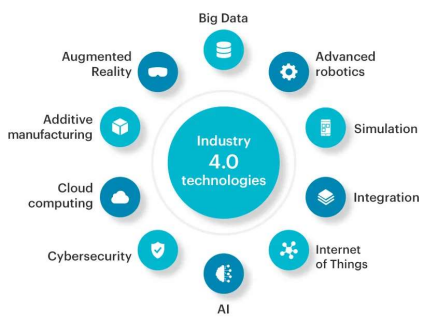
Industry 4.0

1



- Advanced sensing
- Smart Manufacturing
- IIOT
- Digital Twin
- CPS

1

Introduction to Big Data and Big Data Analytics

2

2

contents

3

- Evolution of Big data
- sources of Big Data
- What is Big Data?
- Characteristic of Big Data(5 Vs)
- Tools used in Big Data
- Introduction to Big Data analytics
- Big Data analytics goals
- Applications/use cases of Big Data analytics
- Challenges of Big Data
- How Hadoop solves the Big Data problem

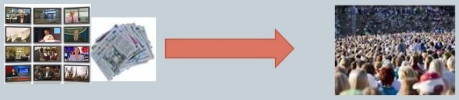
3

Evolution of Big Data

4

- **The Model of Generating/Consuming Data has Changed**

Old Model: Few companies are generating data, all others are consuming data



New Model: all of us are generating data, and all of us are consuming data

4

Unit of Data size	Exact size	Approximate Size	Examples	
KB (kilobyte)	2^{10} or 1024 bytes	(10^3 or one thousand) bytes	A typical joke =1KB	
MB(megabyte)	2^{20} bytes	(10^6 or one million) bytes	Complete work of Shakespeare =5MB	
GB (gigabyte)	2^{30} bytes	(10^9 or one billion) bytes	Ten yards of books on a shelf = 1GB	
TB (terabyte)	2^{40} bytes	(10^{12} or one trillion) bytes	All the X-rays for a large hospital =1TB Tweets; created daily =121TB;	
PB (peta byte)	2^{50} bytes	(10^{15} or one quadrillion) bytes	All U.S. academic research libraries = 2PB Data processed in a day by Google =24PB	BIG DATA
EB (exa byte)	2^{60} bytes	(10^{18} or one Quintillion) bytes	Total global data created in 2006 = 161EB	
ZB (zetta byte)	2^{70} bytes	(10^{21} or one Sextillion) bytes	Total amount of global data created in 2012 = 2.7 ZB and expected 44 ZB by 2020	
YB (yotta byte)	2^{80} bytes	(10^{24} or one Septillion) bytes		

5

6

Petabyte era

1 PB = 1000000000000000B = 1015bytes = 1000terabytes

1000 PB = 1 exabyte (EB)

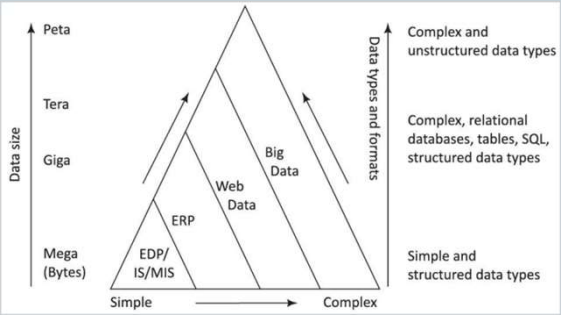
- AT&T transferred about 197 PB of data thorough its network each data (2018)
- processed about 24 petabytes daily (2009)

In fact, we can say that we have already entered the exabyte era.

