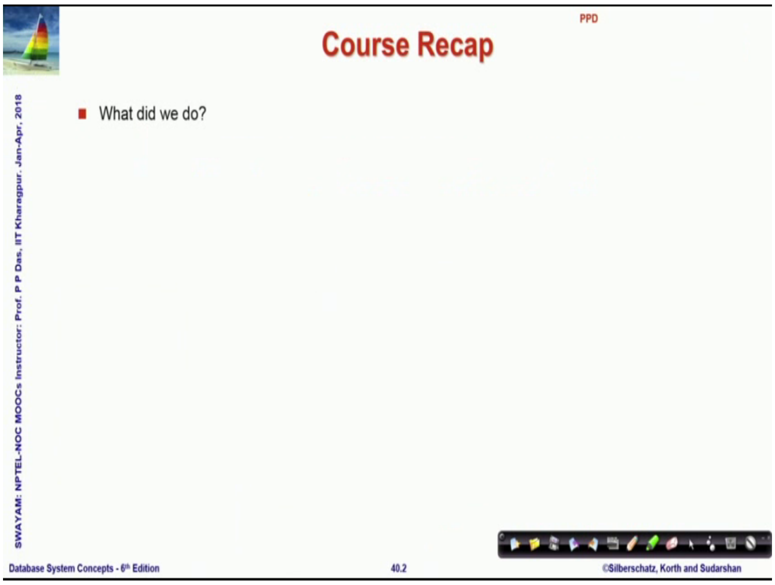
Database Management System
Prof. Partha Pratim Das
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

Lecture – 40
Course Summarization

Welcome to module 40 of Database Management Systems. This is for course summarization this is the last module of the course.

(Refer Slide Time: 00:23)



PPD

Course Recap

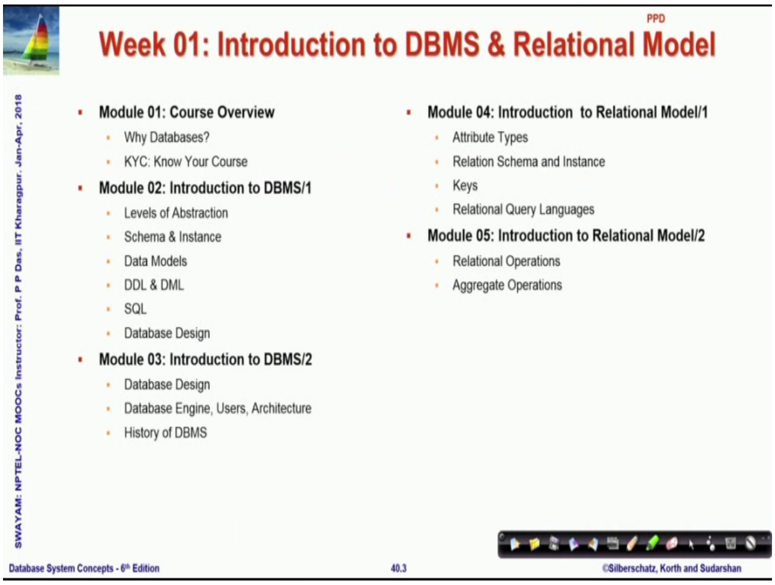
- What did we do?

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.2 ©Silberschatz, Korth and Sudarshan

So, I would just start with doing a quick recap of what all did we cover and what we expectedly learnt.

(Refer Slide Time: 00:32)



PPD

Week 01: Introduction to DBMS & Relational Model

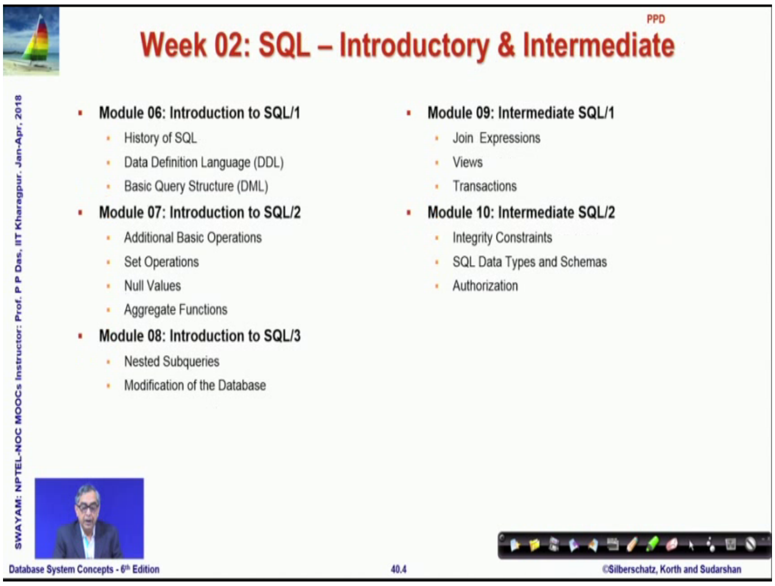
- Module 01: Course Overview
  - Why Databases?
  - KYC: Know Your Course
- Module 02: Introduction to DBMS/1
  - Levels of Abstraction
  - Schema & Instance
  - Data Models
  - DDL & DML
  - SQL
  - Database Design
- Module 03: Introduction to DBMS/2
  - Database Design
  - Database Engine, Users, Architecture
  - History of DBMS
- Module 04: Introduction to Relational Model/1
  - Attribute Types
  - Relation Schema and Instance
  - Keys
  - Relational Query Languages
- Module 05: Introduction to Relational Model/2
  - Relational Operations
  - Aggregate Operations

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.3 ©Silberschatz, Korth and Sudarshan

In the week 1, we talked primarily of Introduction to Database Management System and the Relational Model which is the foundation of a database system.

(Refer Slide Time: 00:42)



PPD

Week 02: SQL – Introductory & Intermediate

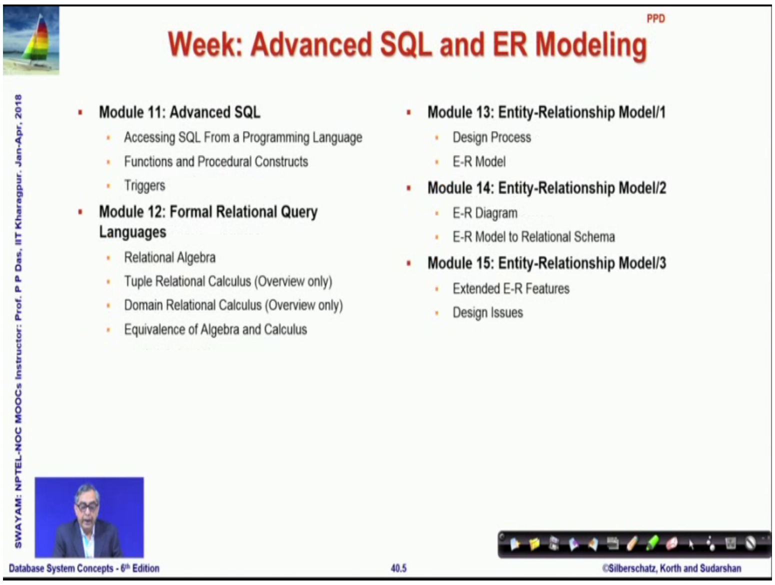
- Module 06: Introduction to SQL/1
  - History of SQL
  - Data Definition Language (DDL)
  - Basic Query Structure (DML)
- Module 07: Introduction to SQL/2
  - Additional Basic Operations
  - Set Operations
  - Null Values
  - Aggregate Functions
- Module 08: Introduction to SQL/3
  - Nested Subqueries
  - Modification of the Database
- Module 09: Intermediate SQL/1
  - Join Expressions
  - Views
  - Transactions
- Module 10: Intermediate SQL/2
  - Integrity Constraints
  - SQL Data Types and Schemas
  - Authorization

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.4 ©Silberschatz, Korth and Sudarshan

In week 2, we started off with query language SQL at an Introductory level and then at an Intermediate level which are really really the first major aspect of a database particularly relational database that a student must master.

(Refer Slide Time: 01:01)



PPD

Week: Advanced SQL and ER Modeling

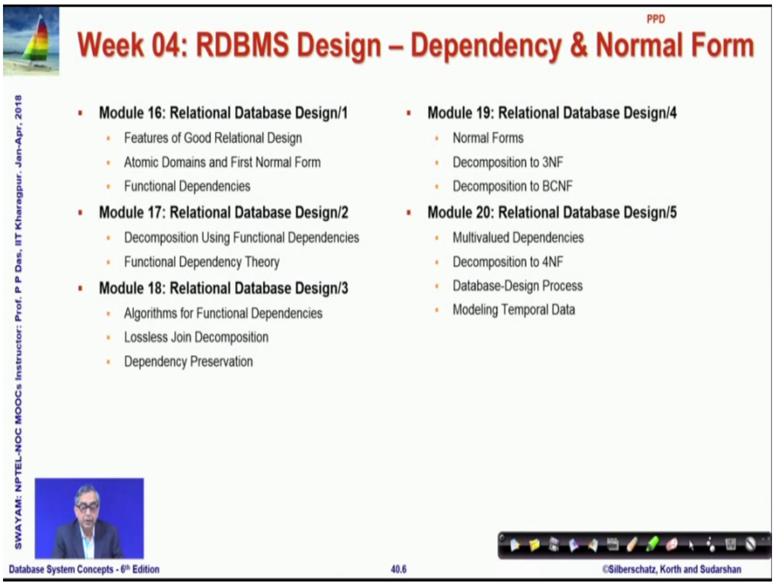
- Module 11: Advanced SQL
  - Accessing SQL From a Programming Language
  - Functions and Procedural Constructs
  - Triggers
- Module 12: Formal Relational Query Languages
  - Relational Algebra
  - Tuple Relational Calculus (Overview only)
  - Domain Relational Calculus (Overview only)
  - Equivalence of Algebra and Calculus
- Module 13: Entity-Relationship Model/1
  - Design Process
  - E-R Model
- Module 14: Entity-Relationship Model/2
  - E-R Diagram
  - E-R Model to Relational Schema
- Module 15: Entity-Relationship Model/3
  - Extended E-R Features
  - Design Issues

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.5 ©Silberschatz, Korth and Sudarshan

In week 3, we continued with the advanced SQL and did the aspects of modeling from specification in terms of Entity Relationship Model.

(Refer Slide Time: 01:12)



PPD

Week 04: RDBMS Design – Dependency & Normal Form

- Module 16: Relational Database Design/1
  - Features of Good Relational Design
  - Atomic Domains and First Normal Form
  - Functional Dependencies
- Module 17: Relational Database Design/2
  - Decomposition Using Functional Dependencies
  - Functional Dependency Theory
- Module 18: Relational Database Design/3
  - Algorithms for Functional Dependencies
  - Lossless Join Decomposition
  - Dependency Preservation
- Module 19: Relational Database Design/4
  - Normal Forms
  - Decomposition to 3NF
  - Decomposition to BCNF
- Module 20: Relational Database Design/5
  - Multivalued Dependencies
  - Decomposition to 4NF
  - Database-Design Process
  - Modeling Temporal Data

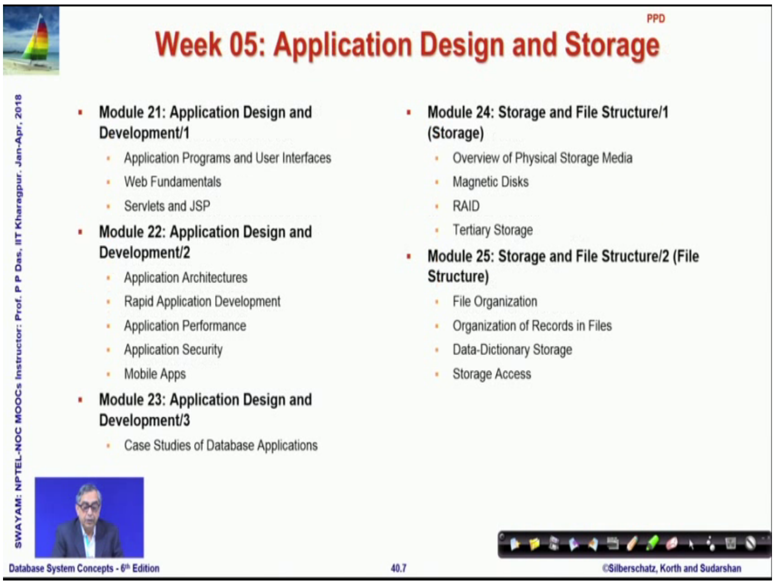
SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.6 ©Silberschatz, Korth and Sudarshan

In week the next week, we did the design issues which was really the involved part and possibly the most important aspect of the relational database design beyond query coding query being able to write queries.

So, this is based on dependency and different normal forms and I am sure you have spent a good time on mastering these.

