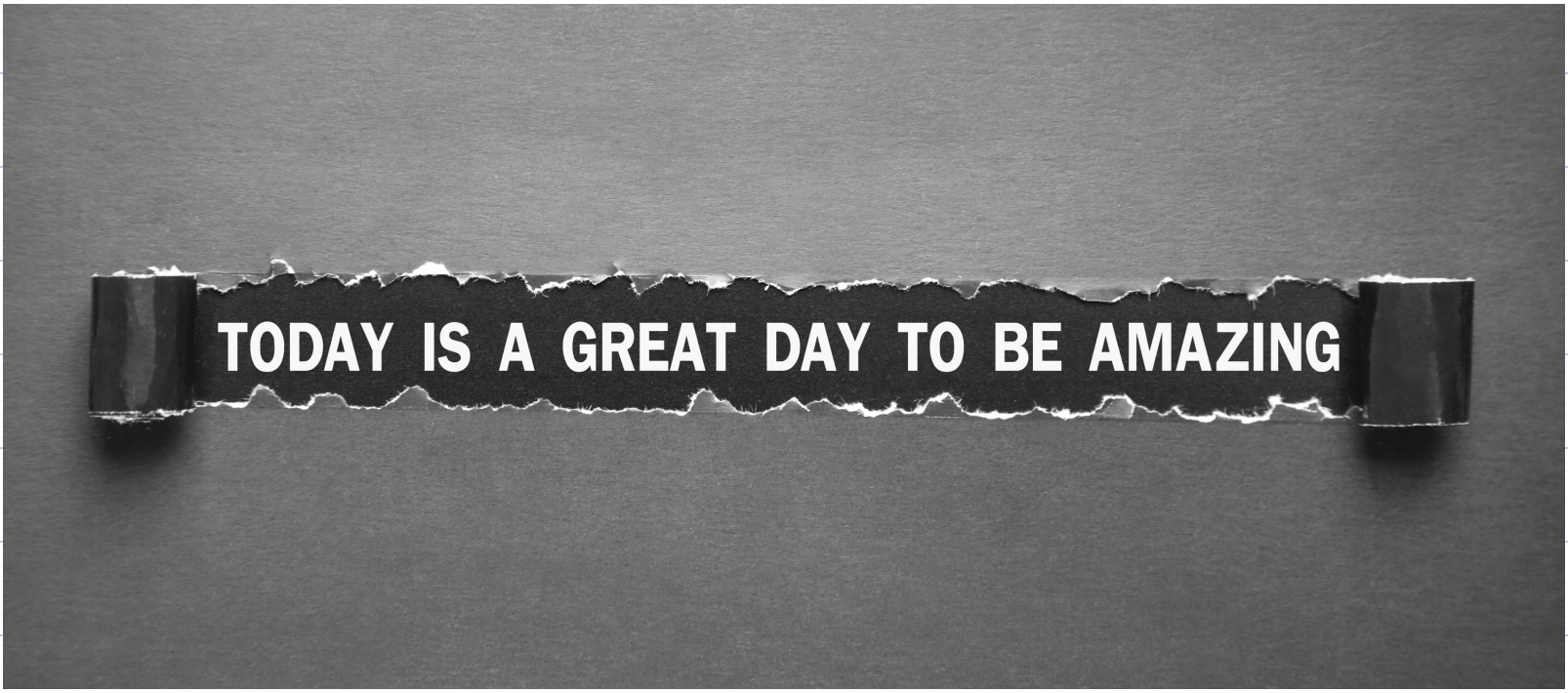