6

Evaluation of Big Data

7

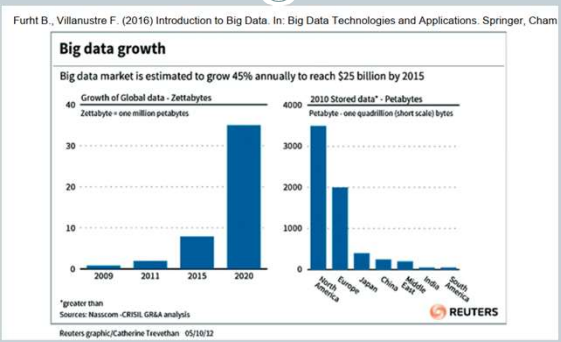


2/13/2025

7

Evaluation of Big Data

8



2/13/2025

8

Evolution of Big Data by technology

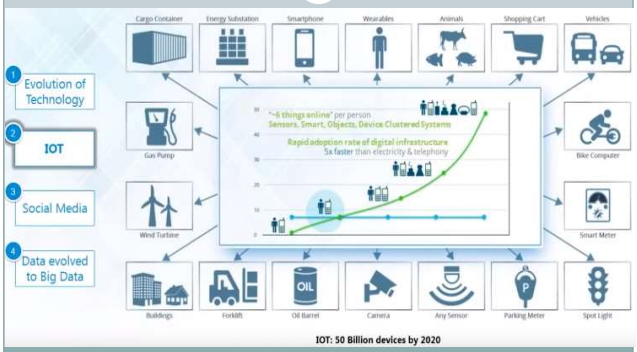
9

2/13/2025

9

Evolution of Big Data by Internet Of Things

10



2/13/2025

10

Evolution of Big Data by Social Media

11

2/13/2025

11

Evolution of Big Data by other factors

12

2/13/2025

12

Big Data sources

13

- Human Generated Data
  - is emails, documents, photos and tweets. We are generating this data faster than ever. Just imagine the number of videos uploaded to You Tube and tweets swirling around. This data can be Big Data too.
- Machine Generated Data
  - is a new breed of data. This category consists of sensor data, and logs generated by 'machines'
  - such as email logs, click stream logs, etc. Machine generated data is orders of magnitude larger than Human Generated Data.

2/13/2025

13

Big Data sources

14

- **Web Data**
  - **Social media data** : Sites like Facebook, Twitter, LinkedIn generate a large amount of data
  - **Click stream data** : when users navigate a website, the clicks are logged for further analysis (like navigation patterns). Click stream data is important in on line advertising and E-Commerce

12+ TBs of tweet data every day

25+ TBs of log data every day

? TBs ofdata every day

2/13/2025

14

Big Data sources

15

sensor data : sensors embedded in roads to monitor traffic and misc.

30 billion RFID tags today (1.3B in 2005)

4.6 billion camera phones world wide

100s of millions of GPS enabled devices sold annually

2+ billion people on the Web by end 2011

76 million smart meters in 2009... 200M by 2014

2/13/2025

15

What is Big Data?

16

Big data

is the term for a collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications.

Real world examples of Big Data

- Facebook : has 40 PB of data and captures 100 TB / day
- Yahoo : 60 PB of data
- Twitter : 8 TB / day
- EBay : 40 PB of data, captures 50TB/ day

2/13/2025

16

What is Big Data?

17

- Big Data is high-volume, high-velocity and/or high-variety information asset that requires new forms of processing for enhanced decision making, insight discovery and process optimization

(Gartner 2012)

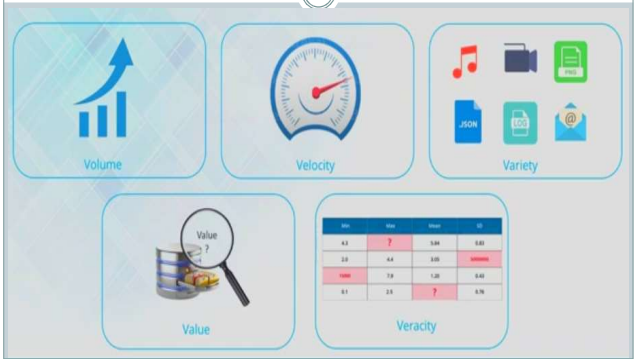
- Term big relates to size of the data and hence the characteristic Size defines the amount or quantity of data, which is generated from an application(s).
- The size determines the processing considerations needed for handling that data

