



Distributed DBMS

tutorialspoint

SIMPLY EASY LEARNING

www.tutorialspoint.com



<https://www.facebook.com/tutorialspointindia>



<https://twitter.com/tutorialspoint>

About this Tutorial

Distributed Database Management System (DDBMS) is a type of DBMS which manages a number of databases hoisted at diversified locations and interconnected through a computer network. It provides mechanisms so that the distribution remains oblivious to the users, who perceive the database as a single database.

This tutorial discusses the important theories of distributed database systems. A number of illustrations and examples have been provided to aid the students to grasp the intricate concepts of DDBMS.

Audience

This tutorial has been prepared for students pursuing either a master's degree or a bachelor's degree in Computer Science, particularly if they have opted for distributed systems or distributed database systems as a subject.

Prerequisites

This tutorial is an advanced topic that focuses of a type of database system. Consequently, it requires students to have a reasonably good knowledge on the elementary concepts of DBMS. Besides, an understanding of SQL will be an added advantage.

Copyright & Disclaimer

© Copyright 2016 by Tutorials Point (I) Pvt. Ltd.

All the content and graphics published in this e-book are the property of Tutorials Point (I) Pvt. Ltd. The user of this e-book is prohibited to reuse, retain, copy, distribute or republish any contents or a part of contents of this e-book in any manner without written consent of the publisher.

We strive to update the contents of our website and tutorials as timely and as precisely as possible, however, the contents may contain inaccuracies or errors. Tutorials Point (I) Pvt. Ltd. provides no guarantee regarding the accuracy, timeliness or completeness of our website or its contents including this tutorial. If you discover any errors on our website or in this tutorial, please notify us at contact@tutorialspoint.com

Table of Contents

About this Tutorial.....	i
Audience.....	i
Prerequisites.....	i
Copyright & Disclaimer.....	i
Table of Contents	ii
PART 1: DDBMS – BASICS.....	1
1. DDBMS – DBMS Concepts.....	2
Database and Database Management System.....	2
Database Schemas.....	3
Types of DBMS.....	3
Operations on DBMS	5
2. DDBMS – Distributed Databases	8
Distributed Database Management System.....	8
Factors Encouraging DDBMS	9
Advantages of Distributed Databases	9
Adversities of Distributed Databases	10
PART 2: DISTRIBUTED DATABASE DESIGN	11
3. DDBMS – Distributed Database Environments.....	12
Types of Distributed Databases.....	12
Distributed DBMS Architectures	13
Architectural Models.....	14
Design Alternatives.....	17
4. DDBMS – Design Strategies.....	19
Data Replication	19
Fragmentation	20
Vertical Fragmentation.....	20
Horizontal Fragmentation	21
Hybrid Fragmentation	21
5. DDBMS – Distribution Transparency	22
Location Transparency	22
Fragmentation Transparency	22
Replication Transparency	22
Combination of Transparencies.....	23
6. DDBMS – Database Control.....	24
Authentication.....	24
Access Rights	24
Semantic Integrity Control	25

