Maktab 52 - Database

By Armin Khoshnamak

Data

Data is nothing but facts and statistics stored or free flowing over a network, generally it's raw and unprocessed.

Data becomes information when it is processed, turning it into something meaningful. (Cookie)

Database

Data becomes information when it is processed, turning it into something meaningful. (Tapes)

So we need a DBMS

DBMS (Database Management System)

A DBMS is a software that allows creation, definition and manipulation of database, allowing users to store, process and analyse data easily. DBMS provides us with an interface or a tool, to perform various operations like creating database, storing data in it, updating data.

DBMS also provides protection and security to the databases. It also maintains data consistency in case of multiple users.

Some Popular DBMS :

- Mysql
- Oracle
- PostgreSQL
- Amazon SimpleDB (cloud Based)

Characteristics of DBMS

1. Data stored into Tables: Data is never directly stored into the database. Data is stored into tables, created inside the database. DBMS also allows to have relationships between tables which makes the data more meaningful and connect.
.
2. Reduced Redundancy: DBMS follows **Normalisation** which divides the data in such a way that repetition is minimum.
.
3. Data Consistency: On Live data, i.e. data that is being continuosly updated and added, maintaining the consistency of data can become a challenge. But DBMS handles it all by itself.
4. Support Multiple user and Concurrent Access:
.
5. Query Language:
.
6. Security:
.
7. DBMS supports **transactions**, which allows us to better handle and manage data integrity in real world applications where multi-threading is extensively used.

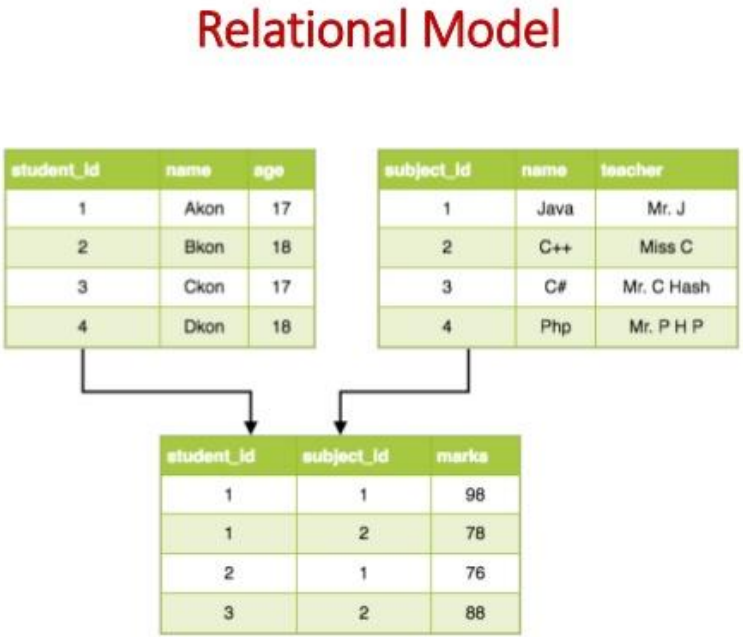
DBMS Database Models

Hierarchical Model

Network Model

Entity-relationship Model

Relational Model



Basic Concepts of ER Model in DBMS

In short, to understand about the ER Model, we must understand about:

- Entity and Entity Set
- What are Attributes? And Types of Attributes.
- Keys
- Relationships

Let's take an example to explain everything. For a School Management Software, we will have to store Student information, Teacher information, Classes, Subjects taught in each class etc.

Entity and Entity Set

Student is an entity, Teacher is an entity, similarly, Class, Subject etc are also entities.

An Entity is generally a real-world object which has characteristics and holds relationships in a DBMS.

complete dataset of all the students will be the Entity Set

Attributes

If a Student is an Entity, then student's roll no., student's name, student's age, student's gender etc will be its attributes.

