Unit I: Big Data Characteristics in Automation Systems

Welcome, Mechatronics Engineering students of SNS College of Technology. This unit explores Big Data's fundamental concepts and its unique impact on modern automation systems. We will delve into how data is transforming the landscape of industrial operations.

by Karthi N,AP/MCT - SNSCT

The 5 Vs of Big Data

1. Volume

Immense quantities of data. Think terabytes or petabytes from sensors, machine logs, and vision systems. A modern factory can generate 1PB annually from PLCs and robotics.

2. Velocity

Speed of data generation and processing. Real-time streams are vital for predictive maintenance (sub-second latency) and robotic control (milliseconds)

5.Value

Transforming data into meaningful, actionable insights is the ultimate goal. This means improving OEE, optimizing energy, reducing downtime and enhancing safety.
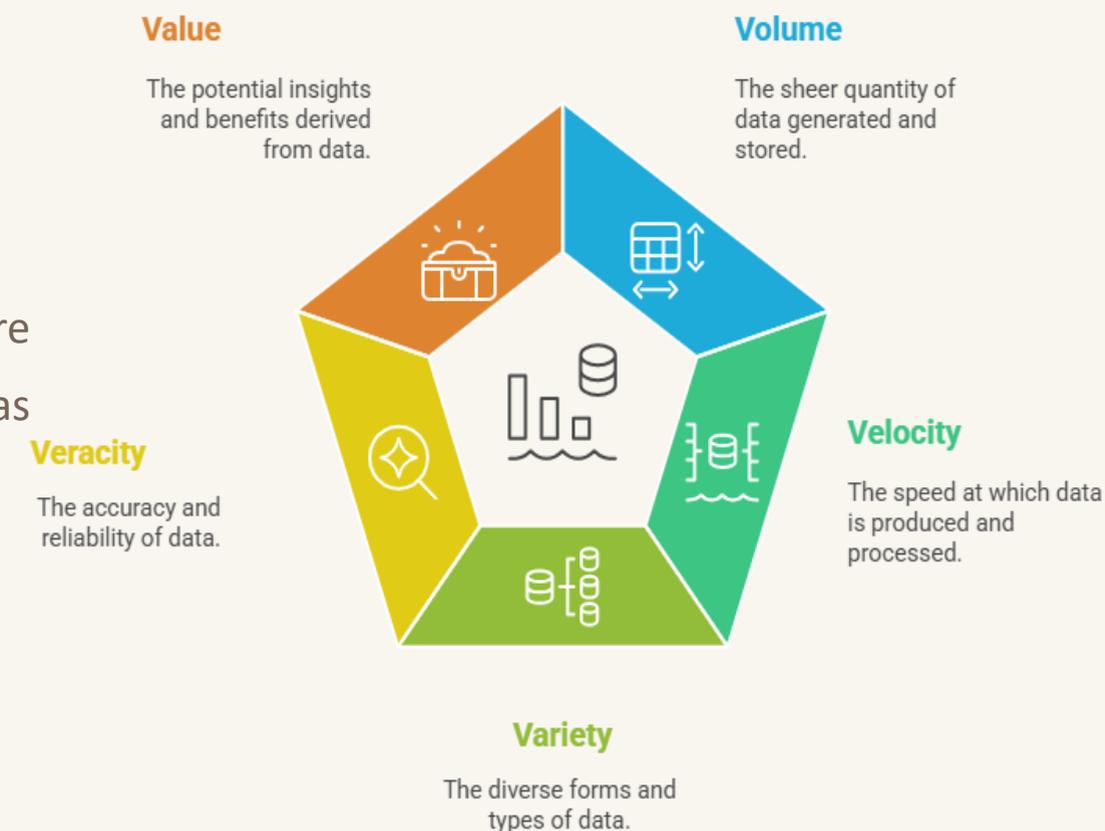
3. Variety

Diverse forms of data. This includes structured sensor readings, semi-structured logs, and unstructured images or audio diagnostics. Combining motor temperature with video feeds illustrates this.

4. Veracity

Data quality, accuracy, and trustworthiness are key. This includes managing noise, errors, bias from sensors and missing values.

Foundations of Big Data

Value
The potential insights and benefits derived from data.

Volume
The sheer quantity of data generated and stored.

Veracity
The accuracy and reliability of data.

Velocity
The speed at which data is produced and processed.