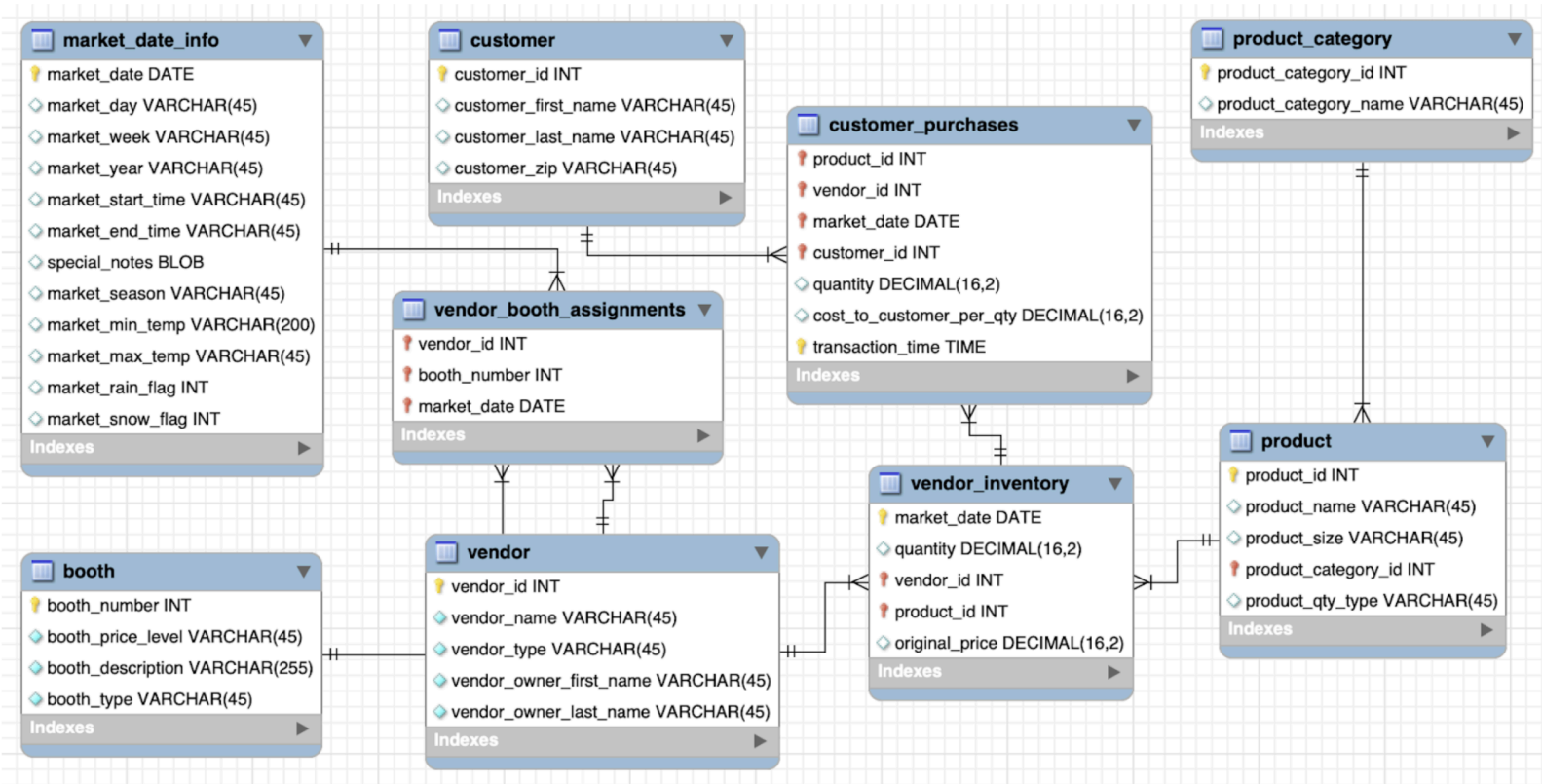


