Learning Objectives

- **SQL Database Installation**
- **Databases Terminology Overview**
- **Creating – Databases, Schemas, Tables**
- **Querying Data**
  - Basic SQL syntax
  - Filtering conditions
  - Alias
  - Searching Patterns
  - Duplicated rows
  - Sorting
  - Grouping
  - Data-type conversion…

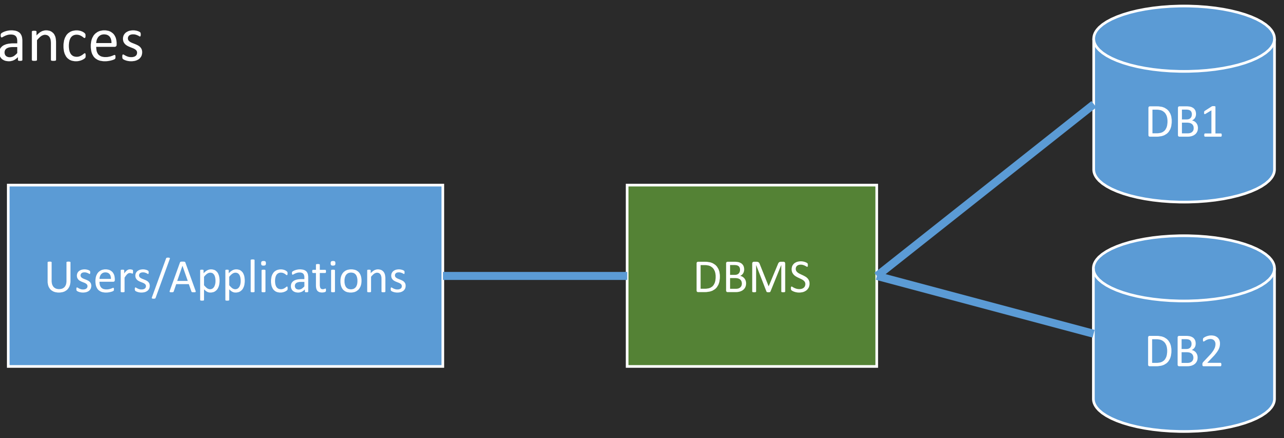
Database, DMBS, SQL

Database, DMBS, SQL

- **What is a Database?**
  - An organized collection of data
  - A logical container for storing data
  - A record – one line combined from multiple attributes
    - Student : name, address, phone number…..
  - Multiple entities that are connected to each other
    - Students
    - Teachers
    - Classes
    - ….

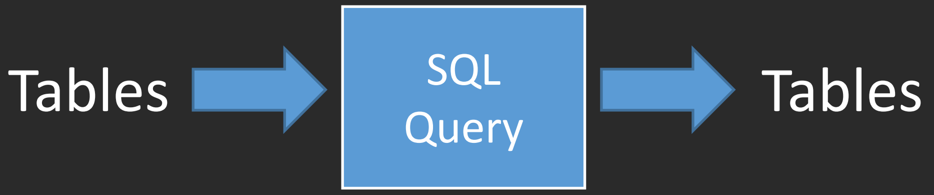
Database, DMBS, SQL

- **DBMS - Database Management System**
  - Interface between users/Apps and databases
  - Translate and process a query request
  - Control access to the data
  - Handle data integrity
  - A unified layer to access the data
  - Create and manage database instances
  - RDBMS – Relational DBMS



