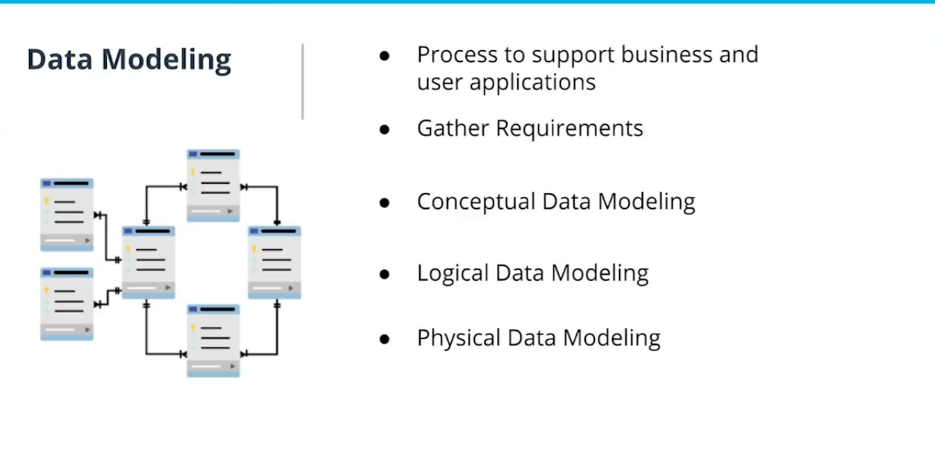
**DATA ENGINEERING**

1. DATA MODEL?

‘An Abstraction that organizes elements of data and how they relate to each other’

‘The processing of creating data modeling for an information system’

‘The Data Modeling easily called Database modeling’



To Begin with the data modeling process, the team must **gather requirements** from the application team, the business user, and our end-user to understand that data must be retained and served to the business or the end-user.

Next, we focused on **conceptual data modeling** with entity mapping. This can be done by hand or with many tools to do this work. This is the mapping of the concepts of the data that you have or will have. Data will be organized in the process.

Next, will be logical data modeling, it will done where the conceptual data models are mapped, using the concept of tables, schemas and columns.

Next, will be Physical data modeling, it is done by transforming the logical data model to the Data Definition Language(DDL).

Who does this type of work? -> Data Scientist, Data Engineer, Software Engineer, and Those Who are working with databases.

1. **DataBase:**

Relation and Non-Relation Database Do Data modeling Differently.

1. **Relation Database:**
2. **Relational Model:**

Chart, diagram

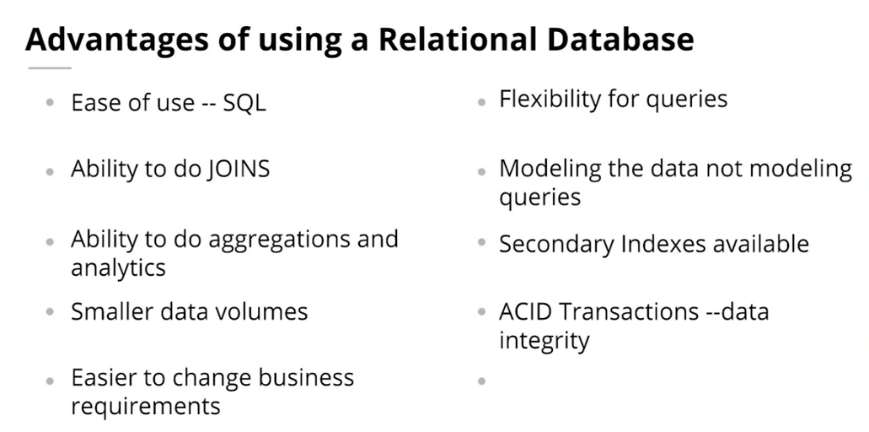
Description automatically generated

The Digital Database based on the relational model of data is called a Relational Database.

And the software that maintains relational databases is a relational Database management system (RDBMS).