2/13/2025

17

Characteristics of Big Data(5 Vs of Big data)

18



2/13/2025

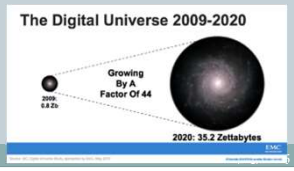
18

Characteristics of Big Data(5 Vs of Big data)

19

- 1st V-volume

- **Data Volume**
  - 44x increase from 2009 to 2020 From 0.8 zettabytes to 35zb
  - Data volume is increasing exponentially

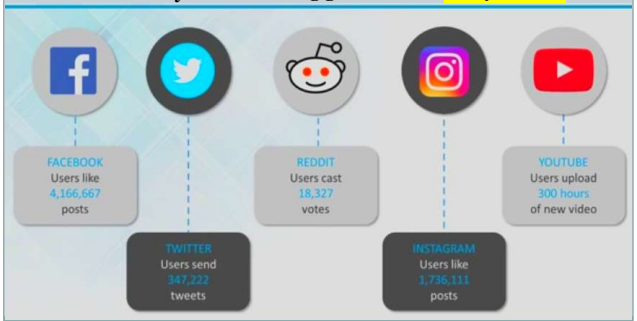


19

Characteristics of Big Data(5 Vs of Big data)

20

- 2nd V-velocity: **Data is being generated at every minute**



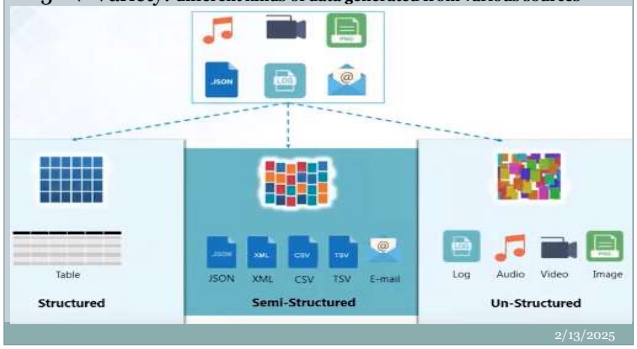
2/13/2025

20

Characteristics of Big Data(5 Vs of Big data)

21

- 3rd V-Variety: **different kinds of data generated from various sources**



2/13/2025

21

Characteristics of Big Data(5 Vs of Big data)

22

- 4th V - Veracity: **uncertainties and inconsistencies in big data**

Min	Max	Mean	SD
4.3	?	5.84	0.83
2.0	4.4	3.05	50000000
15000	7.9	1.20	0.43
0.1	2.5	?	0.76

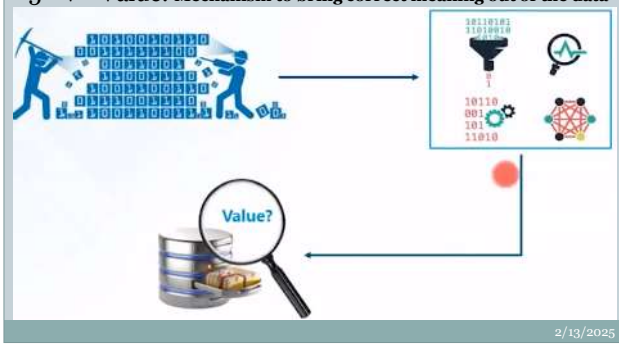
2/13/2025

22

Characteristics of Big Data(5 Vs of Big data)

23

- 5th V - Value: **Mechanism to bring correct meaning out of the data**



2/13/2025

23

Characteristics of Big Data(5 Vs of Big data)

