DATABASE

# What is a Database?

https://www.toptal.com/database

Database

- A database is a collection of data and a method for us to access and miniplate that data

- Data everywhere, in drones, our phones, websites, security cameras and even in cars

- We produce 2.5 quintillion bytes of data every year

- More data has been generated and stored in the past two years then in the entirety of the know human history

- The stored data that we have is 44 zettabytes of data

- Databases have two components, hardware and software

- Hardware is the computer (which has storge)

- Software is what we use to access the data in the hardware

History of Databases

- We first started storing information on clay

- Pen and Paper revolutionized that

- Then the printed press could duplicate the data we store

- We then move to the punch cards in 1890

- 1932 drum memory

- 1951 Magnetic tape

- 1956 HDD

- 1967 floppy disk

- 1982 CD

- 1991 SSD

- 1999 SD and USB Flash

History of Databases
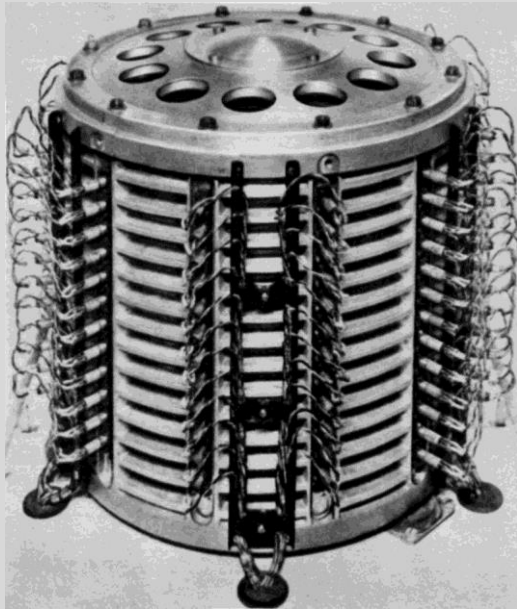
Drum memory. (2023, January 22). In *Wikipedia*. https://en.wikipedia.org/wiki/Drum_memory

Database DBMS

◦ DBMS stands for, Database Management system(Software)

◦ DBMS is a software that is used to manage your Database, it receives instructions from the user, then instructs the system to make retrieve, store or make changes to the database

◦ DBMS uses crud operations

◦ CRUD Stands for are Create Read Update and Delete

RDBMS

◦ RDBMS stands for Relational Database management system

◦ This is a subset of DBMS

◦ RDBMS are the most used type of databases and generally the most useful

◦ They include, SQL, MySQL, SQL Lite, oracle, Microsoft server

SQl

- SQL Stands For Structured Query Language

- SQL is a language that allows us to interact with a Database

- SQL is written as a very simple and English type language that can become simple once you learn its rules

- Once you know SQL Learning things like MySQL, PostgreSQL, Oracle… Will be really ease

- SQL vs Sequel
  - Sequel was the original name for SQL and you may hear them being used interchangeably
  - SEQUEL Stands for Structured English Query Language
  - Sequel had to change the name due to copyright issues and thus we have SQL

- SQL was invented in the 70's but was brought into use in the 80's

- Brought into inception by Edger Codd from IBM in a paper called "A Relational Model Of Data for large shared databanks"

- The paper inspired two other developers from IBM Ray Boyce and Donald Chamberlin to create SQL

- SQL is Standardized

Exercise

- https://www.w3schools.com/sql/trysql.asp?filename=trysql_op_in

Relentless

◦ Relentless Just stared as a new company. They took their first online order and decided to write it with pen a paper.

◦ By the time Relentless gets the 100th order their wrist hurts so they start using an Excel spread sheet.

◦ After order 1000 relentless finds their spreadsheet looking messy, they have orders being returned and people moving from one place to another so they have duplicates all over, they finally decide to switch to a Database

DB vs SS

◦ Why did the Relentless CEO decide to change his data storage model form a spread sheet to a database?

◦ Are they not the same?

◦ Of course they aren't!

◦ Though they may have similarities, Its like comparing a human to a chimpanzee, They are both have bipedal capabilities, but we can really see the differences when we become familiar with them.

◦ Spread sheets tend have no data types for each cell, unlike Databases which are type specific. Ex. A date is a date and will never be a string.