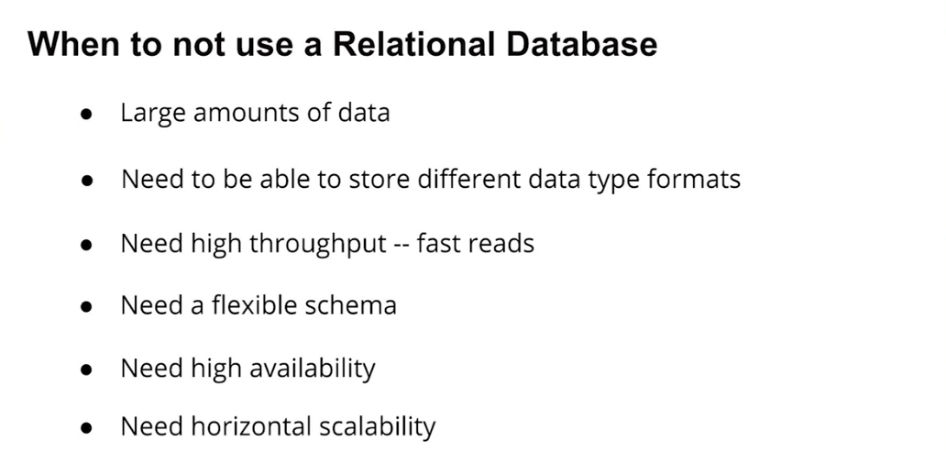
Most Common Type of Relation Databases.

Oracle, TeraData, MySql, PostgreSQL, Sqlite.



1. **ACID Properties:**

Atomicity, Consistency, Isolation, Durability



1. **Non -** **Relation Database: (NO SQL)**

No SQL = Not Only SQL; No SQL and Non Relational are interchangeable terms

Have simple design, simple horizontal scaling, finer control of availability. Different Data Structure are used different than Relational Database are make some operations faster.

Graphical user interface, text, application

Description automatically generated

**APACHE CASSANDRA:**

Apache Cassandra provides scalability and high availability without compromising performance.

Its use own query Language call CQL

Table

Description automatically generated

Table

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

**DataBase:**

A set of related data and the way it is organized.and DBMS consisting of computer software that allows user to interact with database and provide all access to all of the data.

Graphical user interface, application

Description automatically generated Text

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1st Normal Form: Automic Value.

2nd Normal Form: All Column in the able rely on primary key.

3rd Normal Form: No Transitive Dependencies.

Graphical user interface, text, application, email

Description automatically generated

**Facts and Dimension:**

**Fact**: It consist tables consists of the measurements, metrics or facts of a business process.

**Dimension**: A structure that categorizes facts and measures in order to enable users to answer business questions. Dimensions are people, products, place and time.

**Implementing Different Schemas**:

Two of the most popular data mart schema for data warehouses are :

1. Star Schema
2. SnowFlake Schema

Timeline

Description automatically generated

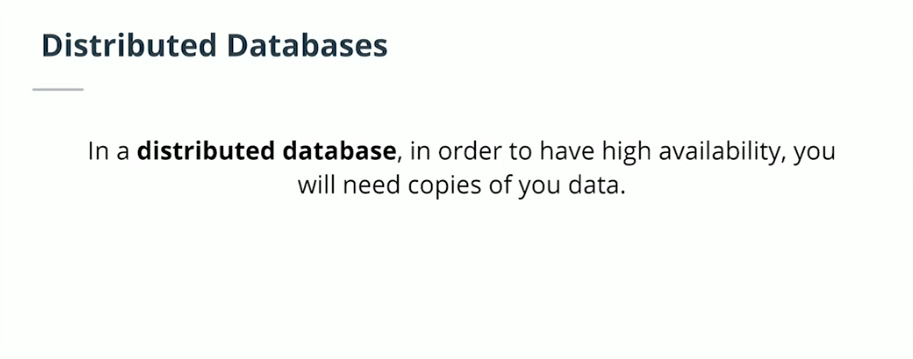
Graphical user interface, text, application