Variety
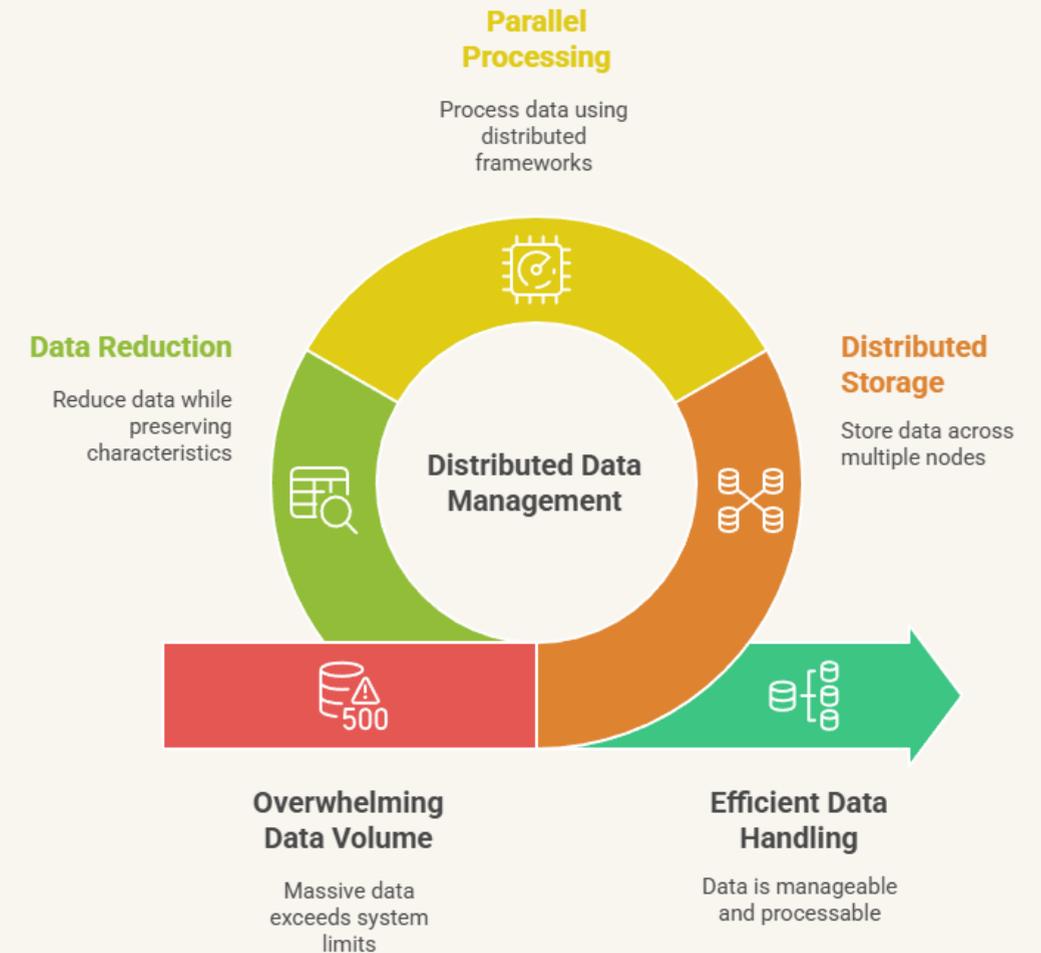The diverse forms and types of data.

Made with Napkin

VOLUME

- Volume refers to the sheer *amount* of data being generated and processed.

- This is perhaps the most commonly associated characteristic of big data.

- The scale of data is enormous, often measured in terabytes (TB), petabytes (PB), exabytes (EB), and even zettabytes (ZB).

Examples:

•Social media platforms generate massive amounts of user-generated content daily, including posts, images, videos, and comments.

•E-commerce websites collect data on customer browsing behavior, purchase history, and product reviews.

•Sensor networks in manufacturing plants generate vast amounts of data on machine performance, environmental conditions, and product quality.