WhatsApp Link

https://chat.whatsapp.com/EvVhJ555AGW1cJaeKI0USe

Dataset



Problem Statement

- It's your first day as a Data Analyst at Reliance Fresh.
- Reliance Fresh has recently decided to open big stores where vendors can directly sell their produce from their stores.
- All the data is stored in a Database
- Your manager reaches out to you and gives you a DB schema of the backend system that looks like this and asks you to derive all the customer and product-related insights from this schema?

First thoughts

1. Where is this data stored? Files? DB?
2. What is a DB schema?
3. What does each of the box represent?
4. What do those lines represent?

Where is this data stored?

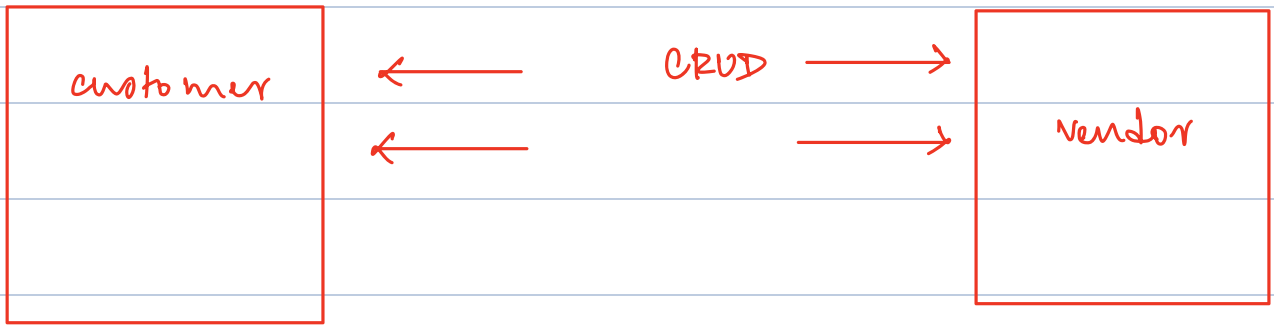
- Relational DBMS called MySQL

What is DBMS?

Database Management System

Database – collection of related data stored as tables

Management system – program that gives us functionality to perform CRUD operation to our data

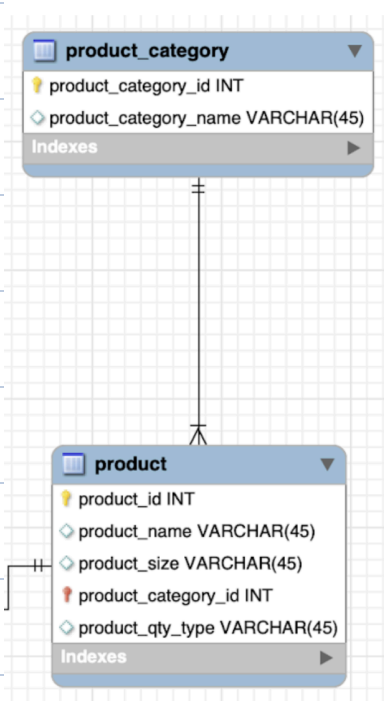


Databases Examples

MySQL, Oracle, MongoDB, Redis, DynamoDB, HBase, ..etc

What is a DB schema ?

- Represents how data is organized and provides information about the relationship between the tables.



① column name, datatypes of column

product_category_id , int

② There is a relationship between product_category and product

this is also part of schema

Concept of Keys

Keys in RDBMS ensure that you can uniquely identify a record in given table.

CID	Name	Phone	age
1	Dhiraj	7299362499	29
2	Sayon	9999999999	99
3	Shreya	88 88 88 88 88	98
4	Dhiraj	777777 7777	28

What are set of columns we can use to uniquely identify the record.

CID, phone, (cid, name),
(CID, phone), (cid, age)

Super Key

Candidate Keys

The minimum set of fields (non-reducible) that can uniquely identify each record in a table.

cid, phone

Candidate Key

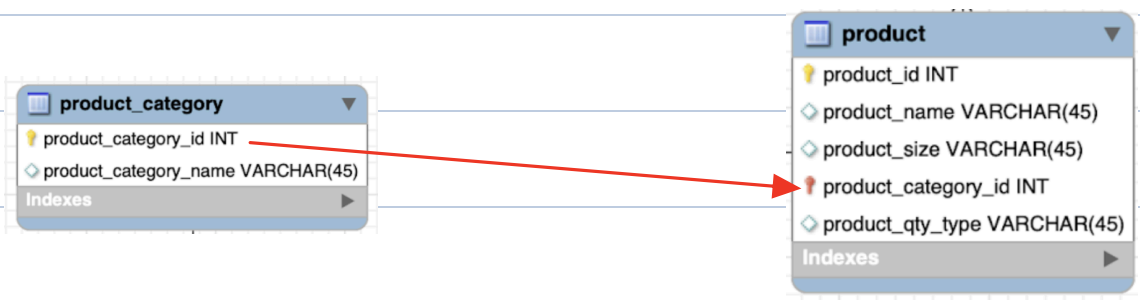
Primary Key

A candidate key selected to uniquely identify rows in the table

cannot be NULL values customer ID

Foreign Key

A reference to another table using other table's primary k



Composite Key

A candidate key consisting of multiple attributes (columns) and not just a single attribute

