🎬 Video 24 – Data Derivation

🧠 What Is Data Derivation?

- **Definition**: The process of combining and processing existing (base) data using an algorithm to create new, derived data.
- **Purpose**: Unlock insights that aren't directly available from raw data.

🔍 Example Scenario

- **Business Need**: Track shoes that have been on warehouse shelves for 30+ days.
- **Solution**: Use the **arrival date-timestamp** to calculate shelf duration.
- **Derived Metric**: Days on shelf → used to generate a daily report for decision-making.

⚙️ How It Works

1. **Start with base data** (e.g., timestamps).
2. **Apply an algorithm** (e.g., date difference calculation).
3. **Generate new data** (e.g., shelf duration).
4. **Use derived data** for reporting, analysis, and visualization.

⚠️ Challenges of Data Derivation

1. Accuracy Risks

- Errors in the algorithm can lead to incorrect derived data.
- Base data may contain errors or change after derivation.

2. Data Privacy

- Derived data involving **PII (Personally Identifiable Information)** must comply with consent and privacy regulations.
- Consent for original data use doesn't automatically apply to derived uses.

✅ Best Practices

- Validate algorithms carefully.
- Monitor changes in base data.
- Respect privacy and ownership policies.
- Use derived data to enhance performance and answer complex questions.

💡 Key Takeaway

Data derivation is a powerful tool for cloud data analysts, enabling deeper insights—but it must be used with care to ensure accuracy and compliance.