Managing Big Data Volume

Parallel Processing
Process data using distributed frameworks

Data Reduction
Reduce data while preserving characteristics

Distributed Data Management

Distributed Storage
Store data across multiple nodes

Overwhelming Data Volume
Massive data exceeds system limits

Efficient Data Handling
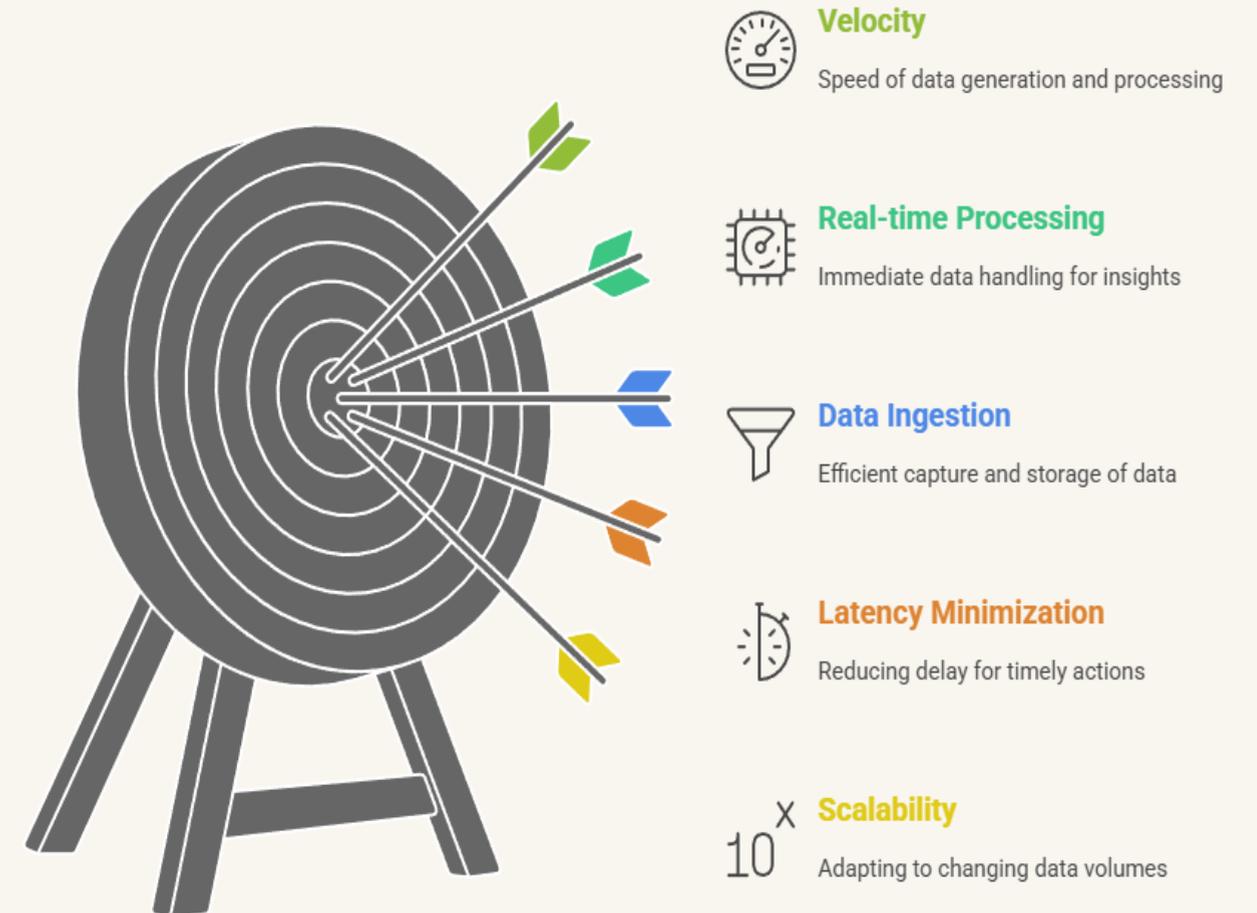Data is manageable and processable

Made with Napkin

VELOCITY

- Velocity refers to the *speed* at which data is generated and processed.

- In many applications, data is not only large but also arrives continuously and rapidly.

- This requires real-time or near real-time processing capabilities.

Examples:

- Financial markets generate high-velocity data on stock prices, trading volumes, and market sentiment.

- Network monitoring systems collect data on network traffic, server performance, and security events in real-time.

- Online advertising platforms process data on user clicks, impressions, and conversions to optimize ad campaigns.

- IoT devices continuously stream data from sensors, actuators, and other connected devices.

Velocity in Big Data Processing

Velocity
Speed of data generation and processing

Real-time Processing
Immediate data handling for insights

Data Ingestion
Efficient capture and storage of data

Latency Minimization
Reducing delay for timely actions

Scalability
Adapting to changing data volumes

Made with Napkin

VARIETY

Data Type Comparison

	Structured Data	**Semi-structured Data**	**Unstructured Data**
Format	Predefined format	Some organizational properties	No predefined format
Examples	Relational databases, spreadsheets	JSON, XML, log files	Text documents, images, audio
Integration	Relatively easy	Requires transformation	Very challenging
Governance	Easier to manage	Requires careful planning	Difficult to manage
Schema	Rigid	Flexible	Schema-on-read
Discovery	Straightforward	Needs metadata analysis	Complex

- Variety refers to the *diversity* of data types and formats.
- Big data encompasses structured, semi-structured and unstructured data.

