



SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)



Approved by AICTE, New Delhi, Affiliated to Anna University, Chennai
Accredited by NAAC-UGC with 'A++' Grade (Cycle III) & Accredited by NBA (B.E - CSE, EEE, ECE, Mech & B.Tech.IT)
COIMBATORE-641 035, TAMIL NADU

DEPARTMENT OF COMPUTER APPLICATIONS

19CAE716 – DATA SCIENCE

UNIT – I: INTRODUCTION TO DATA SCIENCE

TOPIC: BIG DATA & DATA SCIENCE



BIG DATA

1. Volume:

1. Definition: Refers to the sheer size of the data generated or collected.
2. Example: Terabytes, petabytes, or even exabytes of data.

2. Velocity:

1. Definition: Describes the speed at which data is generated, processed, and transmitted.
2. Example: Real-time data streams, continuous data generation.

3. Variety:

1. Definition: Encompasses the different types and formats of data, including structured, unstructured, and semi-structured data.
2. Example: Text, images, videos, sensor data, social media posts.

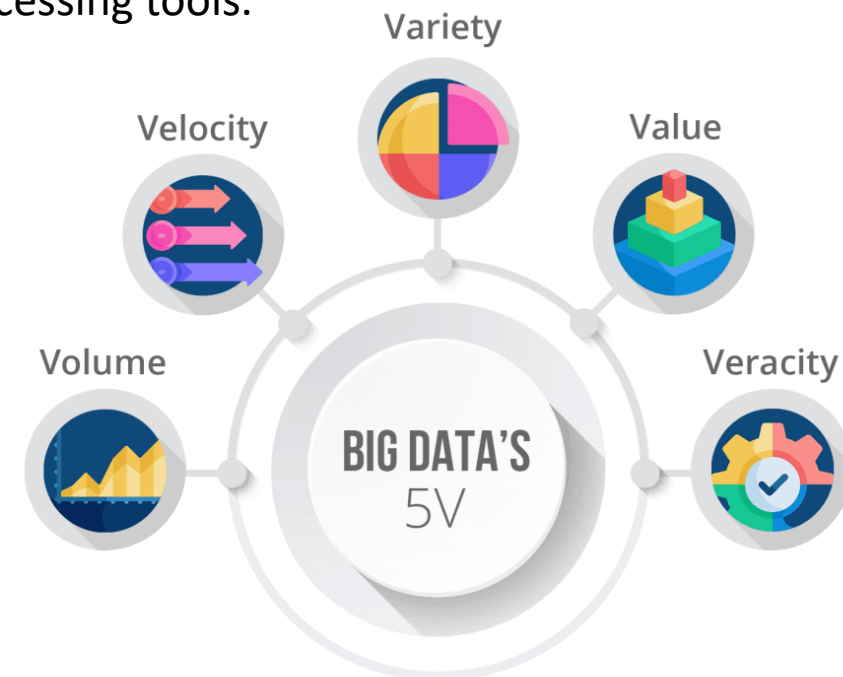
4. Value:

1. Definition: Refers to the inconsistency or variations in the data flow, both in terms of its format and meaning.
2. Example: Fluctuations in data quality, changing data sources.

5. Veracity:

1. Definition: Addresses the quality and reliability of the data, including accuracy, trustworthiness, and consistency.
2. Example: Assessing the trustworthiness of data sources, dealing with data uncertainty.

Big data refers to extremely large and complex sets of data that cannot be easily managed, processed, or analyzed with traditional data processing tools.





Big Data vs Data Science

Definition	Extremely large and complex datasets.	An interdisciplinary field that extracts insights from data using scientific methods.
Focus	Handling, managing, and processing large volumes of data.	Analyzing and interpreting data to derive actionable insights.
Characteristics	Volume, Velocity, Variety, Veracity, Value.	Extraction of insights, patterns, and knowledge from data.
Challenges	Storage, processing, and analysis of massive datasets.	Understanding complex data structures and applying advanced analytics techniques.
Data Types	Structured, semi-structured, and unstructured data.	Structured and unstructured data.
Processes	Data storage, data processing, and distributed computing.	Data cleaning, exploration, modeling, and interpretation.



Big Data vs Data Science

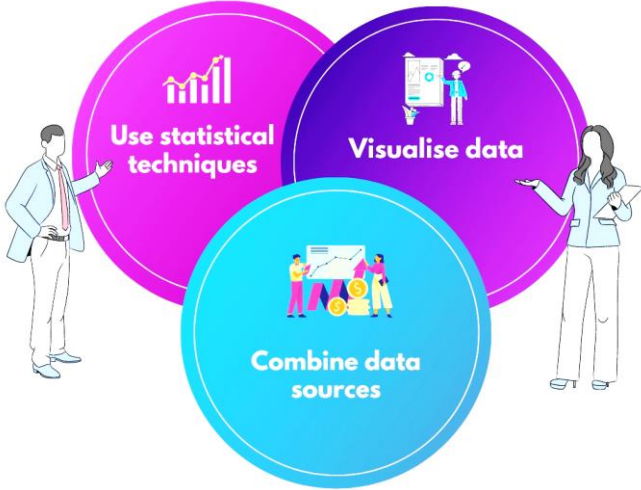
Goal	Efficiently manage and process vast amounts of data.	Extract actionable insights and knowledge for decision-making.
Scale	Dealing with large-scale datasets, often in petabytes or exabytes.	Can work with both small and large datasets, focusing on insights rather than scale.
Applications	Used in various industries for analytics, insights, and decision support.	Applied across industries for predictive modelling, pattern recognition, and business intelligence.
Integration	Data science techniques are applied within big data analytics processes.	Often integrated with big data technologies for efficient analysis.
Outcome	Enabling organizations to store, process, and analyze large datasets efficiently.	Providing actionable insights, predictions, and knowledge to support decision-making.



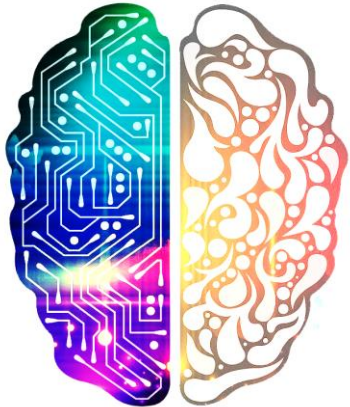
Big Data vs Data Science



STATISTICAL ANALYSIS



MACHINE LEARNING



Datawrapper