1. Simple attribute: The attributes with values that are atomic and cannot be broken down further are simple attributes. For example, student's age.
.
2. Composite attribute: A composite attribute is made up of more than one simple attribute. For example, student's address will contain, house no., street name, pincode etc.
.
3. Derived attribute: These are the attributes which are not present in the whole database management system, but are derived using other attributes. For example, *average age of students in a class*.

Keys

If the attribute id . can uniquely identify a student entity, amongst all the students, then the attribute id. will be said to be a key.

1. Super Key
2. Candidate Key
3. Primary Key

Relationships

When an Entity is related to another Entity, they are said to have a relationship. For example, A Class Entity is related to Student entity, becasue students study in classes, hence this is a relationship.

ERD (LINK)

RDBMS

Relational Database management System

is a DBMS based on the relational model introduced by E.F Codd. In relational model, data is stored in relations(tables) and is represented in form of tuples(rows).

Relational database is a collection of organized set of tables related to each other, and from which data can be accessed easily.

Table and Tuple

In Relational database model, a table is a collection of data elements organised in terms of rows and columns. A table is also considered as a convenient representation of relations. But a table can have duplicate row of data while a true relation cannot have duplicate data. Table is the most simplest form of data storage.

A single entry in a table is called a Tuple or Record or Row. A tuple in a table represents a set of related data. For example, the above Employee table has 4 tuples/records/rows.

ID	Name	Age	Salary
1	Adam	34	13000
2	Alex	28	15000
3	Stuart	20	18000
4	Ross	42	19020

Attribute, Schema, key

Attribute

A table consists of several records(row), each record can be broken down into several smaller parts of data known as Attributes. The above Employee table consist of four attributes, ID, Name, Age and Salary.

Attribute Domain

When an attribute is defined in a relation(table), it is defined to hold only a certain type of values, which is known as Attribute Domain.

Relation Schema

A relation schema describes the structure of the relation, with the name of the relation(name of table), its attributes and their names and type.

Relation Key

A relation key is an attribute which can uniquely identify a particular tuple(row) in a relation(table).

Relational Integrity Constraints

The three main Integrity Constraints are:

- Key Constraints
- Domain Constraints
- Referential integrity Constraints(later)

The Key Constraint specifies that there should be such an attribute(column) in a relation(table), which can be used to fetch data for any tuple(row).

The Key attribute should never be NULL or same for two different row of data.

Domain constraints refers to the rules defined for the values that can be stored for a certain attribute.

Like we explained above, we cannot store Address of employee in the column for Name.

Relational Algebra(LINK)

Relational Calculus

Normalization

Database Normalization is a technique of organizing the data in the database. Normalization is a systematic approach of decomposing tables to eliminate data redundancy(repetition) and undesirable characteristics like Insertion, Update and Deletion Anomalies. It is a multi-step process that puts data into tabular form, removing duplicated data from the relation tables.

Normalization is used for mainly two purposes,

- Eliminating redundant(useless) data.
- Ensuring data dependencies make sense i.e data is logically stored.

1st Normal Form

Rules

Rule 1: Single Valued Attributes

Rule 2: Attribute Domain should not change

Rule 3: Unique name for Attributes/Columns

Rule 4: Order doesn't matters

2nd Normal Form

For a table to be in the Second Normal Form, it must satisfy two conditions:

The table should be in the First Normal Form.
There should be no Partial Dependency.

Functional Dependency.

