1 Introduction to Databases

Objectives

After completing this course, you should be able to:

- Introduction to Database Systems, Architectures and Types
- Overview of Data Model
- Exploration of Entity Attribute Relationships and Notations
- Normalization and De-Normalization

Course Roadmap

RDBMS Concepts

- **Lesson 1: Introduction to Databases** — You are here!
- Lesson 2: Overview of Data Model
- Lesson 3: Exploration of Entity Attribute Relationships and Notations
- Lesson 4: Normalization and De-Normalization

Lesson Agenda

- Course objectives and agenda
- Introduction to Databases and its terminologies
- Types of Databases
- Table, Row, Column and Cardinalities

What is Database ?

- The database is a collection of inter-related data which is used to retrieve, insert and delete the data efficiently.
- It is also used to organize the data in the form of a table, schema, views, and reports, etc.
- The **main purpose** of the database is to operate a large amount of information by storing, retrieving, and managing data.
  - **Banking**: for storing customer information, account activities, deposits, payment details, loans, etc.
  - **Manufacturing**: for supply chain management, production tracking and inventory management
  - **Finance**: for storing information about stocks, sales, and purchases of financial instruments like stocks and bonds
  - **Universities**: for student information, course registrations, payroll and grades
  - **Airlines**: for reservations, ticket booking and schedule information

Database Management System

- Database management system is a software which is used to manage the database. For example: MySQL, Oracle, Sybase, MongoDB, Informix, PostgreSQL, SQL Server, etc. are a very popular commercial database which is used in different applications.
- DBMS provides an interface to perform various operations like database creation, storing data in it, updating data, creating a table in the database and a lot more.
- It provides protection and security to the database. In the case of multiple users, it also maintains data consistency.

Cont …

- Modern databases are managed by the database management system (DBMS).
- **SQL** or Structured Query Language is used to operate on the data stored in a database. SQL depends on relational algebra and tuple relational calculus.
- A cylindrical structure is used to display the image of a database.

DBMS allows users the following tasks:

- **Data Definition:** It is used for creation, modification, and removal of definition that defines the organization of data in the database.
- **Data Updation:** It is used for the insertion, modification, and deletion of the actual data in the database.
- **Data Retrieval:** It is used to retrieve the data from the database which can be used by applications for various purposes.
- **User Administration:** It is used for registering and monitoring users, maintain data integrity, enforcing data security, dealing with concurrency control, monitoring performance and recovering information corrupted by unexpected failure.

What is Data?

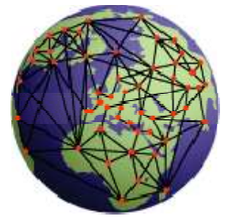
- Data is a collection of a distinct small unit of information. It can be used in a variety of forms like text, numbers, media, bytes, etc. it can be stored in pieces of paper or electronic memory, etc.
- Word 'Data' is originated from the word 'datum' that means 'single piece of information.' It is plural of the word datum.
- In computing, Data is information that can be translated into a form for efficient movement and processing. Data is interchangeable.

Focus Areas : Oracle Database as a example

Information Management

Application Development

Cloud



Infrastructure Grids

Oracle Database 12c

High Availability

Manageability

Performance

Security

Information Integration

MentorLabs℠

Characteristics of DBMS

- It uses a digital repository established on a server to store and manage the information.
- It can provide a clear and logical view of the process that manipulates data.
- DBMS contains automatic backup and recovery procedures.
- It contains ACID properties which maintain data in a healthy state in case of failure.
- It can reduce the complex relationship between data.
- It is used to support manipulation and processing of data.
- It is used to provide security of data.
- It can view the database from different viewpoints according to the requirements of the user.

Advantages of DBMS

- **Controls database redundancy:** It can control data redundancy because it stores all the data in one single database file and that recorded data is placed in the database.
- **Data sharing:** In DBMS, the authorized users of an organization can share the data among multiple users.
- **Easily Maintenance:** It can be easily maintainable due to the centralized nature of the database system.
- **Reduce time:** It reduces development time and maintenance need.
- **Backup:** It provides backup and recovery subsystems which create automatic backup of data from hardware and software failures and restores the data if required.
- **multiple user interface:** It provides different types of user interfaces like graphical user interfaces, application program interfaces

Disadvantages of DBMS

- **Cost of Hardware and Software:** It requires a high speed of data processor and large memory size to run DBMS software.
- **Size:** It occupies a large space of disks and large memory to run them efficiently.
- **Complexity:** Database system creates additional complexity and requirements.
- **Higher impact of failure:** Failure is highly impacted the database because in most of the organization, all the data stored in a single database and if the database is damaged due to electric failure or database corruption then the data may be lost forever.

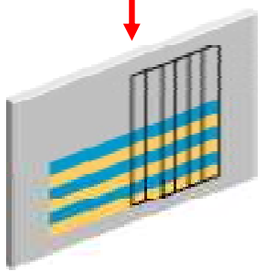
Evolution of Databases

- The database has completed more than 50 years of journey of its evolution from flat-file system to relational and objects relational systems. It has gone through several generations.