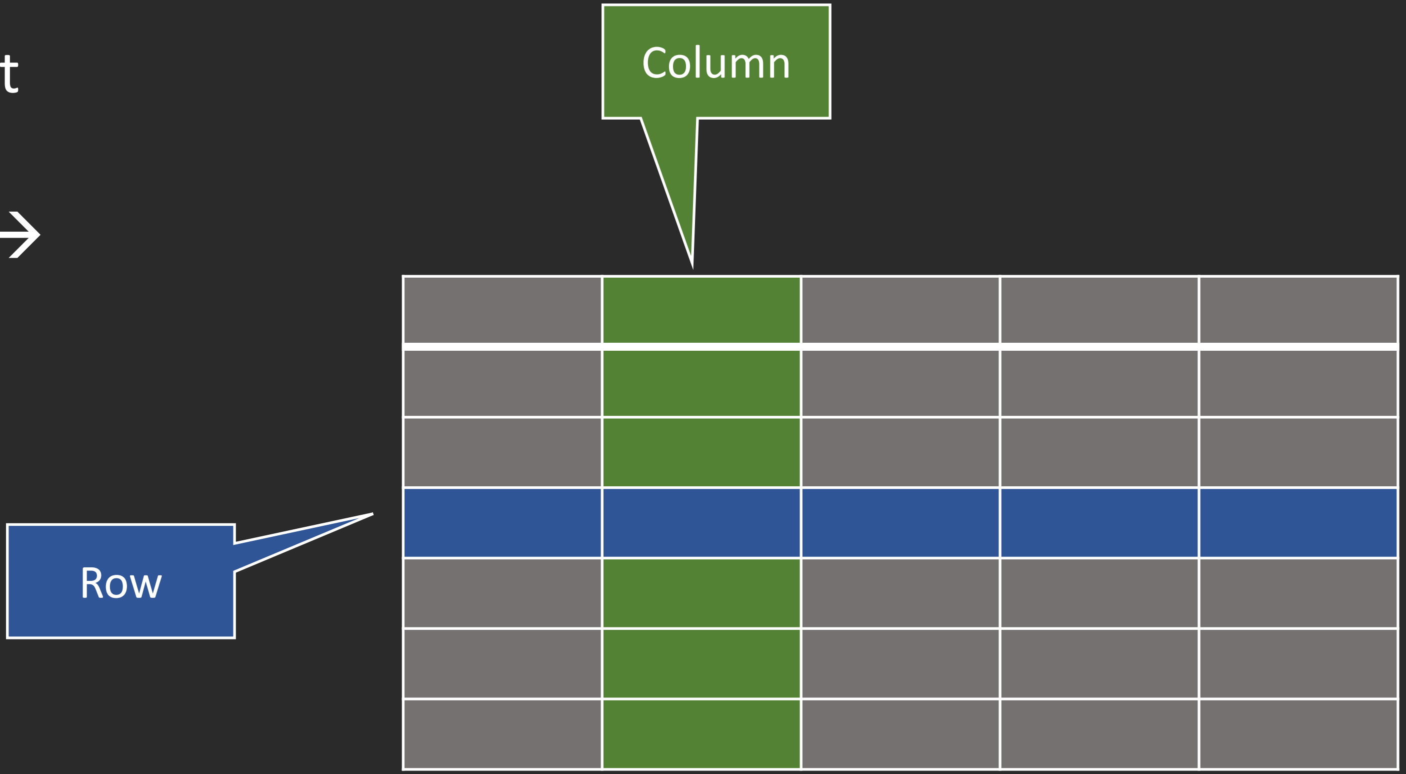
Database, DMBS, SQL

- **SQL - Structured Query Language**
  - Query, update, and modify data
  - Stored and managed in relational databases
  - The most popular language to query data
  - Syntax - descriptive words (SELECT, FROM, WHERE….)
  - Input and output are tables
  - CRUD
    - C – Create
    - R – Read
    - U – Update
    - D – Delete