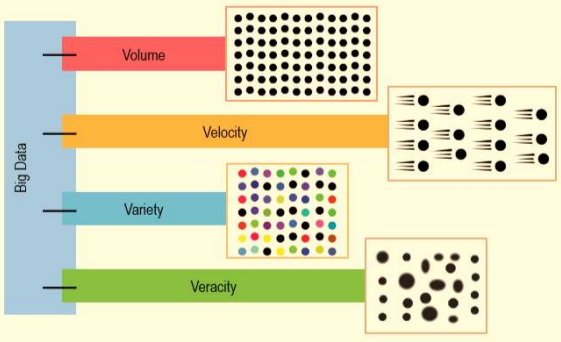
24

2/13/2025

24

Characteristics of Big Data(5 Vs of Big data)

25



2/13/2025

25

Traditional DB vs Big Data

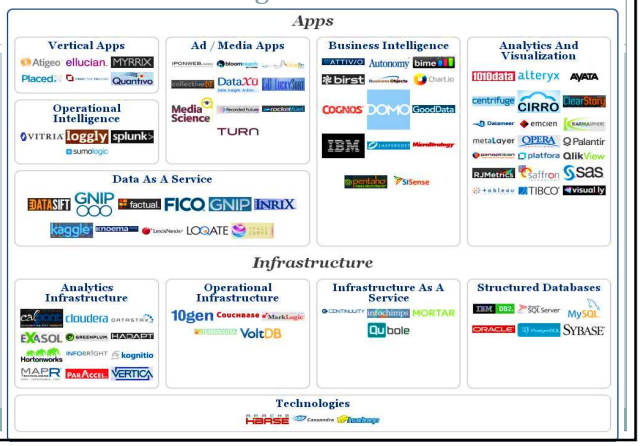
26

Traditional data base/ data warehouse	Big Data
• Data ○ TB to PB ○ Only structured • Hardware ○ big central servers ○ Expensive ○ Hardware reliability ○ Limited scalability • Software ○ Centralized ○ Schema based ○ Oracle/mysql/sql server	• Data ○ PB to ZB ○ structured and unstructured • Hardware ○ computer clusters ○ Cost effective ○ Unreliable HW ○ Scales further • Software ○ Distributed ○ Not schema based ○ Hadoop

2/13/2025

26

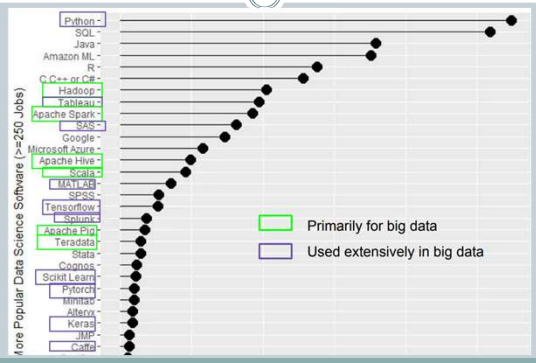
Big data tools



27

Big data tools

28



2/13/2025

28

What is Big data analytics

29

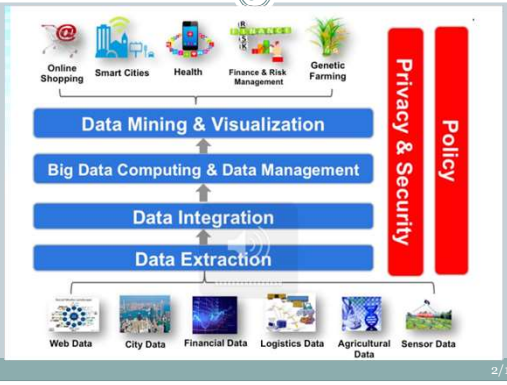
"Big data analytics examines large and different types of data to uncover hidden patterns, correlations and other insights"

2/13/2025

29

Big Data Analytics Objectives

30



2/13/2025

30

31

2/13/2025

31

Stages in Big data analytics

32

2/13/2025

32

Big data analytics goals

33

2/13/2025

33

Big data analytics goals

34

1.Making organizations more smarter and efficient

2/13/2025

34

Big data analytics goals

35

2 Optimize Business Operations by analysing customer behaviour

Analysing all the clicks of every visitor on a website

Studying the paths leading them to buy products

Customer Satisfaction

Amazon uses customer click-stream data and historical purchase data of more than 300 million customers and each user is shown customized results on customized web pages.