Data Storage on Different Media

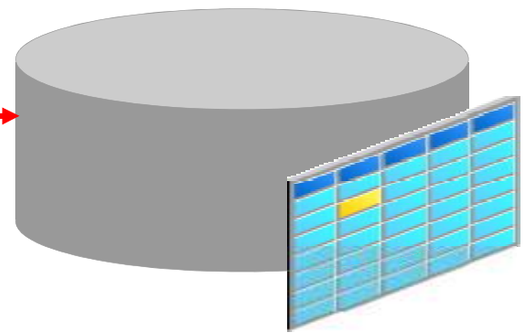
	DEPARTMENT_ID	DEPARTMENT_NAME	MANAGER_ID	LOCATION_ID
1	10	Administration	200	1700
2	20	Marketing	201	1800
3	50	Shipping		
4	60	IT		
5	80	Sales		
6	90	Executive		
7	110	Accounting		
8	190	Contracting		

	GRADE_LEVEL	LOWEST_SAL	HIGHEST_SAL
1	A	1000	2999
2	B	3000	5999
3	C	6000	9999
4	D	10000	14999
5	E	15000	24999
6	F	25000	40000



Electronic spreadsheet

Filing cabinet

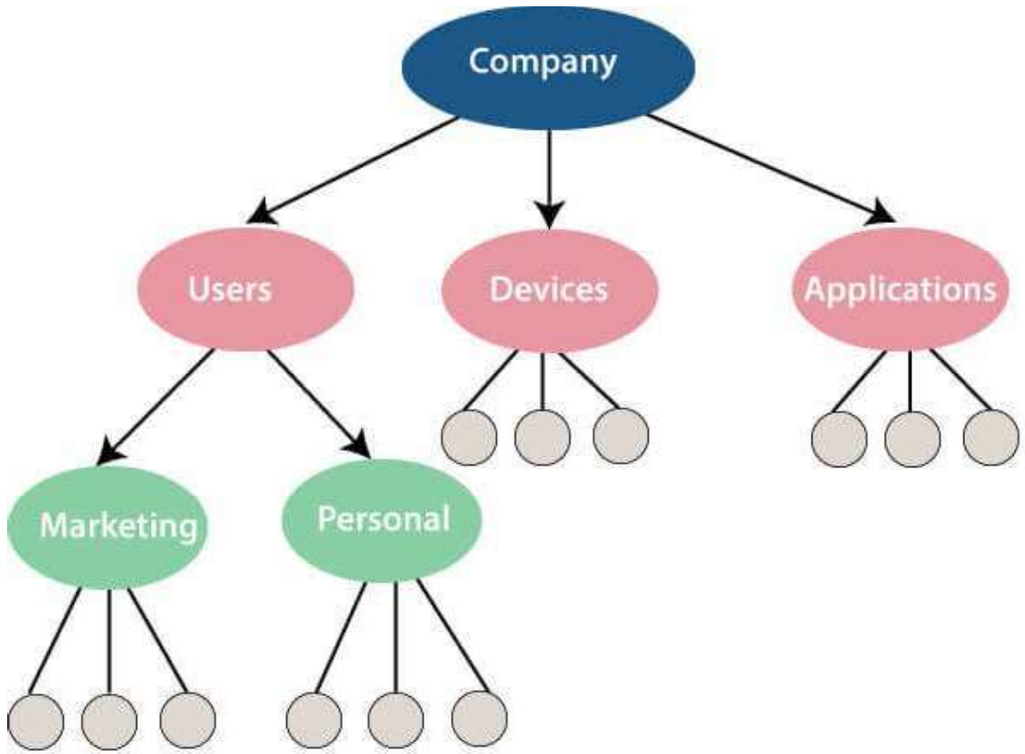


Database

MENTORLABS℠

Hierarchical Data Model

- 1968-1980 was the era of the Hierarchical Database. Prominent hierarchical database model was IBM's first DBMS. It was called IMS (Information Management System).



- Like file system, this model also had some limitations like complex implementation, lack structural independence, can't easily handle a many-many relationship, etc.

Relational Database

- 1970 - Present: It is the era of Relational Database and Database Management. In 1970, the relational model was proposed by E.F. Codd.
- Relational database model has two main terminologies called instance and schema.
- The instance is a table with rows or columns
- Schema specifies the structure like name of the relation, type of each column and name.
- This model uses some mathematical concept like set theory and predicate logic.
- The first internet database application had been created in 1995.
- During the era of the relational database, many more models had introduced like object-oriented model, object-relational model, etc.

Database on Cloud Services

Cloud provides three types of services:

- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Infrastructure as a Service (IaaS)
- Examples
  - AWS (Amazon Web Services)
  - Snowflake Computing
  - Oracle Database Cloud Services
  - Microsoft SQL server
  - Google cloud spanner

NoSQL Database

- A NoSQL database is an approach to design such databases that can accommodate a wide variety of data models. NoSQL stands for "not only SQL." It is an alternative to traditional relational databases in which data is placed in tables, and data schema is perfectly designed before the database is built.
- NoSQL databases are useful for a large set of distributed data.
  - MongoDB, CouchDB, Cloudant **(Document-based)**
  - Memcached, Redis, Coherence **(key-value store)**
  - HBase, Big Table, Accumulo **(Tabular)**

The Object-Oriented Databases

- The object-oriented databases contain data in the form of object and classes. Objects are the real-world entity, and types are the collection of objects.
- An object-oriented database is a combination of relational model features with objects oriented principles. It is an alternative implementation to that of the relational model.

Object-oriented programming properties

- Objects
- Classes
- Inheritance
- Polymorphism
- Encapsulation

Relational database properties

- Atomicity
- Consistency
- Integrity
- Durability
- Concurrency
- Query processing

Graph Databases

- A graph database is a NoSQL database. It is a graphical representation of data. It contains nodes and edges. A node represents an entity, and each edge represents a relationship between two edges. Every node in a graph database represents a unique identifier.
- Graph databases are beneficial for searching the relationship between data because they highlight the relationship between relevant data.

