



U4 Big Data

Mr. J. Mishra
MGCUB, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References

Emerging Technology: Big Data

Course: Master of Arts (Hindi)

Course Name: Computer and Information Technology

Course Code: HIND4014

Semester: II

Session: 2019-20



Mr. Joynath Mishra

Assistant Professor (Guest)

Department of Computer Science and Information Technology

Mahatma Gandhi Central University
Bihar, INDIA

May 31, 2020



Outline

U4 Big Data

Mr. J. Mishra
MGCU, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References

- 1 Objectives
- 2 Introduction
- 3 Characteristics
- 4 Classification
- 5 Challenges
- 6 Sources
- 7 Generation
- 8 Handling Tools
- 9 Disadvantage
- 10 Relation with Cloud Computing
- 11 Conclusion
- 12 Exercise
- 13 References



Objectives

U4 Big Data

Mr. J. Mishra
MGCUB, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with Cloud Computing

Conclusion

Exercise

References

Objectives

- Context of Big Data
- Characteristics of Big Data
- Big data generation and handling tools
- Big data relation with cloud computing



Introduction[1]

U4 Big Data

Mr. J. Mishra
MGCUB, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References

Introduction

- Data and information is the currency of new digital world
- Different software firms and digital movement generates huge amount of data
- These are exponential increase the amount of data with time
- These data demands to be saved as well as analysis to predict future events
- Large amount of data difficult to solve using traditional database system
- Big Data generation includes stock exchanges, social media sites etc
- Big Data is a phrase used to mean a massive volume of data and their processing style beyond the traditional way

Benefits of Big Data Processing

- Businesses can utilize outside intelligence while taking decisions
- Improved customer service
- Early identification of risk to the product/services, if any
- Better operational efficiency



Characteristics

U4 Big Data

Mr. J. Mishra
MGCUB, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with Cloud Computing

Conclusion

Exercise

References

Characteristics

- 1 Velocity: speed of generation of data (2 million queries on Google, 200 million mails are sent)
- 2 Volume: amount of data (social media data in Zettabyte)
- 3 Variety: structured and unstructured data
- 4 Value: learning, recognizing, prediction
- 5 Veracity: meaningful result in autonomous space
- 6 Variability: data variance over time
- 7 Visualization: complex graph with large number of variables

Concept of 3V's of Big Data

- Volume: data quantity
- Velocity: data speed
- Variety: data type



Characteristics (Contd...)

U4 Big Data

Mr. J. Mishra
MGCUB, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References

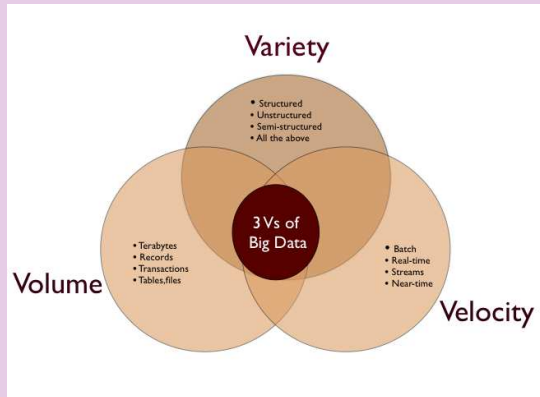


Figure 1: Concept of 3V's of Bigdata[Source: Internet]



Classification

U4 Big Data

Mr. J. Mishra
MGCUB, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References

Structured Data:

- Defined format/pattern of data
- Eg. numbers, dates, and groups of words and numbers called strings
- Generally stored in databases

Unstructured Data

- No pattern of data fields
- Massive data eg. Newspaper
- Heteroginous files: application, music, movie, pictures, medical images

Semi-Structured Data

- The data which do not have a proper formate attched to it
- Eg. data within an email, data in doc File



Big Data Challenges

U4 Big Data

Mr. J. Mishra
MGCU, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References

Big Data Challenges

- 1 Capture
- 2 Search
- 3 Storage
- 4 Sharing
- 5 Curation
- 6 Transfer
- 7 Analysis
- 8 Visualization



