



SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)
COIMBATORE – 35

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (UG & PG)

Subject Code & Name: 19ITE305-Big Data& Analytics

Classification of big data:

Structured Data:

Structured data refers to highly organized information that is easily searchable and typically stored in relational databases or spreadsheets. It adheres to a rigid schema, meaning each data element is clearly defined and accessible in a fixed field within a record or file. Examples of structured data include:

- Customer names and addresses in a [customer relationship management](#) (CRM) system
- Transactional data in financial records, such as sales figures and account balances
- Employee data in human resources databases, including job titles and salaries

Structured data's main advantage is its simplicity for entry, search and analysis, often using straightforward database queries like SQL. However, the rapidly expanding universe of big data means that structured data represents a relatively small portion of the total data available to organizations.

Unstructured Data:

Unstructured data lacks a pre-defined data model, making it more difficult to collect, process and analyze. It comprises the majority of data generated today, and includes formats such as:

- Textual content from documents, emails and social media posts
- Multimedia content, including images, audio files and videos
- Data from IoT devices, which can include a mix of sensor data, log files and time-series data

The primary challenge with unstructured data is its complexity and lack of uniformity, requiring more sophisticated methods for indexing, searching and analyzing. NLP, machine learning and advanced analytics platforms are often employed to extract meaningful insights from unstructured data.

Semi-structured data:

Semi-structured data occupies the middle ground between structured and unstructured data. While it does not reside in a relational database, it contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. Examples include:

- JSON (JavaScript Object Notation) and XML (eXtensible Markup Language) files, which are commonly used for web data interchange
- Email, where the data has a standardized format (e.g., headers, subject, body) but the content within each section is unstructured
- NoSQL databases, can store and manage semi-structured data more efficiently than traditional relational databases

Semi-structured data is more flexible than structured data but easier to analyze than unstructured data, providing a balance that is particularly useful in web applications and data integration tasks.