



SNS COLLEGE OF ENGINEERING



Kurumbapalayam(Po), Coimbatore – 641 107

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Artificial Intelligence and Data Science

19IT601- Data Science and Analytics

III Year / VI Semester

Unit 1 – Introduction

Topic 3- Types of Digital Data





Types of Digital Data

1. Structured Data

- It owns a dedicated data model. It also has a well defined structure, it follows a consistent order and it is designed in such a way that it can be easily accessed and used by person or a computer. Structured data is usually stored in well defined columns and databases.
- – Structured Schema
- – Tables with rows and columns of data
- – Example : DBMS,RDBMS

- When a data conforms to a pre-defined schema / structure we say it is structured data.

Sources of structured data

Databases: Oracle Corp-Oracle, IBM-DB2, Microsoft-Microsoft SQL Server, EMCGreenplum, Teradata-Teradata, MySQL, PostgreSQL.

Spreadsheets : MS Excel, Google sheets

On-Line Transaction Processing (OLTP) Systems



Types of Digital Data

2.Semi-Structured Data

It is considered as another form of structured data. It inherits few properties of structured data, but major parts of this kind of data fail to have a definitive structure and also it does not obey the formal structure of data models such as RDBMS.

This is the data which does not conform to a data model but has some structure

- Schema is not defined properly
- JSON, XML, CSV, RSS
- Ex: Transactional history file, Log file

Semi-structured data is also referred to as self-describing structure. It has the following features

- It does not conform to the data models that one typically associates with relational databases or any other form of data tables.
- It uses tags to segregate semantic elements.
- Tags are also used to enforce hierarchies of records and fields within data.



Types of Digital Data

3. Unstructured Data

Unstructured data is completely different of which neither has a structure nor obeys to follow formal structural rules of data models. It does not even have a consistent format and it found to be varying all the time. But rarely it has information related to data and time.

- Heterogeneous Data
- Text file, Images, Videos, Audio.

Sources of Unstructured Data

Web Pages, Images, Free-Form Text, Audios, Videos, Body of Email, Text, Messages, Chats, Social Media data, Word Document.



THANK YOU