

SNS COLLEGE OF TECHNOLOGY



Coimbatore-35 An Autonomous Institution

Department of Information Technology



19ITE305 – BIG DATA ANALYTICS

III B.Tech. IT/ VI SEMESTER

UNIT I : INTRODUCTION TO BIG DATA AND ANALYTICS

Topic 1 : Classification of Digital Data, Structured and Unstructured Data

Classification of Digital Data, Structured and Unstructured Data - Introduction to Big Data: Characteristics – Evolution – Definition - Challenges with Big Data - Other Characteristics of Data - Why Big Data -Traditional Business Intelligence versus Big Data - Data Warehouse and Hadoop Environment **Big Data Analytics:** Classification of Analytics – Challenges - Big Data Analytics important - Data Science -Data Scientist - Terminologies used in Big Data Environments.



What is data and information?



Data is raw, unorganized, unprocessed information. E.g., the information collected for writing a research paper is data until it is presented in an organized manner.

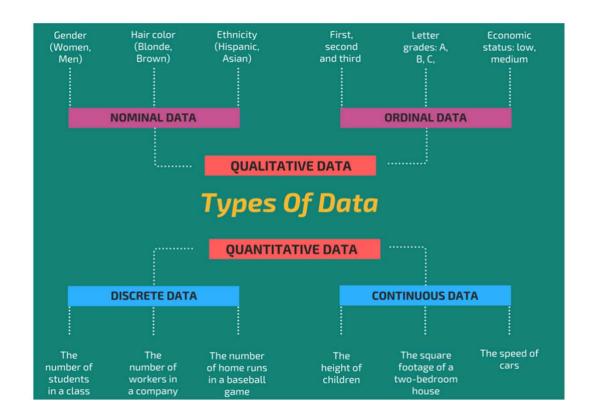


Data generates information and from information we can draw valuable insight.

Information is the processed, organized data that is beneficial in providing useful knowledge. For eg., the data compiled in an organized way in a research paper provides information about a particular concept/ topic.

19ITE305 Big Data Analytics/ Unit 1

Types of data



QUANTITATIVE DATA QUALITATIVE DATA Quantitative data can be expressed as a Qualitative data can't be expressed as a number or can be quantified. Simply put, number and can't be measured. quantitative data can be measured by Oualitative data consist of words. pictures, and symbols, not numbers. 👥 👠 ી **EXAMPLES** - Colors e.g. the color of the sea - Your favorite holiday destination such as Hawaii, New Zealand. - Names as John, Patricia,.... - Scores on tests and exams e.g. 85, - Ethnicity such as American Indian, Asian, etc. - The weight of a person or a

QUALITATIVE DATA

subject. - Your shoe size.

67, 90 and etc.

numerical variables.

- The temperature in a room.

EXAMPLES

19ITE305 Big Data Analytics/ Unit 1

INSTITUTIONS

VS

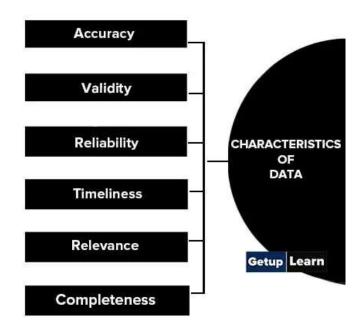
QUANTITATIVE



Characteristics of Data

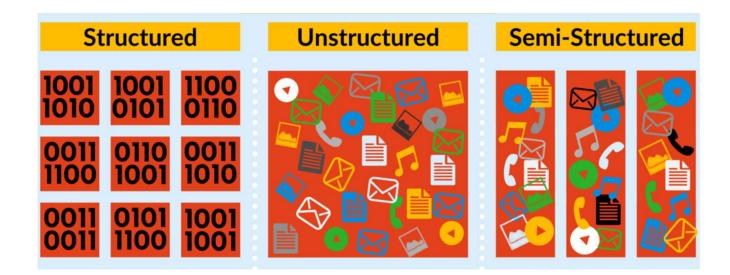
The following are six key characteristics of data

- Accuracy
- Validity
- Reliability
- Timeliness
- Relevance
- Completeness





Classification of Digital Data



19ITE305 Big Data Analytics/ Unit 1





STRUCTURED

Structured Data

Structured Data refers to the data that has a proper structure associated with it. For example, the data that is present within the databases, the CSV files, and the excel spreadsheets can be referred to as Structured Data.

Employee_ID	Employee_Name	Gender	Department	Salary_In_lacs
2365	Rajesh Kulkarni	Male	Finance	650000
3398	Pratibha Joshi	Female	Admin	650000
7465	Shushil Roy	Male	Admin	500000
7500	Shubhojit Das	Male	Finance	500000
7699	Priya Sane	Female	Finance	550000

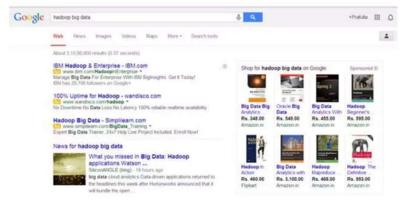
19ITE305 Big Data Analytics/ Unit 1





Unstructured Data

Un-Structured Data refers to the data that does not have any structure associated with it at all. For example, the image files, the audio files, and the video files can be referred to as Un-Structured Data.



19ITE305 Big Data Analytics/ Unit 1



Semi-structured Data



Semi-Structured Data refers to the data that does not have a proper structure associated with it. For example, the data that is present within the emails, the log files, and the word documents can be referred to as Semi-Structured Data.

<rec><name>Prashant Rao</name><sex>Male</sex><age>35</age></rec> <rec><name>Seema R.</name><sex>Female</sex><age>41</age></rec> <rec><name>Satish Mane</name><sex>Male</sex><age>29</age></rec> <rec><name>Subrato Roy</name><sex>Male</sex><age>26</age></rec> <rec><name>Jeremiah J.</name><sex>Male</sex><age>35</age></rec>



TEXT BOOKS

Seema Acharya, Subhashini Chellappan, "Big Data and Analytics", Wiley Publications, First Edition, 2015

REFERENCES

- 1. Judith Huruwitz, Alan Nugent, Fern Halper, Marcia Kaufman, "Big data for dummies", John Wiley & Sons, Inc. (2013)
- 2. Tom White, "Hadoop The Definitive Guide", O'Reilly Publications, Fourth Edition, 2015
- 3. Dirk Deroos, Paul C.Zikopoulos, Roman B.Melnky, Bruce Brown, Rafael Coss, "Hadoop For Dummies", Wiley Publications, 2014
- 4. Robert D.Schneider, "Hadoop For Dummies", John Wiley & Sons, Inc. (2012)
- 5. Paul Zikopoulos, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill, 2012



19ITE305 Big Data Analytics/ Unit 1