BIG DATA OVERVIEW(CHALLENGES AND ADVANTAGES)

ABDULLAH ABDULABBAS NAHI AL-RABEEAH

COMPUTER SCIENCE DEPARTMENT

CIHAN UNIVERSITY

ACADEMIC YEAR 2018/2019

BIG DATA OVERVIEW

Extremely large data sets that may be analyses computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions.

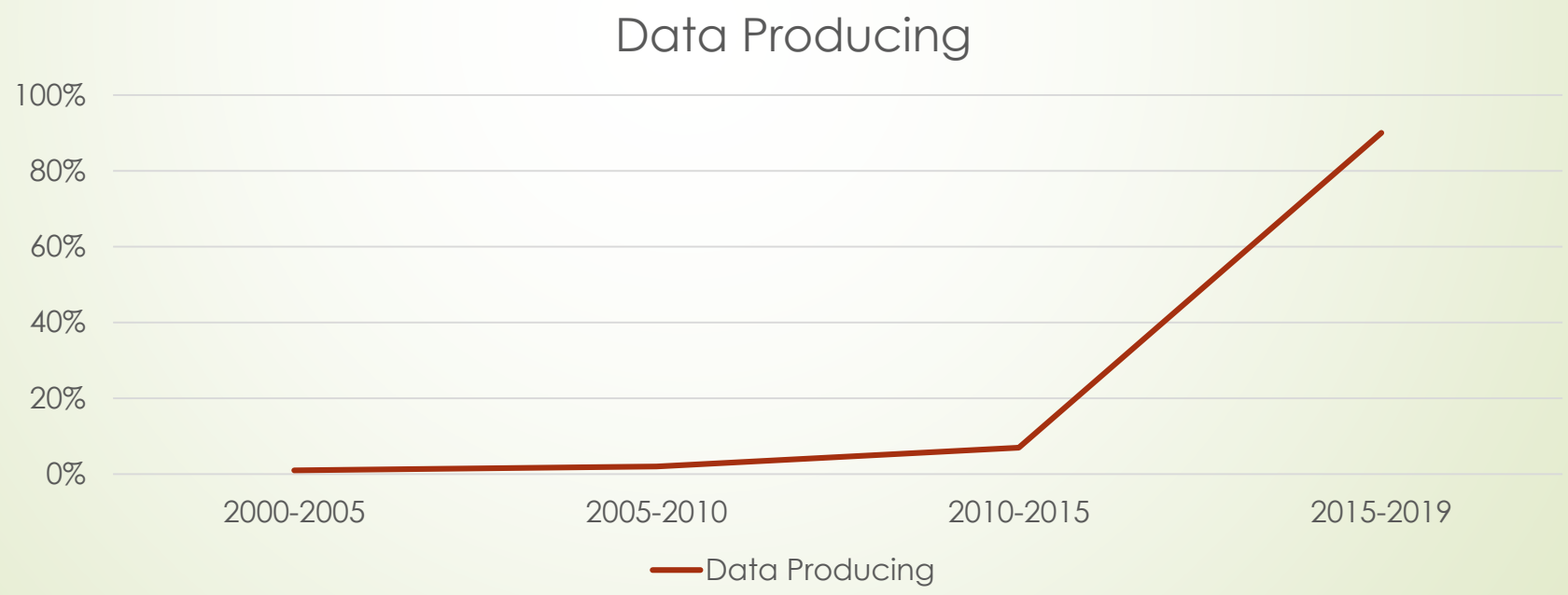
Examples :

1. Our daily activities
2. Schools and Universities
3. Hospitals and Restaurants
4. Social Network Sites
5. IoT devices
6. Organizations