RDBMS (Relational Database Management System)

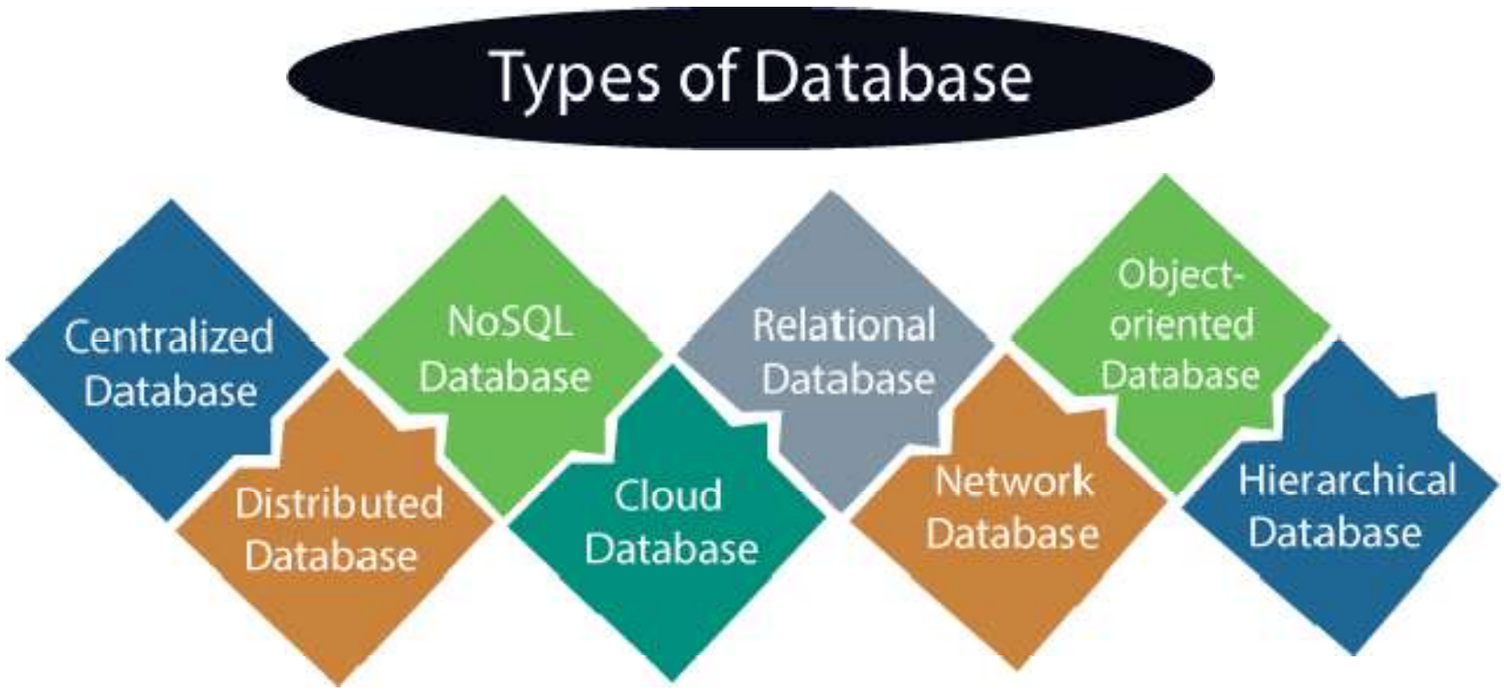
- The word RDBMS is termed as 'Relational Database Management System.' It is represented as a table that contains rows and column.
- RDBMS is based on the Relational model; it was introduced by E. F. Codd.

A relational database contains the following components:

- Table
- Record/ Tuple
- Field/Column name /Attribute
- Instance
- Schema
- Keys

- An RDBMS is a tabular DBMS that maintains the security, integrity, accuracy, and consistency of the data.

Types of Databases


Types of Database
Centralized Database
Distributed Database
NoSQL Database
Cloud Database
Relational Database
Network Database
Object-oriented Database
Hierarchical Database

Centralized Database

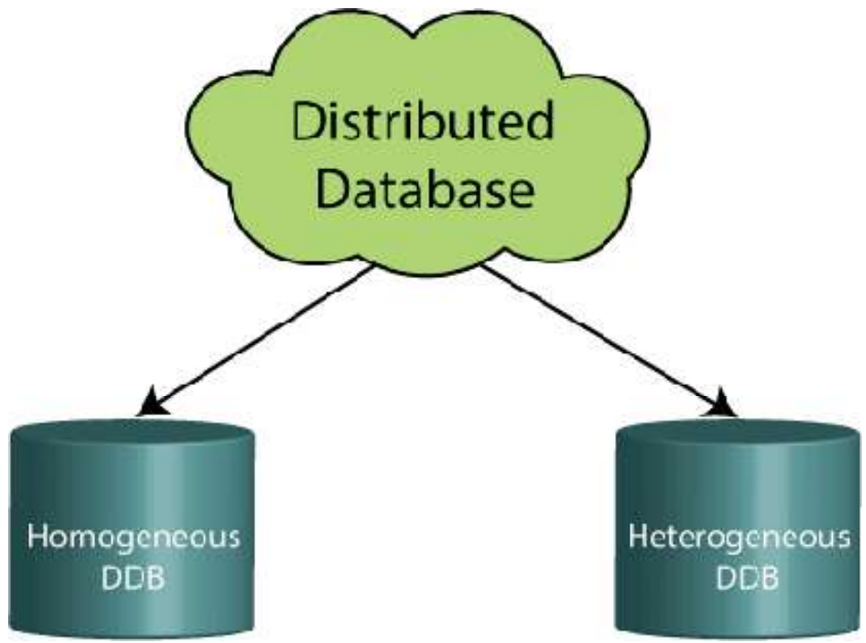
- It is the type of database that stores data at a centralized database system. It comforts the users to access the stored data from different locations through several applications.
- These applications contain the authentication process to let users access data securely. An example of a Centralized database can be Central Library that carries a central database of each library in a college/university.