2/13/2025

35

Big data analytics goals

36

3 Cost Reduction

Parkland Hospital uses analytics and predictive modelling to identify high-risk patients and predict likely outcomes once patients are sent home. As a result, Parkland reduced 30-day readmissions for patients with heart failure, by 31 percent, saving $500,000 annually.

2/13/2025

36

Big data analytics goals

37

4 Next Generation Products

Big Data tools are used to operate Google's Self Driving Cars. The Toyota Prius is fitted with cameras, GPS as well as powerful computers and sensors to safely drive on the road without the intervention of human beings.

Netflix launched the seasons of its TV show House of Cards based on the user reviews, ratings and viewership.

NETFLIX

A smart yoga mat has sensors embedded in the mat will be able to provide feedback on your postures, score your practice, and even guide you through an at-home practice.

2/13/2025

37

Big data analytics application domains

38

Domains using Big Data Analytics

Healthcare
Telecom
Insurance
Government
Big Data Analytics
Finance
Automobile
Education
Retail

2/13/2025

38

Big data analytics use cases

39

Use Case 1 - Starbucks

STARBUCKS COFFEE

Starbucks uses behavioural analytics to cater to its customers

Starbucks gather a lot of info about their customers' coffee-buying habits from their preferred drinks to what time of day they're usually ordering

The company directs exciting offers and coupons to their customers and ensures to maintain their interest

2/13/2025

39

Big data analytics use cases

40

Use Case 2 – Procter & Gamble

P&G
Procter&Gamble

P&G uses Market Basket Analysis and price optimization to optimize their products

Market Basket Analysis, analyses customer buying habits by finding associations between the different items that customers place in their 'shopping baskets'

The company uses simulation models and predictive analysis in order to create the best design for its products.

2/13/2025

40

Big data analytics use cases

41

Walmart boosted its sales by leveraging the power of Big Data

While forecasting the demand for emergency supplies for approaching Hurricane Sandy, they gain some amazing insights:

Walmart

Extra supplies of Strawberry Pop Tarts were dispatched to stores in Hurricane Sandy's path in 2012, and sold extremely well

Along with flashlights and emergency equipment, they found an upsurge in sales of strawberry Pop Tarts

HURRICANE SEASON

2/13/2025

41

Big data analytics use cases

42

Big Data helped Donald Trump to win against Hillary Clinton in the US election

Collect Personal data from various resources like club cards, newspaper Subscription, social media, etc.

Messages were targeted based on voter profiles using platforms such as Facebook, Snapchat, Pandora radio, etc.

Build an algorithm that generated top cities to reach the highest concentration of persuadable voters

2/13/2025

42

Big data analytics use cases

43

Apixio uses big data analytics to improve healthcare decision

APIXIO

80% of medical and clinical information about patients is in unstructured format, such as written physician notes

Analysis of medical data using variety of different methodologies & algorithms that are machine learning based and have NLP capabilities

The patient data model generated is aggregated across population to derive larger insights like disease prevalence, treatment patterns, etc.

2/13/2025

43

Big data analytics use cases

44

- IBM Big data analytics – Big data collected by smart meters

Data was collected in 1 Month

Earlier

Data is collected in 15 Minutes

Now

Managing the large volume and velocity of information generated by short-interval reads of smart meter data can overwhelm existing IT resources

96 million reads per day for every million meters

Big Data generated by Smart Meter

IBM

2/13/2025

44

Big data analytics use cases

45

- IBM Big data analytics – problem with smart meter big data

To manage and use this information to gain insight, utility companies must be capable of high-volume data management and advanced analytics designed to transform data into actionable insights.