◦ Formatting your database isn't something you can do, like in spread sheets where you can change the font size and color. Database are only focus on the data

◦ You can't do a lot of calculations in your database unlike in spread sheets

◦ Now this may sound like you would prefer to use spread sheets, because it can do more! But that's the issue the more functions spread sheets have the less integrity and consistency our data has.

Database Vs Spreadsheets

◦ What is Data integrity

◦ Data integrity means you can't store different types of data in the same field, and it is unlikely someone will mistake a data value for an outcome of a calculation, especially in large data sets. Data integrity is a strong advantage when working with databases.

◦ Databases have tables as we have seen earlier, and these tables are designed to easily interact with one another. Spreadsheets have worksheets that can do similar actions but are limited in their capabilities

◦ Speadsheets are also limited in the amount of rows they can handle, you can barely make it over a million (Excel can only handle 1,048,567)

◦ Databases are more focused on storing the data so they can easily store huge amounts and tend to be more limited by the hardware vs the software
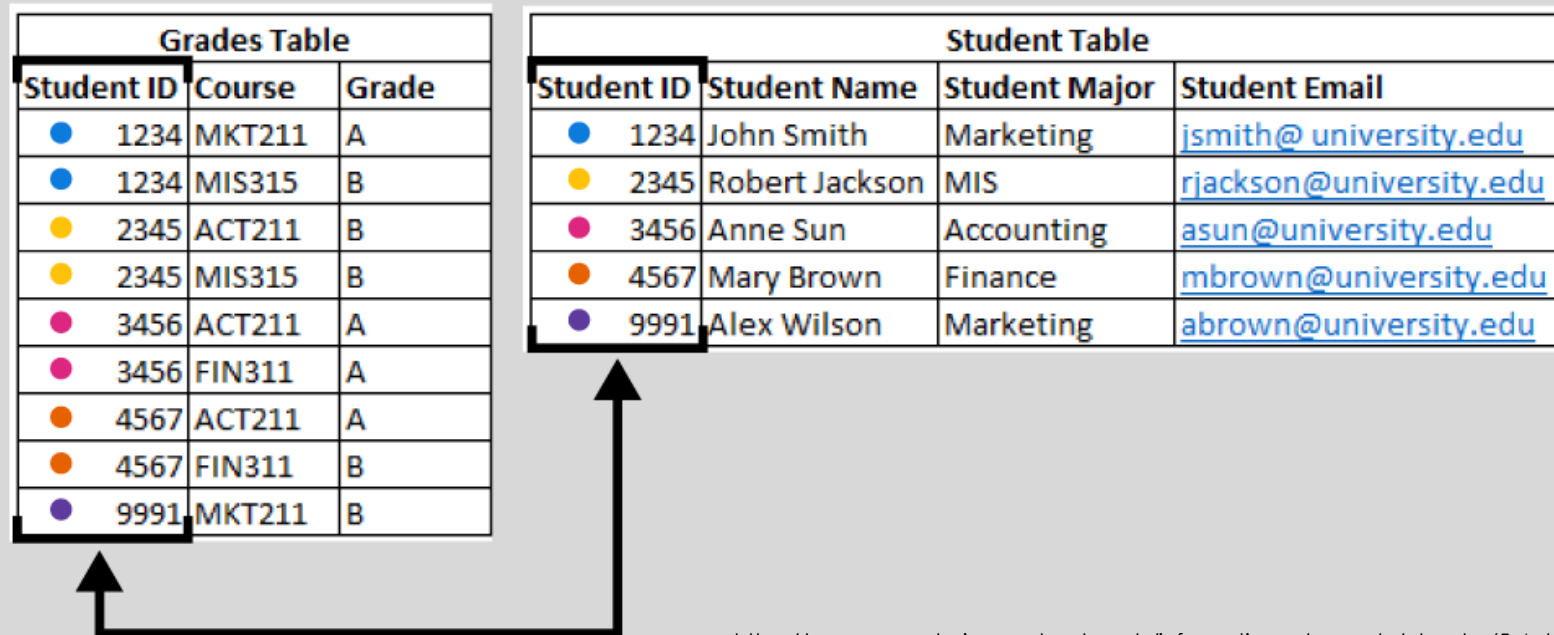
Five types of Databases