Advantages & Disadvantages

- **Advantages of Centralized Database**
  - It has decreased the risk of data management, i.e., manipulation of data will not affect the core data.
  - Data consistency is maintained as it manages data in a central repository.
  - It provides better data quality, which enables organizations to establish data standards.
  - It is less costly because fewer vendors are required to handle the data sets.
- **Disadvantages of Centralized Database**
  - The size of the centralized database is large, which increases the response time for fetching the data.
  - It is not easy to update such an extensive database system.
  - If any server failure occurs, entire data will be lost, which could be a huge loss.

Distributed Database

- Unlike a centralized database system, in distributed systems, data is distributed among different database systems of an organization. These database systems are connected via communication links. Such links help the end-users to access the data easily. **Examples** of the Distributed database are Apache Cassandra, HBase, Ignite, etc.



Relational Database

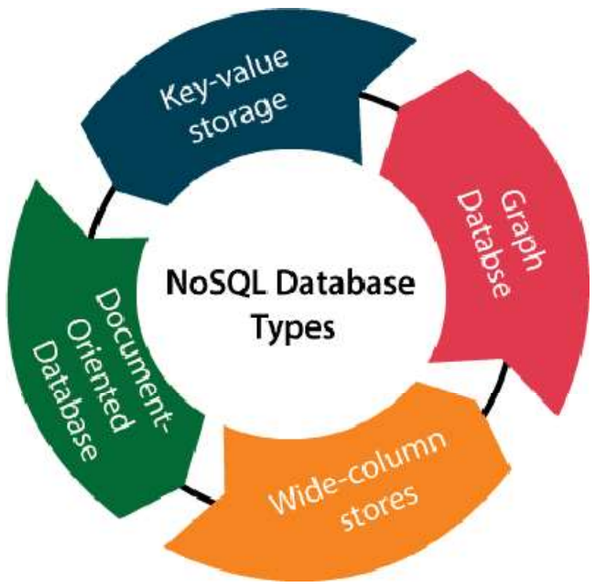
- This database is based on the relational data model, which stores data in the form of rows(tuple) and columns(attributes), and together forms a table(relation).
- A relational database uses SQL for storing, manipulating, as well as maintaining the data. E.F. Codd invented the database in 1970.
- Each table in the database carries a key that makes the data unique from others. **Examples** of Relational databases are MySQL, Microsoft SQL Server, Oracle, etc.

Properties of Relational Database

- **A means Atomicity:** This ensures the data operation will complete either with success or with failure. It follows the 'all or nothing' strategy. For example, a transaction will either be committed or will abort.
- **C means Consistency:** If we perform any operation over the data, its value before and after the operation should be preserved. For example, the account balance before and after the transaction should be correct, i.e., it should remain conserved.
- **I means Isolation:** There can be concurrent users for accessing data at the same time from the database. Thus, isolation between the data should remain isolated. For example, when multiple transactions occur at the same time, one transaction effects should not be visible to the other transactions in the database.
- **D means Durability:** It ensures that once it completes the operation and commits the data, data changes should remain permanent.

NoSQL Database

- Non-SQL/Not Only SQL is a type of database that is used for storing a wide range of data sets.
- It is not a relational database as it stores data not only in tabular form but in several different ways.
- It came into existence when the demand for building modern applications increased. Thus, NoSQL presented a wide variety of database technologies in response to the demands.



- **Key-value storage:** It is the simplest type of database storage where it stores every single item as a key (or attribute name) holding its value, together.
- **Document-oriented Database:** A type of database used to store data as JSON-like document. It helps developers in storing data by using the same document-model format as used in the application code.
- **Graph Databases:** It is used for storing vast amounts of data in a graph-like structure. Most commonly, social networking websites use the graph database.
- **Wide-column stores:** It is similar to the data represented in relational databases. Here, data is stored in large columns together, instead of storing in rows.

Advantage of NoSQL

- **High Scalability**
  - NoSQL can handle an extensive amount of data because of scalability. If the data grows, NoSQL database scale it to handle that data in an efficient manner.
- **High Availability**
  - NoSQL supports auto replication. Auto replication makes it highly available because, in case of any failure, data replicates itself to the previous consistent state.
- It is a better option for managing and handling large data sets.
- Users can quickly access data from the database through key-value.

Disadvantage of NoSQL

- **Open source**
  - NoSQL is an open-source database, so there is no reliable standard for NoSQL yet.
- **Management challenge**
  - Data management in NoSQL is much more complicated than relational databases. It is very challenging to install and even more hectic to manage daily.
- **GUI is not available**
  - GUI tools for NoSQL database are not easily available in the market.
- **Backup**
  - Backup is a great weak point for NoSQL databases. Some databases, like MongoDB, have no powerful approaches for data backup.