(Refer Slide Time: 01:40)



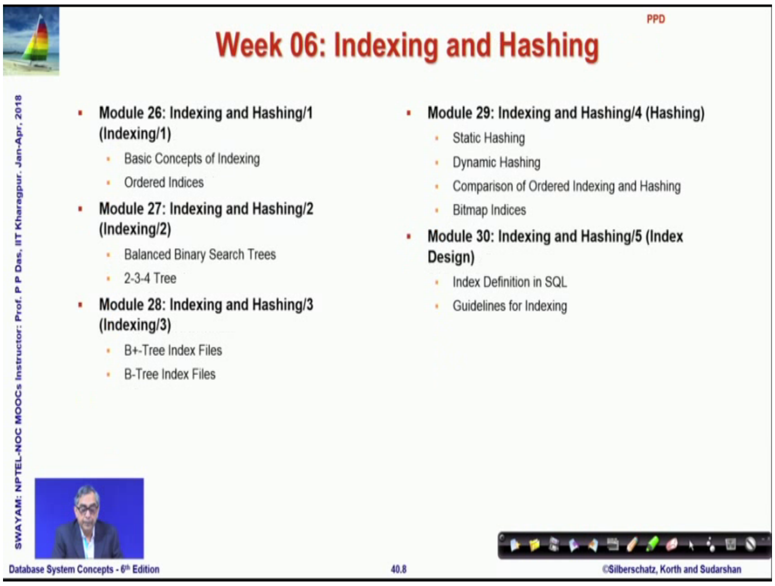
Week 05: Application Design and Storage

- Module 21: Application Design and Development/1
  - Application Programs and User Interfaces
  - Web Fundamentals
  - Servlets and JSP
- Module 22: Application Design and Development/2
  - Application Architectures
  - Rapid Application Development
  - Application Performance
  - Application Security
  - Mobile Apps
- Module 23: Application Design and Development/3
  - Case Studies of Database Applications
- Module 24: Storage and File Structure/1 (Storage)
  - Overview of Physical Storage Media
  - Magnetic Disks
  - RAID
  - Tertiary Storage
- Module 25: Storage and File Structure/2 (File Structure)
  - File Organization
  - Organization of Records in Files
  - Data-Dictionary Storage
  - Storage Access

PPD
SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018
Database System Concepts - 6th Edition
40.7
©Silberschatz, Korth and Sudarshan

We followed up in week 5 with application design and discussing aspects of storage structure, how will actually the items, data items be stored in the different memory and disk structure.

(Refer Slide Time: 01:56)



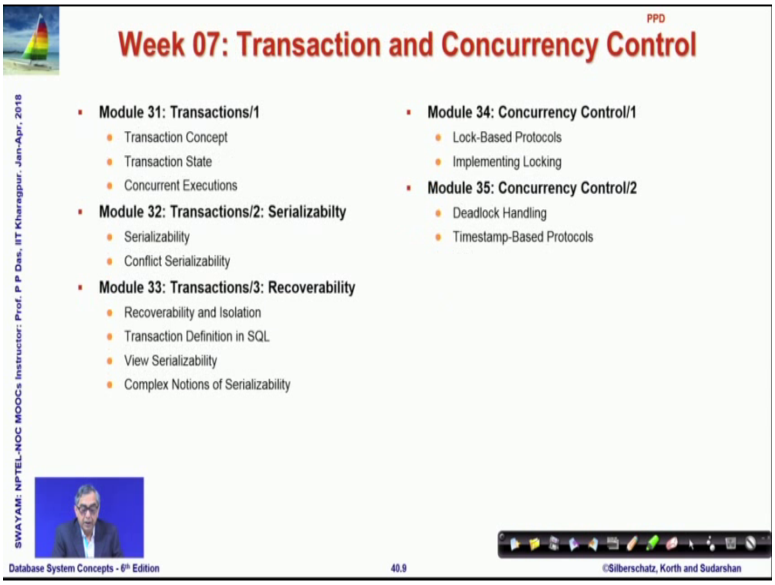
Week 06: Indexing and Hashing

- Module 26: Indexing and Hashing/1 (Indexing/1)
  - Basic Concepts of Indexing
  - Ordered Indices
- Module 27: Indexing and Hashing/2 (Indexing/2)
  - Balanced Binary Search Trees
  - 2-3-4 Tree
- Module 28: Indexing and Hashing/3 (Indexing/3)
  - B+-Tree Index Files
  - B-Tree Index Files
- Module 29: Indexing and Hashing/4 (Hashing)
  - Static Hashing
  - Dynamic Hashing
  - Comparison of Ordered Indexing and Hashing
  - Bitmap Indices
- Module 30: Indexing and Hashing/5 (Index Design)
  - Index Definition in SQL
  - Guidelines for Indexing

PPD
SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018
Database System Concepts - 6th Edition
40.8
©Silberschatz, Korth and Sudarshan

In the following week, in week 6, we discussed about indexing and hashing to for making the accesses really efficient.

(Refer Slide Time: 02:05)



PPD

Week 07: Transaction and Concurrency Control

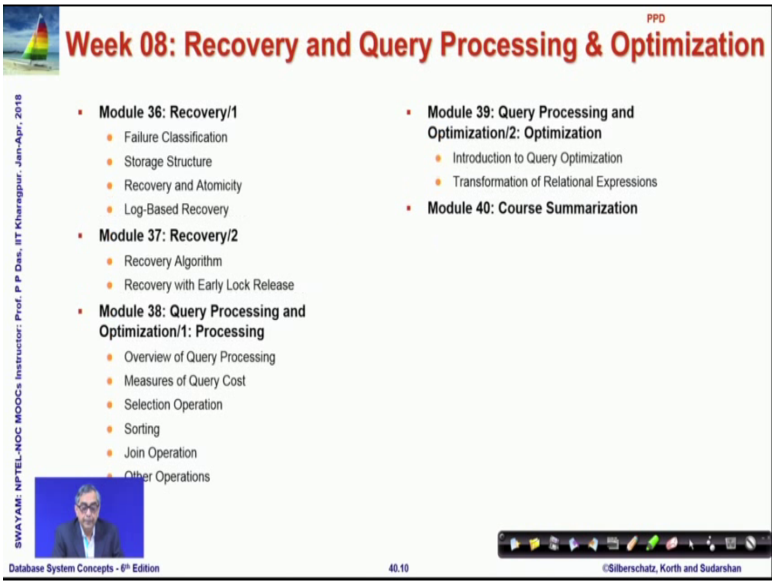
- Module 31: Transactions/1
  - Transaction Concept
  - Transaction State
  - Concurrent Executions
- Module 32: Transactions/2: Serializabilty
  - Serializability
  - Conflict Serializability
- Module 33: Transactions/3: Recoverability
  - Recoverability and Isolation
  - Transaction Definition in SQL
  - View Serializability
  - Complex Notions of Serializability
- Module 34: Concurrency Control/1
  - Lock-Based Protocols
  - Implementing Locking
- Module 35: Concurrency Control/2
  - Deadlock Handling
  - Timestamp-Based Protocols

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.9 ©Silberschatz, Korth and Sudarshan

In week 7, we did another critical aspect of database systems that is how to make transactions work concurrently. So, we defined transactions and define what is Concurrency and then we took in 2 different critical aspects of Serializability that it is possible that we can execute transactions in a manner so that their instructions are intermixed, but then even in that case, they actually produce a result which is as if these transactions could have been executed in the serial order and we talked about the issues of recoverability in this respect and we specifically looked at different protocols, particularly 2 phase locking protocol for managing this kind of concurrency and the evils of deadlock that may happen when you do it concurrency and how simple protocols like time based protocol can handle that.

(Refer Slide Time: 03:03)



PPD

Week 08: Recovery and Query Processing & Optimization

- Module 36: Recovery/1
  - Failure Classification
  - Storage Structure
  - Recovery and Atomicity
  - Log-Based Recovery
- Module 37: Recovery/2
  - Recovery Algorithm
  - Recovery with Early Lock Release
- Module 38: Query Processing and Optimization/1: Processing
  - Overview of Query Processing
  - Measures of Query Cost
  - Selection Operation
  - Sorting
  - Join Operation
  - Other Operations
- Module 39: Query Processing and Optimization/2: Optimization
  - Introduction to Query Optimization
  - Transformation of Relational Expressions
- Module 40: Course Summarization

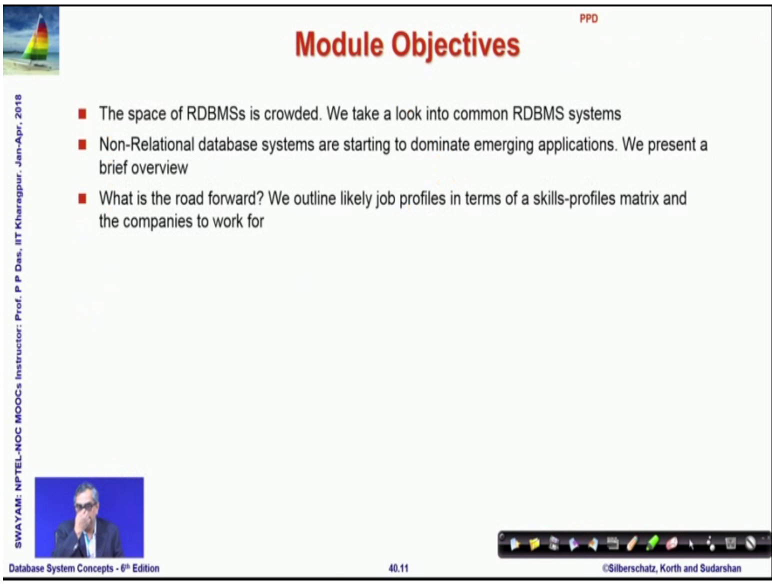
SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.10 ©Silberschatz, Korth and Sudarshan

And in the current week, we have dwelt with different strategies of recovery, particularly log based recovery and we have touched upon query processing and optimization.

So, this is you have got a very first level overview of this course. This is in a limited time and with limited number of assignments. So, you will just get a first level idea, this is not to make you really an expert of database systems, but this will certainly get you started well in terms of the database management programs, in terms of you are taking up advanced courses later on or in terms of actually taking up a job in different database area.

(Refer Slide Time: 03:48)



PPD

Module Objectives

- The space of RDBMSs is crowded. We take a look into common RDBMS systems
- Non-Relational database systems are starting to dominate emerging applications. We present a brief overview
- What is the road forward? We outline likely job profiles in terms of a skills-profiles matrix and the companies to work for

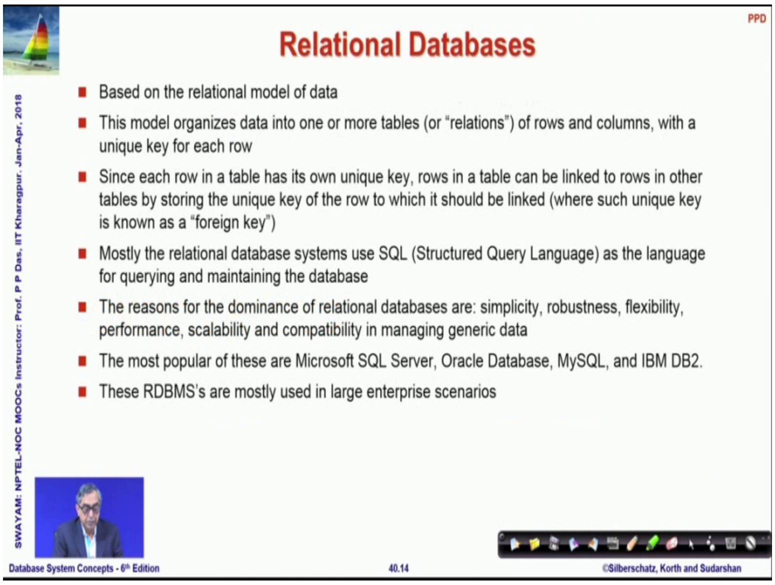
SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.11 ©Silberschatz, Korth and Sudarshan

So, given that given that in this current module, I would discuss about few things beyond the textbook actually, the space of databases RDBMS bases are quite crowded, the lot of RDBMS bases you will see. So, I will just take a quick look in terms of the common RDBMS systems and there had been a number of queries on the forum and in the live session about that. I will also touch upon we would like to discuss about non relational database systems which was not a part of the curriculum that we did here, but will just present a brief overview.

And then finally, I would like to conclude with what should be the road forward from you, presenting you a kind of a skill profile matrix so that what skills you must pick up to actually get a job off certain profile and what are the companies that you might look at working for.

(Refer Slide Time: 04:55)



PPD

Relational Databases

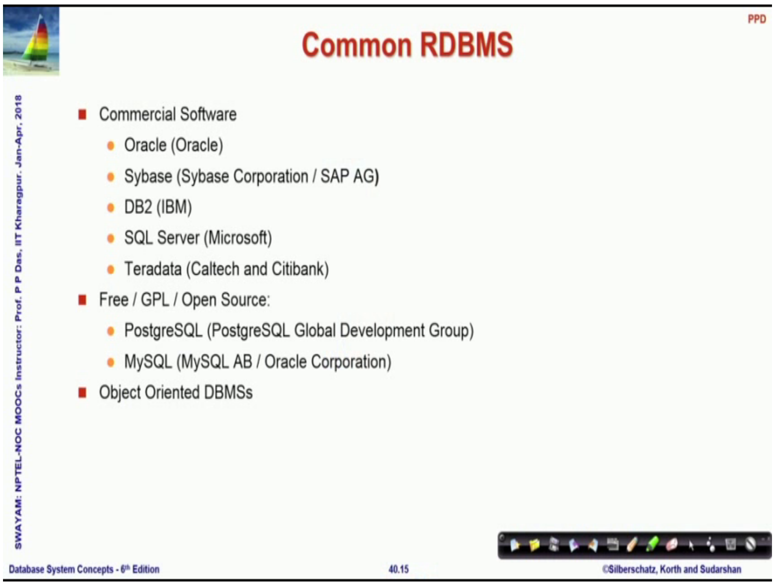
- Based on the relational model of data
- This model organizes data into one or more tables (or "relations") of rows and columns, with a unique key for each row
- Since each row in a table has its own unique key, rows in a table can be linked to rows in other tables by storing the unique key of the row to which it should be linked (where such unique key is known as a "foreign key")
- Mostly the relational database systems use SQL (Structured Query Language) as the language for querying and maintaining the database
- The reasons for the dominance of relational databases are: simplicity, robustness, flexibility, performance, scalability and compatibility in managing generic data
- The most popular of these are Microsoft SQL Server, Oracle Database, MySQL, and IBM DB2.
- These RDBMS's are mostly used in large enterprise scenarios

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.14 ©Silberschatz, Korth and Sudarshan

So, starting with the common databases, there are several relational database systems. So, in this slide you summarize basically what are the different aspects of relational database systems which you have been discussing so far. So, this is just a summary of that.