- Relational Model databases are the most popular(PostgreSQL, MySQL)

- Document model database (MongoDB, CouchDB, Firebase)// this looks more like a document vs rows and columns, these are good for scalability

- Key Value(Redis, Riak, NoSQL)// this is a way to access data via a key system, you ask for 1 you get 1 back

- Graph model Databases(Neo4j, AWS Neptune)//this is more complex and less common, you can see data connected in graphs, social media might use graphs to see how users are all connected to one another

- Wide Columnar model Databases(Apache Cassandra)//these are newer databases, and were pioneered by google

Relating Our Database

The benefit database relationships is that we can connect a row of data to another data base with a number

Relational Database

Grades Table

Student ID	Course	Grade
🔵 1234	MKT211	A
🔵 1234	MIS315	B
🟡 2345	ACT211	B
🟡 2345	MIS315	B
🔴 3456	ACT211	A
🔴 3456	FIN311	A
🟠 4567	ACT211	A
🟠 4567	FIN311	B
🟣 9991	MKT211	B

Student Table

Student ID	Student Name	Student Major	Student Email
🔵 1234	John Smith	Marketing	jsmith@ university.edu
🟡 2345	Robert Jackson	MIS	rjackson@university.edu
🔴 3456	Anne Sun	Accounting	asun@university.edu
🟠 4567	Mary Brown	Finance	mbrown@university.edu
🟣 9991	Alex Wilson	Marketing	abrown@university.edu

https://ecampusontario.pressbooks.pub/informationsystemscdn/chapter/5-4-database-models/

SQL Playgound

- https://www.db-fiddle.com/f/ogAiTgZPjwvDxwVHiVK3Ek/0

What Is A Query

◦ In English a query is a question

◦ In SQL query is an instruction

◦ In the mind if I ask you a question your brain has to search for, create or remove things from your mind

◦ Query == SQL Statement

◦ SELECT * FROM USERS // this is an sql statement

SQL Statement

◦ SELECT is the used for all queries that return values, It is the "R" in CRUD and the most common

◦ SELECT * FROM Customer

◦ INSERT is used to add a row to a table this is the C in CRUD

◦ INSERT INTO Customer (name, city, state)

       VALUES ('John Smith', 'Renton', WA);

◦ Update Is the used to change data, This Is the "U" in crud

◦ UPDATE Customer

  SET

     Address = '123 Music Avenue',

     Zip = '98056'

   WHERE id = 5;

The DELETE Statement is used to remove rows, This is the D in CRUD

DELETE FROM Customer WHERE id = 4;

Query components

SELECT Name } this is a clause

FROM Users } this is a clause

WHERE Role = 'Manager'; } this is a clause

- SELECT Name FROM Users WHERE Role = 'Manager';

- Keyword: Select, From, Where
- Identifier: Name, Users, Role
- Condition: Role = 'Manager'

- Each line is a clause
- An Identifier is what identifies a part of the data

Practice

○ https://www.db-fiddle.com/f/ogAiTgZPjwvDxwVHiVK3Ek/0

SELECT name

FROM "User"

where role = 'manager'

-- is used for comments

Now try to select certain fields.

Declarative vs Imperative language

◦ WHAT will happen is declarative statement, we state what we want to happen but we don't know HOW it's going to happen

◦ HOW will it happen, In this case we are telling the program how to do something.

◦ Example

◦ You go to a restaurant and request someone makes you a sandwich, then a sandwich appears. That would be equivalent to declarative programming. You declare you want a sandwich and it appears. It could be made in a black hole and pulled out from another dimension, but you don't know because you just ask for the sandwich and it appears.

◦ Imperative is when you own the restaurant and instruct someone to make a sandwich you then teach them how to open the black hole and how to pull the bread from the second dimension and the lettuce from the forth. And then they instruct you to butter the bread and assemble the sandwich.

◦ So far we have mostly done Imperative, with the exception of python.

◦ Declarative is more simple

Practice

- https://www.w3schools.com/sql/exercise.asp?filename=exercise_select1

SQL Standards

◦ If Microsoft, Oracle, Postgres and all supply Databases then how can we make sure that they are using SQL in the correct way?