Examples:

•A healthcare organization collects structured data on patient demographics and medical history, semi-structured data on lab results and clinical notes, and unstructured data on medical images and doctor's reports.

•A marketing company collects structured data on customer purchases, semi-structured data on website activity, and unstructured data on social media posts.

•A scientific research project collects structured data on experimental parameters, semi-structured data on simulation results, and unstructured data on research papers.

VARACITY

How to ensure data veracity in big data analytics?

Data Quality
Implement data cleaning and validation techniques to address errors and inconsistencies.
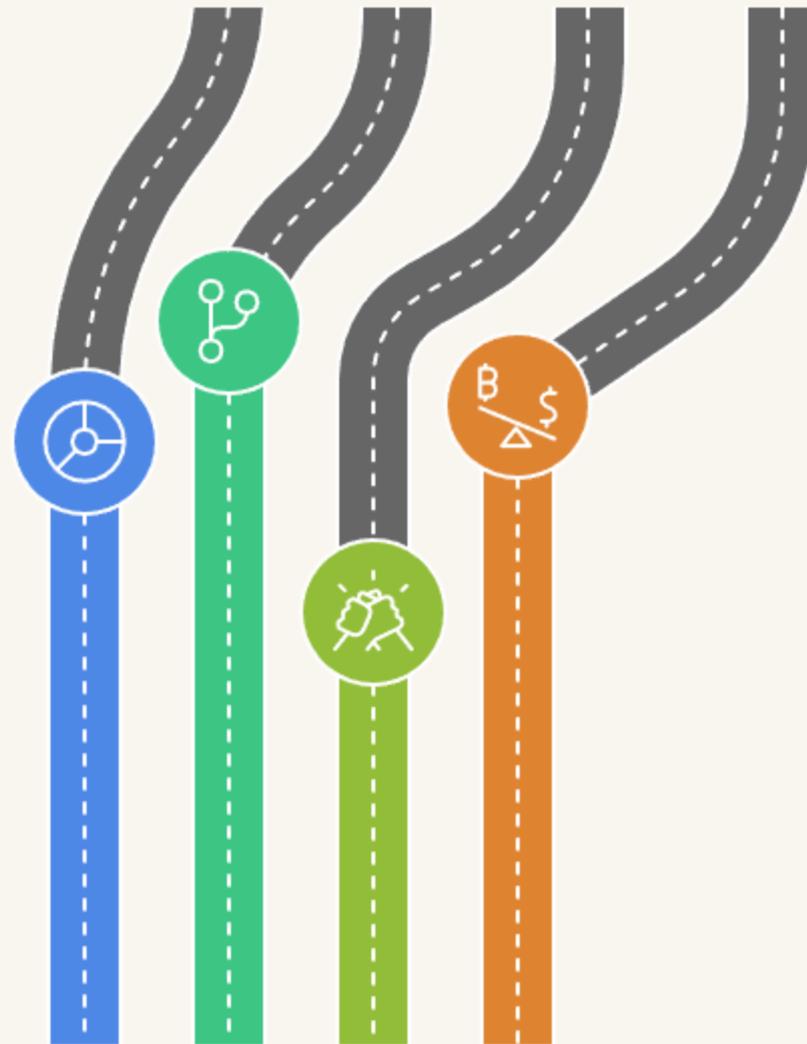
Data Provenance
Track data origin and lineage to understand reliability.

Data Trust
Build trust in data for widespread adoption.