Database on Cloud Services

Cloud provides three types of services:

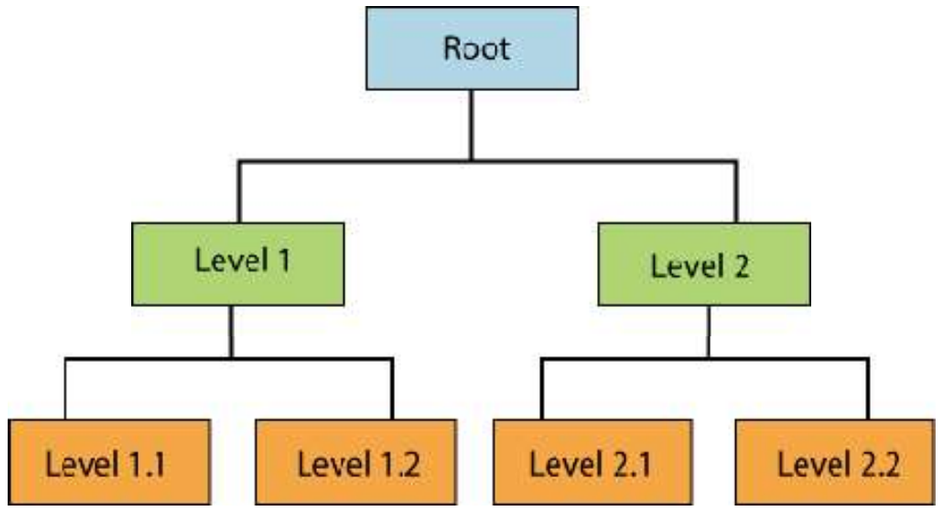
- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Infrastructure as a Service (IaaS)
- Examples
  - AWS (Amazon Web Services)
  - Snowflake Computing
  - Oracle Database Cloud Services
  - Microsoft SQL server
  - Google cloud spanner

Object-Oriented Databases

- The type of database that uses the object-based data model approach for storing data in the database system.
- The data is represented and stored as objects which are similar to the objects used in the object-oriented programming language.

Hierarchical Databases

- It is the type of database that stores data in the form of parent-children relationship nodes. Here, it organizes data in a tree-like structure.



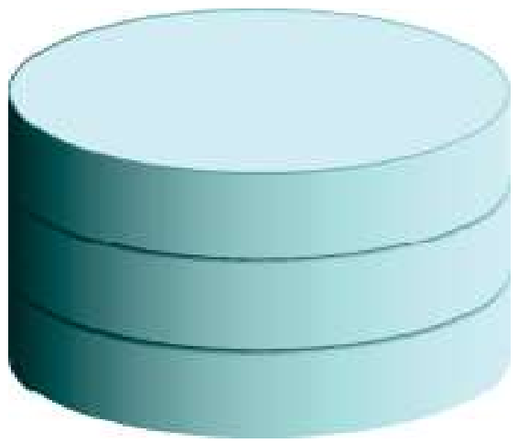
Hierarchical Database

Network Databases

- It is the database that typically follows the network data model. Here, the representation of data is in the form of nodes connected via links between them.
- Unlike the hierarchical database, it allows each record to have multiple children and parent nodes to form a generalized graph structure.

Relational and Object Relational

- Relational model and object relational model
- User-defined data types and objects
- Fully compatible with relational database
- Supports multimedia and large objects
- High-quality database server features



Relational Database Concept

- Dr. E. F. Codd proposed the relational model for database systems in 1970.
- It is the basis for the relational database management system (RDBMS).
- The relational model consists of the following:
  - Collection of objects or relations
  - Set of operators to act on the relations
  - Data integrity for accuracy and consistency

Definition of a Relational Database

- A relational database is a collection of relations or two-dimensional tables controlled by the Oracle server.

Database Server

Table name: EMPLOYEES

EMPLOYEE_ID	FIRST_NAME	LAST_NAME	EMAIL
100	Steven	King	SKING
101	Neena	Kochhar	NKOCHHAR
102	Lex	De Haan	LDEHAAN

...

Table name: DEPARTMENTS

DEPARTMENT_ID	DEPARTMENT_NAME	MANAGER_ID
10	Administration	200
20	Marketing	201
50	Shipping	124

...

What is a row or record?

- A row of a table is also called a record or tuple. It contains the specific information of each entry in the table. It is a horizontal entity in the table. For example, The above table contains 5 records.
- **Properties of a row:**
  - No two tuples are identical to each other in all their entries.
  - All tuples of the relation have the same format and the same number of entries.
  - The order of the tuple is irrelevant. They are identified by their content, not by their position.

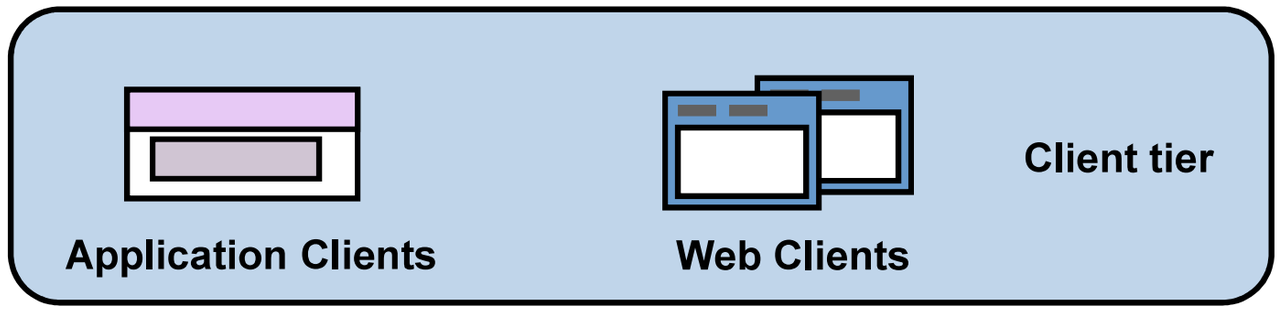
ID	Name	Address	Contact
1	Rahul	#23, Jayanagar Bangalore	9876543210