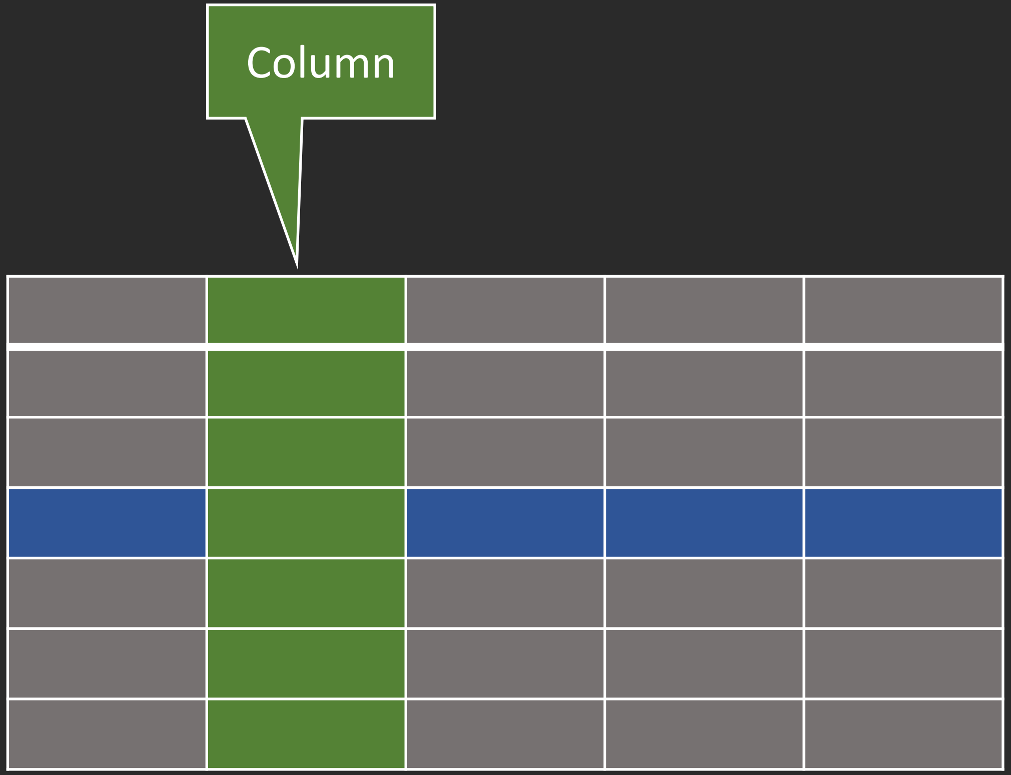
Tables, Columns, Rows

- **Database Table**
  - Looks like an Excel spreadsheet
  - Specific one type of entity
  - Database → multiple entities → multiple tables
  - Rows and columns



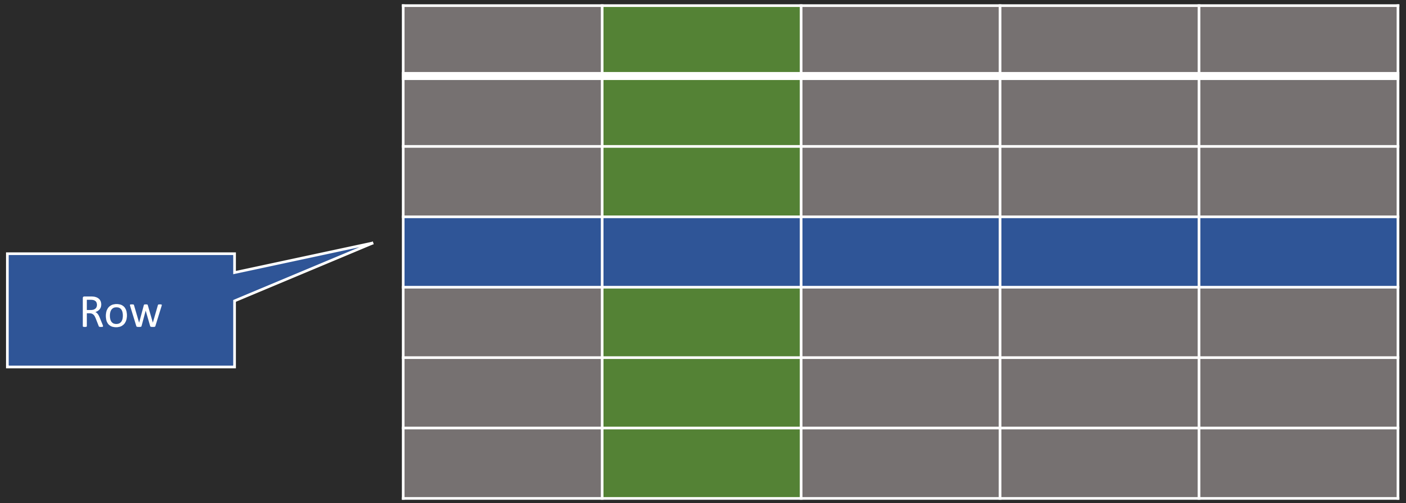
Tables, Columns, Rows

- **A Column**
  - Related to specific attribute
  - A pre-defined data type
    - String, numeric, datetime….
  - Columns are also called attributes
  - Data inside those columns are called values
  - Attribute – Speed
    - Data type – numerical
    - Unit – KM/h