Big Data Sources

U4 Big Data

Mr. J. Mishra
MGCU, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References

Big Data Sources

- 1 Users
- 2 Application
- 3 System
- 4 Sensors



Big Data Generation

U4 Big Data

Mr. J. Mishra
MGCUB, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References

Big Data Generation

- 1 Smart Phones: mobile captured images, GPS data (sensor data)
- 2 Internet: searching content in Zetabyte
- 3 Emails: large amount of emails are sent
- 4 Blogs: large amount of content is added
- 5 Social Media: Facebook, Twitter, LinkedIn, Instagram
- 6 Google: Searching content in Petabyte
- 7 Youtube: Large amount of video watcher



Big Data Handling

U4 Big Data

Mr. J. Mishra
MGCUB, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References

Traditional System

- 1 Traditional system is not able to handle large amount of data
- 2 Eg. RDBMS is failure to handle such data

Big Data Tools

- 1 Some tools has been developed to handle big data
- 2 Eg. Hadoop is capable to handle big data



Big Data Handling

Hadoop Tools

U4 Big Data

Mr. J. Mishra
MGCUB, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References

Hadoop Tools

- 1 Data Extraction Tool: Talend, Pentaho
- 2 Data Storing Tool: MongoDB, Hive, Sqoop
- 3 Data Mining Tool: Oracle Data Mining
- 4 Data Analysis Tool: HBase, Pig
- 5 Data Integrating Tool: Zookeeper



Disadvantage

U4 Big Data

Mr. J. Mishra
MGCU, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References

Disadvantages

- 1 It requires special computer power
- 2 Using real-time insights requires a different way of working within your organisation
- 3 Difficult to assess accuracy and uncertainty
- 4 Privacy and confidentiality issues
- 5 Bad Data: Sometimes huge irrelevant data may store
- 6 Cost: Data collection, aggregation, storage, analysis and reporting all have cost
- 7 Data Privacy: It's hard to ensure people's personal data are safe from criminals



Relation with Cloud Computing[2]

U4 Big Data

Mr. J. Mishra
MGCUB, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with Cloud Computing

Conclusion

Exercise

References

Comparison between Big Data and Cloud Computing

	Cloud Computing	Big Data
Definition	provides resources on demand	provides way to handle huge volume of data
Reference	Services referes as SaaS, PaaS, IaaS	it is related to data structured, unstructured and semi-structured form
Method	connect data over internet	discover data pattern and actions on it
Formats	computing resource	kind of data
Use for	store data and information on remote server	describe huge volume of data and information



Conclusion

U4 Big Data

Mr. J. Mishra
MGCU B, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with Cloud Computing

Conclusion

Exercise

References

Conclusion

- Big data improved customer service, operational efficiency
- Faster & better decision making over large volume of data
- Analysis of data gives ideas for new products and demanded services
- Easy to manage large amount of user comments and variety of data
- Real-time monitoring and prediction of demands/events

Future of Big Data

- Big Data increases its demands from private as well as government sectors
- It reduces the cost of storage of data and analysis
- Prescriptive analytics are the future of Big Data (Eg. Google's self-driving car uses prescriptive analytics)



Exercise

U4 Big Data

Mr. J. Mishra
MGCUB, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References

Exercise

- What is big data? Explain with example.
- How big data is generating regularly?
- What are the characteristics of big data concept?
- Explain 3V's concept of big data.



References I

U4 Big Data

Mr. J. Mishra
MGCUB, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References



R. Buyya, *Big Data- Principles and Paradigms*. Todd Green, 2016.



Data flair tutorial. [Online]. Available: <https://data-flair.training>



U4 Big Data

Mr. J. Mishra
MGCUB, INDIA

Objectives

Introduction

Characteristics

Classification

Challenges

Sources

Generation

Handling Tools

Disadvantage

Relation with
Cloud
Computing

Conclusion

Exercise

References

Get in touch via...



+91 9046174189



jaynath4025@gmail.com

Thank You...