Bias Detection
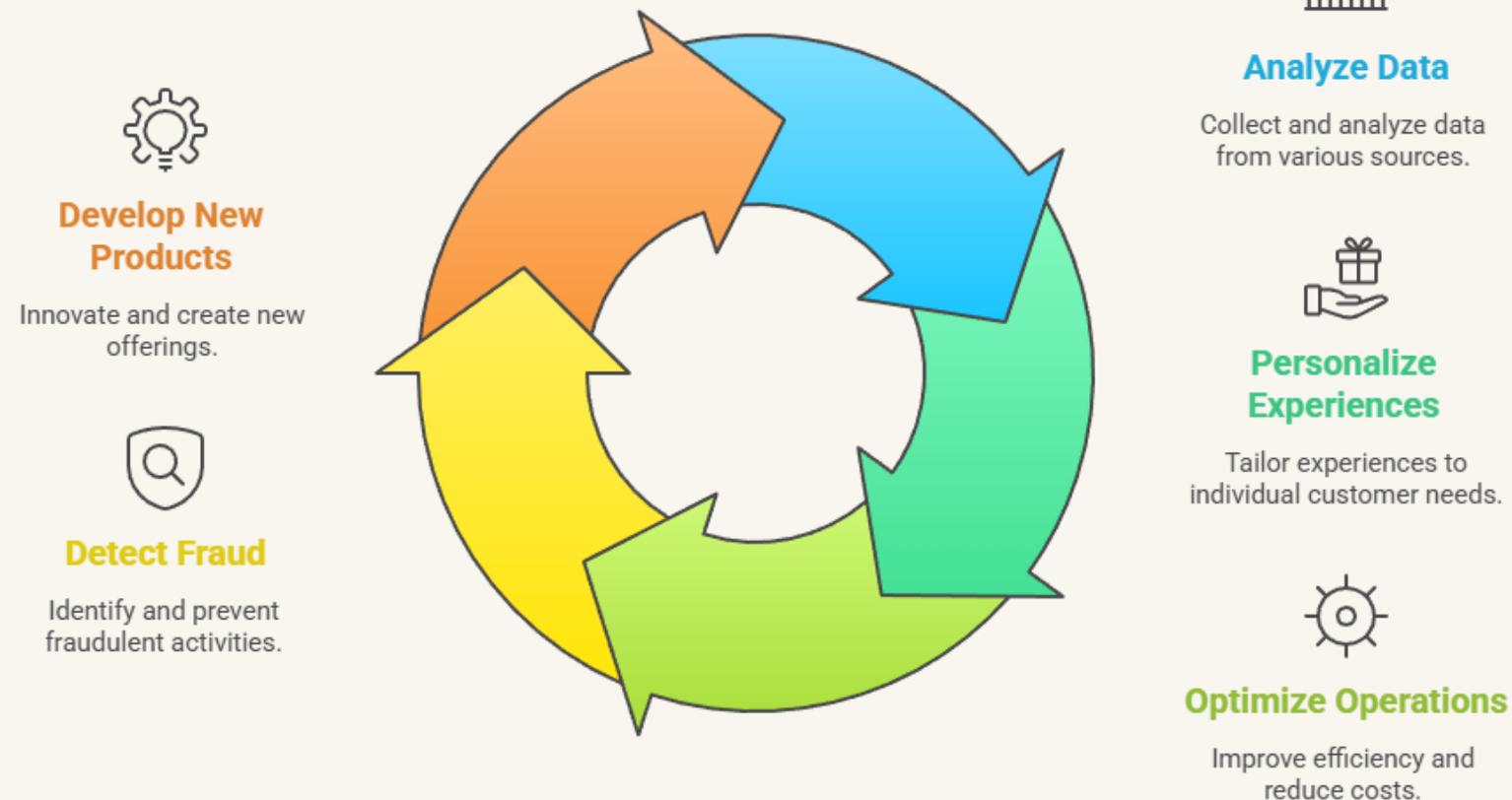Identify and mitigate biases to ensure fairness.

Made with Napkin

- Veracity refers to the *accuracy* and *reliability* of data.
- Big data often comes from multiple sources, and the quality of data can vary significantly.
- Inaccurate or inconsistent data can lead to misleading insights and poor decision-making.

Examples:

•Social media data can be noisy and contain misinformation.

•Sensor data can be affected by calibration errors and environmental factors.

•Customer data can be incomplete or outdated.

•Web scraping data can be inaccurate or biased.

VALUE

Big Data Value Cycle

Develop New Products
Innovate and create new offerings.

Detect Fraud
Identify and prevent fraudulent activities.

Analyze Data
Collect and analyze data from various sources.

Personalize Experiences
Tailor experiences to individual customer needs.

Optimize Operations
Improve efficiency and reduce costs.

Made with Napkin

- Value refers to the business value that can be derived from big data.

- Ultimately, the goal of big data analytics is to extract meaningful insights that can improve decision-making, optimize operations, and create new business opportunities.

Examples:

- Using big data to personalize customer experiences and increase sales.

- Using big data to optimize supply chain operations and reduce costs.

- Using big data to detect fraud and prevent financial losses.

- Using big data to develop new products and services.

Conclusion & References

Big Data in automation is defined by its Volume, Velocity, Variety, Veracity, and Value. Understanding these characteristics is crucial for Mechatronics engineers. It transforms traditional automation into intelligent, adaptive, and predictive systems. Embracing Big Data analytics is essential for careers in Industry 4.0 and advanced manufacturing.

> **Reference:** Loshin, D. (2013). *Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph*. Morgan Kaufmann.