Tables, Columns, Rows

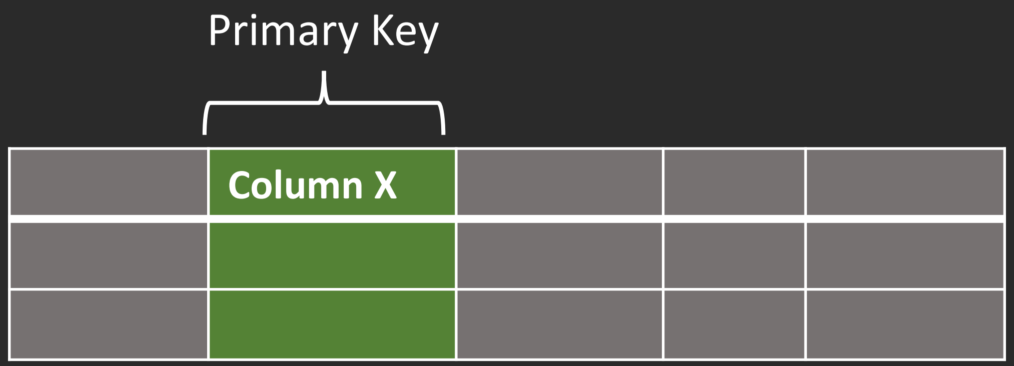
- A Row
  - Collection of attributes related to an entity
    - Customers table: one row = one customer
  - The number of rows is dynamic
    - 0 → Millions of records…
  - Also called a record/tuple