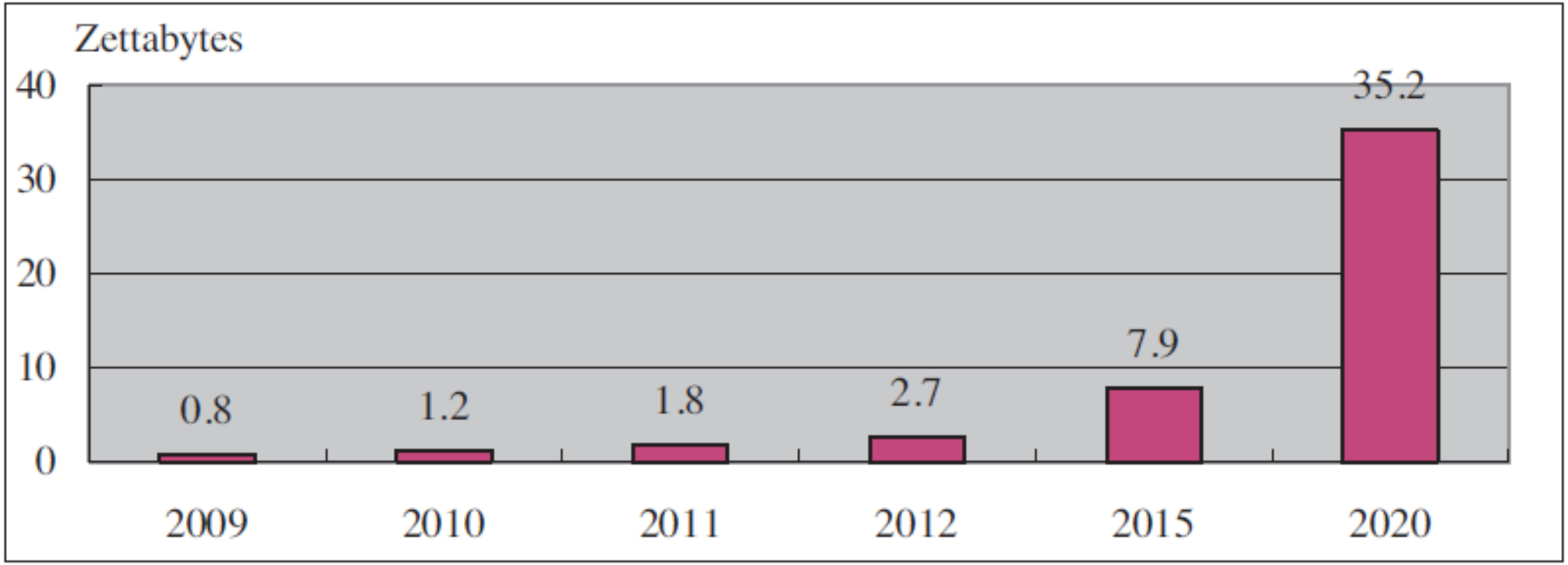
DATA PRODUCING STATISTICS

According to IBM 2019 statistics :

There are 2.5 quintillion bytes of data created each day at our current pace, but that pace is only accelerating with the growth of the Internet of Things (IoT). Over the last two years alone 90 percent of the data in the world was generated.