Description automatically generated

**NO SQL:**

Graphical user interface, application

Description automatically generated

****

**Eventual Consistency:**

If the system is distributed and my node will go down this means I need copies of my data. My data can not live at one place with the fact that my data is copies throughtout my system leads to with the fact that my data is copied throughtout my system leads to the fact my data may not be up to date in all locations.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Diagram

Description automatically generated

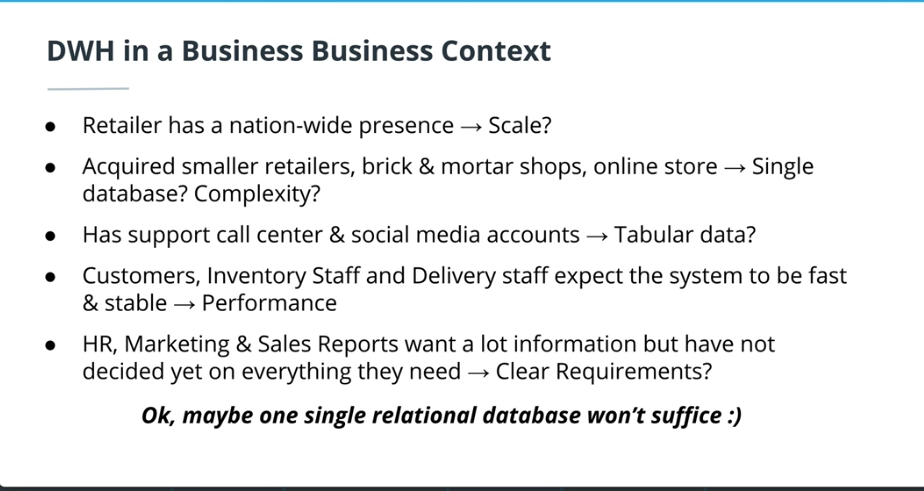
**CQL:** In CQL we have no JOIN , GROUPBY and Sub-Query are not in CQL(Cassandra Query Language) and not supported by CQL

**DATA WAREHOUSE:**

**DATA WAREHOUSE BUSINESS PERSPECTIVE:**

**Graphical user interface, text, application

Description automatically generated**

****

**Graphical user interface, text

Description automatically generated**

**Graphical user interface, text

Description automatically generated**

**Graphical user interface

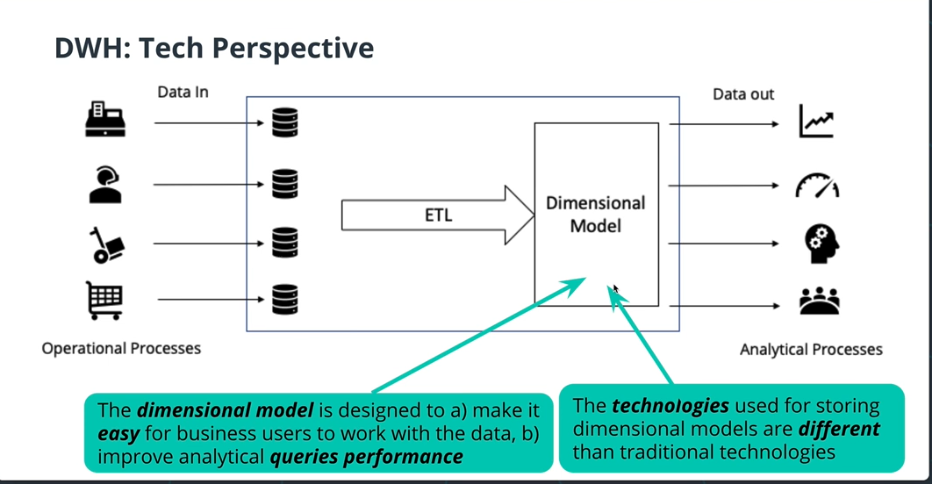
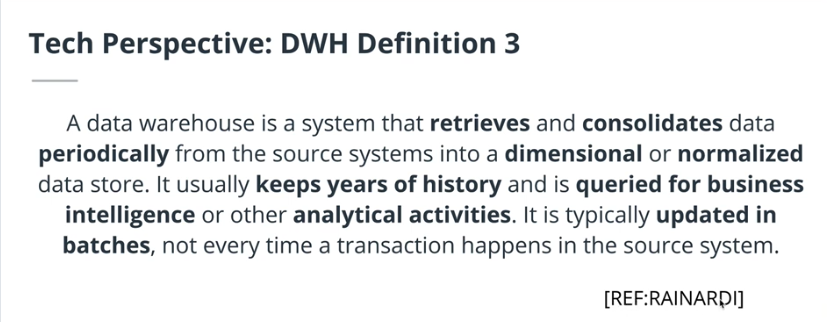
Description automatically generated with medium confidence**