Partial Dependency

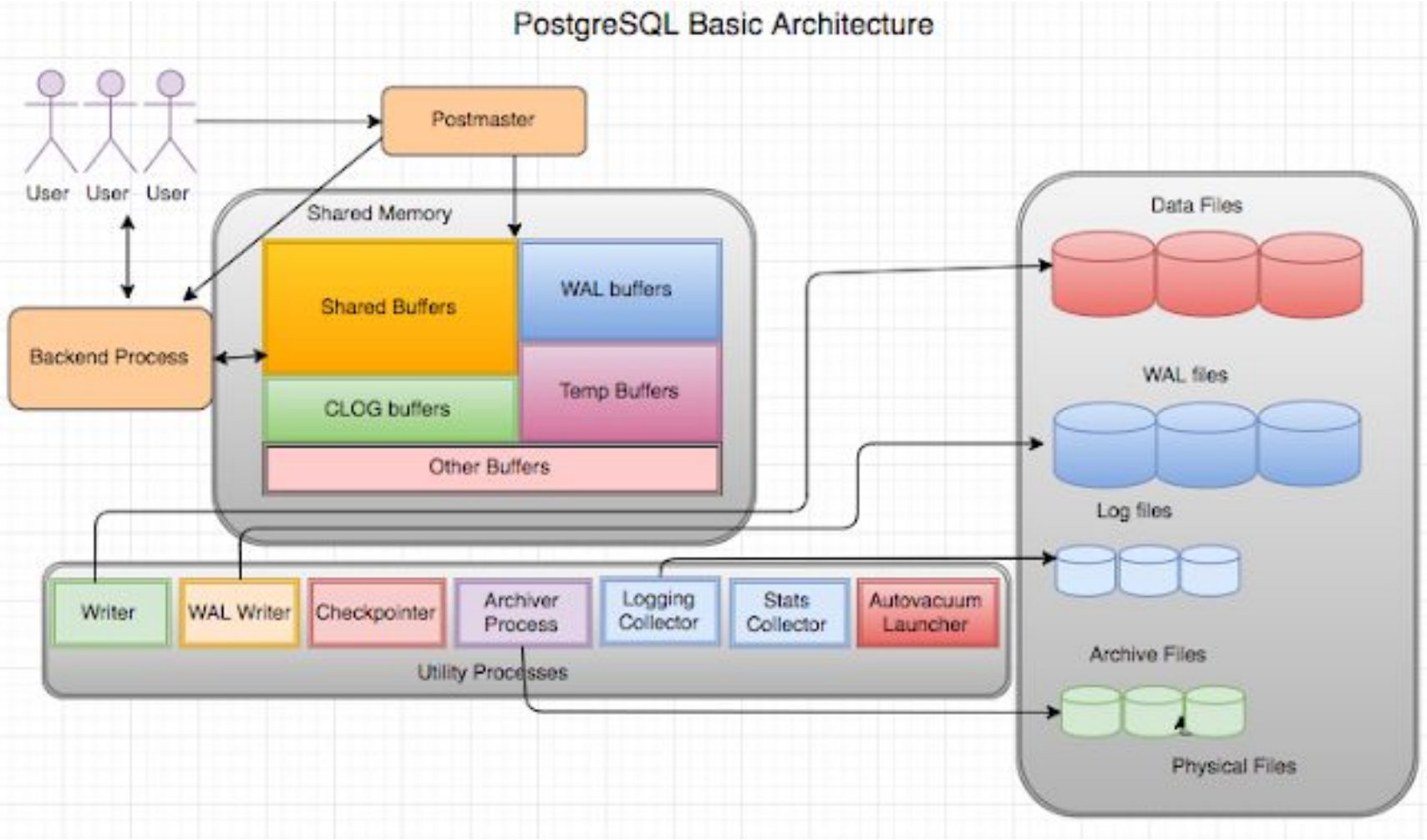
To remove Partial dependency, we can divide the table, remove the attribute which is causing partial dependency, and move it to some other table where it fits in well.

3rd Normal Form

Transitive Dependency

When a non-prime attribute depends on other non-prime attributes rather than depending upon the prime attributes or primary key..

PostgreSQL


PostgreSQL Basic Architecture
User
User
User
Postmaster
Shared Memory
Shared Buffers
WAL buffers
Backend Process
CLOG buffers
Temp Buffers
Other Buffers
Data Files
WAL files
Log files
Writer
WAL Writer
Checkpointer
Archiver Process
Logging Collector
Stats Collector
Autovacuum Launcher
Utility Processes
Archive Files
Physical Files