PART 3: QUERY OPTIMIZATION	27
7. DDBMS – Relational Algebra for Query Optimization	28
Query Optimization Issues in DDBMS	28
Query Processing.....	28
Relational Algebra	29
Translating SQL Queries into Relational Algebra.....	32
Computation of Relational Algebra Operators.....	33
Computation of Selection.....	34
Computation of Joins.....	34
8. DDBMS – Query Optimization in Centralized Systems.....	36
Query Parsing and Translation	36
Approaches to Query Optimization.....	38
9. DDBMS – Query Optimization in Distributed Systems.....	39
Distributed Query Processing Architecture.....	39
Mapping Global Queries into Local Queries.....	39
Distributed Query Optimization	40
PART 4: CONCURRENCY CONTROL.....	43
10. DDBMS – Transaction Processing Systems	44
Transactions	44
Transaction Operations	44
Transaction States	45
Desirable Properties of Transactions.....	45
Schedules and Conflicts	46
Serializability.....	47
11. DDBMS – Controlling Concurrency.....	48
Locking Based Concurrency Control Protocols.....	48
Timestamp Concurrency Control Algorithms	48
Optimistic Concurrency Control Algorithm	49
Concurrency Control in Distributed Systems.....	50
12. DDBMS – Deadlock Handling	52
What are Deadlocks?.....	52
Deadlock Handling in Centralized Systems.....	52
Deadlock Handling in Distributed Systems.....	54
PART 5: FAILURE AND RECOVERY.....	57
13. DDBMS – Replication Control.....	58
Synchronous Replication Control	58
Asynchronous Replication Control	59
Replication Control Algorithms	59
14. DDBMS – Failure & Commit	62
Soft Failure	62
Hard Failure	62
Network Failure	62

Commit Protocols	63
Transaction Log	63
15. DDBMS – Database Recovery	65
Recovery from Power Failure	65
Recovery from Disk Failure	65
Checkpointing	66
Transaction Recovery Using UNDO / REDO	67
16. DDBMS – Distributed Commit Protocols	69
Distributed One-phase Commit	69
Distributed Two-phase Commit	69
Distributed Three-phase Commit	70
PART 6: DISTRIBUTED DBMS SECURITY	71
17. DDBMS – Database Security & Cryptography	72
Database Security and Threats	72
Measures of Control	72
What is Cryptography?	72
Conventional Encryption Methods	73
Public Key Cryptography	73
Digital Signatures	74
18. DDBMS – Security in Distributed Databases	75
Communications Security	75
Data Security	75
Data Auditing	76

Part 1: DDBMS – Basics

1. DDBMS – DBMS CONCEPTS

For proper functioning of any organization, there's a need for a well-maintained database. In the recent past, databases used to be centralized in nature. However, with the increase in globalization, organizations tend to be diversified across the globe. They may choose to distribute data over local servers instead of a central database. Thus, arrived the concept of **Distributed Databases**.

This chapter gives an overview of databases and Database Management Systems (DBMS). A database is an ordered collection of related data. A DBMS is a software package to work upon a database. A detailed study of DBMS is available in our tutorial named "Learn DBMS". In this chapter, we revise the main concepts so that the study of DDBMS can be done with ease. The three topics covered are database schemas, types of databases and operations on databases.

Database and Database Management System

A **database** is an ordered collection of related data that is built for a specific purpose. A database may be organized as a collection of multiple tables, where a table represents a real world element or entity. Each table has several different fields that represent the characteristic features of the entity.

For example, a company database may include tables for projects, employees, departments, products and financial records. The fields in the Employee table may be Name, Company_Id, Date_of_Joining, and so forth.

A **database management system** is a collection of programs that enables creation and maintenance of a database. DBMS is available as a software package that facilitates definition, construction, manipulation and sharing of data in a database. Definition of a database includes description of the structure of a database. Construction of a database involves actual storing of the data in any storage medium. Manipulation refers to the retrieving information from the database, updating the database and generating reports. Sharing of data facilitates data to be accessed by different users or programs.

Examples of DBMS Application Areas

- Automatic Teller Machines
- Train Reservation System
- Employee Management System
- Student Information System

Examples of DBMS Packages

- MySQL
- Oracle
- SQL Server
- dBASE
- FoxPro
- PostgreSQL, etc.

Database Schemas

A database schema is a description of the database which is specified during database design and subject to infrequent alterations. It defines the organization of the data, the relationships among them, and the constraints associated with them.

