DATABASE BASICS

HOW ARE DATA ORGANIZED?

1. **STRUCTURED**
   - Follows a schema, which defines the organization & relationships of the data = efficient queries.
   - Ex: relational data, SQL

2. **SEMI-STRUCTURED**
   - Not constrained by a fixed schema = flexible & portable, but harder to query & optimize
   - Ex: XML, JSON, NoSQL

3. **UNSTRUCTURED**
   - No inherent structure = tricky to extract & search. Also, requires more storage space.
   - Ex: Online forums, voice recordings, PDFs

HOW ARE DATA STORED?

OLD SCHOOL
- Traditional OLTP
  - Real-time querying

NEW SCHOOL
- Data Warehouse
  - Analysis of structured data
  - Curated data ("source of truth")
  - Schema - on - write

- Data Lake
  - Storage of all types of data
  - Very large, low cost
  - Schema - on - read

HOW DO DATA FLOW?

**ETL**
- Extract
- Transform
- Load
- Data Warehouse

**ELT**
- Extract
- Load
- Transform
- Data Lake

**ETL** ensures that data adhere to database schemas. It is reliable but ingestion is slower.
**ELT** allows for speedier loading of data & ad-hoc transformations at query time, but setup can be complex.