



Semester : _____

Subject : _____

Academic Year: 20 - 20

Types of Big Data:-

Big data is essentially classified into three types:-

— Structured Data —

Structured data is highly organized and thus, is the easiest to work with. Its dimensions are defined by set parameters. Every piece of information is grouped into rows and columns like spreadsheets. Structured data has quantitative data such as age, contact, address, billing, expense, debit or credit card number etc.

Due to structured data's quantitative nature, it is easy for programs to sort through and collect data. It requires little to no preparation to process structured data. The data only needs to be cleaned and pared down to the relevant points. The data does not need to be converted or interpreted too deeply to perform inquiry.

Unstructured Data:-

Any set of data that is not structured or well-defined is called unstructured data. This kind of data is unorganized and difficult to handle, understand and analyze. It does not follow a consistent format and vary at different point of time. Most of the data you encounter comes under this category.

For example, unstructured data are your comments, tweets, shares, posts and likes on social media. The videos we watch on Youtube and Text messages we send via WhatsApp all pile up as a huge heap of unstructured data.

Semi-Structured Data

Semi-structured data falls somewhere between structured data and unstructured data. It mostly translates to unstructured data that has metadata attached to it. Semi-structured data can be inherited such as location, time, email-address, or device ID stamp. It can even be a semantic tag attached to data later.



PARSHWANATH CHARITABLE TRUST'S

A.P. SHAH INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering
Data Science



Semester : VII

Subject : Big Data Analytics

Academic Year: 2024-2025





Difference between Structured, Unstructured and Semi structured Data

Aspect	Structured Data	Unstructured Data	Semi-Structured Data
Definition	Organized	Lack of predefined structure	Mix of organization and flexibility
Examples	Sales transactions in a relational database, student records in spreadsheets.	Social media posts, customer reviews, medical images, and audio recordings.	XML documents, JSON data, NoSQL databases.
Storage Efficiency	Efficient storage and retrieval due to organized format.	Varied storage efficiency based on content types, can be challenging to manage.	Balances storage efficiency with flexibility, optimized for complex data structures.
Querying	Well-suited for structured query languages (SQL), efficient querying.	Challenging for querying, requires advanced techniques like natural language processing.	Requires specialized querying techniques, adaptable to complex relationships.
Data Complexity	Well-organized and straightforward to manage.	Chaotic and challenging to organize due to lack of structure.	Balances flexibility with some level of organization, and moderate complexity.
Flexibility	Limited flexibility, data must adhere to a predefined structure.	Highly flexible, can capture diverse content, but may lack uniformity.	Offers flexibility while maintaining some level of structure, adaptable to changes.
Integration	Well-suited for traditional relational databases and structured applications.	May require advanced integration techniques due to diverse formats.	Adaptable for web applications, APIs, and systems with varying data sources.
Analysis Difficulty	Easier to analyze, suitable for quantitative analysis and reporting.	Requires advanced techniques for sentiment analysis, pattern recognition, etc.	Complex analysis may involve specialized techniques but accommodates diverse structures.
Scalability	Efficient for managing large volumes due to structured format.	Scalability challenges due to data diversity and potential noise.	Scalable, but complexity might increase with data volume and structure.