Databases are often represented through the **three-schema architecture** or **ANSI-SPARC architecture**. The goal of this architecture is to separate the user application from the physical database. The three levels are:

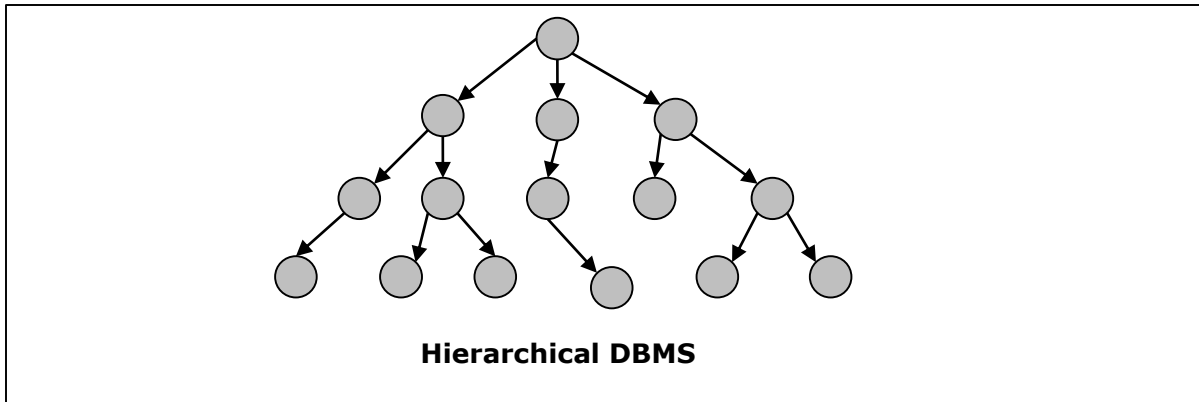
- **Internal Level having Internal Schema** – It describes the physical structure, details of internal storage and access paths for the database.
- **Conceptual Level having Conceptual Schema** – It describes the structure of the whole database while hiding the details of physical storage of data. This illustrates the entities, attributes with their data types and constraints, user operations and relationships.
- **External or View Level having External Schemas or Views** – It describes the portion of a database relevant to a particular user or a group of users while hiding the rest of database.

Types of DBMS

There are four types of DBMS.

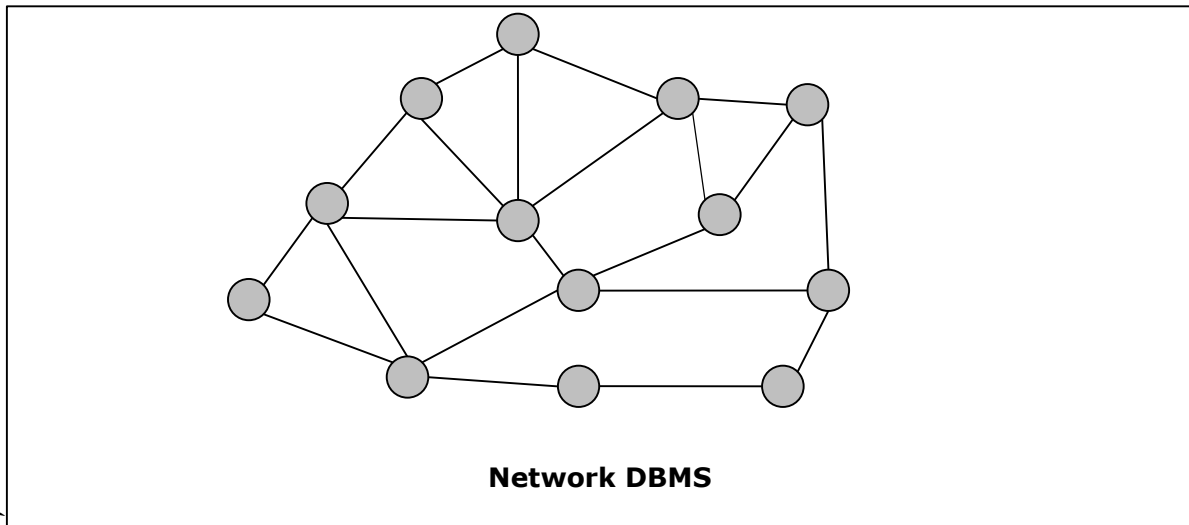
Hierarchical DBMS

In hierarchical DBMS, the relationships among data in the database are established so that one data element exists as a subordinate of another. The data elements have parent-child relationships and are modelled using the “tree” data structure. These are very fast and simple.