◦ We have a gold Standard for SQL. also know as rules that are to be followed

◦ This means that there is a committee of people that implement these rules

◦ This is super beneficial to companies because it means you have consistency.

◦ Back in the early part of the industrial revolution they didn't have standardized machine parts, this lead to the fabrication of very specific parts that would only work on one machine. This was a rough time for people because if a gear broke in their machine then they had to replace it with a one of a kind part.

◦ The standardization of these parts led to a huge surge in production because machines could be fixed with ease. Just like many computers these days, I can put the same RAM in an ASUS, MSI, Macs, and more, because the Ram is Standardized

◦ Now with SQL, They have the same standardization, but companies have the ability to add unique qualities to their versions to make them more useful to the end user.

◦ Just like our ASUS may have the same RAM Capabilities as the MAC, the mother board may have some unique properties that utilize the RAM in a more effective way

◦ MySQL and Oracle have the same basics for SQL and both have extras that make their product unique

The what and why of Databases?

◦ Before Databases we had file processing systems and before that we had files and cabinets.

◦ File processing systems had no relationship between data. IF I had a file processing system for a real estate firm and I have one file for Listings, one for Customers, and one for Sales. When I go and change the address for the Customer address, It wont update on all the other files, which means I have to go in and manually update all the past sales with that customer. This could be tedious and add redundant data.

◦ Now imagine we have a school where I have a student ID, Exam and Library. All three database systems are independent from one another. If my student takes out a book then I have to copy the data from one file to another, then if my student takes and Exam I then have to add the student ID information to the exam file as well. This adds up to a bunch of redundant data. The use of the relational database systems was to allow for no redundancy of data, and conglomerate all the individual pieces and have them in a related form

Quiz!

- https://www.jetpunk.com/user-quizzes/1336712/sql-starter-quiz

Database Models

◦ Database models are a way to organize and store data

◦ Types of DB models
  ◦ Hierarchical // replaced by relational
  ◦ Networking //replaced by relational
  ◦ Entity-Relationship
  ◦ Relational
  ◦ Object Oriented
  ◦ Flat
  ◦ Semi-Structures
  ◦ …And so much more!

