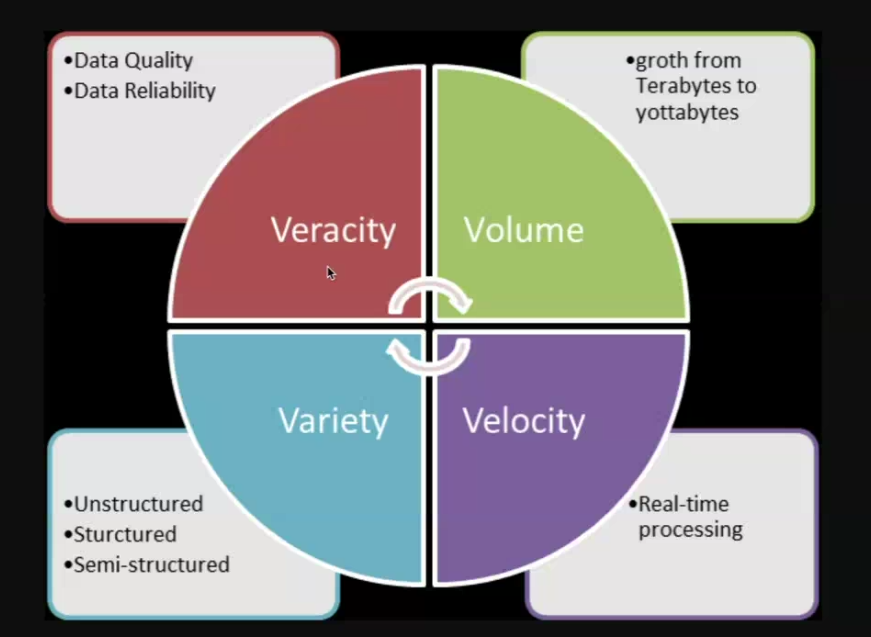
Introduction to Big Data and the 4Vs

**What is Big Data?**

Big Data refers to extremely large datasets that are complex and grow rapidly, making them difficult to manage and process using traditional data processing tools. The term doesn't only refer to the size of data but also its variety, speed, and trustworthiness. To better understand the nature of Big Data, the concept of the **4Vs** is used:

* **Volume**
* **Velocity**
* **Variety**
* **Veracity**

Each “V” represents a characteristic of data that makes it "big.”