Store

Analyze

IBM

2/13/2025

45

Big data analytics use cases

46

- IBM Big data analytics – how smart meter big data analysed

Before analyzing Big Data

Energy utilization and billing has increased

After analyzing Big Data

APPLIANCE

APPLIANCE

During peak-load the users require more energy

During off-peak times the users required less energy

Time-of-use pricing encourages cost-savvy retail like industrial heavy machines to be used at off-peak times

2/13/2025

46

Big data analytics use cases

47

- IBM Big data analytics – IBM smart meter solution

IBM offers an integrated suite of products designed to enable IT to leverage big data in a variety of ways that can contribute to the success of energy companies

Analytics

Data Analysis

Data Mining

Data Warehousing

User Data Security

Reporting

IBM Solution

1. Managing smart meter data
2. Monitoring the distribution grid
3. Optimizing unit commitment
4. Optimizing energy trading
5. Forecasting and scheduling loads

2/13/2025

47

Big data analytics use cases

48

ONCOR

Oncor Electric Delivery has incorporate IBM Smart Meter service

1. Instrumented — Utilizes smart electricity meters to accurately measure the electricity usage of a household
2. Interconnected — Unprecedented access to detailed information about their electricity use
3. Intelligent — Consumers monitor and control their electricity usage through near-real time readings of electricity meters

BENEFITS

Customers in Oncor's service territory showed last year during the company's biggest energy saver contest that by using the information from Oncor's advanced meter

Users reduced their electric usage and bills by 25 percent or more

2/13/2025

48

Types of Big data analytics

49

1 Descriptive Analysis
2 Predictive Analysis
3 Prescriptive Analysis
4 Diagnostic Analytics

Types of Big Data Analytics

What is happening now based on incoming data.

Google Analytics Tool is the best example for descriptive analysis. A business gets result from the web server through the tool which help understand what actually happened in the past and validate if a promotional campaign was successful or not based on basic parameters like page views.

2/13/2025

49

Types of Big data analytics

50

1 Descriptive Analysis
2 Predictive Analysis
3 Prescriptive Analysis
4 Diagnostic Analytics

Types of Big Data Analytics

What might happen in the future

For example, Southwest Airlines analyses sensor data on their planes in order to identify patterns that indicate a potential malfunction, thus allowing the airlines to the necessary repairs before its schedule.

2/13/2025

50

Types of Big data analytics

51

1 Descriptive Analysis
2 Predictive Analysis
3 Prescriptive Analysis
4 Diagnostic Analytics

Types of Big Data Analytics

What action should be taken.

Google's self-driving car is a perfect example of prescriptive analytics. It analyses the environment and decides the direction to take based on data.

2/13/2025

51

Types of Big data analytics

52

1 Descriptive Analysis
2 Predictive Analysis
3 Prescriptive Analysis
4 Diagnostic Analytics

Types of Big Data Analytics

Why did it happen

For a Social Media marketing campaign, you can use diagnostic analytics to assess the number of posts, mentions, followers, fans, page views, reviews, pins, etc. and analyse the failure and success rate of the campaign at a fundamental level.

2/13/2025

52

Challenges/problems with Big data

53

Problem 1: Storing exponentially growing huge datasets

- Data generated in past **2 years** is more than the previous history in total
- By 2020, total digital data will grow to **44 Zettabytes** approximately
- By 2020, about **1.7 MB** of new info will be created every second for every person

2/13/2025

53

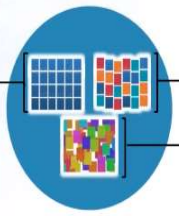
Challenges/problems with Big data

54

Problem 2: Processing data having complex structure

Structured

- Organized data format
- Data schema is fixed
- Ex: RDBMS data, etc.



Semi – Structured

- Partial organized data
- Lacks formal structure of a data model
- Ex: XML & JSON files, etc.

Unstructured

- Un-organized data
- Unknown schema
- Ex: multi-media files, etc.