DATA PRODUCING STATISTICS

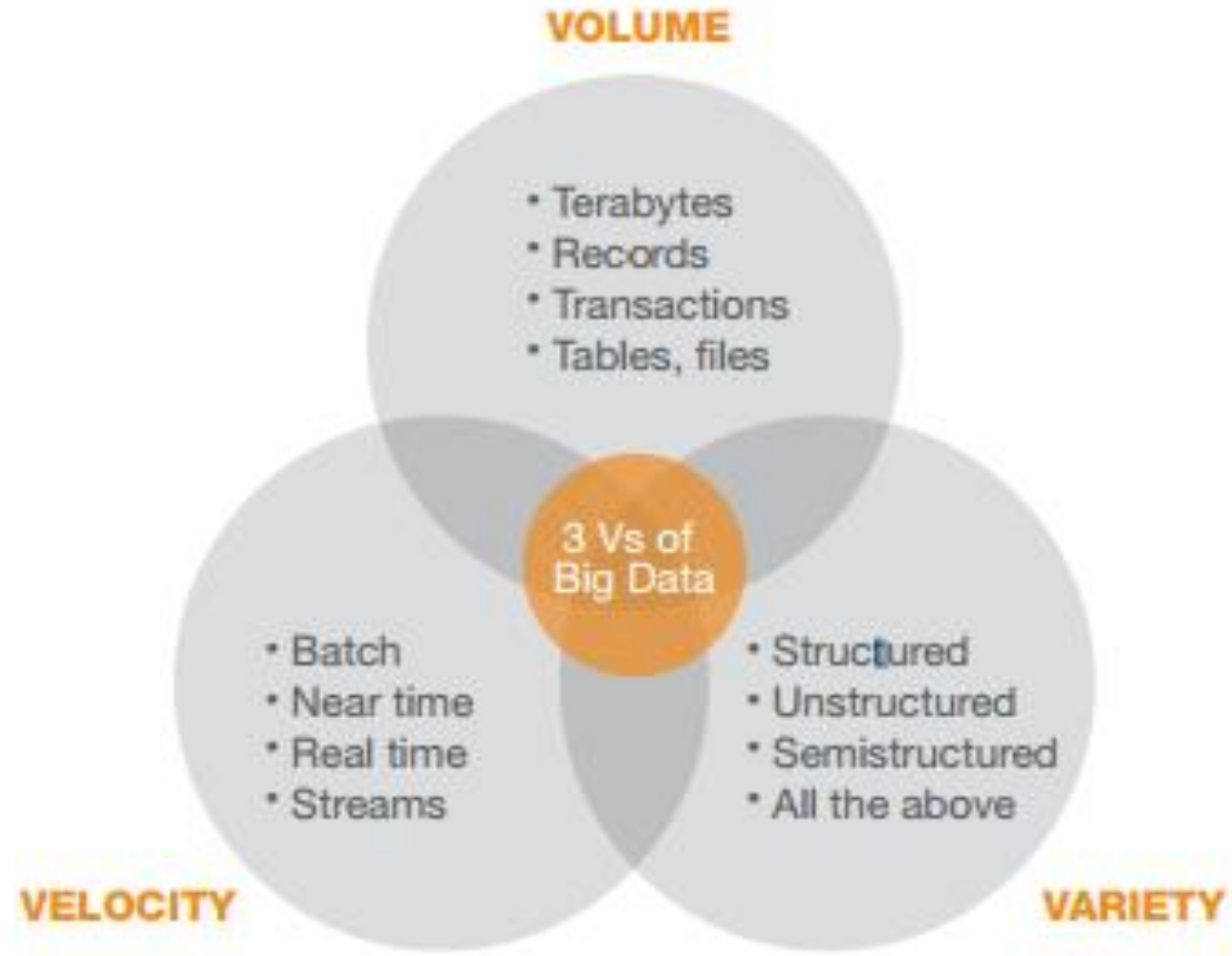


The Three Vs of Big Data

Volume

Velocity

Variety


VOLUME
• Terabytes
• Records
• Transactions
• Tables, files
3 Vs of Big Data
• Batch
• Near time
• Real time
• Streams
• Structured
• Unstructured
• Semistructured
• All the above
VELOCITY
VARIETY

BIG DATA CHALLENGES

1. The most obvious challenge associated with big data is simply storing and analyzing all that information.

2. According to the NewVantage Partners survey, the most common goals associated with big data projects included the following:

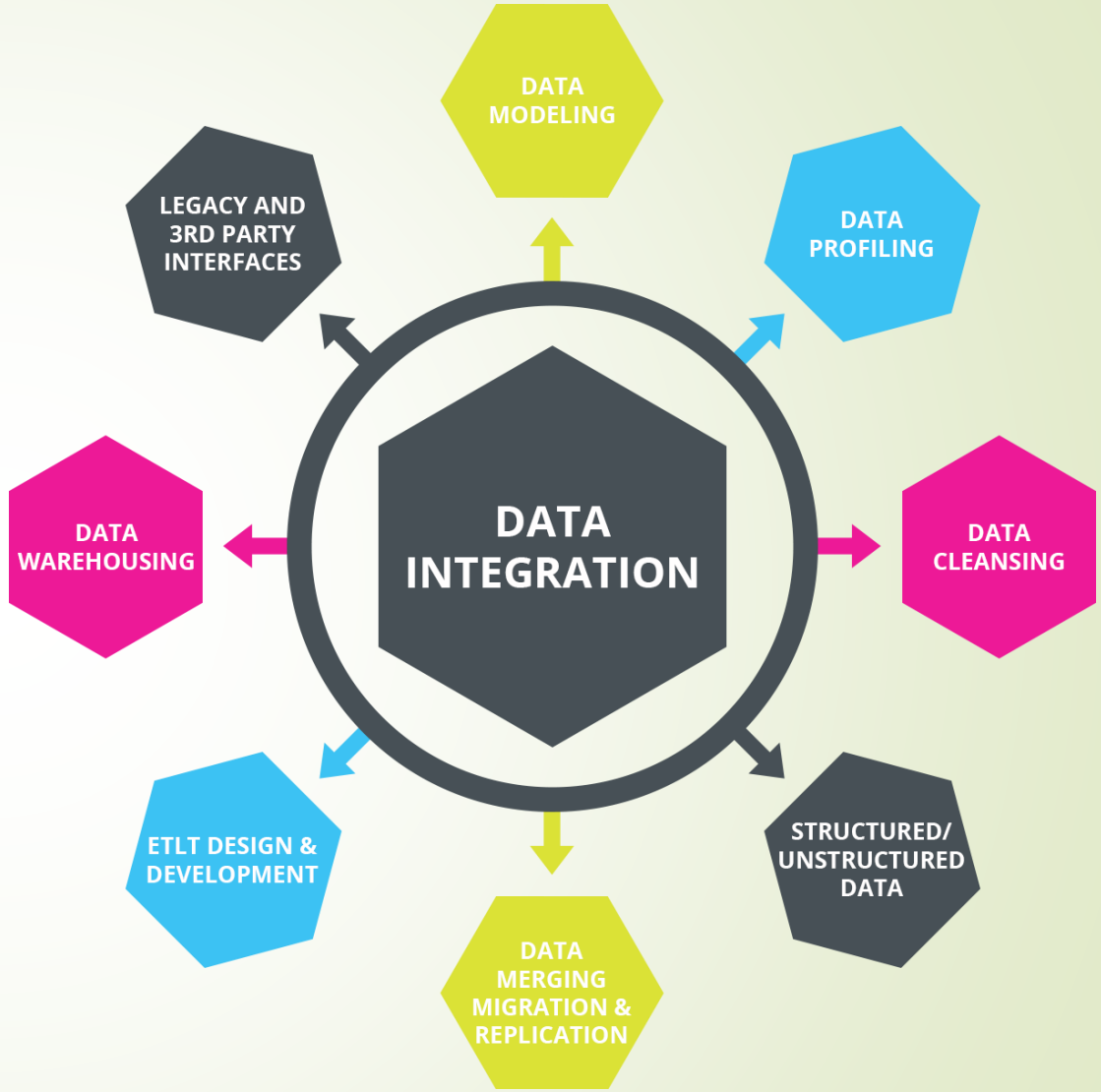
- Decreasing expenses through operational cost efficiencies
- Establishing a data-driven culture
- Creating new avenues for innovation and disruption
- Accelerating the speed with which new capabilities and services are deployed
- Launching new product and service offerings

