**Sources (Data Producers)**

These are the systems where raw data originates. Sources can be:

* **Databases:** MySQL, PostgreSQL, MongoDB, Cassandra
* **Files & Blobs:** CSV, JSON, Parquet, Avro (stored in AWS S3, Google Cloud Storage, Azure Blob)
* **APIs & Web Services:** REST APIs, SOAP APIs, GraphQL
* **Streaming Sources:** Kafka, AWS Kinesis, Google Pub/Sub
* **Application Logs & Event Data:** Server logs, IoT data

**2. Data Ingestion Methods**

Defines how data is extracted and loaded into the warehouse:

* **Batch Ingestion:** Periodic data transfer (e.g., using Apache NiFi, Airflow, Dataflow)
* **Streaming Ingestion:** Real-time data ingestion (e.g., Kafka, Flink, Spark Streaming)
* **CDC (Change Data Capture):** Captures and syncs changes from source databases (e.g., Debezium, Fivetran,