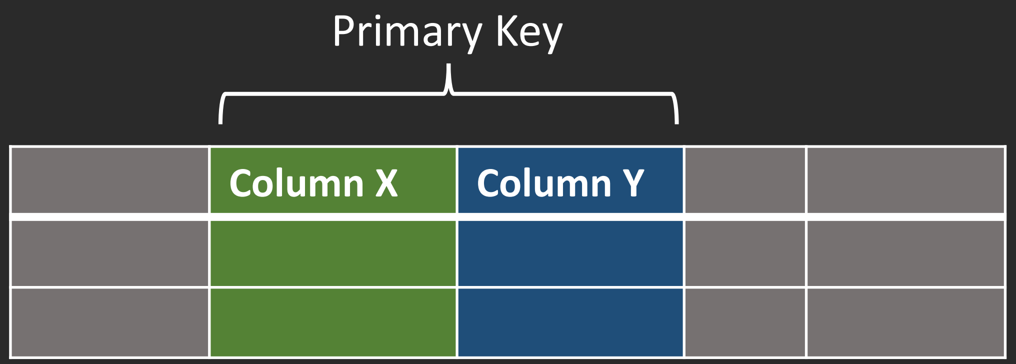


Row

Primary and Foreign Keys

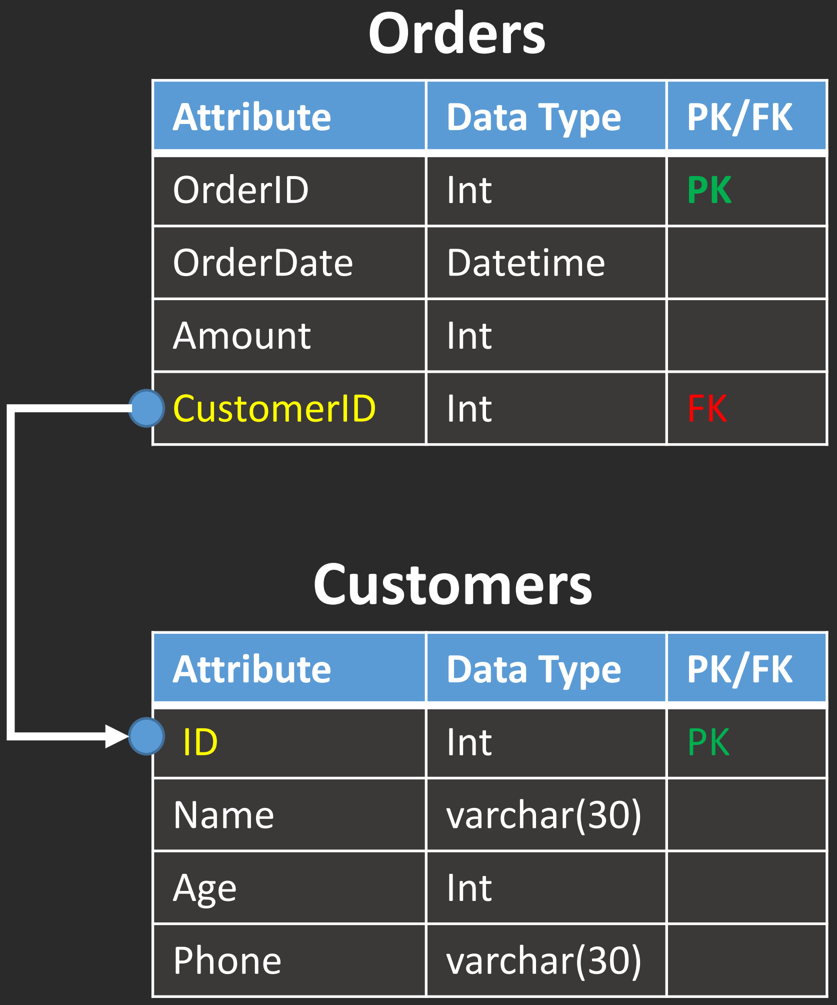
- A **Primary Key**
  - Uniquely identify a single row in table
  - Created from a single column/group of columns
    - Customers table → "Customer ID" as a primary key
  - Defined when creating a table





Primary and Foreign Keys

- A **Foreign Key**
  - Used to navigate from one table to another
  - References a primary key in another table
  - Created from a column or group of columns
  - Doesn’t have to be unique



Orders

Attribute	Data Type	PK/FK
OrderID	Int	PK
OrderDate	Datetime	
Amount	Int	
CustomerID	Int	FK

Customers

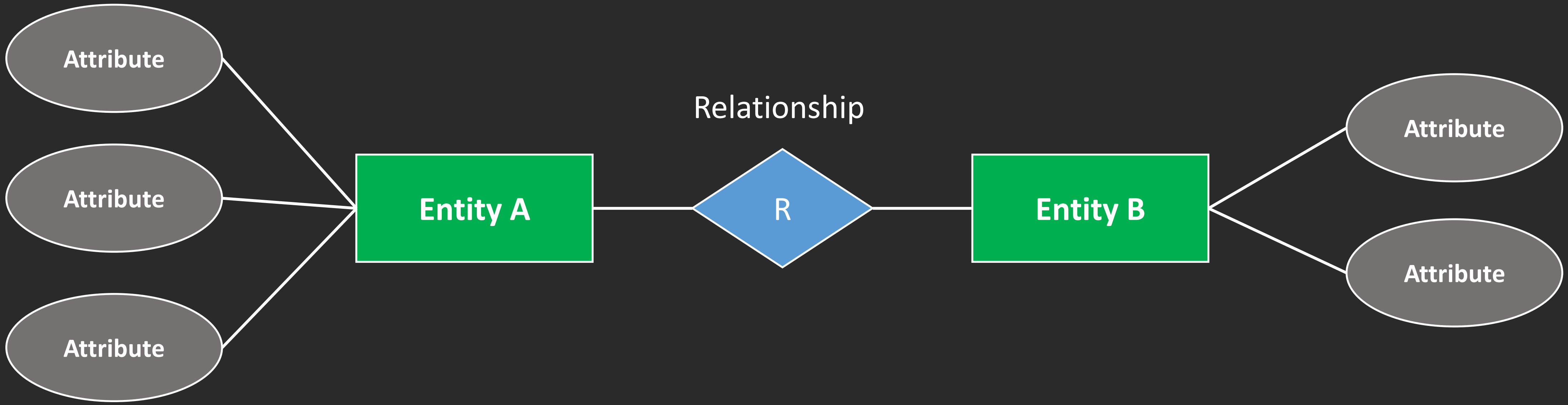
Attribute	Data Type	PK/FK
ID	Int	PK
Name	varchar(30)	
Age	Int	
Phone	varchar(30)	