- Column/attribute
  - A column is a vertical entity in the table which contains all information associated with a specific field in a table.
- Data item/Cells
  - The smallest unit of data in the table is the individual data item. It is stored at the intersection of tuples and attributes.
- Cardinality
  - The total number of tuples at any one time in a relation is known as the table's cardinality. The relation whose cardinality is 0 is called an empty table.
  - **For example, the Employee table has 5 rows, and its cardinality is 5.**

Difference between DBMS and RDBMS

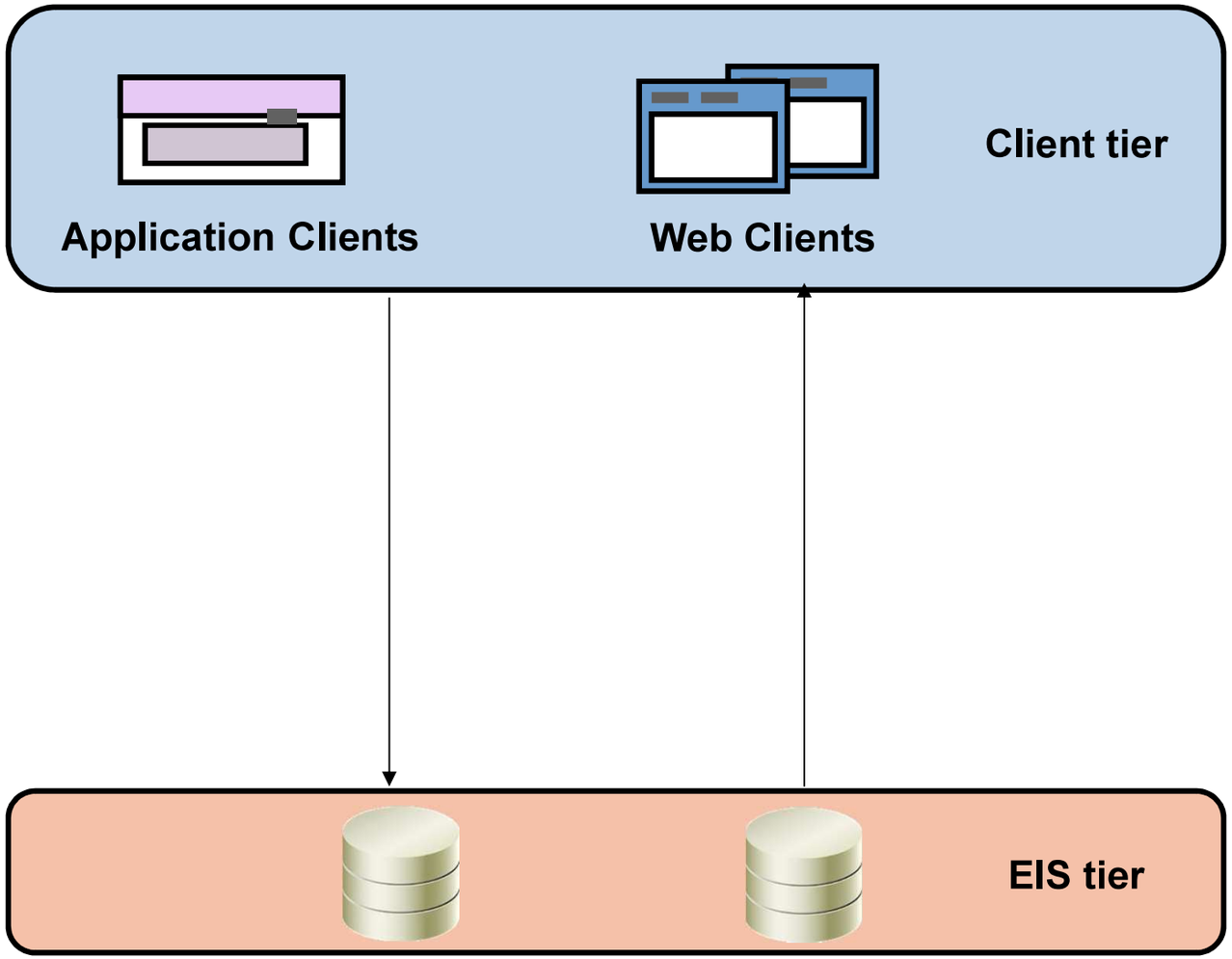
No.	DBMS	RDBMS
1)	DBMS applications store **data as file**.	RDBMS applications store **data in a tabular form**.
2)	In DBMS, data is generally stored in either a hierarchical form or a navigational form.	In RDBMS, the tables have an identifier called primary key and the data values are stored in the form of tables.
3)	**Normalization is not** present in DBMS.	**Normalization is** present in RDBMS.
4)	DBMS does **not apply any security** with regards to data manipulation.	RDBMS **defines the integrity constraint** for the purpose of ACID (Atomocity, Consistency, Isolation and Durability) property.
5)	DBMS uses file system to store data, so there will be **no relation between the tables**.	in RDBMS, data values are stored in the form of tables, so a **relationship** between these data values will be stored in the form of a table as well.
6)	DBMS has to provide some uniform methods to access the stored information.	RDBMS system supports a tabular structure of the data and a relationship between them to access the stored information.
7)	DBMS **does not support distributed database**.	RDBMS **supports distributed database**.
8)	DBMS is meant to be for small organization and **deal with small data**. it supports **single user**.	RDBMS is designed to **handle large amount of data**. it supports **multiple users**.
9)	Examples of DBMS are file systems, **xml** etc.	Example of RDBMS are **mysql**, **postgre**, **sql server**, **oracle** etc.