(Refer Slide Time: 05:07)



PPD

Common RDBMS

- Commercial Software
  - Oracle (Oracle)
  - Sybase (Sybase Corporation / SAP AG)
  - DB2 (IBM)
  - SQL Server (Microsoft)
  - Teradata (Caltech and Citibank)
- Free / GPL / Open Source:
  - PostgreSQL (PostgreSQL Global Development Group)
  - MySQL (MySQL AB / Oracle Corporation)
- Object Oriented DBMSs

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

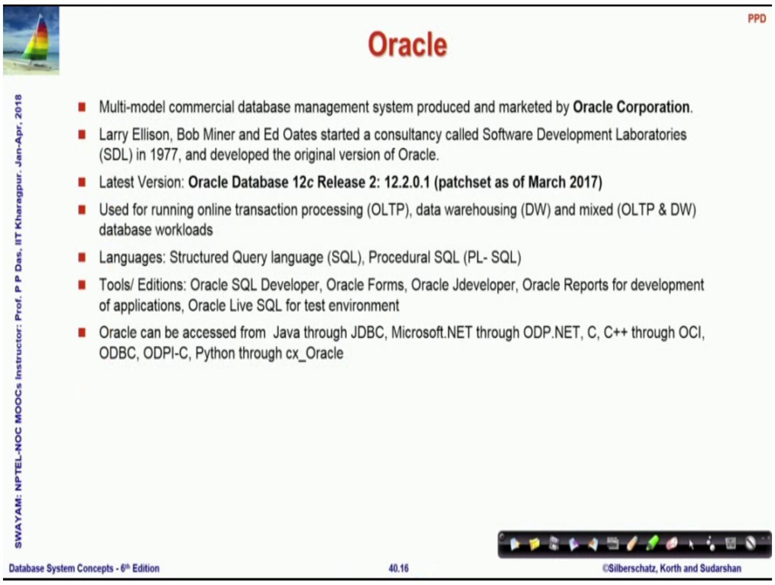
Database System Concepts - 6th Edition 40.15 ©Silberschatz, Korth and Sudarshan

Now, these are the common database systems. So, I have chosen the ones which are most widely used, most easily accessible and kind of large companies use them, large databases exist on them. So, there are primarily 2 classifications; one is a set of database systems are commercial Oracle from the Oracle corporation, Sybase from Sybase corporation which is now SAP AG, DB2 from IBM, SQL Server from Microsoft and the recent entrant to that who is making a regular ripples is Teradata which is a you know joint database systems from Caltech and certain group of Citibank. So, I if you are working for a company who subscribes to any of these database software, then you should be able to use them and understand what all you can do, but if you are working with smaller companies or you are working as a student, then you will need to use some of the database systems which are free or are on the GPL licensing or open source.

So, most prominent amongst them is PostgreSQL which is from a Postgres global development group. So, these are non commercial the software in the sense that you do not need to pay for them and they are on the GPL and some of some part of that would be open source as well and a very commonly used is my SQL which is was originally from a Swedish company called my SQL AB, but now it is acquired by Oracle corporation, but it still does not you do not need to pay for that. So, these are the databases and systems to primarily look for and besides that there are some other database systems which use certain object oriented features on top of the relational features.

So, if you look in through these, then in most cases you will find in terms of the gross functionality of the kind of SQL that you can write, a large subset of the SQL that you can write on databases maintained through these database systems will be same. So, what you have learnt here would be applicable irrespective of which database which of these database systems you are using, but of course there are specifics which would be different amongst them. So, in the next couple of slides, I have 4 on each one slide, I have given a brief background about the particular database system.

(Refer Slide Time: 07:48)



PPD

Oracle

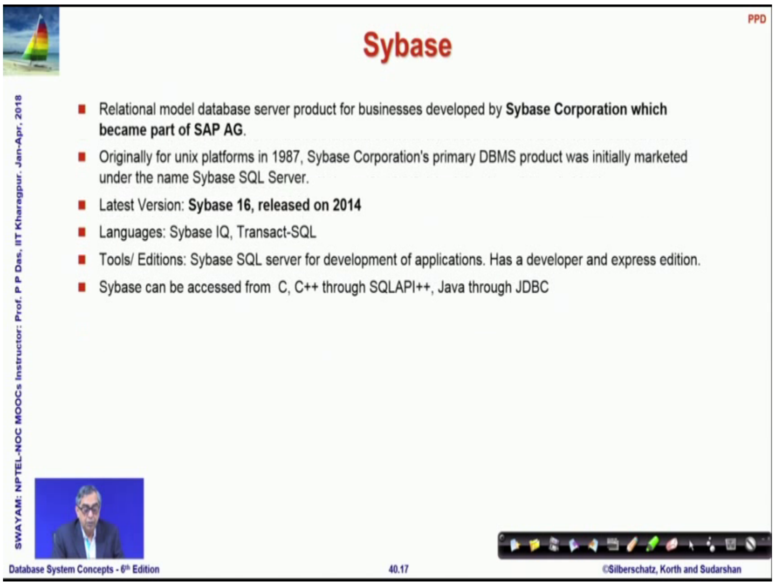
- Multi-model commercial database management system produced and marketed by **Oracle Corporation**.
- Larry Ellison, Bob Miner and Ed Oates started a consultancy called Software Development Laboratories (SDL) in 1977, and developed the original version of Oracle.
- Latest Version: **Oracle Database 12*c* Release 2: 12.2.0.1 (patchset as of March 2017)**
- Used for running online transaction processing (OLTP), data warehousing (DW) and mixed (OLTP & DW) database workloads
- Languages: Structured Query language (SQL), Procedural SQL (PL- SQL)
- Tools/ Editions: Oracle SQL Developer, Oracle Forms, Oracle Jdeveloper, Oracle Reports for development of applications, Oracle Live SQL for test environment
- Oracle can be accessed from Java through JDBC, Microsoft.NET through ODP.NET, C, C++ through OCI, ODBC, ODPI-C, Python through cx_Oracle

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.16 ©Silberschatz, Korth and Sudarshan

So that you know you know how stable, how old or you know what are the basic nuances of that database system for example, Oracle started in 77. So, you can say, it is a it is a 40 year old database system. The latest version is 12 C and these are the different supports that it has.

(Refer Slide Time: 08:13)



PPD

Sybase

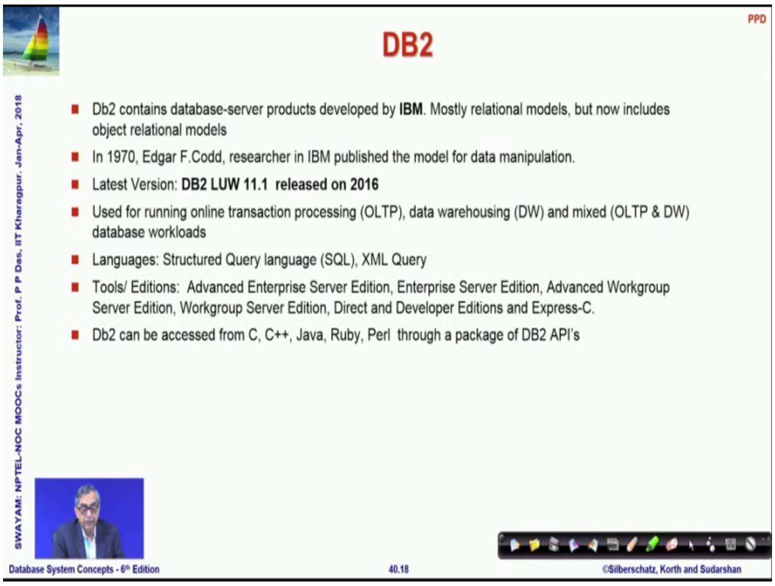
- Relational model database server product for businesses developed by **Sybase Corporation which became part of SAP AG**.
- Originally for unix platforms in 1987, Sybase Corporation's primary DBMS product was initially marketed under the name Sybase SQL Server.
- Latest Version: **Sybase 16, released on 2014**
- Languages: Sybase IQ, Transact-SQL
- Tools/ Editions: Sybase SQL server for development of applications. Has a developer and express edition.
- Sybase can be accessed from C, C++ through SQLAPI++, Java through JDBC

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.17 ©Silberschatz, Korth and Sudarshan

Sybase also started in 1987. So, that is almost 30 years, but it is less you know less vibrant right now, the last stable released happen in 2014 about nearly more than 3 and a half or 4 years ago and, but Sybase is a has been a very good database systems for programming through API’s and it has really good support for that.

(Refer Slide Time: 08:42)



PPD

DB2

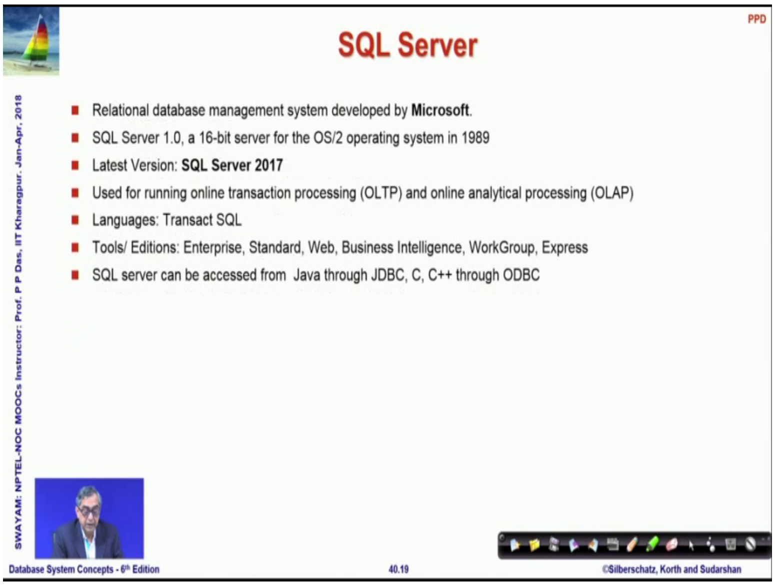
- Db2 contains database-server products developed by **IBM**. Mostly relational models, but now includes object relational models
- In 1970, Edgar F.Codd, researcher in IBM published the model for data manipulation.
- Latest Version: **DB2 LUW 11.1 released on 2016**
- Used for running online transaction processing (OLTP), data warehousing (DW) and mixed (OLTP & DW) database workloads
- Languages: Structured Query language (SQL), XML Query
- Tools/ Editions: Advanced Enterprise Server Edition, Enterprise Server Edition, Advanced Workgroup Server Edition, Workgroup Server Edition, Direct and Developer Editions and Express-C.
- Db2 can be accessed from C, C++, Java, Ruby, Perl through a package of DB2 API’s

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.18 ©Silberschatz, Korth and Sudarshan

DB 2 is also a very old possibly the oldest surviving database systems which started in 1970. So, when almost the E. F Codd of the Boyce Codd normal form published the data manipulation schemes from IBM. So, this is also a widely used, last release 2016.

(Refer Slide Time: 09:07)



PPD

SQL Server

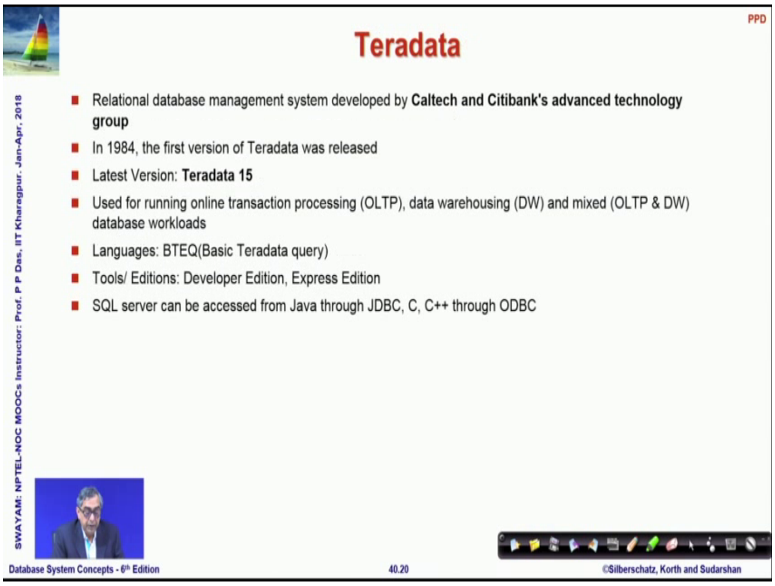
- Relational database management system developed by **Microsoft**.
- SQL Server 1.0, a 16-bit server for the OS/2 operating system in 1989
- Latest Version: **SQL Server 2017**
- Used for running online transaction processing (OLTP) and online analytical processing (OLAP)
- Languages: Transact SQL
- Tools/ Editions: Enterprise, Standard, Web, Business Intelligence, WorkGroup, Express
- SQL server can be accessed from Java through JDBC, C, C++ through ODBC

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.19 ©Silberschatz, Korth and Sudarshan

Microsoft the started database systems in 1989, last release happened last year and that is very widely used if you are particularly on windows system, it is one of the most popular one in terms of the windows operating system.