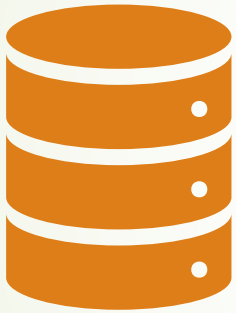
3. Recruiting and retaining big data talent

But in order to develop, manage and run those applications that generate insights, organizations need professionals with big data skills. That has driven up demand for big data experts — and big data salaries have increased dramatically as a result.

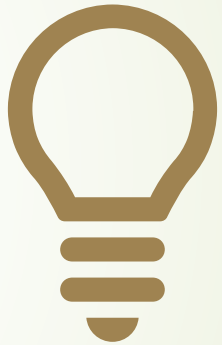
4. Integrating disparate data sources

The variety associated with big data leads to challenges in data integration. Big data comes from a lot of different.





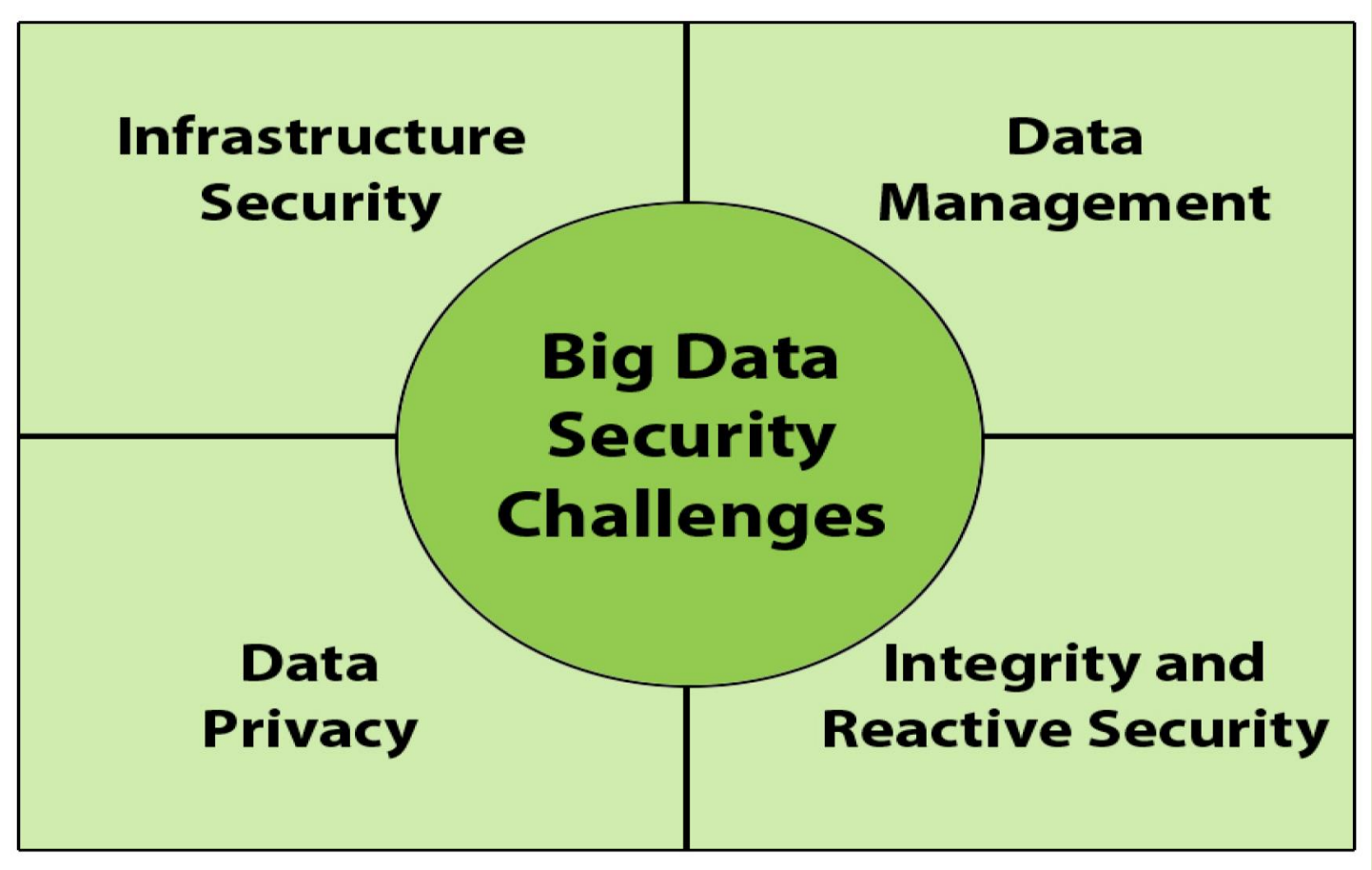
5. Validating data



Closely related to the idea of data integration is the idea of data validation.

6. Securing big data

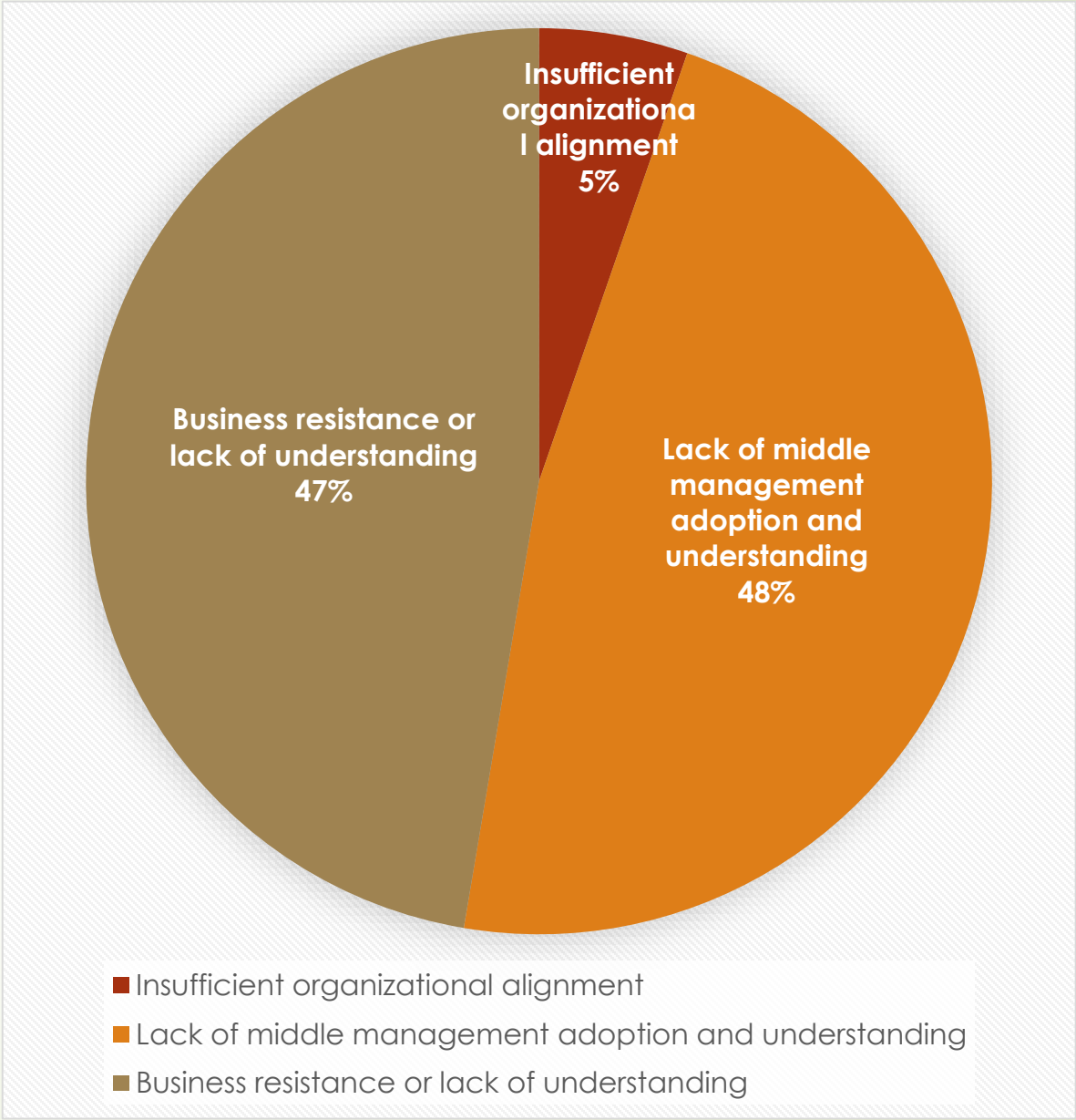
Security is also a big concern for organizations with big data stores. After all, some big data stores can be attractive targets for hackers or advanced persistent threats (APTs).