Single Tier Architecture

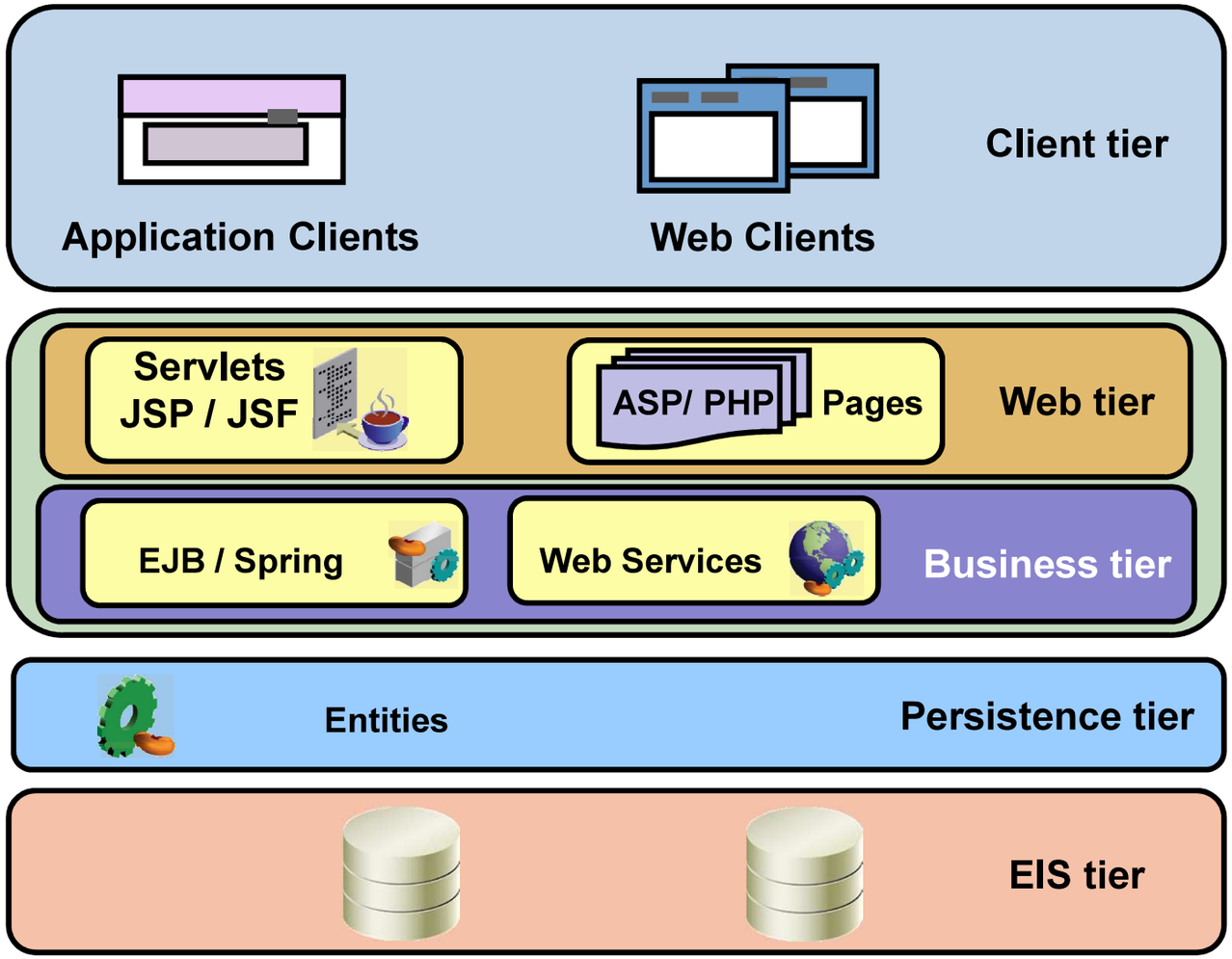


- Simple, Faster, Secured
- No N/W, Modification is Difficult, Data Lost is Lost

Two Tier Architecture



Distributed Multi Tier Architecture



Relating Multiple Tables

- Each row of data in a table can be uniquely identified by a primary key.
- You can logically relate data from multiple tables using foreign keys.