(Refer Slide Time: 09:24)



PPD

Teradata

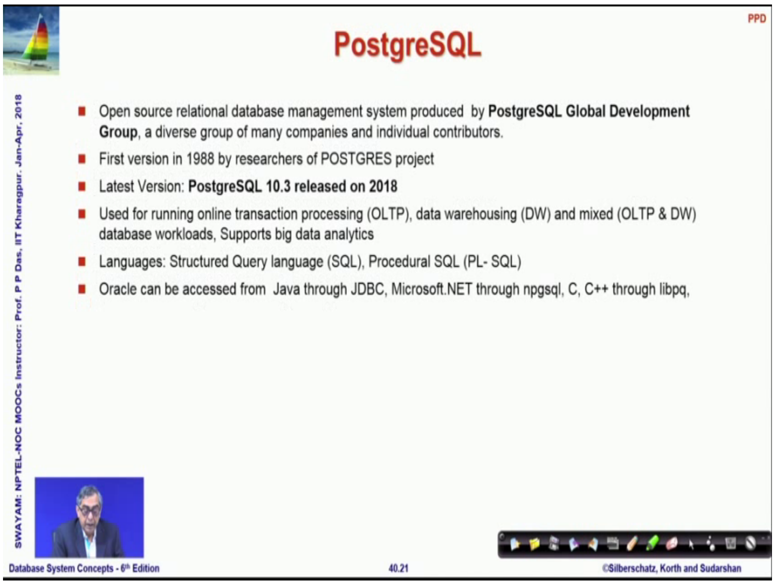
- Relational database management system developed by **Caltech and Citibank's advanced technology group**
- In 1984, the first version of Teradata was released
- Latest Version: **Teradata 15**
- Used for running online transaction processing (OLTP), data warehousing (DW) and mixed (OLTP & DW) database workloads
- Languages: BTEQ(Basic Teradata query)
- Tools/ Editions: Developer Edition, Express Edition
- SQL server can be accessed from Java through JDBC, C, C++ through ODBC

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.20 ©Silberschatz, Korth and Sudarshan

Teradata is relatively new. It was released in 1984 and it is, but it is it is one where lot of new developments are still happening and new experiments keep on happening and the current version is a Teradata 15.

(Refer Slide Time: 09:43)



PPD

PostgreSQL

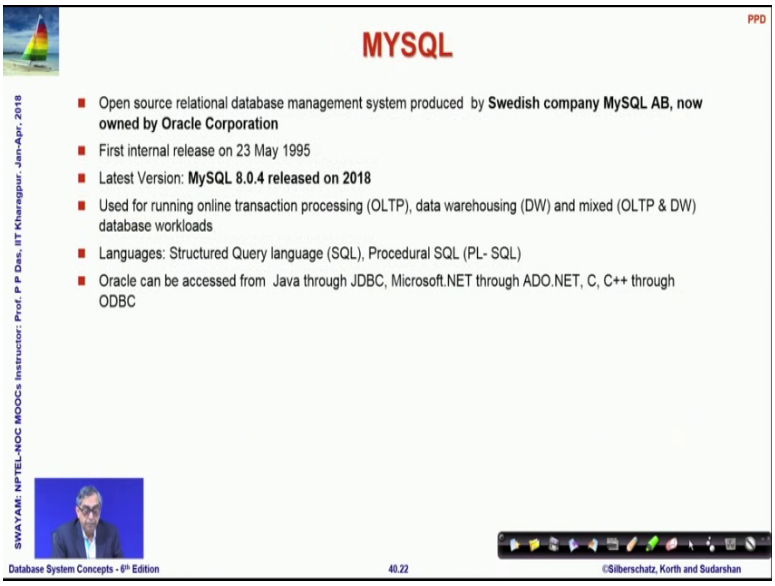
- Open source relational database management system produced by **PostgreSQL Global Development Group**, a diverse group of many companies and individual contributors.
- First version in 1988 by researchers of POSTGRES project
- Latest Version: **PostgreSQL 10.3 released on 2018**
- Used for running online transaction processing (OLTP), data warehousing (DW) and mixed (OLTP & DW) database workloads, Supports big data analytics
- Languages: Structured Query language (SQL), Procedural SQL (PL- SQL)
- Oracle can be accessed from Java through JDBC, Microsoft.NET through npgsql, C, C++ through libpq,

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.21 ©Silberschatz, Korth and Sudarshan

And in terms of the ma free or open source GPL databases Postgres started in 1988 about 30 years back and as I said, this is this is has a release even half of this year.

(Refer Slide Time: 10:04)



PPD

MYSQL

- Open source relational database management system produced by **Swedish company MySQL AB, now owned by Oracle Corporation**
- First internal release on 23 May 1995
- Latest Version: **MySQL 8.0.4 released on 2018**
- Used for running online transaction processing (OLTP), data warehousing (DW) and mixed (OLTP & DW) database workloads
- Languages: Structured Query language (SQL), Procedural SQL (PL- SQL)
- Oracle can be accessed from Java through JDBC, Microsoft.NET through ADO.NET, C, C++ through ODBC

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

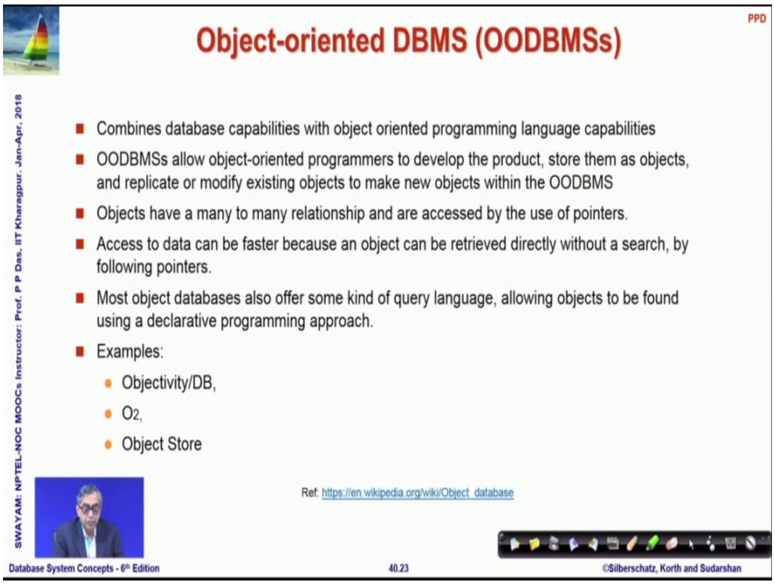
Database System Concepts - 6th Edition

40.22

©Silberschatz, Korth and Sudarshan

So, it is a very vibrant system. MySQL probably most widely used amongst the free community among the open source community also where the first internal release happened in 1995. The recent releases happened this year. So, these are the common database systems that you will come across. So, I mean given the organization that you are working with, first find out which database system it uses and then look into the specific manual for that and specific features.

(Refer Slide Time: 10:34)



PPD

Object-oriented DBMS (OODBMSs)

- Combines database capabilities with object oriented programming language capabilities
- OODBMSs allow object-oriented programmers to develop the product, store them as objects, and replicate or modify existing objects to make new objects within the OODBMS
- Objects have a many to many relationship and are accessed by the use of pointers.
- Access to data can be faster because an object can be retrieved directly without a search, by following pointers.
- Most object databases also offer some kind of query language, allowing objects to be found using a declarative programming approach.
- Examples:
  - Objectivity/DB,
  - O2,
  - Object Store

Ref: https://en.wikipedia.org/wiki/Object_database

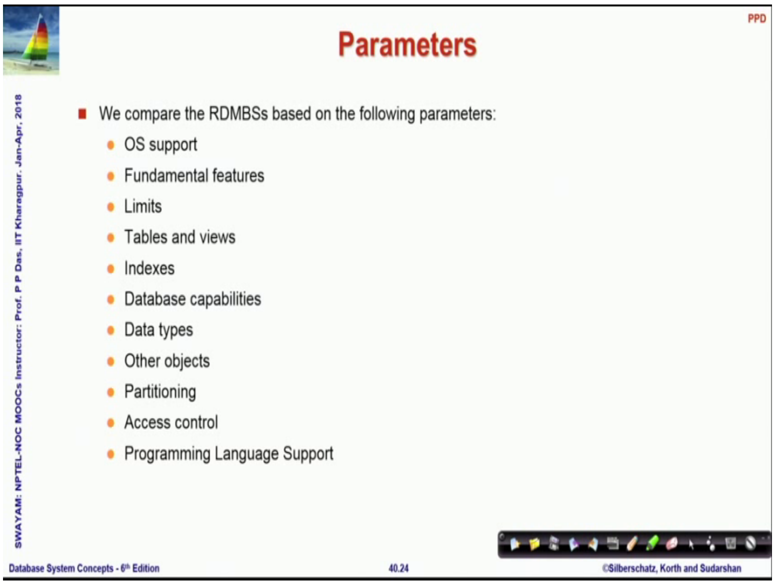
SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.23 ©Silberschatz, Korth and Sudarshan

Beyond these a Relational Database Systems also, there are certain database systems which use Object oriented notions in that. So, if you are familiar with object orientation then you would have understood that the relational approach does not make keep things object oriented because you are always flattening out in terms of attributes and you are trying to look at the attributes, but it went for example, when you want to model the same thing in terms of a C plus plus or java program, you would like to look at a course as an as an object, as a class you would like to look at instructor as a class, you would like to look at teaches as a as a kind of class and their instances. So, there has been attempts to make give a object orientation kind of layer on top of relational databases or define things in that way. Objectivity DBO 2 objects store are some of the examples, but unfortunately this is have not been as popular as a regular relational databases.

So, if you happen to use any one of them, then you should be cause a cautious that you know you really know why you are using it and you would be able to go a long way in terms of that.

(Refer Slide Time: 11:53)



PPD

Parameters

- We compare the RDMBSs based on the following parameters:
  - OS support
  - Fundamental features
  - Limits
  - Tables and views
  - Indexes
  - Database capabilities
  - Data types
  - Other objects
  - Partitioning
  - Access control
  - Programming Language Support

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.24 ©Silberschatz, Korth and Sudarshan

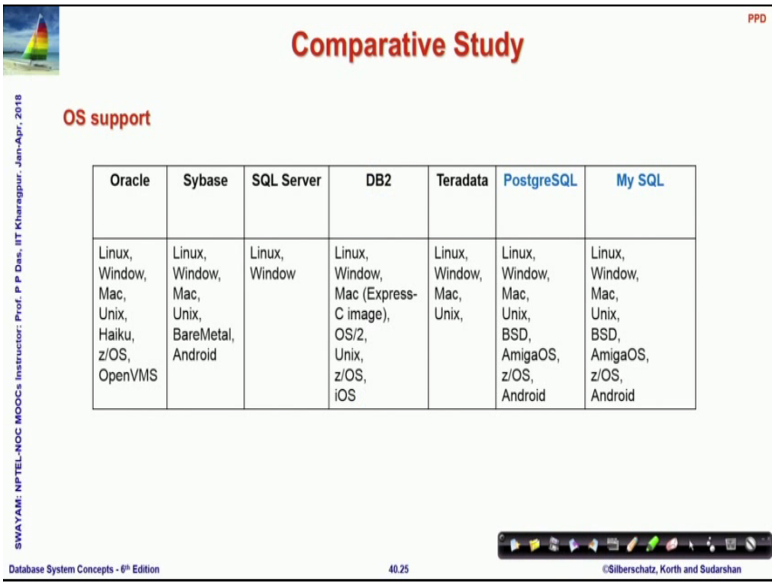
Now, when you come across a particular system that your company or your university needs to use and you would like to choose, then it will be good to look at the different aspects of that system. These are here are some of the parameters on which these database systems vary in a in a minimal to a very large extent, in terms of what operating systems it supports, what are the fundamental features, what are the limits for example, every database sets a number of limits in terms of the index size, in terms of the table size and whole lot of that.

How are the tables and views created what the kind of restrictions you have that, what kind of indexes has support the capabilities the data types, the different databases support. We have talked about a very limited set of data types in terms of SQL, but in an actual commercial or even you know open source database, the data types could be wider than that what kind of other objects partitioning access control mechanism. Access control is very important for ensuring security and finally, what kind of programming language support do you have.

Because as we have seen in the application development module, that it is not enough to just have a you know the database firing SQL queries, no application user will actually fire SQL queries. The application user needs and in GUI possibly or a text interface through which it will put queries in a different form and that needs to be processed by taking it to the database engine. So, you need possibly an interface which is in terms of

C, C plus plus, Java, Python, this kind of programming language. So, how do you connect to or embed such embed your relational query into different languages that differ between different database systems. So, these are the parameters that you must look at. In the next series of slides, which I will not you know discuss really because the these are more like data.