Network DBMS

Network DBMS is one where the relationships among data in the database are of type many-to-many in the form of a network. The structure is generally complicated due to the existence of numerous many-to-many relationships. Network DBMS is modelled using "graph" data structure.



Relational DBMS

In relational databases, the database is represented in the form of relations. Each relation models an entity and is represented as a table of values. In the relation or table, a row is called a tuple and denotes a single record. A column is called a field or an attribute and denotes a characteristic property of the entity. RDBMS is the most popular database management system.

For example: A Student Relation→

Field			
S_Id	Name	Year	Stream

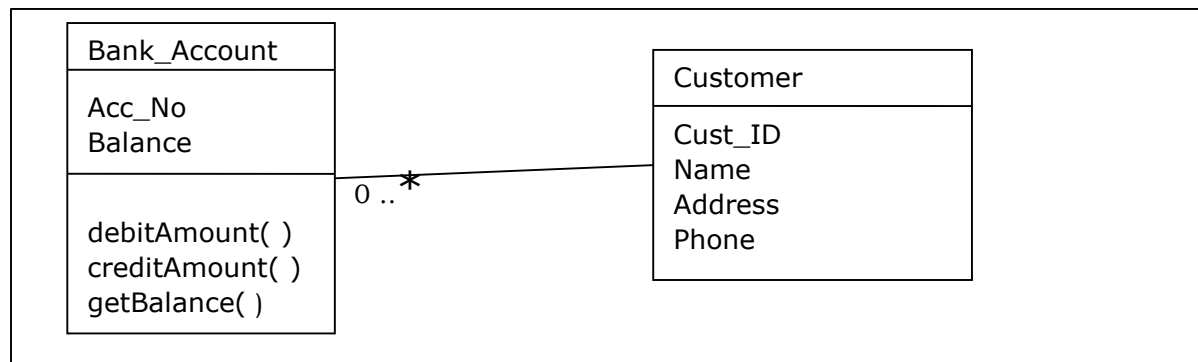
Tuple →

1	Ankit Jha	1	Computer Science
2	Pushpa Mishra	2	Electronics
5	Ranjini Iyer	2	Computer Science

Object Oriented DBMS

Object-oriented DBMS is derived from the model of the object-oriented programming paradigm. They are helpful in representing both consistent data as stored in databases, as well as transient data, as found in executing programs. They use small, reusable elements called objects. Each object contains a data part and a set of operations which works upon the data. The object and its attributes are accessed through pointers instead of being stored in relational table models.

For example: A simplified Bank Account object-oriented database –



Distributed DBMS

A distributed database is a set of interconnected databases that is distributed over the computer network or internet. A Distributed Database Management System (DDBMS) manages the distributed database and provides mechanisms so as to make the databases transparent to the users. In these systems, data is intentionally distributed among multiple nodes so that all computing resources of the organization can be optimally used.

Operations on DBMS

The four basic operations on a database are Create, Retrieve, Update and Delete.

- **CREATE** database structure and populate it with data – Creation of a database relation involves specifying the data structures, data types and the constraints of the data to be stored.

Example: SQL command to create a student table:

```
CREATE TABLE STUDENT
(
```

```
ROLL INTEGER PRIMARY KEY,  
NAME VARCHAR2(25),  
YEAR INTEGER,  
STREAM VARCHAR2(10)  
);
```

Once the data format is defined, the actual data is stored in accordance with the format in some storage medium.

End of ebook preview
If you liked what you saw...
Buy it from our store @ <https://store.tutorialspoint>