7. Organizational resistance

It is not only the technological aspects of big data that can be challenging — people can be an issue too.

In the NewVantage Partners survey, 85.5 percent of those surveyed said that their firms were committed to creating a data-driven culture, but only 37.1 percent said they had been successful with those efforts. When asked about the impediments to that culture shift, respondents pointed to three big obstacles within their organizations.



BIGDATA ADVANTAGES

- **Cost Saving**
- **Time Reductions**
- **New Product Development**
- **market conditions**
- **online reputation**

BIG DATA ADVANTAGES

Cost Savings : Some tools of Big Data like Hadoop and Cloud-Based Analytics can bring cost advantages to business when large amounts of data are to be stored and these tools also help in identifying more efficient ways of doing business.

BIG DATA ADVANTAGES

Time Reductions : The high speed of tools like Hadoop and in-memory analytics can easily identify new sources of data which helps businesses analyzing data immediately and make quick decisions based on the learnings.

BIG DATA ADVANTAGES

New Product Development :

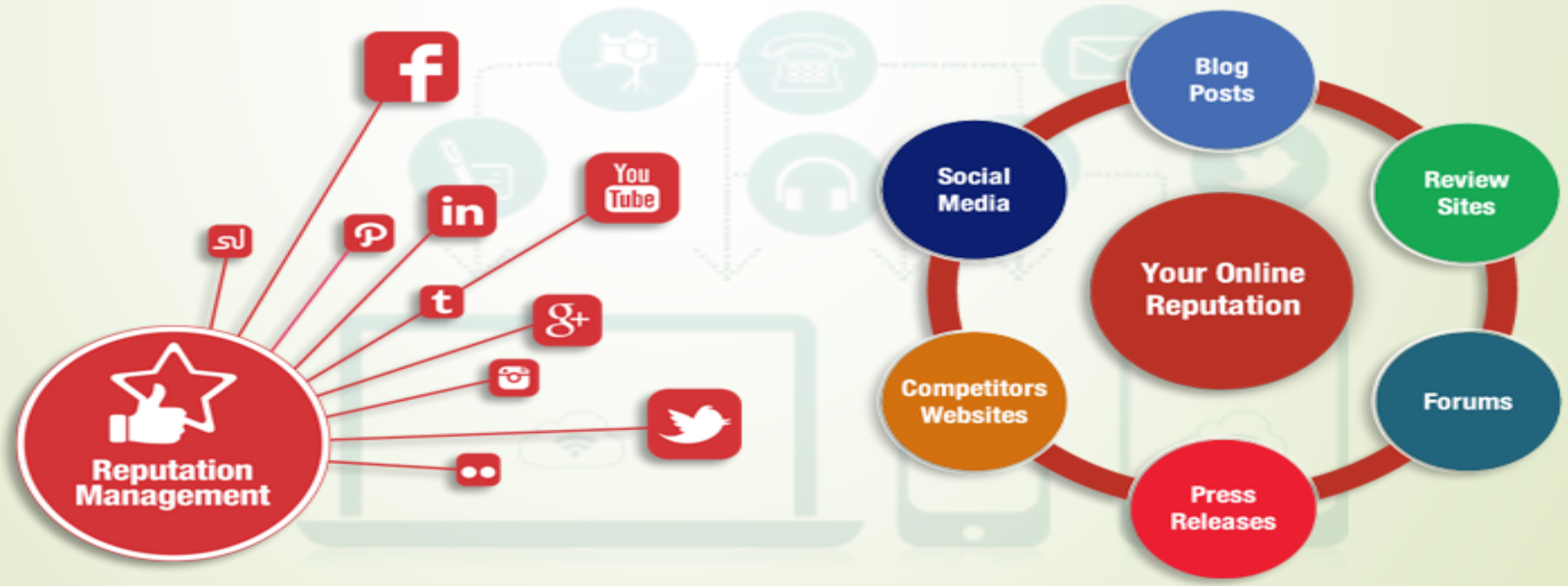
By knowing the trends of customer needs and satisfaction through analytics you can create products according to the wants of customers.

BIG DATA ADVANTAGES

Understand the market conditions : By analyzing big data you can get a better understanding of current market conditions. For example, by analyzing customers' purchasing behaviors, a company can find out the products that are sold the most and produce products according to this trend. By this, it can get ahead of its competitors.

BIG DATA ADVANTAGES

Control online reputation: Big data tools can do sentiment analysis. Therefore, you can get feedback about who is saying what about your company. If you want to monitor and improve the online presence of your business, then, big data tools can help in all this.