Relational Model, ER Diagram

- **Relational Model**
  - **Entity**
    - Distinguishable, unique, and distinct
    - All entities from the same entity type will be managed under the same table
  - **Attribute**
    - A characteristic of an entity
    - Stored under a dedicated column (relevant data type)
  - **Relationship**
    - Describes association between different entities
      - One-to-many
      - Many-to-many
      - One-to-one

Relational Model, ER Diagram

- **ERD – Entity Relationship Diagram**
  - Illustrates how “entities” are related to each other
  - Model and design relational databases



Attribute

Attribute

Attribute

Relationship

Entity A

R

Entity B

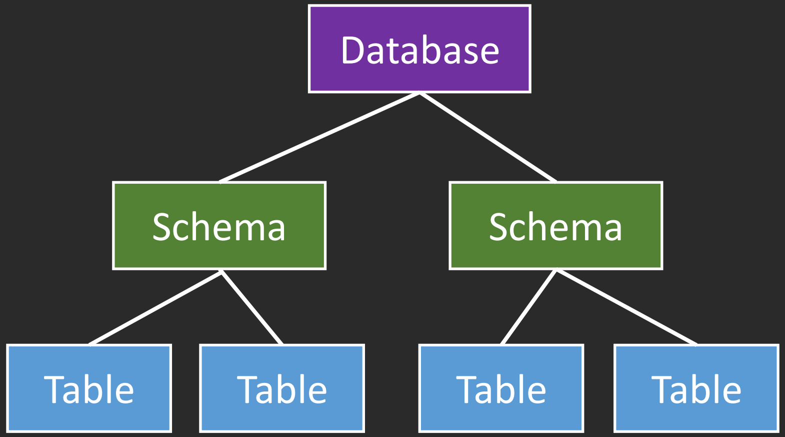
Attribute

Attribute

Schema, Metadata, Data Dictionary

- **Database Logical Schema**
  - The logical structure, constraints and associations of a group of related tables
  - The “blueprint”- how the data must be stored and accessed
  - Database instance → multiple schemas
    - DB1.Schema1
    - DB1.Schema2
  - Each schema will have a unique namespace
  - Qualify the table name