◦ We are primarily going to be looking at Relational

Hierarchal Model

◦ This is base on a tree like structure where every child has one parent

Hierarchical Model

Hierarchical database model. (2015, October 14). In *Wikipedia*. https://simple.wikipedia.org/wiki/Hierarchical_database_model
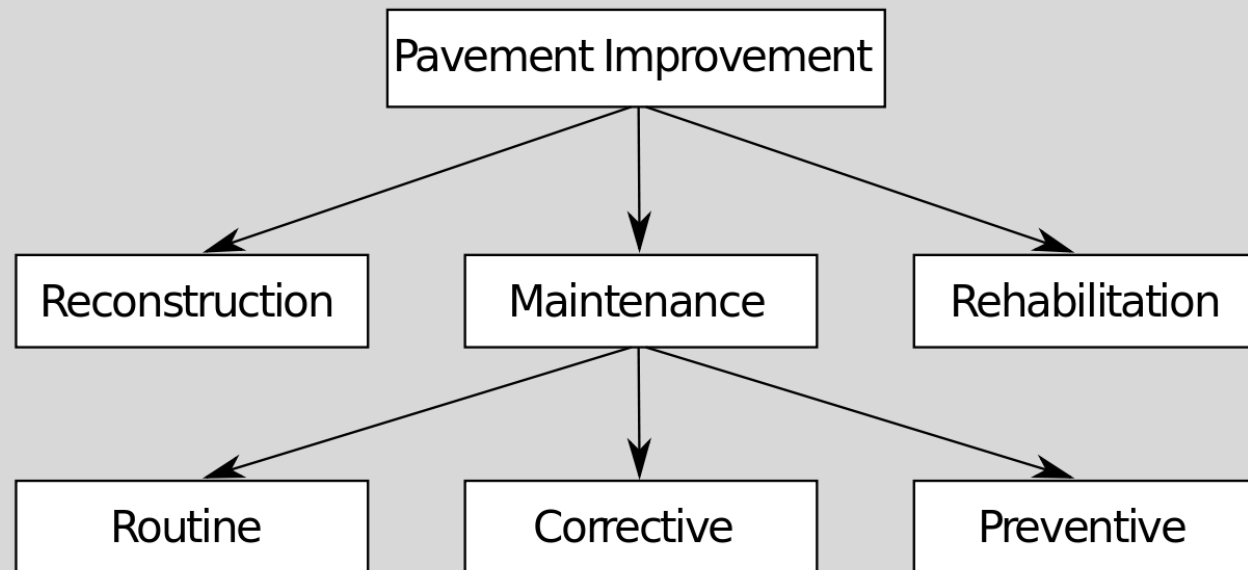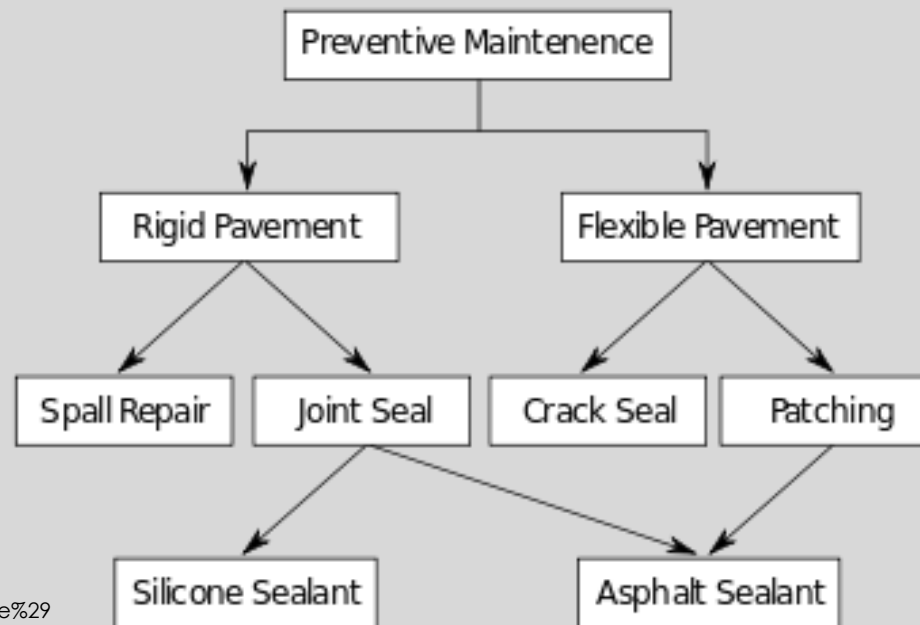
Network Model