2/13/2025

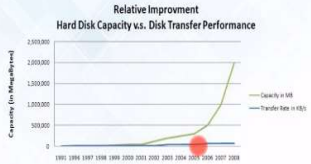
54

Challenges/problems with Big data

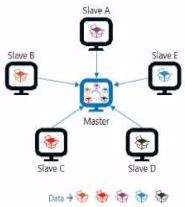
55

Problem 3: Processing data faster

The data is growing at much faster rate than that of disk read/write speed



Bringing huge amount of data to computation unit becomes a bottleneck



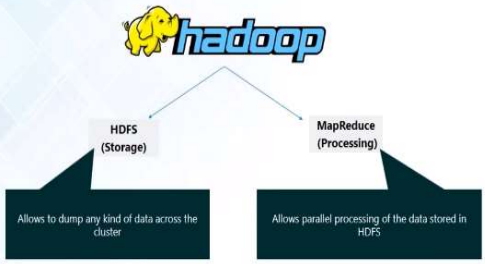
2/13/2025

55

HADOOP is solution to Big data problems

56

Hadoop is a framework that allows us to store and process large data sets in parallel and distributed fashion



2/13/2025

56

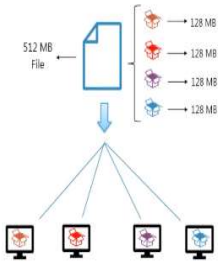
HADOOP is solution to Big data problems

57

Problem 1: Storing exponentially growing huge datasets

Solution: HDFS

- Storage unit of Hadoop
- It is a Distributed File System
- Divide files (input data) into smaller chunks and stores it across the cluster
- Scalable as per requirement



2/13/2025

57

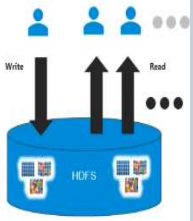
HADOOP is solution to Big data problems

58

Problem 2: Storing unstructured data

Solution: HDFS

- Allows to store any kind of data, be it structured, semi-structured or unstructured
- Follows WORM (Write Once Read Many)
- No schema validation is done while dumping data



2/13/2025

58

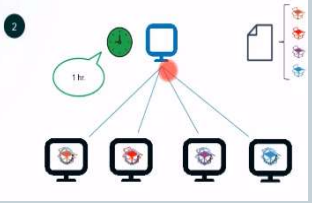
HADOOP is solution to Big data problems

59

Problem 3: Processing data faster

Solution: Hadoop MapReduce

- Provides parallel processing of data present in HDFS
- Allows to process data locally i.e. each node works with a part of data which is stored on it



2/13/2025

59

Big Data in Industry 4.0

60

- **Role in Industry 4.0**
  ✓ Smart factories optimize production with sensor data.
  ✓ Predictive maintenance prevents failures & downtime.
  ✓ Real-time data enables automated decision-making.

- **Key Benefits**
  ✅ Increased production efficiency
  ✅ Reduced operational costs
  ✅ Enhanced automation & real-time insights

2/13/2025

60

Cases of Big Data in Industry4.0

61

📦 **Warehouse Optimization**
➡️ Sensors & portable devices detect errors, perform quality checks, and optimize workflows.
🚧 **Bottleneck Elimination**
➡️ Identifies performance issues and suggests improvements at no extra cost.
📊 **Predictive Demand**
➡️ Uses internal & external analysis for better forecasting and product optimization.
🛠️ **Predictive Maintenance**
➡️ Sensors detect failure patterns and send alerts before breakdowns occur.
🔍 **Other Benefits:** Improved security, load optimization, supply chain management, and non-conformity analysis.

2/13/2025

61

How Businesses Use Big Data Analytics

62

- 🚀 **Industry 4.0 Applications**
  ✓ **Supply Chain Optimization** – Identifies patterns to improve logistics.
  ✓ **Predictive Maintenance** – Cuts downtime by 25% with AI-driven monitoring.
  ✓ **Production Management Automation** – Uses historical & real-time data for self-adjusting processes.

- 🛠️ **Smart Factory Technologies**
  ➡️ **Self-Service Systems** – Real-time analytics for decision-making (e.g., Intel's smart factory).
  ➡️ **Automated Production** – Robots & actuators adjust equipment settings autonomously.
  ➡️ **Faster Fault Detection** – Reduces reaction time from 4 hours to 30 seconds.

2/13/2025

62