🎓 Video 13 Summary: Data Ingestion Techniques – Batch vs. Streaming

📥 What Is Data Ingestion?

Data ingestion is the process of **collecting data from multiple sources** and moving it into a **staging area** for further processing and analysis.

🌐 Why It Matters

Organizations rely on diverse data sources:

- Websites
- Point-of-sale systems
- Social media
- Machinery sensors

Each source provides **partial insights**, so combining them is essential for a **complete picture**.

⏱️ Time Sensitivity in Data Ingestion

Choosing the right ingestion technique depends on whether the data is **time-sensitive**:

- If data **must be acted on quickly**, use **streaming ingestion**.
- If data can **wait**, use **batch ingestion**.

🚍 Analogy: Bus vs. Taxi

- **Bus (Batch)**: Waits for scheduled processing — economical and efficient for large volumes.
- **Taxi (Streaming)**: Immediate processing — ideal for real-time needs.

🔁 Two Main Ingestion Techniques

1. Batch Ingestion

- **Definition**: Collects and processes data in groups at scheduled intervals.
- **Best for**: High-volume data that isn't time-critical.
- **Benefits**:
  - More economical
  - Requires less computing power
- **Example**:
  A non-profit collects donation data daily from its website and phone system, then processes it once per day.

2. Streaming Ingestion

- **Definition**: Processes data **as soon as it becomes available**.
- **Best for**: Real-time monitoring and fast response.
- **Example**:
  A pharmaceutical company monitors dryer temperatures via sensors and reacts immediately to unsafe changes.

🧠 Takeaway

Choose **batch ingestion** for efficiency and cost savings when timing isn't critical. Choose **streaming ingestion** for real-time responsiveness when data must be acted on immediately.