**1. Volume**

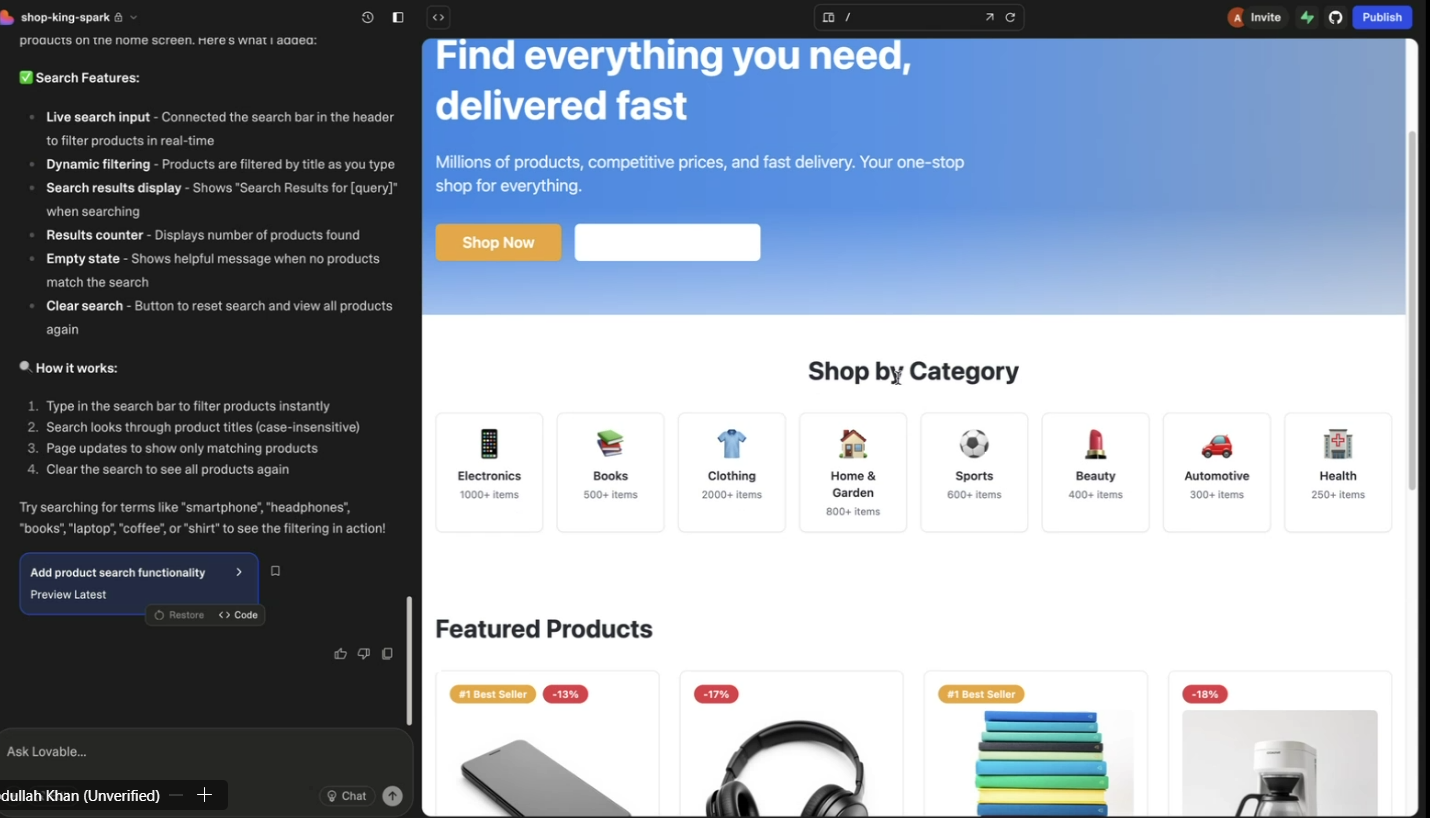
**Definition:** Volume refers to the quantity of data generated and stored. The scale of the data determines whether it is considered Big Data. (Growth from terabyte to yottabytes).

**Lovvable AI:**  
Lovvable AI is an AI-driven marketing platform that collects massive volumes of customer behaviour data from multiple touchpoints – including websites, social media, and purchase history. This volume helps them build customer personas and personalize marketing campaigns.

Businesses generate terabytes or even petabytes of data daily.

Social media platforms (e.g., Instagram, Facebook) deal with large volumes of user-generated content.

So being in touch with the emerging technologies



**2.Velocity**

**Definition:** Velocity refers to the speed at which data is generated, collected, and processed. Real-time processing is often essential in Big Data environments. (Real-time processing )

**AGENTS AI:**  
Agents AI uses live customer interaction data in real-time to help agents respond intelligently. For example, chatbots powered by Agents AI analyze input instantly and provide responses in milliseconds.

**3. Variety**

**Definition:** Variety represents the different types of data – structured, semi-structured, and unstructured – from various sources.(Different forms of data).

**Dmart:**  
Dmart handles various forms of data such as:

* Structured: Billing records, inventory, Online ordering system.
* Semi-structured: Customer feedback in forms, some .csv files.
* Unstructured: When people visit to Dmart on their own and pay the bills through pos.

**4. Veracity**

**Definition:**  
Veracity refers to the trustworthiness, accuracy, and quality of data. In the world of Big Data, not all data is clean or reliable. Low veracity means the data might be incomplete, inconsistent, outdated, or biased — leading to wrong conclusions and ineffective decisions. ( Data quality, Data Reliability)

**Survey on Women’s Rights:**  
Imagine a survey was conducted to understand **public opinion on women's rights policies**. However, the sampling was flawed:

* **80% of the responses came from men**, and only **20% from women**.

As a result, the data **does not accurately reflect women’s perspectives**, which is critical for such a topic. This leads to **bias** and **low veracity** of the dataset. Any conclusions drawn from this survey would be misleading and could result in poor policy-making.