**Diagram

Description automatically generated**

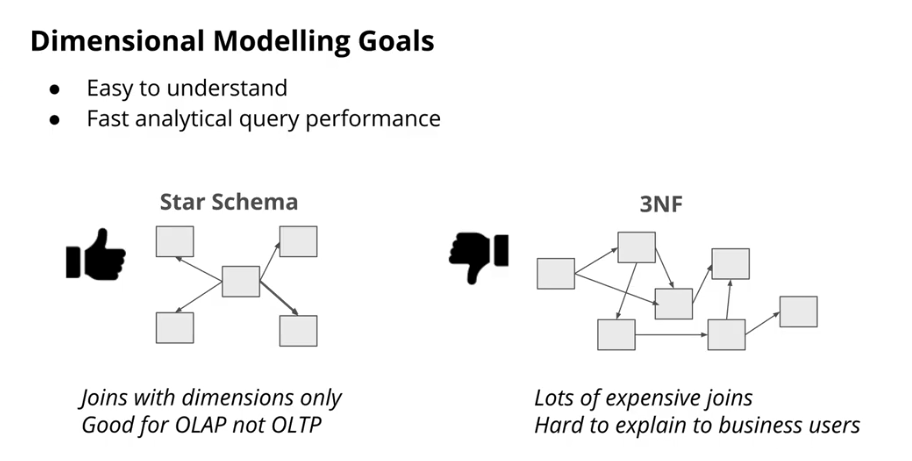
**DATA WAREHOUSE TECH PERSPECTIVE:**

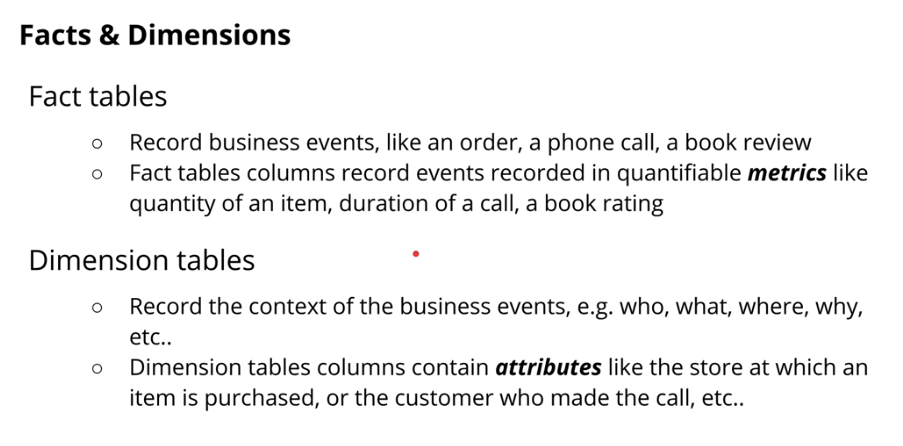
**Graphical user interface, text, application, email