Table name: `DEPARTMENTS`

DEPARTMENT_ID	DEPARTMENT_NAME	MANAGER_ID	LOCATION_ID
10	Administration	200	1700
20	Marketing	201	1800
50	Shipping	124	1500
60	IT	103	1400
80	Sales	149	2500
90	Executive	100	1700
110	Accounting	205	1700
190	Contracting	(null)	1700

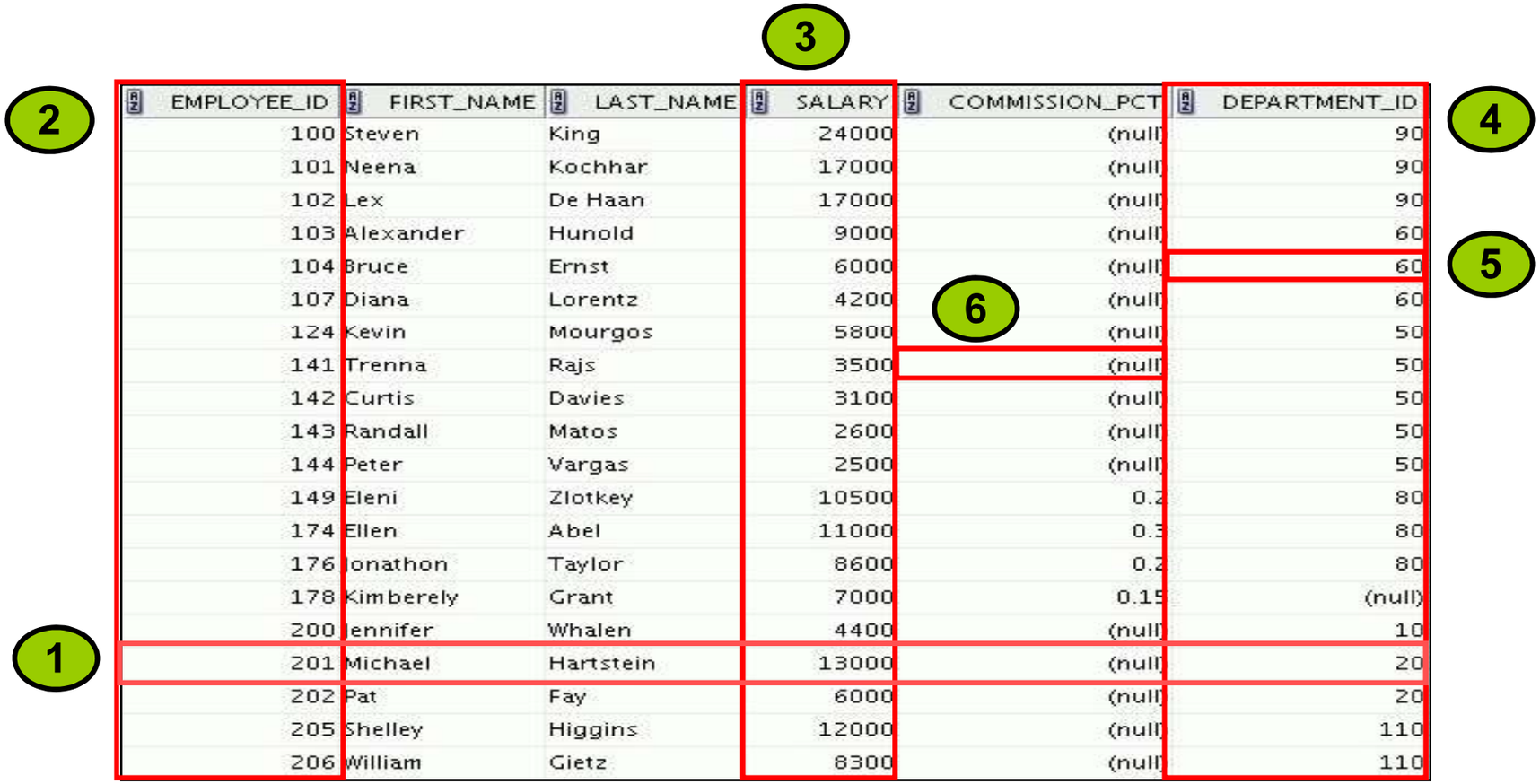
Primary key

Table name: `EMPLOYEES`

EMPLOYEE_ID	FIRST_NAME	LAST_NAME	DEPARTMENT_ID
100	Steven	King	90
101	Neena	Kochhar	90
102	Lex	De Haan	90
103	Alexander	Hunold	60
104	Bruce	Ernst	60
107	Diana	Lorentz	60
124	Kevin	Mourgos	50
141	Trenna	Rajs	50
142	Curtis	Davies	50

Primary key **Foreign key**

Relational Database Terminology



EMPLOYEE_ID	FIRST_NAME	LAST_NAME	SALARY	COMMISSION_PCT	DEPARTMENT_ID
100	Steven	King	24000	(null)	90
101	Neena	Kochhar	17000	(null)	90
102	Lex	De Haan	17000	(null)	90
103	Alexander	Hunold	9000	(null)	60
104	Bruce	Ernst	6000	(null)	60
107	Diana	Lorentz	4200	(null)	60
124	Kevin	Mourgos	5800	(null)	50
141	Trenna	Rajs	3500	(null)	50
142	Curtis	Davies	3100	(null)	50
143	Randall	Matos	2600	(null)	50
144	Peter	Vargas	2500	(null)	50
149	Eleni	Zlotkey	10500	0.2	80
174	Ellen	Abel	11000	0.3	80
176	Jonathon	Taylor	8600	0.2	80
178	Kimberely	Grant	7000	0.15	(null)
200	Jennifer	Whalen	4400	(null)	10
201	Michael	Hartstein	13000	(null)	20
202	Pat	Fay	6000	(null)	20
205	Shelley	Higgins	12000	(null)	110
206	William	Gietz	8300	(null)	110

Summary

In this lesson, you should have learned that:

- Course objectives and agenda
- Introduction to Databases and its terminologies
- Types of Databases
- Table, Row, Column and Cardinalities