A	B	C	D

→ (A, B) combined can uniquely identify a single

Unique Key

It is unique for all the records of the table

It can have NULL values

Aadhar

9:57 ⟶ 10:05 pm

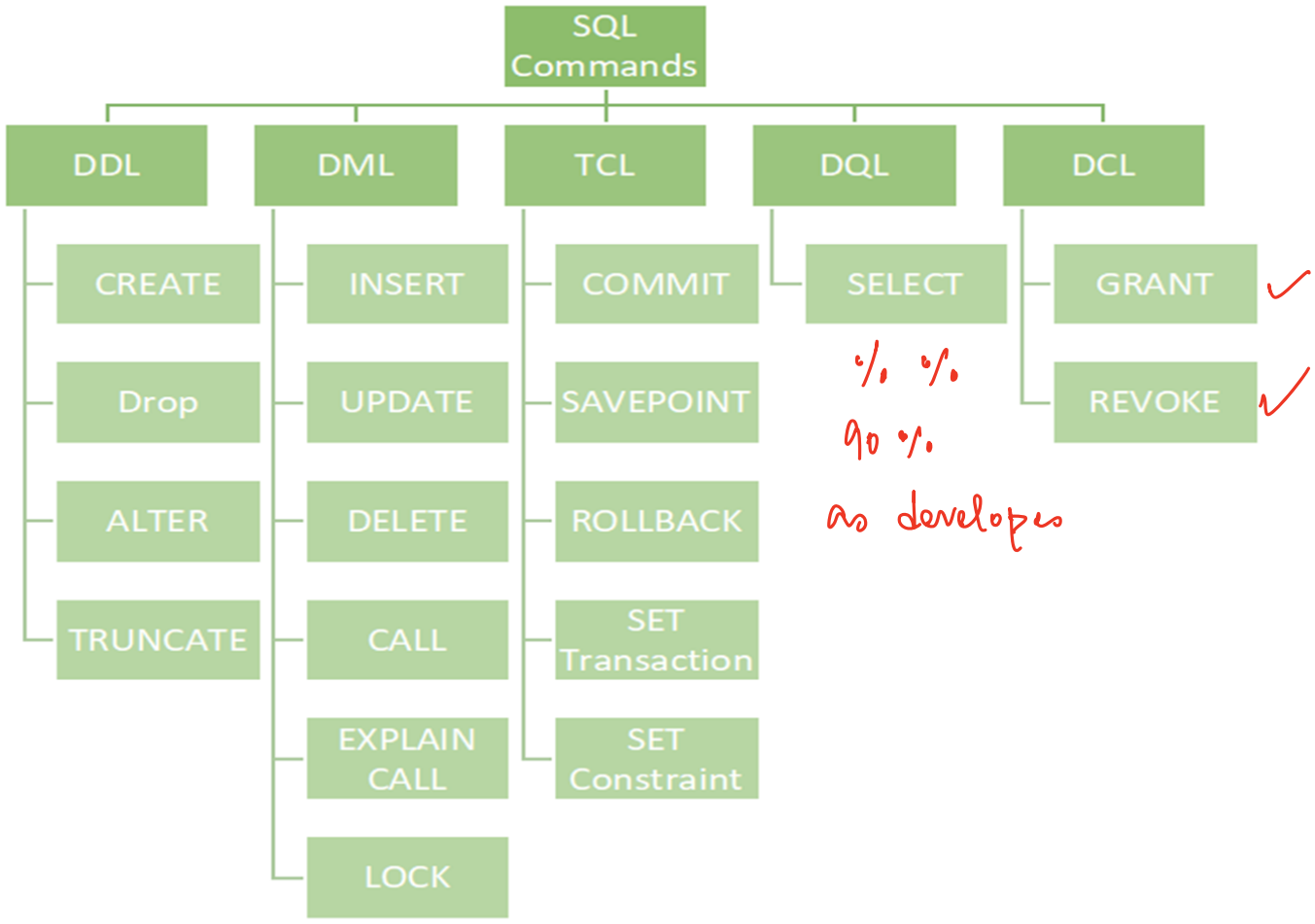
Relationships in a schema

Type	Visualize	Description
one to one	─┼──┼─	one record in A related to one in B
one to many	─┼──<─	one record in A related to many in B
many to one	>──┼─	many records in A related to one in B
many to many	>──<─	many records in A related to many in B

Home work

Use only two tables to achieve many to many relationship

SQL Language



% %
90 %
as developes

DDL - Data definition language

DML - Data manipulation language

TCL - Transaction control language

DQL - Data Query language

DCL - Data Control language

Setup Big Query

DataSet - https://drive.google.com/drive/folders/1p9AvqWcMdyLqZL8BAcjVPQMwfeo9TWjH

How to Setup BigQuery - https://drive.google.com/file/d/1x9IKhdj-wWj0FgYVkPjzbhaGPBIprZda/view