Description automatically generated**

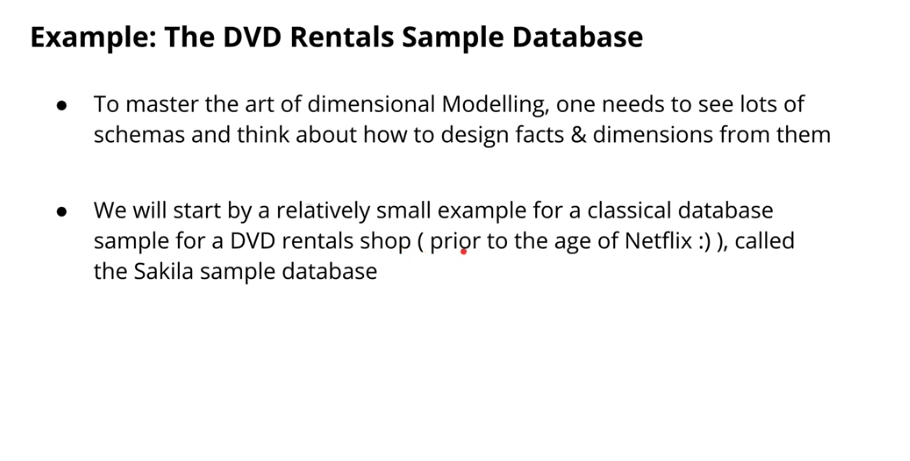
**Goal of Data WareHouse:**

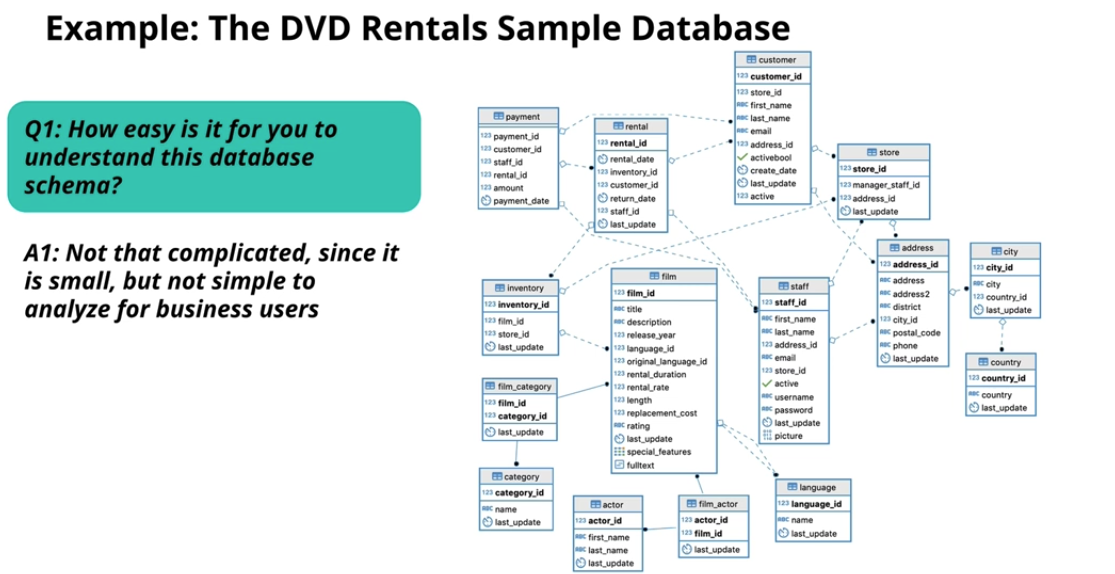
1. Simple to understand
2. Performant (Performance)
3. Quality Assured
4. Handle new questions well
5. Secure











Q2: Can You spot the candidate for a fact table?

A2: rentals and payment

Q3: Can You spot the candidate for a dimensional table?

A3: Customer, Store, Staff, film what about the rest.

