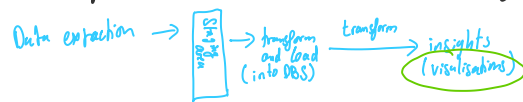Data pipelines
Monday, 9 February 2026   09:14

Data pipelines

: is a set of tools and processes for collecting, processing and delivering data
- gathers it from the source and transforms it for suitability for target location
- extracts data from various sources
- A well designed data pipeline is crucial to leverage data effectively and gain insights.

Data extraction → [single data] → transform and load (into DBS) → transform → insights (visualisations)

two Common data pipeling:

ETL — extract, transform, load

ELT - extract, Load, transform.

Common challenges to creating data pipelines:

- Connection
- Flexibility
- Centralisation
- Latency

Components of Data pipeline:

1. Source
2. Destination
3. Data flow
4. Processing
5. Workflow
6. monitoring.

Data pipeline architecture

1. Data ingestion
2. Pre-processing
3. Information Extraction
4. ETL
5. presentation area.

Typical pipeline

data ingestion ⟶ Data processing → Data Storage → Data workflow orchestration → Data visualisation

+ve

- if dealing with multiple data sources.
- Big data
- Real time analytics
- Complex data processing

ETL / ELT tools

- Mattillion, Talend, Apache Nifi

Data Storage: Snowflake, Google Big Query, Amazon redshift