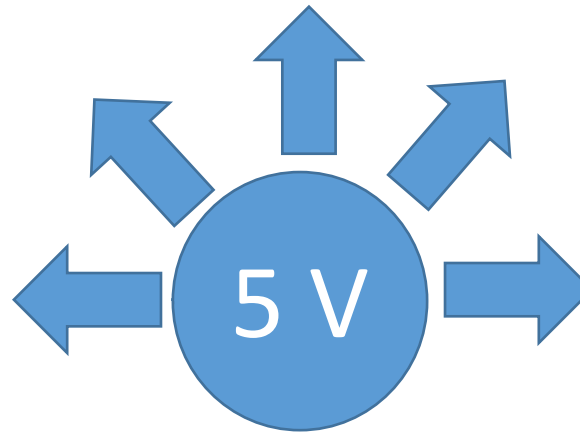
Variety



Velocity

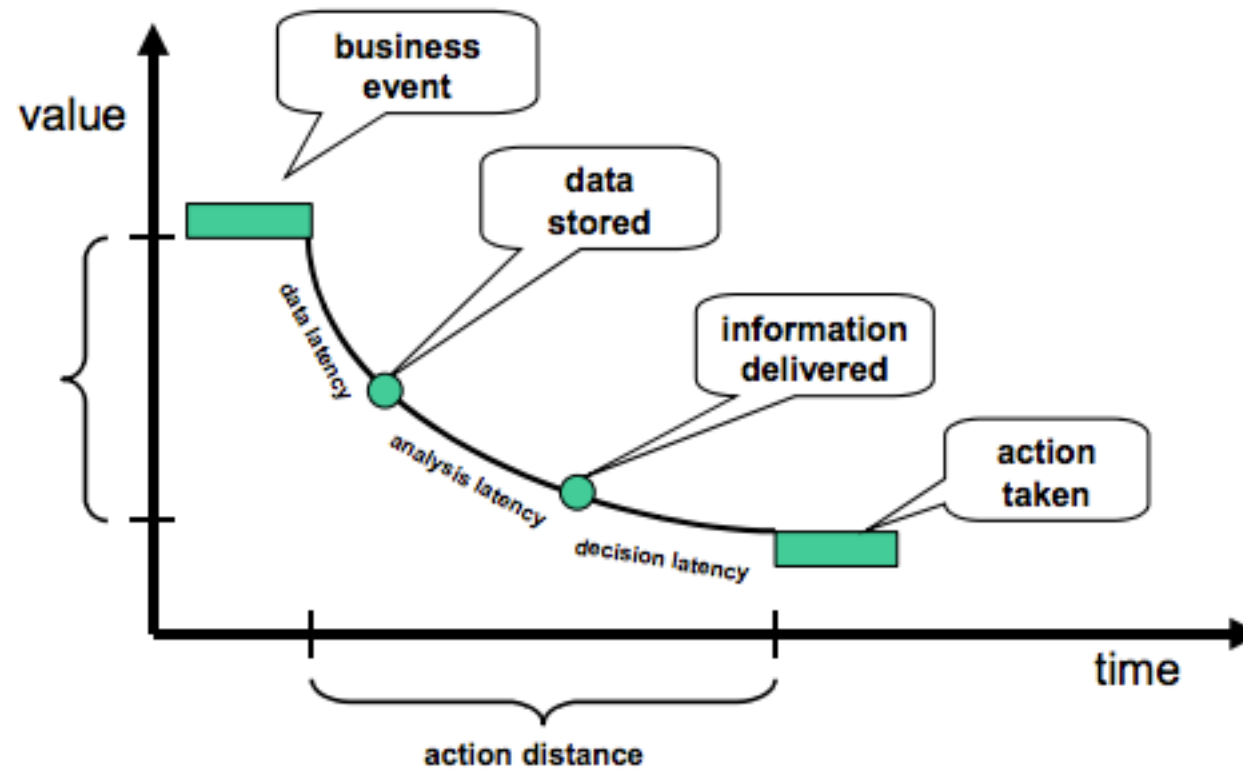


Variability

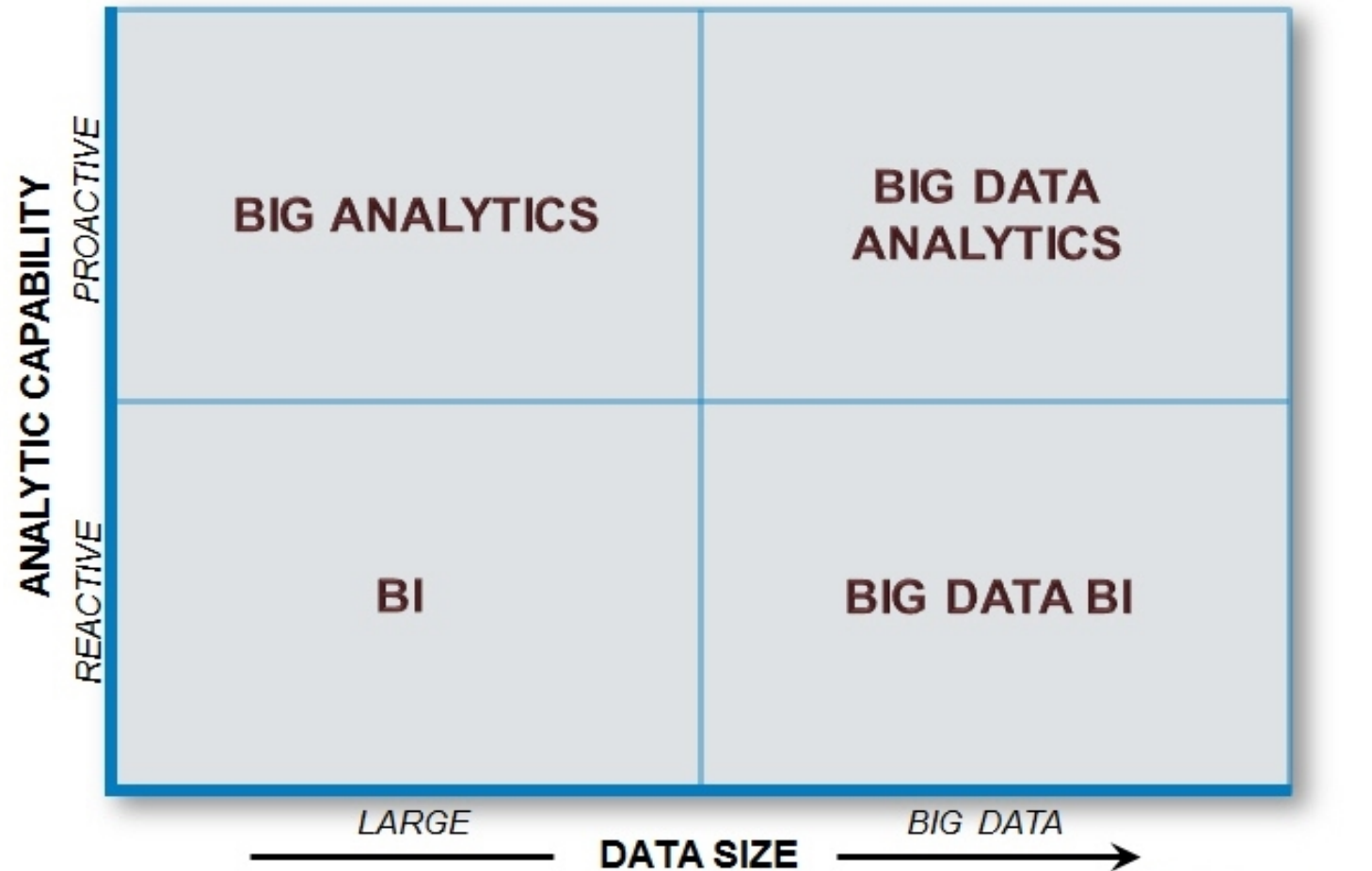


Veracity

Velocity- Time to action vs. Value *Hackathorn, 2002*



Kinds of Big Data Problems *Davis, 2012*



Name	Short scale (U.S., Canada and modern British)
Million	10^6
Milliard	
Billion	10^9
Trillion	10^{12}
Quadrillion	10^{15}
Quintillion	10^{18}

Source: wikipedia

The total amount of digital data will reach [180 zettabytes](#) in 2025. Approximately 80 percent of this data will be unstructured...

Megabyte (1 000 000 Bytes)

Gigabyte (1 000 000 000 Bytes)

Terabyte (1 000 000 000 000 Bytes)

Petabyte (1 000 000 000 000 000 Bytes)

Exabyte (1 000 000 000 000 000 000 Bytes)

Zettabyte (1 000 000 000 000 000 000 000 Bytes)

Yottabyte (1 000 000 000 000 000 000 000 000 Bytes)

Opportunities

- Analyzing Big Data helps in informed decision-making
- Advantages of Big Data
 - Understanding and targeting customers
 - Understanding and optimizing business process
 - Improve science and research
 - Improving healthcare
 - Optimizing machine and device performance