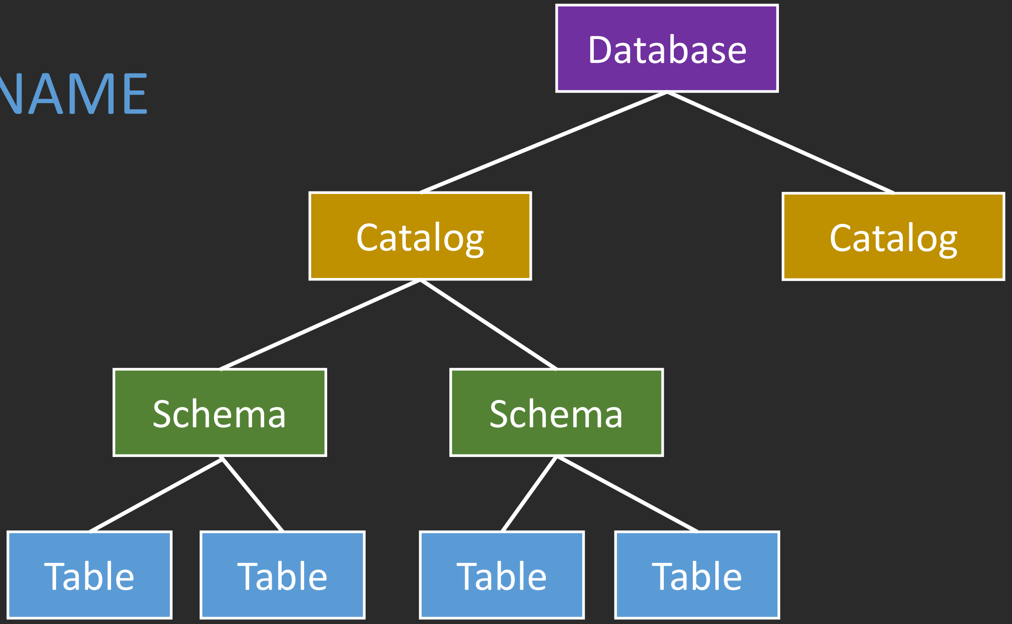
    SCHEMA_NAME.TABLE_NAME



Schema, Metadata, Data Dictionary

- **Database Catalog**
  - When using multiple schemas may not be enough
  - Add another level to the overall hierarchy
  - Qualify the table name

    CATALOG_NAME.SCHEMA_NAME.TABLE_NAME



Metadata and Data Dictionary

- **Metadata**
  - Data about the database structure and database objects
  - For example –
    - List of tables in a schema
    - List of columns in a table
    - Column data type
- **Data Dictionary**
  - Metadata information is managed in dedicated tables stored in the data dictionary

Null Values

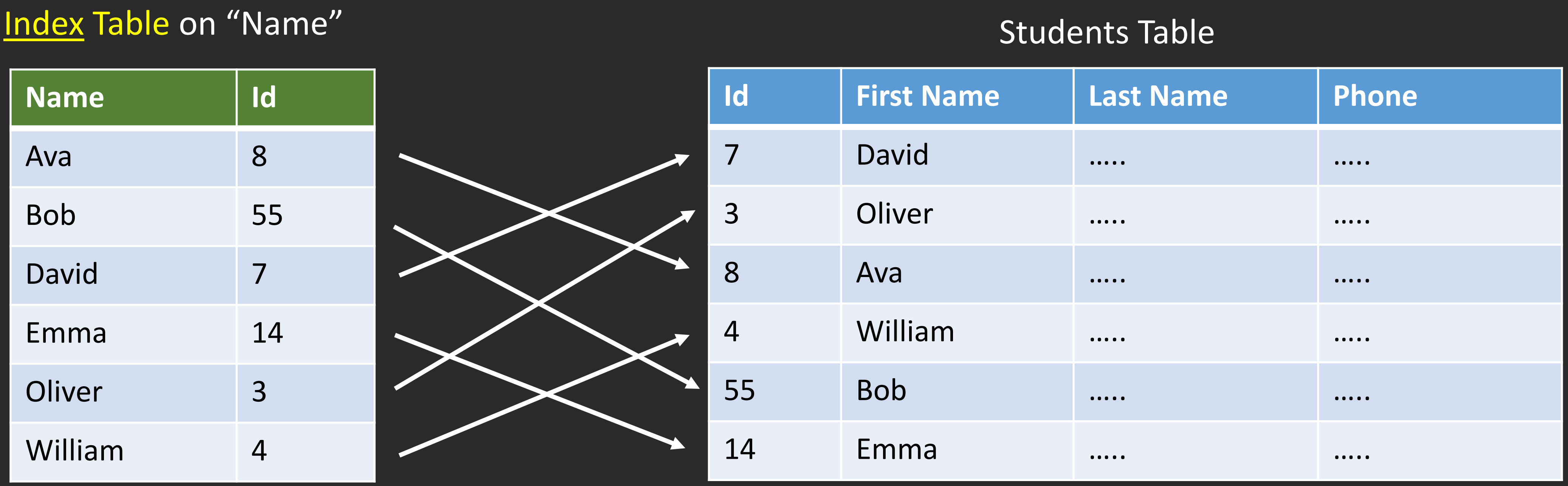
- **What is a Null Value?**
  - "nothing" - the absence of any value
  - Used to mark missing data in columns
- **When inserting a Row (with missing data)**
  - Two Options
    1. Reject the row
    2. Insert new row and mark missing values as null
  - Based on the column definition
  - NULL value has a special marker

Indexes

- **What is an index?**
  - Rows are NOT saved in a specific order based on columns
  - DBMS can provide sorted data
    - List of students ordered by name
    - 100 records → 10,000,000 records
      - Requires a complete table scan
      - Inefficient and time-consuming process
  - An index - a mechanism to locate rows in a table