In the network model our child can have two parents
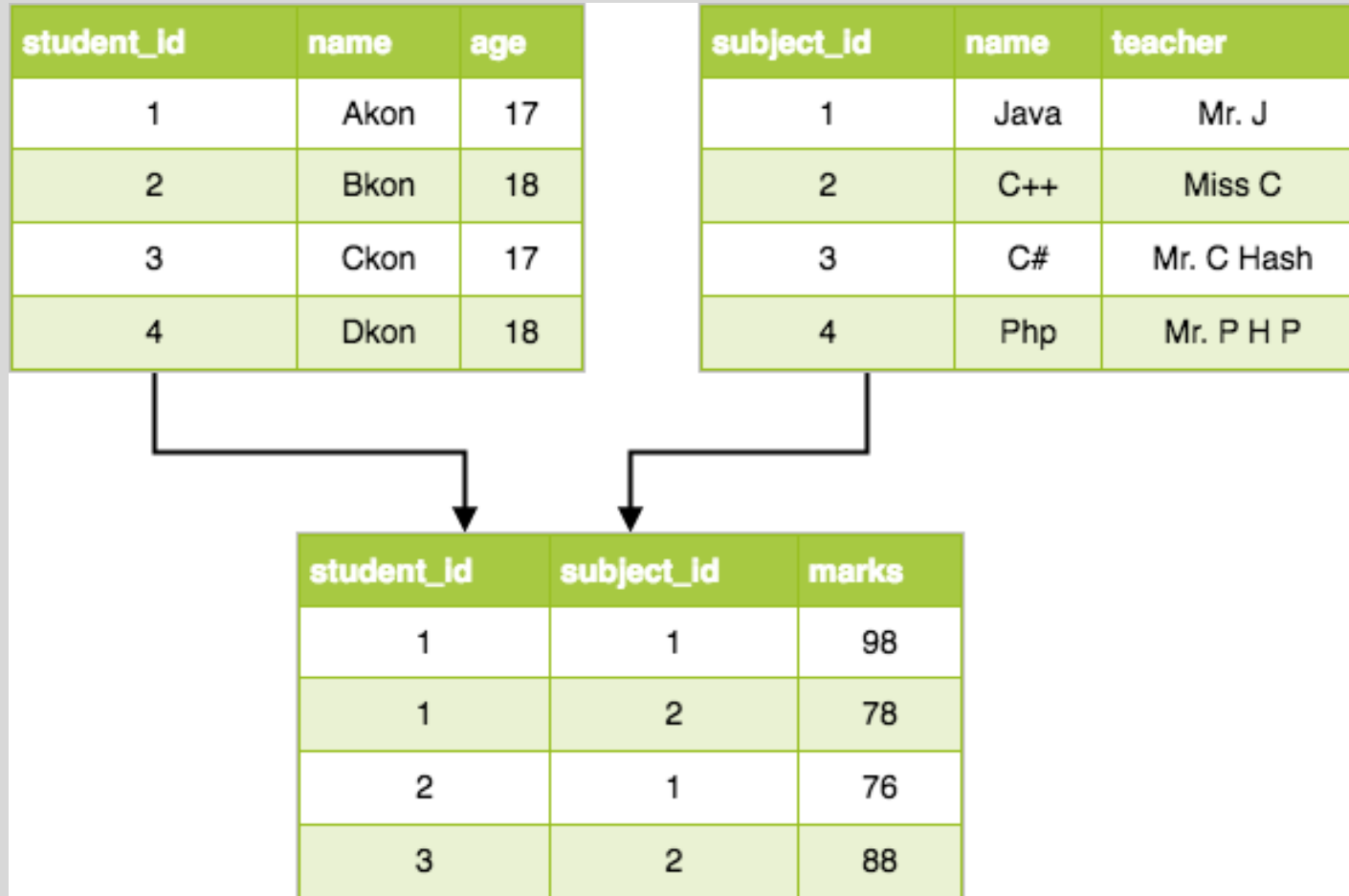
The relational model replaced this

Network Model

Relational Model

◦ It is the first model that had a set of rules to get the most out of the data structure

◦ Is based on Tables

◦ Tables have specific names like FirstName, LastName, and So on

◦ We also have a unique Identifier for our data

◦ Their may be someone with the same first and last name, so its better to use an uniqueID

◦ We can relate these through a separate table

◦ This is managed by the DBMS

Relational Model

student_Id	name	age
1	Akon	17
2	Bkon	18
3	Ckon	17
4	Dkon	18

subject_Id	name	teacher
1	Java	Mr. J
2	C++	Miss C
3	C#	Mr. C Hash
4	Php	Mr. P H P

student_Id	subject_Id	marks
1	1	98
1	2	78
2	1	76
3	2	88

DBMS

◦ Database management Software (System) is vital to the relationship model

◦ This interacts with SQL

◦ DBMS is a program that follows the rule

◦ Its like a smarter Excel sheet

◦ What can a DBMS do?

  ◦ CRUD (Create, Read, Update and Delete)

  ◦ Manage your data(save it in the right place and make sure it goes on the right hard drive)

  ◦ Secure your data by putting in the right security measures and the right permissions

  ◦ Transaction management(this makes sure things like two people changing the same thing at the same time doesn't happen, or when you have a purchase made that a person doesn't get double charged)

◦ Who makes these?

  ◦ Microsoft

  ◦ IBM

  ◦ MySQL

  ◦ Oracle

  ◦ POSTGres

◦ All of these software use the foundation of SQL

DBMS

◦ Lets say your DBMS is a chef, SQL is a type of cooking(Pizza)

◦ All Chefs DO things different when they make pizza, If you get Pizza at 10 different restaurants you will get different pizza, but it is still pizza

◦ SQL the pizza and when we order from MySQL the "pizza" will have a slightly different flavor than Oracle.