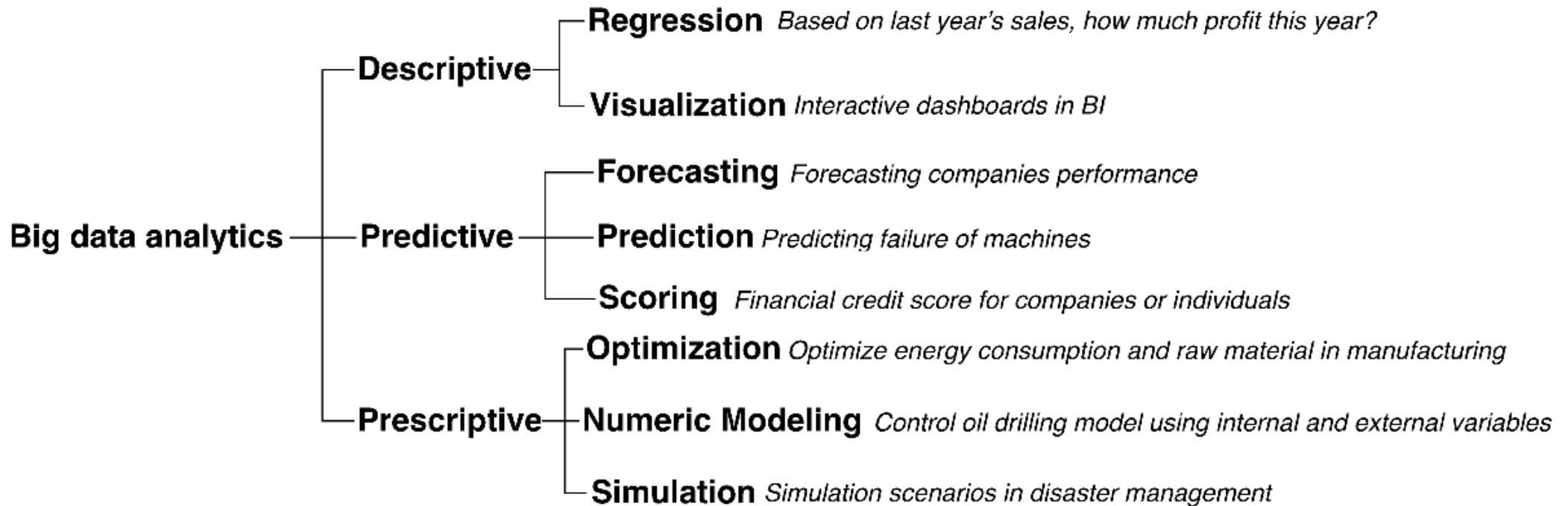
Applications

- Southwest analyses sensor data on their planes in order to identify patterns that indicate a potential malfunction or safety issue
- Oil and Gas exploration industries optimize drilling process by using earth's sedimentation characteristics, temperature, pressure, soil type, depth, chemical composition, molecular structures, seismic activity, machine data
- Aurora Health care Center saved \$6 Million by using analysis on patient's history, allergies, medications, environmental conditions, etc.,
- Predictive analysis also helps to change price dynamically to attract consumers and maximize profits

Challenges

- Processing Data
 - Processing data at speed it's been generated
- Storing Data
 - Scalable, Reliable and Faster storage
- Interpreting Data
 - Understanding Heterogonous and untagged data
- Privacy and Security

Types of Big Data Analytics *Pusala, 2016*



Big Data Analytics Platforms

- Big Data Analytics Platforms
 - MapReduce
 - Apache Hadoop
 - Apache Spark
- Data Management Systems for Big Data Analytics
 - Hadoop Distributed File System
 - NoSQL Databases
 - Graph Databases

What is Data Science

- Just having platforms and big data are not enough, we need to have “Data Science” to fill the gap between platforms and data.

- Data science employs techniques and theories drawn from many fields such as statistics and machine learning to extract knowledge and insights from big data by leveraging big data platform

- Turing award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational and now data-driven) and asserted that "everything about science is changing because of the impact of information technology and the data deluge".

How is Big Data Related to Me?

- There will be a shortage of talent necessary for organizations to take advantage of big data. By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions. (Source: McKinsey Global Institute; [Big data: The next frontier for innovation, competition, and productivity](#))

#16

Highest Paying Job
in Demand

3,433

Number of Job
Openings

\$105,395

Average Base
Salary

#1

Best Job in America
for 2016

Sources: [25 Best Jobs in America](#) and [25 Highest Paying Jobs in America for 2016](#)

References:

- Davis, J. (2012). What Kind of Big Data Problem Do You Have? SAS Blogs Home. Retrieved December 16, 2012, from <http://blogs.sas.com/content/corneroffice/2012/10/08/what-kind-of-big-data-problem-do-you-have/>
- M. K. Pusala, M. Amini Salehi, J. R. Katukuri, Y. Xie, and V. Raghavan, “Massive Data Analysis: Tasks, Tools, Applications, and Challenges BT - Big Data Analytics: Methods and Applications,” S. Pyne, B. L. S. P. Rao, and S. B. Rao, Eds. New Delhi: Springer India, 2016, pp. 11–40.
- Hackathorn, R. (2002). Current practices in active data warehousing. available: <http://www.dmreview.com/whitepaper/WID489.pdf>

Thank You

Q&A