Indexes

- **What is an index?**
  - An index is a special support table that is keeping rows in a specific order
    - Organizes the data by a specified key



Index Table on “Name”

Name	Id
Ava	8
Bob	55
David	7
Emma	14
Oliver	3
William	4

Students Table

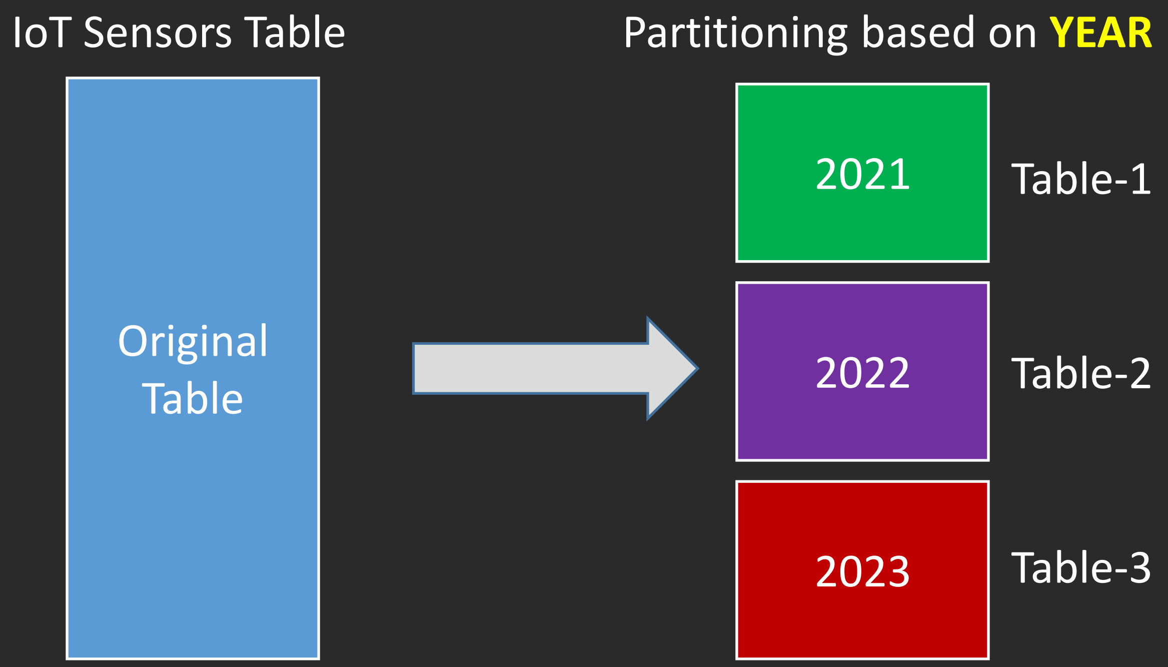
Id	First Name	Last Name	Phone
7	David	…..	…..
3	Oliver	…..	…..
8	Ava	…..	…..
4	William	…..	…..
55	Bob	…..	…..
14	Emma	…..	…..

Partitions

- **The Big Data Era**
  - More and More DATA
  - HUGE challenge for databases
    - Collect, Store, Process…
- **Partitioning**
  - A strategy to handle the growing amount of data
  - Breaking a large table into smaller pieces called partitions

Partitions

- **Partitioning**
  - Faster query - scan the required partition/s instead of the complete table
  - Based on the partition key



IoT Sensors Table

Original Table

Partitioning based on **YEAR**

2021 Table-1

2022 Table-2

2023 Table-3

Partitions

- **Optimized Queries**
  - How a table was partitioned by the DB Admin?
  - Include filtering logic on the partitioning columns
  - Faster exaction time!
  - Cost-effective

Creating Databases, Schemas, and Tables

Retrieving Data with Queries

Query Data (SELECT)

Section	Purpose
SELECT	Determines which columns to include in the query’s result set
FROM	Identifies the tables from which to retrieve data and how the tables should be joined
WHERE	Filters out unwanted data
GROUP BY	Used to group rows together by common column values
HAVING	Filters out unwanted groups
ORDER BY	Sorts the rows of the result set by one or more columns