EIGHT OPEN SOURCE BIG DATA TOOLS TO USE IN 2018

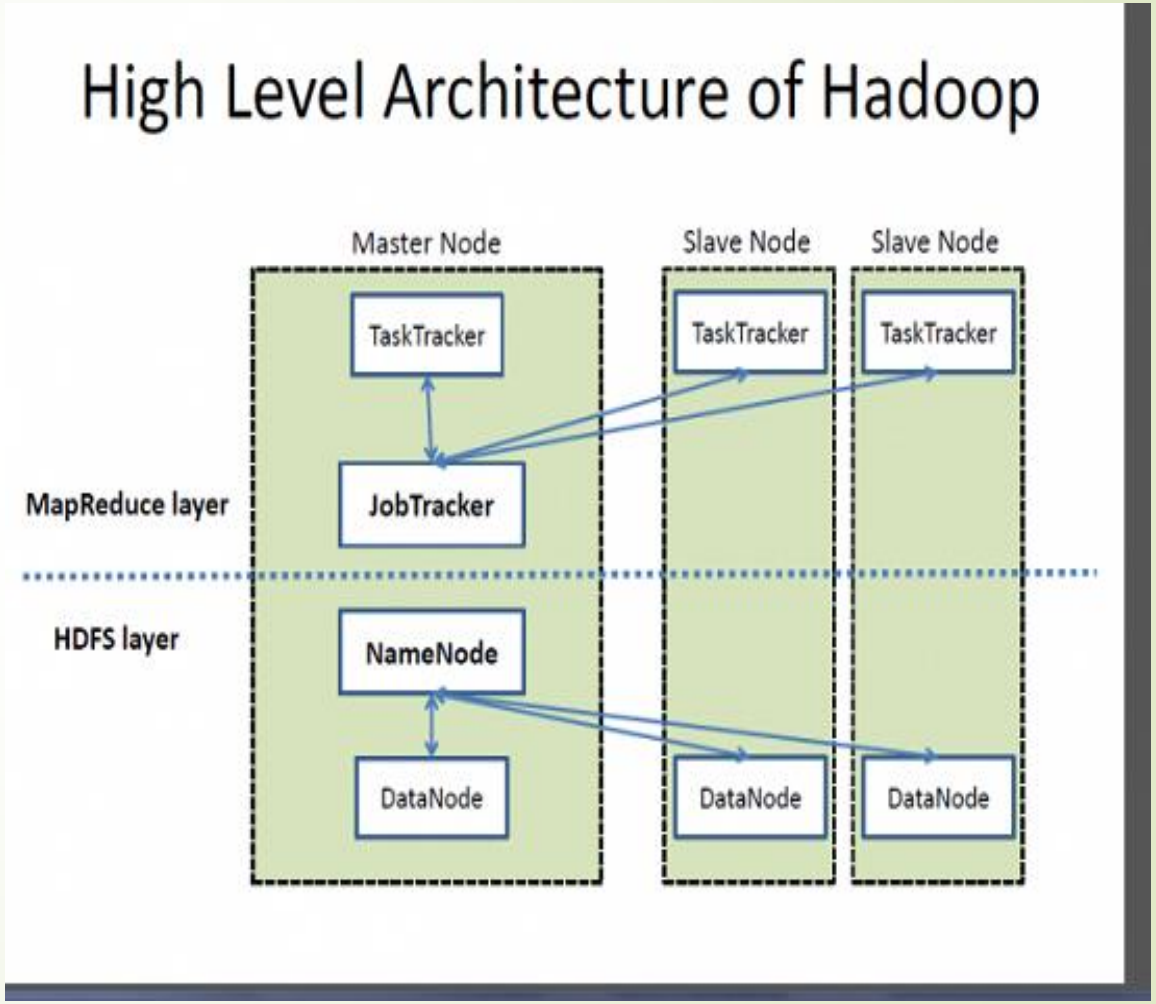
1. Apache Hadoop
2. Apache Spark
3. Apache Storm
4. Apache Cassandra
5. MongoDB
6. R Programming Environment
7. Neo4j
8. Apache SAMOA