(Refer Slide Time: 14:03)



PPD

Comparative Study

OS support

Oracle	Sybase	SQL Server	DB2	Teradata	PostgreSQL	My SQL
Linux, Window, Mac, Unix, Haiku, z/OS, OpenVMS	Linux, Window, Mac, Unix, BareMetal, Android	Linux, Window	Linux, Window, Mac (Express-C image), OS/2, Unix, z/OS, iOS	Linux, Window, Mac, Unix,	Linux, Window, Mac, Unix, BSD, AmigaOS, z/OS, Android	Linux, Window, Mac, Unix, BSD, AmigaOS, z/OS, Android

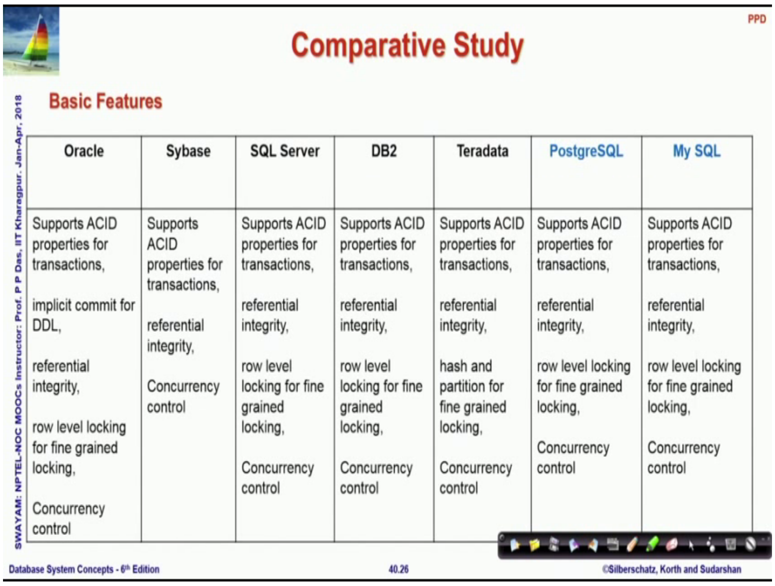
SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.25 ©Silberschatz, Korth and Sudarshan

But here after given a compilation of different you know on different aspects, how do these common database RDMS systems agree or differ. So, this is like a slide which shows what are the operating system support for different databases. So, if you are for example, working on android, then you can easily make out that you do not have a choice to use SQL server or to use Oracle, but you can use Sybase.

But you can use Postgres and MySQL actually, if you look into these two columns, right most columns which are for the open source databases, you will find that they have the widest choice in terms of operating system. In many aspects you will find that these free database systems have a better you know options for you; obviously, when it comes to you know really the core, core of database systems in terms of really really supporting very large databases, really really supporting very fast operations, really really supporting very secure applications, you might need to only work with commercial software because they offer that, but otherwise for a large number of common you know medium scale or low cost applications the free database systems, Postgres and MySQL are really good options there are different such features.

(Refer Slide Time: 15:23)



PPD

Comparative Study

Basic Features

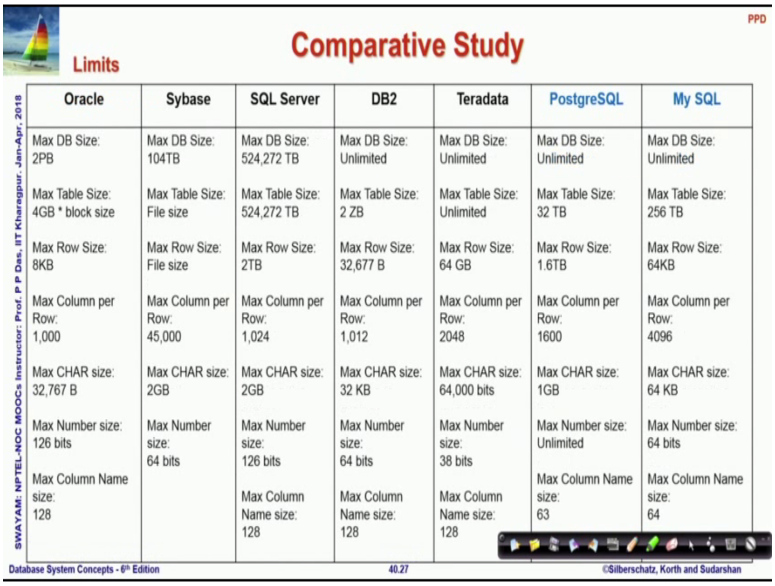
Oracle	Sybase	SQL Server	DB2	Teradata	PostgreSQL	My SQL
Supports ACID properties for transactions, implicit commit for DDL, referential integrity, row level locking for fine grained locking, Concurrency control	Supports ACID properties for transactions, referential integrity, Concurrency control	Supports ACID properties for transactions, referential integrity, row level locking for fine grained locking, Concurrency control	Supports ACID properties for transactions, referential integrity, row level locking for fine grained locking, Concurrency control	Supports ACID properties for transactions, referential integrity, hash and partition for fine grained locking, Concurrency control	Supports ACID properties for transactions, referential integrity, row level locking for fine grained locking, Concurrency control	Supports ACID properties for transactions, referential integrity, row level locking for fine grained locking, Concurrency control

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.26 ©Silberschatz, Korth and Sudarshan

The basic features are compared to here.

(Refer Slide Time: 15:27)



PPD

Comparative Study

Limits

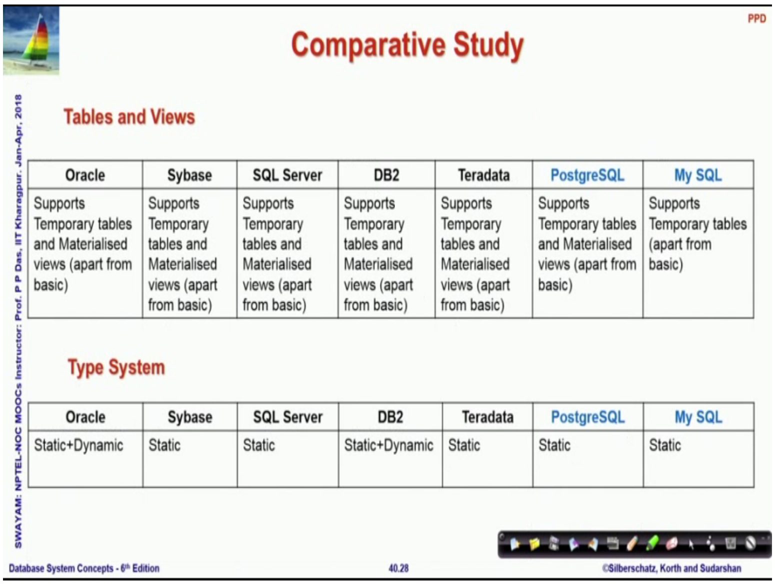
Oracle	Sybase	SQL Server	DB2	Teradata	PostgreSQL	My SQL
Max DB Size: 2PB	Max DB Size: 104TB	Max DB Size: 524,272 TB	Max DB Size: Unlimited	Max DB Size: Unlimited	Max DB Size: Unlimited	Max DB Size: Unlimited
Max Table Size: 4GB * block size	Max Table Size: File size	Max Table Size: 524,272 TB	Max Table Size: 2 ZB	Max Table Size: Unlimited	Max Table Size: 32 TB	Max Table Size: 256 TB
Max Row Size: 8KB	Max Row Size: File size	Max Row Size: 2TB	Max Row Size: 32,677 B	Max Row Size: 64 GB	Max Row Size: 1.6TB	Max Row Size: 64KB
Max Column per Row: 1,000	Max Column per Row: 45,000	Max Column per Row: 1,024	Max Column per Row: 1,012	Max Column per Row: 2048	Max Column per Row: 1600	Max Column per Row: 4096
Max CHAR size: 32,767 B	Max CHAR size: 2GB	Max CHAR size: 2GB	Max CHAR size: 32 KB	Max CHAR size: 64,000 bits	Max CHAR size: 1GB	Max CHAR size: 64 KB
Max Number size: 126 bits	Max Number size: 64 bits	Max Number size: 126 bits	Max Number size: 64 bits	Max Number size: 38 bits	Max Number size: Unlimited	Max Number size: 64 bits
Max Column Name size: 128		Max Column Name size: 128	Max Column Name size: 128	Max Column Name size: 128	Max Column Name size: 63	Max Column Name size: 64

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.27 ©Silberschatz, Korth and Sudarshan

At compile different limits of different database systems here. So, you can see in terms of maximum row size, columns per row and so on and so forth, how do they differ. So, if you are making a choice in terms of what database I will use.

(Refer Slide Time: 15:42)



PPD

Comparative Study

Tables and Views

Oracle	Sybase	SQL Server	DB2	Teradata	PostgreSQL	My SQL
Supports Temporary tables and Materialised views (apart from basic)	Supports Temporary tables and Materialised views (apart from basic)	Supports Temporary tables and Materialised views (apart from basic)	Supports Temporary tables and Materialised views (apart from basic)	Supports Temporary tables and Materialised views (apart from basic)	Supports Temporary tables and Materialised views (apart from basic)	Supports Temporary tables (apart from basic)

Type System

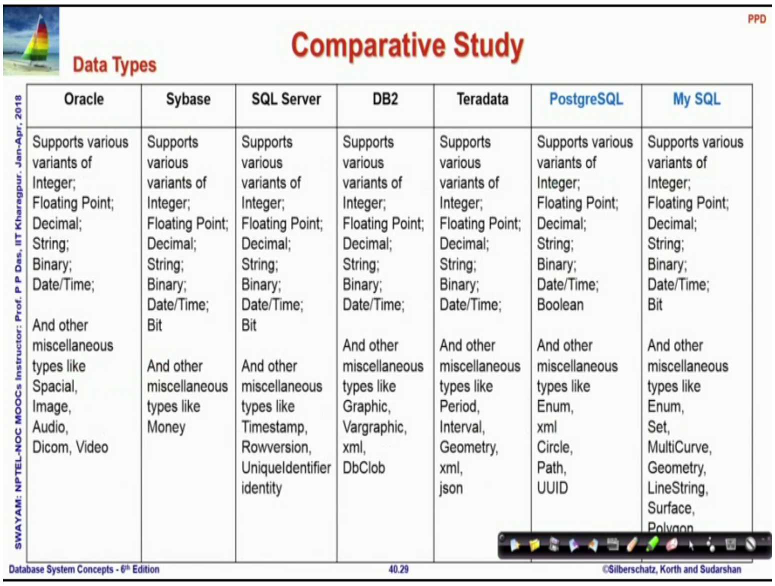
Oracle	Sybase	SQL Server	DB2	Teradata	PostgreSQL	My SQL
Static+Dynamic	Static	Static	Static+Dynamic	Static	Static	Static

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.28 ©Silberschatz, Korth and Sudarshan

You can use these information. This talks about tables we use the type systems, what kind of typing is used.

(Refer Slide Time: 15:48)



PPD

Comparative Study

Data Types

Oracle	Sybase	SQL Server	DB2	Teradata	PostgreSQL	My SQL

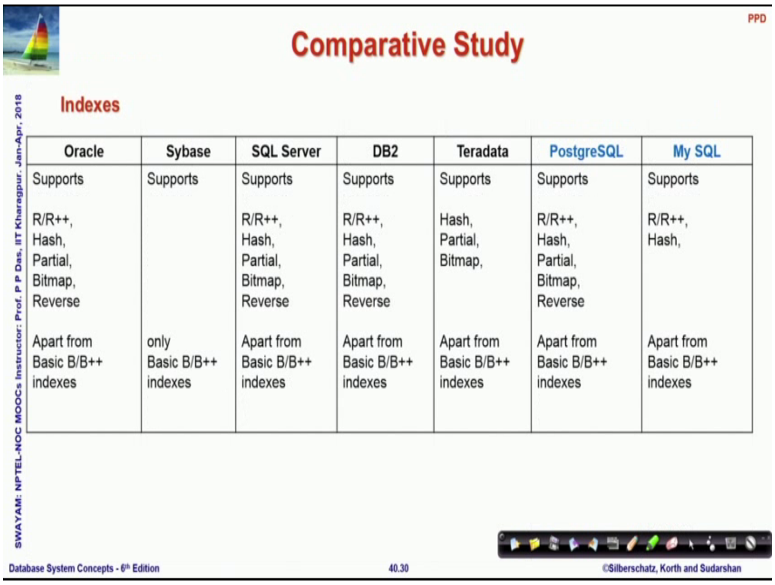
Supports various variants of Integer;
Floating Point;
Decimal;
String;
Binary;
Date/Time;
And other miscellaneous types like Spacial, Image, Audio, Dicom, Video
	Supports various variants of Integer;
	Floating Point;
	Decimal;
	String;
	Binary;
	Date/Time;
	Bit
	And other miscellaneous types like Money
		Supports various variants of Integer;
		Floating Point;
		Decimal;
		String;
		Binary;
		Date/Time;
		Bit
		And other miscellaneous types like Timestamp, Rowversion, Uniqueldentifier identity
			Supports various variants of Integer;
			Floating Point;
			Decimal;
			String;
			Binary;
			Date/Time;
			And other miscellaneous types like Graphic, Vargraphic, xml, DbClob
				Supports various variants of Integer;
				Floating Point;
				Decimal;
				String;
				Binary;
				Date/Time;
				And other miscellaneous types like Period, Interval, Geometry, xml, json
					Supports various variants of Integer;
					Floating Point;
					Decimal;
					String;
					Binary;
					Date/Time;
					Boolean
					And other miscellaneous types like Enum, xml Circle, Path, UUID
						Supports various variants of Integer;
						Floating Point;
						Decimal;
						String;
						Binary;
						Date/Time;
						Bit
						And other miscellaneous types like Enum, Set, MultiCurve, Geometry, LineString, Surface, Polygon

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.29 ©Silberschatz, Korth and Sudarshan

The different data types that are used are given here.

