dotnotes

DATABASE BASICS

HOW ARE DATA ORGANIZED?

TYPES OF DATABASE SYSTEMS

OLTP
"Online Transaction Processing"

vs

OLAP
"Online Analytical Processing"

- Frequent updates
- Simple, standard queries
- Fast processing, real-time usage
- Critical for operational processes

- Created from archived OLTP data
- Supports complex, ad-hoc queries
- Very large, indexing required
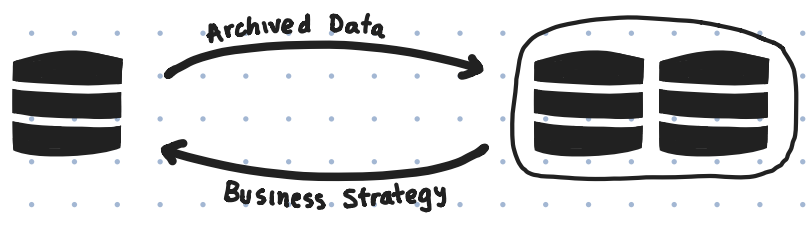- Enables high-level decision making

Archived Data

Business Strategy

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.
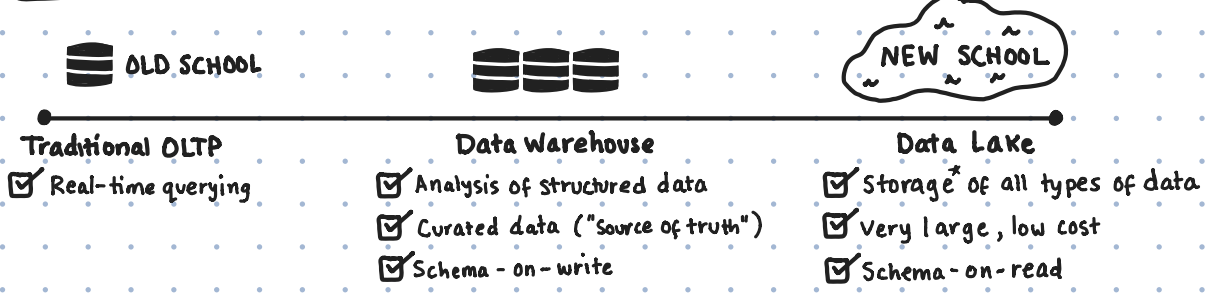- Ex: Online forums, voice recordings, PDFs

---

DATABASE DESIGN => How will data be stored, read, updated & queried?

At a high level, **DATABASE MODELS** represent the logical structure, including relationships & constraints. Ex: relational model, NoSQL model

At a granular level, **SCHEMAS** define the actual implementation, like blueprints.
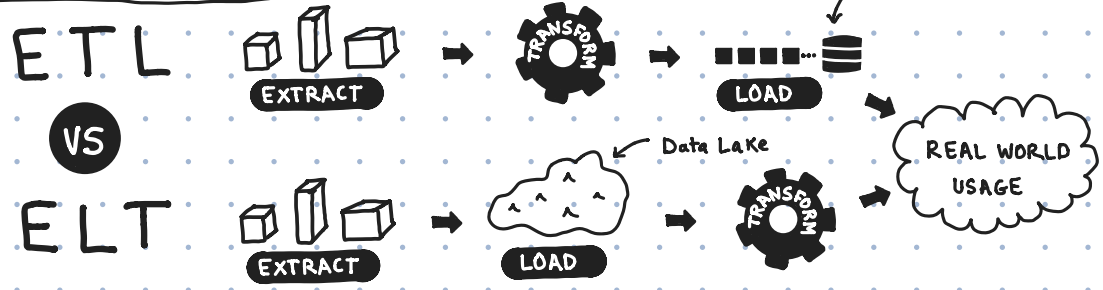Ex: what fields? what datatypes?

Diagrams are used to illustrate both models & schemas.

HOW ARE DATA STORED?

OLD SCHOOL ――――――――――― **NEW SCHOOL**

Traditional OLTP
- ☑ Real-time querying

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

> *New services like Spark & Hadoop enable analytics to be performed on data lake stores.

HOW DO DATA FLOW?

ETL
vs
ELT

EXTRACT → TRANSFORM → LOAD → *Data Warehouse*

EXTRACT → *Data Lake* → LOAD → TRANSFORM → REAL WORLD USAGE

> *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.