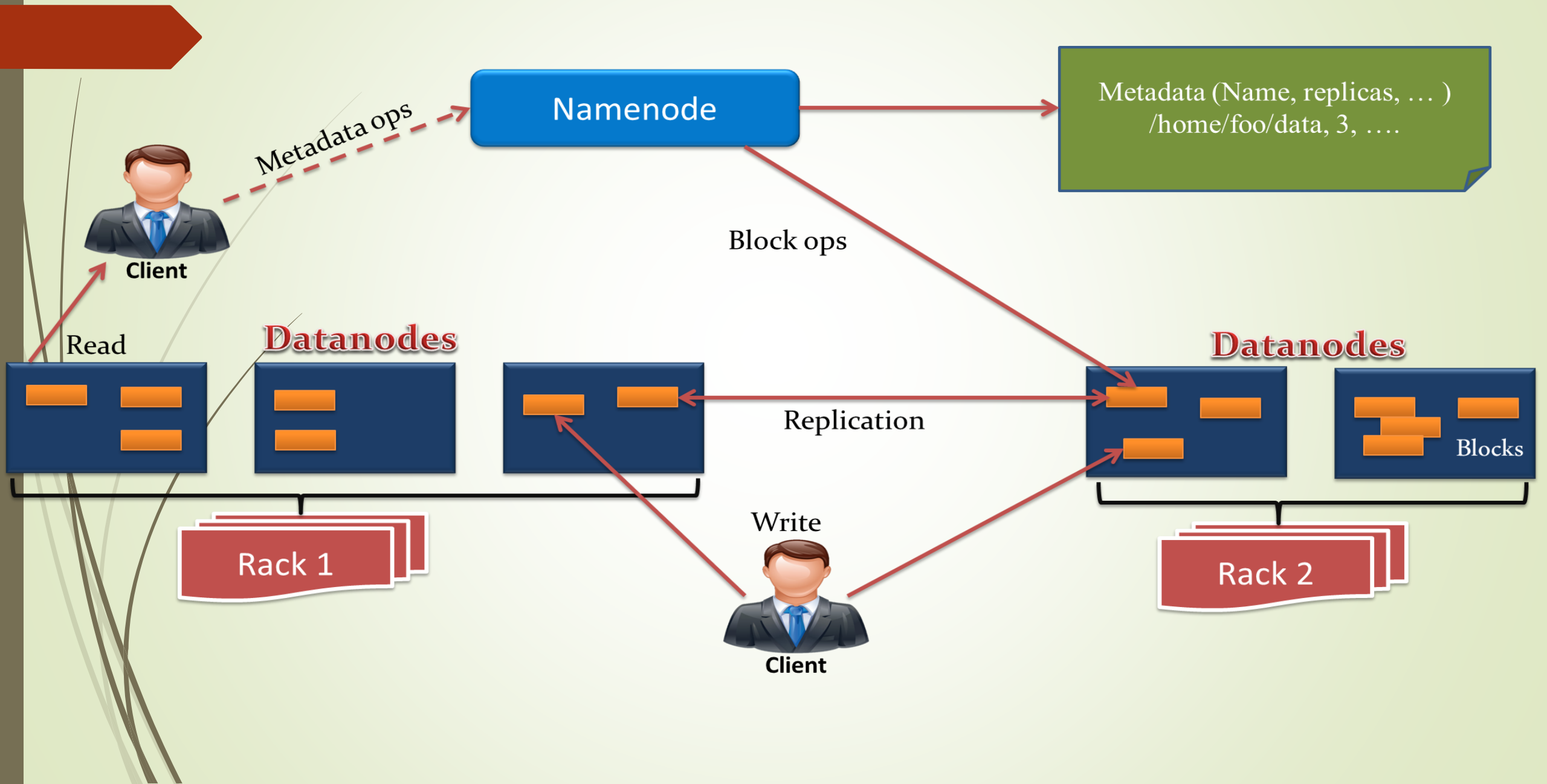
HADOOP

Hadoop is an open source distributed processing framework that manages data processing and storage for big data applications running in clustered systems. It is at the center of a growing ecosystem of big data technologies that are primarily used to support advanced analytics initiatives.

HADOOP ARCHITECTURE

- Hadoop Distributed File System (HDFS) – Patterned after the UNIX file system
- Hadoop MapReduce
- Yet Another Resource Negotiator (YARN)




Namenode
Metadata (Name, replicas, …)
/home/foo/data, 3, ….
Metadata ops
Client
Block ops
Read
Datanodes
Datanodes
Replication
Blocks
Rack 1
Write
Rack 2
Client

CONCLUSION: BIG DATA-A COMPETITIVE ADVANTAGE FOR BUSINESSES

The use of Big Data is becoming common these days by the companies to outperform their peers. In most industries, existing competitors and new entrants alike will use the strategies resulting from the analyzed data to compete, innovate and capture value.

Big Data helps the organizations to create new growth opportunities and entirely new categories of companies that can combine and analyze industry data. These companies have ample information about the products and services, buyers and suppliers, consumer preferences that can be captured and analyzed.

It also understands and optimizes business processes. Retailers can easily optimize their stock based on predictive models generated from the social media data, web search trends and weather forecasts.


Q&A
How?
Why?
When?
Why?
How?
What?
WHAT?
Why?
When?
WHAT?
WHERE?
WHEN?
HOW?
WHEN?
When?
WHERE?
When?
WHERE?
WHAT?
Why?
When?
WHO?
WHAT?
WHEN?
WHERE?
What?
HOW?
Why?
HOW?
Where?
Where?
WHERE?
When?
What?
WHERE?
WHERE?
When?
When?
What?
Why?
WHEN?
HOW?
When?
When?
HOW?
WHAT?
Why?
WHEN?