(Refer Slide Time: 15:52)



PPD

Comparative Study

Indexes

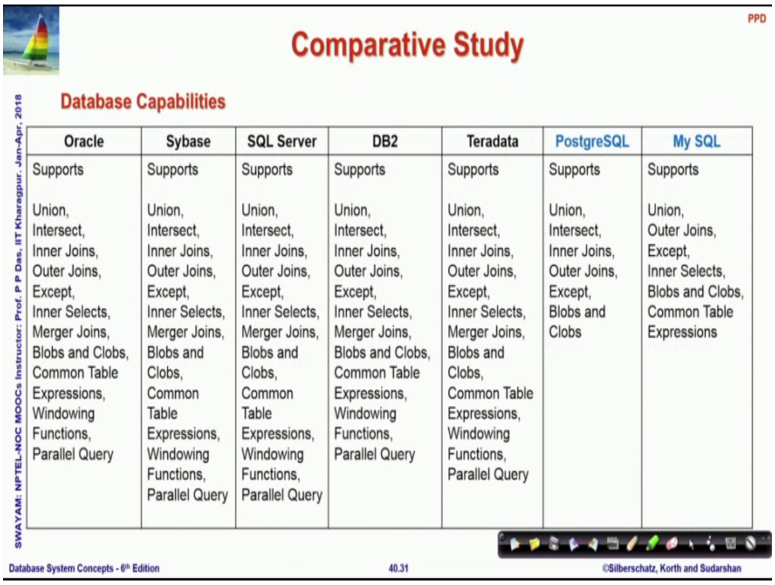
Oracle	Sybase	SQL Server	DB2	Teradata	PostgreSQL	My SQL
Supports R/R++, Hash, Partial, Bitmap, Reverse Apart from Basic B/B++ indexes	Supports only Basic B/B++ indexes	Supports R/R++, Hash, Partial, Bitmap, Reverse Apart from Basic B/B++ indexes	Supports R/R++, Hash, Partial, Bitmap, Reverse Apart from Basic B/B++ indexes	Supports Hash, Partial, Bitmap, Apart from Basic B/B++ indexes	Supports R/R++, Hash, Partial, Bitmap, Reverse Apart from Basic B/B++ indexes	Supports R/R++, Hash, Apart from Basic B/B++ indexes

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.30 ©Silberschatz, Korth and Sudarshan

The index mechanisms are discussed here.

(Refer Slide Time: 15:57)



PPD

Comparative Study

Database Capabilities

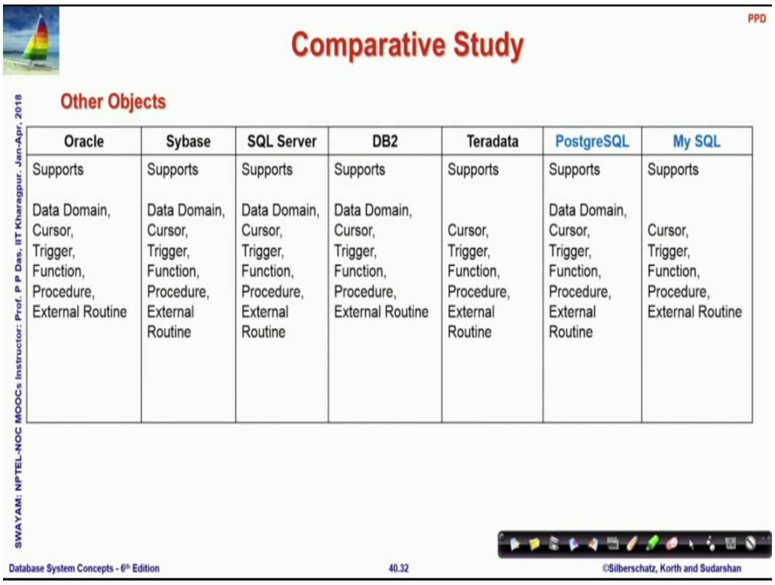
Oracle	Sybase	SQL Server	DB2	Teradata	PostgreSQL	My SQL
Supports Union, Intersect, Inner Joins, Outer Joins, Except, Inner Selects, Merger Joins, Blobs and Clobs, Common Table Expressions, Windowing Functions, Parallel Query	Supports Union, Intersect, Inner Joins, Outer Joins, Except, Inner Selects, Merger Joins, Blobs and Clobs, Common Table Expressions, Windowing Functions, Parallel Query	Supports Union, Intersect, Inner Joins, Outer Joins, Except, Inner Selects, Merger Joins, Blobs and Clobs, Common Table Expressions, Windowing Functions, Parallel Query	Supports Union, Intersect, Inner Joins, Outer Joins, Except, Inner Selects, Merger Joins, Blobs and Clobs, Common Table Expressions, Windowing Functions, Parallel Query	Supports Union, Intersect, Inner Joins, Outer Joins, Except, Inner Selects, Merger Joins, Blobs and Clobs, Common Table Expressions, Windowing Functions, Parallel Query	Supports Union, Intersect, Inner Joins, Outer Joins, Except, Blobs and Clobs	Supports Union, Outer Joins, Except, Inner Selects, Blobs and Clobs, Common Table Expressions

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.31 ©Silberschatz, Korth and Sudarshan

The capabilities of the database, what it can do overall.

(Refer Slide Time: 16:01)



PPD

Comparative Study

Other Objects

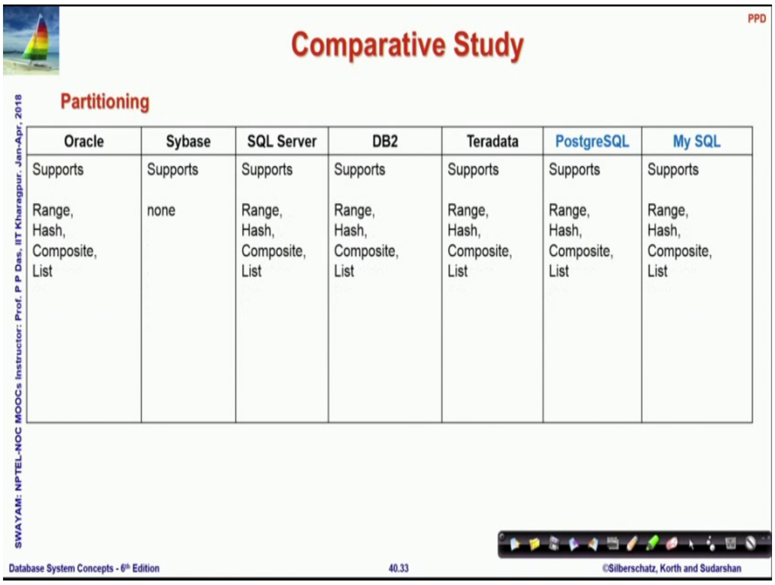
Oracle	Sybase	SQL Server	DB2	Teradata	PostgreSQL	My SQL
Supports	Supports	Supports	Supports	Supports	Supports	Supports
Data Domain, Cursor, Trigger, Function, Procedure, External Routine	Data Domain, Cursor, Trigger, Function, Procedure, External Routine	Data Domain, Cursor, Trigger, Function, Procedure, External Routine	Data Domain, Cursor, Trigger, Function, Procedure, External Routine	Cursor, Trigger, Function, Procedure, External Routine	Data Domain, Cursor, Trigger, Function, Procedure, External Routine	Cursor, Trigger, Function, Procedure, External Routine

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.32 ©Silberschatz, Korth and Sudarshan

Other kinds of objects that it supports.

(Refer Slide Time: 16:05)



PPD

Comparative Study

Partitioning

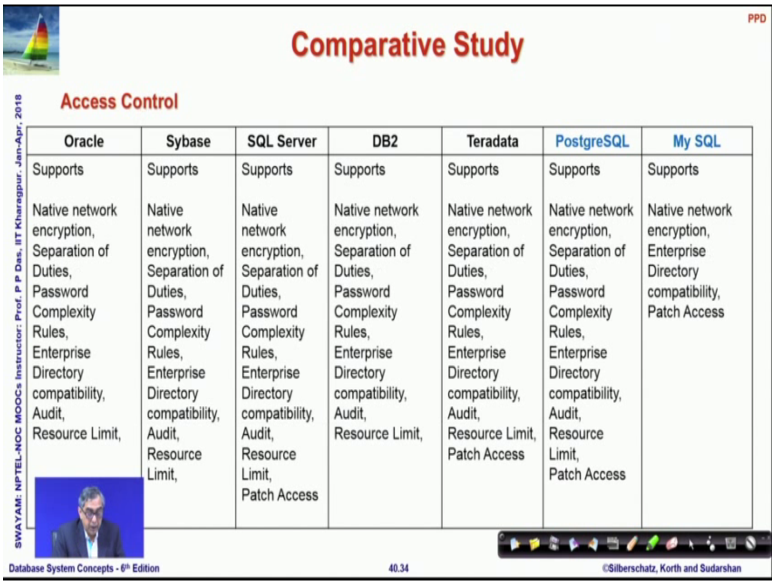
Oracle	Sybase	SQL Server	DB2	Teradata	PostgreSQL	My SQL
Supports	Supports	Supports	Supports	Supports	Supports	Supports
Range, Hash, Composite, List	none	Range, Hash, Composite, List	Range, Hash, Composite, List	Range, Hash, Composite, List	Range, Hash, Composite, List	Range, Hash, Composite, List

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.33 ©Silberschatz, Korth and Sudarshan

The different partitioning mechanism.

(Refer Slide Time: 16:08)



PPD

Comparative Study

Access Control

Oracle	Sybase	SQL Server	DB2	Teradata	PostgreSQL	My SQL
Supports Native network encryption, Separation of Duties, Password Complexity Rules, Enterprise Directory compatibility, Audit, Resource Limit,	Supports Native network encryption, Separation of Duties, Password Complexity Rules, Enterprise Directory compatibility, Audit, Resource Limit,	Supports Native network encryption, Separation of Duties, Password Complexity Rules, Enterprise Directory compatibility, Audit, Resource Limit, Patch Access	Supports Native network encryption, Separation of Duties, Password Complexity Rules, Enterprise Directory compatibility, Audit, Resource Limit,	Supports Native network encryption, Separation of Duties, Password Complexity Rules, Enterprise Directory compatibility, Audit, Resource Limit, Patch Access	Supports Native network encryption, Separation of Duties, Password Complexity Rules, Enterprise Directory compatibility, Audit, Resource Limit, Patch Access	Supports Native network encryption, Enterprise Directory compatibility, Patch Access

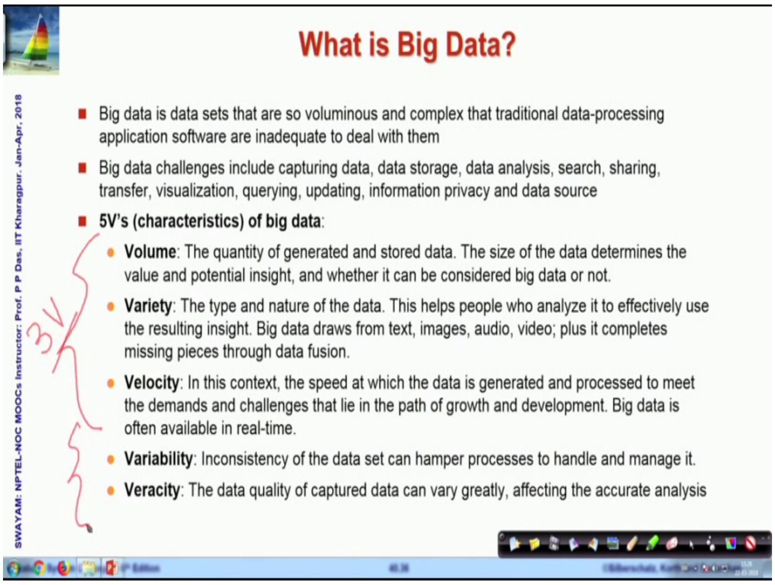
SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.34 ©Silberschatz, Korth and Sudarshan

The access control which is critical for ensuring good security and good management of that database are given here.

So, these are kind of the different across different features, this is parameters, this is how they compare. So, well this is for your you know reference only, this is this is not for your assignment or examination, but I just wanted to have a view that with all the theory and you know a small hands on that you have seen, when you go to the real life what are the expected things what are the expected system this will have to work with. Now let us just move on and let us let me just briefly talk about the a group of database systems which are non relational and of course I would warn you that the basis for these DBMS is are not covered in this course, is beyond the course, but just for your information and to keep in keep you in tune with what is happening frequently around the industry today.

(Refer Slide Time: 17:16)



What is Big Data?

- Big data is data sets that are so voluminous and complex that traditional data-processing application software are inadequate to deal with them
- Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source
- **5V's (characteristics) of big data**:
  - **Volume**: The quantity of generated and stored data. The size of the data determines the value and potential insight, and whether it can be considered big data or not.
  - **Variety**: The type and nature of the data. This helps people who analyze it to effectively use the resulting insight. Big data draws from text, images, audio, video; plus it completes missing pieces through data fusion.
  - **Velocity**: In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development. Big data is often available in real-time.
  - **Variability**: Inconsistency of the data set can hamper processes to handle and manage it.
  - **Veracity**: The data quality of captured data can vary greatly, affecting the accurate analysis

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

So, the non-relational database systems have arisen from what they all have must have heard of is the whole aspect of Big Data. Big Data as a name suggests is certainly voluminous data, complex data and now the question that you might have is if I i have done a good relational design, if I have a good RDMS, can I not handle big data. The question here really is, big data is not only about volume, the volume is only one aspect which is really large. So, big data typically are characterized by certain Vs.

So, these are not any very standardized characterization, but these are more commonly accepted once. So, one is volume, that the quantity of data when a for a big data situation has to be very very large. Now again what is a very very large is again a subjective question, some you can say that a million record is large, someone else would say no million record is small, it is actually 10 million is large; some might say that it is a database need to be petabytes to be large and so on.