Close look at tables

◦ What is a table?

◦ A table is a representation of an object, So we could say student, Teacher, User, Customer, these are all our objects

◦ We give them a name, and the name relates to the concept of the data we are going to store

Database Organization

Row (record)	Column (field)		Data Value		
Position Title	**Education Requirements**	**Functional Area**		**Max Pay**	**Min Pay**
Executive Assistant	Associate degree	Human Resources		60,000	40,000
Recruiter	Bachelor's degree	Human Resources		110,000	85,000
SW Engineer	Bachelor's degree	Engineering		140,000	110,000
SQA Engineer	Bachelor's degree	Engineering		140,000	110,000

Table (object)

Columns

◦ A column can be called an "Attribute"

◦ A collection of columns are known as a "Degree"

◦ When we talk about what a column can contain we call it the "Domain" or "Constraint"

◦ For example, If you have a date of birth (DOB) column you will have a constraint that only allows you to use 1991-05-18 formats. Or for firstName You will only have strings

◦ You may hear people say "The Degree of the relation contains these Domains" or "The table Attributes have these constraints"

◦ These are all interchangeably

◦ You can lock down the information that can be put into a column by using Data types

Row

- Rows are used to store the data

- Rows can be called a "Tuple"

- A tuple is a singular row of data

- "Tuples" is a plural for multiple rows

- Tuples must follow the column constraints

Key

◦ The key is the foundation of the relational model, It's the ability to link relationships between different types of data

◦ We have a primary key, which uniquely identifies each row of data in that table

◦ There is also a Foreign key which identifies a row from another table.

◦ We inject a new column into the table that has the foreign key

Keys

EMPLOYEE_INFO TABLE

EmployeeID	EmployeeName	Emergency ContactName	PhoneNumber	Address	Insurance Number	Pan Number
01	Shanaya	Abhinay	9898765612	Oberoi Street 23	DWJB3182	ABCKQ11213
02	Anay	Soumya	9432156783	Marathalli House No 23	WJBD8271	FEWEF34531
03	Preeti	Rohan	9764234519	Queens Road 45	SJDN8219	HWKLF21211
04	Vihaan	Akriti	9966442211	Brigade Road Block 4	XCWK9128	JAEWA92167
05	Manasa	Shourya	9543176246	Mayo Road 23	NKAS9829	WEBKQ73992

Super Key
Primary Key
Candidate Key
Unique Key
Alternate Key
Foreign Key
Composite Key

EmployeeID	Month_Year_of_Salary	Salary
01	Jan_2019	40000
02	Feb_2019	60000
01	Jan_2019	40000
02	Feb_2019	60000

EMPLOYEE_SALARY TABLE

edureka!

https://www.edureka.co/blog/sql-commands

OLTP vs OLAP

◦ There are two different uses for our types of databases

◦ One supports the day to day operations of a company

◦ The other supports the analysis of a company

◦ For example with something like amazon their day to day would look like a log of their orders, what credit cards they will use, if they have new customers... all things the supports the day to day business

◦ Support Analysis would be, Now they have the data on you, how can we use it to get you to buy more, thinks like what do you like. Target had used this type of data analysis to determine if a woman is pregnant. This way then could mail things to them that would increase their sales and if they could determine it before the person started showing they would have a huge advantage.

◦ Day to day to day is called OLTP(Online Transaction Processing)

◦ Support Analysis is called OLAP(Online Analytical Processing)

Exercise!

◦ Guess which is which:

◦ A database is being used to log orders and Customers

◦ A database is being used to figure out what new products we should offer

◦ A database is being used to derive statistics for reporting to the executives

◦ A database is being used to keep track of logged in users

Exercise!

◦ Guess which is which:

◦ A database is being used to log orders and Customers; OLTP

◦ A database is being used to figure out what new products we should offer; OLAP

◦ A database is being used to derive statistics for reporting to the executives; OLAP

◦ A database is being used to keep track of logged in users; OLTP

Relational Model Quiz!

- https://www.jetpunk.com/user-quizzes/1336712/the-relational-model

Thank You!