But these are all subjectivity, but it is large has a voluminous existent in certain sense, there has to be different variety different types of data. So, all that we have seen in the relational database is basically your you know strings and numbers if you look at it in different ways we are seeing, whatever we have dealt with in all these through all this 39 modules so far, they are primarily about strings and numbers, but nothing else we have not, but big data can be about free text, it could be about natural language comments your regularly writing comments on your Facebook.

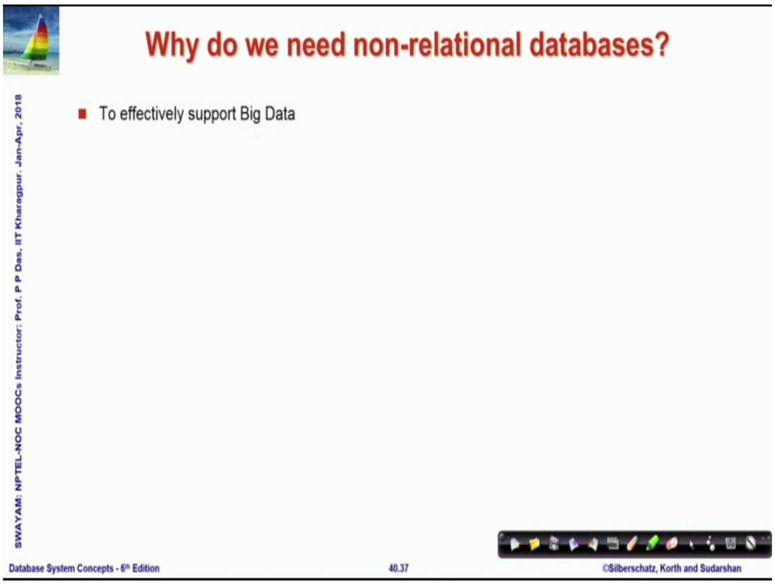
So, those Facebook comments are phrases and if I want to make certain query based on that, if I want to make a query that, how many Facebook users has commented on the success of Virat Kohli as a captain of India if I want to make such a query, then the question is how do I do that? Because it is not something where you have a at the information in a very structured way this is no there is no relational schema which says that well the values are put in terms of Virat Kohli having done very good, moderately or marginally or you know the captain Indian captains are successful, not successful and so on.

These are this happen in terms of various texts, phrases, clauses that we write. So, variety is a major issue then, it could have word your images video. So, big data includes all of that. The third V is about velocity that the processing speed may need to be really really fast, often in big data often we say that the processing has to be real time which means what is real time. Real time is basically that from the time I fire the query and to the time I get the result, there is a fixed time limit within which it has to happen.

So, if I if I if I really want to do a railway reservation that also is a kind of real time, but that is not very critical because it is if I get the reservation done in one minute, it is also ok, if it takes 5 minutes, it is good if we can happen in 10 seconds, but I do not ever need it in say 20 millisecond. So, but in when you talk about real time, it could really be about getting all these processing done in millisecond, microsecond, nanosecond and so on. And those kind of real time systems with a large volume of varied data, it is a big challenge.

So, those are the challenges of big data, then there could be variability inconsistency of the data that you are because maintaining integrity is a big problem; there could be issues in terms of quality of the data that is called veracity. So, actually these things characteristics that I have put, there are lot of debates in terms of that or many people take these 3 and say that there these are the 3 V s of big data. There is a 3 main characteristics, but off let more and more people are also considering variability and veracity are as the characteristics of big data. Now as it happens, is if you look into these requirements and what you have you have a fairly good idea of relational databases now what they can do, how to design them, how to query them, how to implement them, you will understand that it is not easy to meet these requirements using the relational model.

(Refer Slide Time: 22:07)



Why do we need non-relational databases?

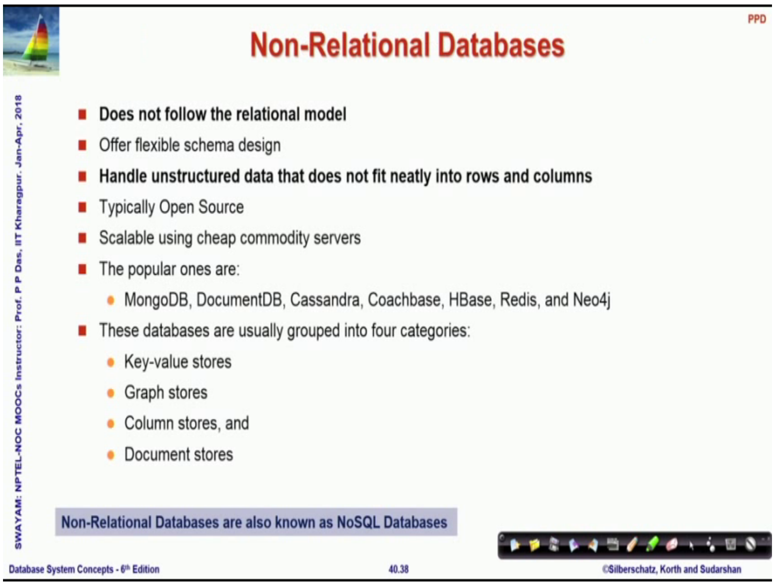
- To effectively support Big Data

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.37 ©Silberschatz, Korth and Sudarshan

So, we need their non-relational databases to effectively support big data. That is that is a one major reason that you need big data.

(Refer Slide Time: 22:16)



PPD

Non-Relational Databases

- **Does not follow the relational model**
- Offer flexible schema design
- **Handle unstructured data that does not fit neatly into rows and columns**
- Typically Open Source
- Scalable using cheap commodity servers
- The popular ones are:
  - MongoDB, DocumentDB, Cassandra, Coachbase, HBase, Redis, and Neo4j
- These databases are usually grouped into four categories:
  - Key-value stores
  - Graph stores
  - Column stores, and
  - Document stores

Non-Relational Databases are also known as NoSQL Databases

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

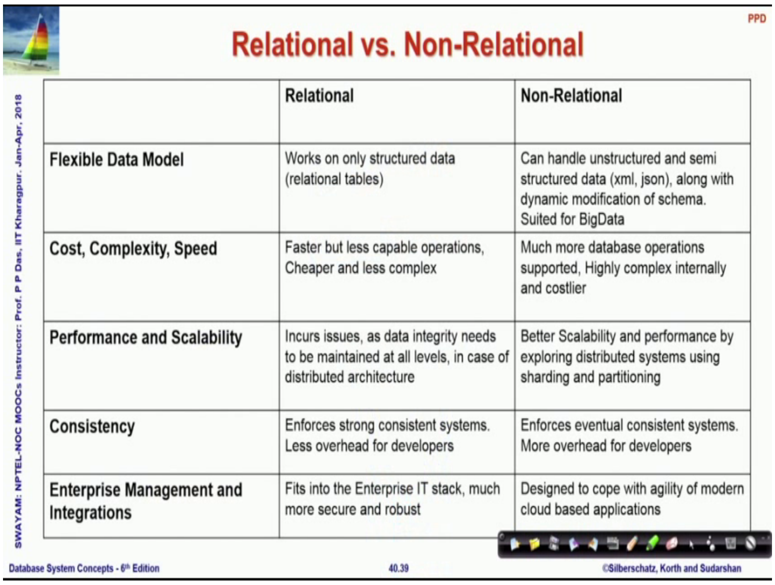
Database System Concepts - 6th Edition 40.38 ©Silberschatz, Korth and Sudarshan

So, that in non-relational databases certainly as the name suggests, do not follow relational model, they offer flexible schema design. The schema may itself change while the database is evolving which is not the case in the relational schema is fixed, only the data can change, but here the schema itself can change. It may be able to handle unstructured data, make natural language comments like images like audio coming in which do not fit nearly into your you know table structure, some of the other feature sites they are typically open source because you know still in an experimental stage and needs to be scalable and some of the popular ones are, these are the names you must have heard about at least some of them like MongoDB like Cassandra like HBase and so on.

Now again in terms of the non relational database, it does differ in terms of all non relational database error are not of the same type, there are there have been 4 different styles or strategies to actually generate realize these non relational databases, they are called key value store graph, store columns stores and document store. They are also known as no SQL databases. I, I personally find the name no SQL a little misnomer, it no SQL does not mean that you are strictly prohibited from not using SQL in these databases.

But I would rather like to read it more as no SQL means that it is not only SQL like in a relational database, you can use only SQL and solve problems; here you need to do lot of other things beyond that.

(Refer Slide Time: 24:00)



PPD

Relational vs. Non-Relational

	Relational	Non-Relational
Flexible Data Model	Works on only structured data (relational tables)	Can handle unstructured and semi structured data (xml, json), along with dynamic modification of schema. Suited for BigData
Cost, Complexity, Speed	Faster but less capable operations, Cheaper and less complex	Much more database operations supported, Highly complex internally and costlier
Performance and Scalability	Incurs issues, as data integrity needs to be maintained at all levels, in case of distributed architecture	Better Scalability and performance by exploring distributed systems using sharding and partitioning
Consistency	Enforces strong consistent systems. Less overhead for developers	Enforces eventual consistent systems. More overhead for developers
Enterprise Management and Integrations	Fits into the Enterprise IT stack, much more secure and robust	Designed to cope with agility of modern cloud based applications

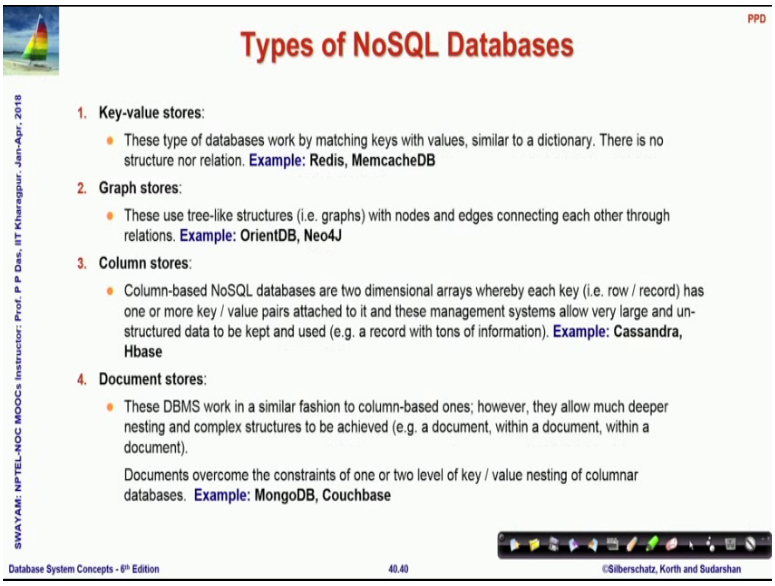
SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.39 ©Silberschatz, Korth and Sudarshan

So, here is a quick comparison between the relational and the non-relational, in terms of the flexibility of the data model, relational is very structured when on relational has to have handle unstructured data, semi structured data and therefore it has to be flexible in terms of the data model, cost complexity and speed faster less capable, but cheaper and less complex, but in non-relational, you are talking about much more database operations highly complex in internal structure usually costlier, performance and scalability certainly non-relational ones need to be better scalable, consistency have a very strict consistency rules in relation, but in non-relational you use some kind of you know eventual consistent system. So, maybe not always not everything is consistent in that way, enterprise management and integration, relational fits very well into because it is been around for as you have seen the little bit of history of all these common databases, it is more than 40 years that they have been around.

So, they easily fit into the IT stack whereas, non-relational is still on the in the in the agile form of development that is becoming more and more common it fits into the cloud based development and so on. So, these are some of the distinctions that exist.

(Refer Slide Time: 25:27)



PPD

Types of NoSQL Databases

1. **Key-value stores**:
   - These type of databases work by matching keys with values, similar to a dictionary. There is no structure nor relation. **Example: Redis, MemcacheDB**
2. **Graph stores**:
   - These use tree-like structures (i.e. graphs) with nodes and edges connecting each other through relations. **Example: OrientDB, Neo4J**
3. **Column stores**:
   - Column-based NoSQL databases are two dimensional arrays whereby each key (i.e. row / record) has one or more key / value pairs attached to it and these management systems allow very large and un-structured data to be kept and used (e.g. a record with tons of information). **Example: Cassandra, Hbase**
4. **Document stores**:
   - These DBMS work in a similar fashion to column-based ones; however, they allow much deeper nesting and complex structures to be achieved (e.g. a document, within a document, within a document).

     Documents overcome the constraints of one or two level of key / value nesting of columnar databases. **Example: MongoDB, Couchbase**

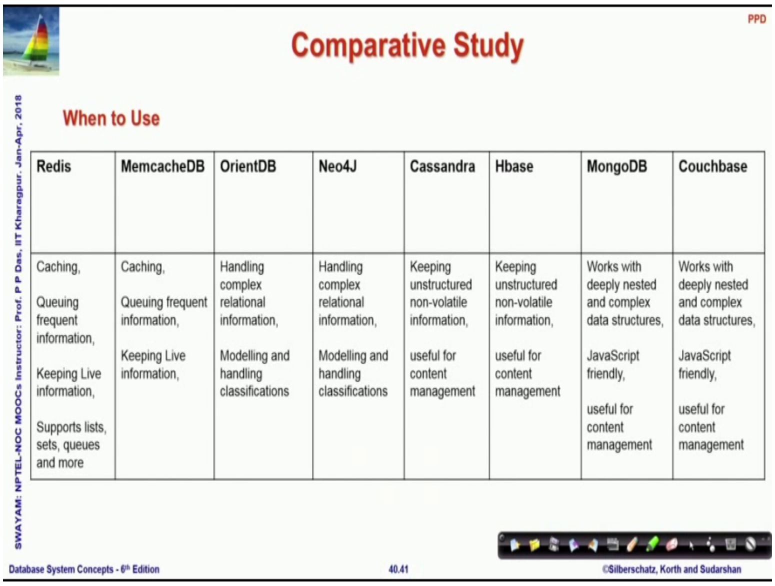
SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.40 ©Silberschatz, Korth and Sudarshan

And these are the different types of no SQL databases, there is a key value store strategy Redis and MemcacheDB follow this strategy, graph store is used by orient DB and Neo4J; column store is used by Cassandra and HBase document store, MongoDB, Couchbase, they use document store. So, these are I am in this is not just about going a deeper into what they are or how they are distinguished, I just want you to have an idea that well. These are different from the relational databases they can do lot of structured, handling of unstructured data they can actually use a scalability of volume a variety which is relationships cannot do and, but they have there are different principles for actually implementing them and there is a deeper.

So, if you are interested, you can take specific courses which deal with the big data and prepare yourself for the bigger challenges ahead.

(Refer Slide Time: 26:29)



PPD

Comparative Study

When to Use

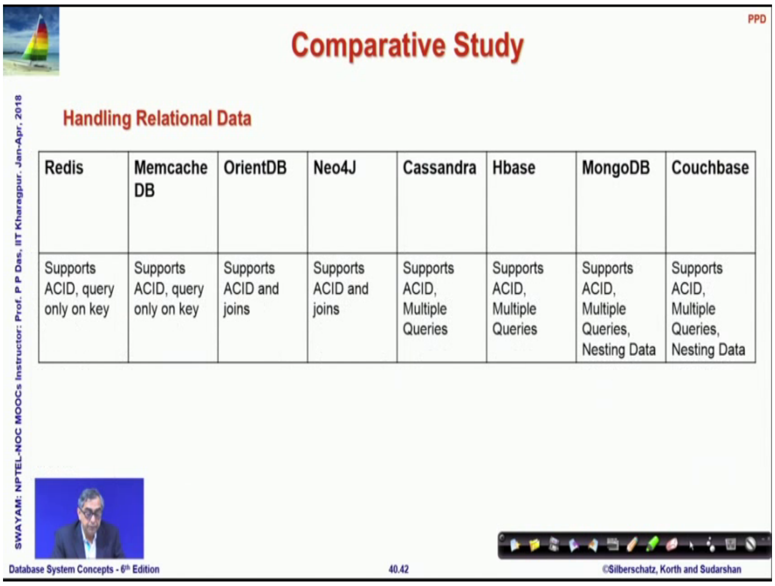
Redis	MemcacheDB	OrientDB	Neo4J	Cassandra	Hbase	MongoDB	Couchbase
Caching, Queuing frequent information, Keeping Live information, Supports lists, sets, queues and more	Caching, Queuing frequent information, Keeping Live information,	Handling complex relational information, Modelling and handling classifications	Handling complex relational information, Modelling and handling classifications	Keeping unstructured non-volatile information, useful for content management	Keeping unstructured non-volatile information, useful for content management	Works with deeply nested and complex data structures, JavaScript friendly, useful for content management	Works with deeply nested and complex data structures, JavaScript friendly, useful for content management

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.41 ©Silberschatz, Korth and Sudarshan

I have also done similar to the relational database, I have presented here a tentative comparative study between these different non no SQL databases in terms of what is the context in which you use them.

(Refer Slide Time: 26:46)



PPD

Comparative Study

Handling Relational Data

Redis	Memcache DB	OrientDB	Neo4J	Cassandra	Hbase	MongoDB	Couchbase
Supports ACID, query only on key	Supports ACID, query only on key	Supports ACID and joins	Supports ACID and joins	Supports ACID, Multiple Queries	Supports ACID, Multiple Queries	Supports ACID, Multiple Queries, Nesting Data	Supports ACID, Multiple Queries, Nesting Data

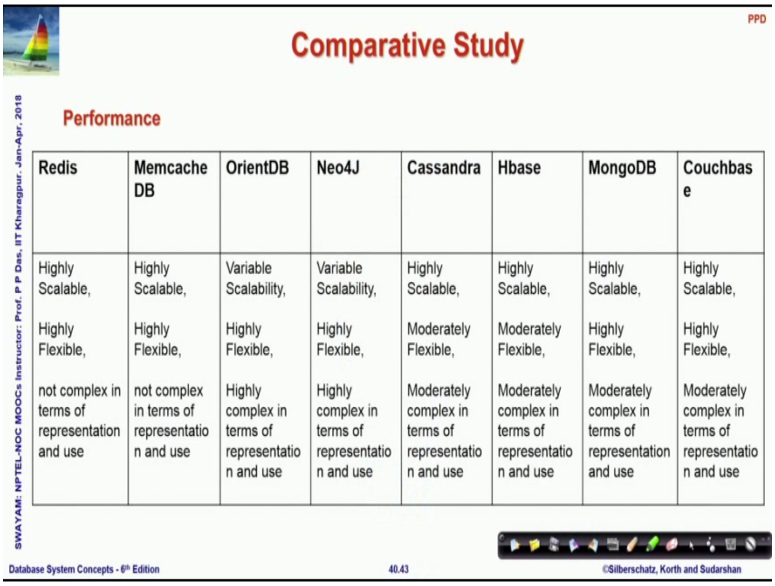
SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.42 ©Silberschatz, Korth and Sudarshan

Or now while you are doing this unstructured data handling, there will be lot of data which is also structured.

So, how do you, along with this know SQL, how do you handle the relational data with these?

(Refer Slide Time: 26:59)



PPD

Comparative Study

Performance

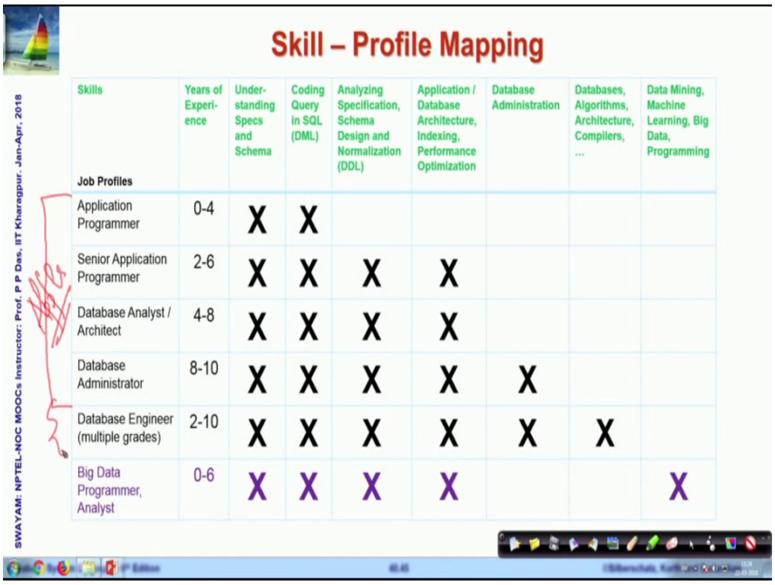
Redis	MemcacheDB	OrientDB	Neo4J	Cassandra	Hbase	MongoDB	Couchbase
Highly Scalable, Highly Flexible, not complex in terms of representation and use	Highly Scalable, Highly Flexible, not complex in terms of representation and use	Variable Scalability, Highly Flexible, Highly complex in terms of representation and use	Variable Scalability, Highly Flexible, Highly complex in terms of representation and use	Highly Scalable, Moderately Flexible, Moderately complex in terms of representation and use	Highly Scalable, Moderately Flexible, Moderately complex in terms of representation and use	Highly Scalable, Highly Flexible, Moderately complex in terms of representation and use	Highly Scalable, Highly Flexible, Moderately complex in terms of representation and use

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Database System Concepts - 6th Edition 40.43 ©Silberschatz, Korth and Sudarshan

Databases what how do how do the performance compare between these different no SQL databases and based on all that you can make some judgment and it is it is very important to in today’s time naturally knowing relational databases the foundational ones are very important, but it is always good to look forward be with the time and I will urge that if you have started growing interest in handling of data do take specific courses on big data and no SQL databases. I will end this discussion with a very simple skill job profile matrix which Will give you some idea in terms of, it will you can use it for a certain kind of self-assessment as well.

(Refer Slide Time: 27:37)



SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

Skill – Profile Mapping

Skills Job Profiles	Years of Experi-ence	Under-standing Specs and Schema	Coding Query in SQL (DML)	Analyzing Specification, Schema Design and Normalization (DDL)	Application / Database Architecture, Indexing, Performance Optimization	Database Administration	Databases, Algorithms, Architecture, Compilers, …	Data Mining, Machine Learning, Big Data, Programming
Application Programmer	0-4	X	X					
Senior Application Programmer	2-6	X	X	X	X			
Database Analyst / Architect	4-8	X	X	X	X			
Database Administrator	8-10	X	X	X	X	X		
Database Engineer (multiple grades)	2-10	X	X	X	X	X	X	
Big Data Programmer, Analyst	0-6	X	X	X	X			X

So, let me just explain the structure of this matrix, what I have tried to do is here I am sorry. Here on the left, I have given different typical job profiles this is if you look into LinkedIn, Naukri and all that you will find these kind of profiles being. So, then at the lowest level there are application programmers for which typically 0 to 4 years of experience are asked for.

Then the next level, this is so, this is your kind of your career progression also. If you if you choose to take up databases as your primary job profession, this next level is a senior application programmer which requires 2 to 6 years of experience depending on the organization and depending on your skills. Then you move on to database analyst or architect which you happen in 4 to 8 years of time and on a little different track because these are these are primarily in terms of application development and hierarchy on that and the other is an administrator track who actually administers the database in an organization, controls all the all that is happening in different database applications, typically 8 to 10 year's experience is required.

And some of that, so this these are the about actually in terms of you know profiles that are related to applications and this is a profile which is related to, if you really want to become a database engineer in a sense that you want to you know make changes in Oracle, you want to make changes in say MySQL, you want to make changes in say MongoDB or say that relation will say the Sybase.

So, if you want to become a database engineer who changes the database system itself or develops the database system itself, then this is the kind of background you will need. There is a kind of number of years, you would need and of course it is not a single grid there are multiple grades, you know junior and mid levels in here and those kind and last which have shown in different color are the whole set of profiles which relate to programming the big data, analyzing the big data and so on.

We are at present, it is companies are typically asking for 0 to 6 years of experience depending on actual skills that you have. On this side, I have shown a whole grouping of skills, this is the first basic level that you must understand specs and schema, without that you cannot do any of this and you must have a skill for coding in inquire in SQL, the DML part significantly, without that as you can see you cannot pick up any of these profiles whereas, if you go a little if you go little senior you know gaining experience and you should be able to analyze specs that is, you should be able to design schema do normalization and get into this.

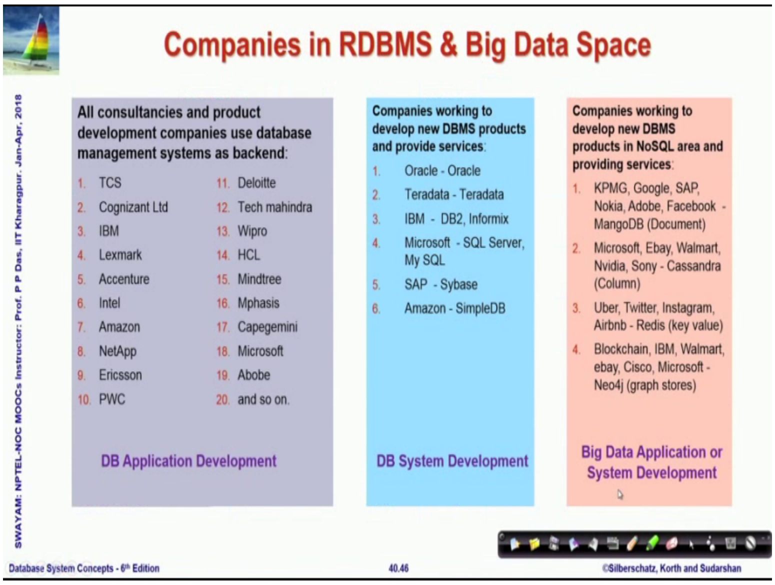
So, at a very initial level, you may not be expected to do all of that schema design and normalization by yourself, but it would be good to be able to do that, but well there will be seniors to help you, but if you once you become a senior application programmer that becomes onward that becomes a critical skill to have. Then the next level would be in terms of application or database architecture management deciding on how to index, performance optimization.

So, between these two, there will be certain overlaps a senior application programmer in addition to doing this might do some of these optimization techniques depending on how competent he or she is. Or some database architect may focus only on this, but these are the typical skills that you need. But to be a database administrator, you need all of these skills, but you are specifically administering a certain organizations enterprises whole database system. So, it is just not one database application, but a whole lot of databases and whole lot of user groups, security, network connectivity and all that.

So, that needs certainly bigger experience it can, you can see that experience level is much higher and the skill sets. If you want to become a database engineer, that is not focus only on the application side, but also have some more understanding in terms of actually doing working in the internals of the database systems, then you need whole lot of additional skills like good knowledge and algorithms, in architecture, in compiler all of that; only then and coupled with coupled with all the database knowledge, then you will be able to work as a database system engineer. And in the emerging areas of what is big data where, you need to have now of course, the I am saying this is 0 to 6, it could be 0 to 8 kind of, not more than that because it did not exist quite a long time ago, but you need to have a basic level of at least this much of the relational database understanding and knowledge, but what is critical is a whole set of other skills like, you must be aware with big data the data mining, warehousing strategies machine learning or is often very useful in this kind of big data applications, python programming, tensor flow all these become critical. You have to be a good programmer in any case I mean not only just an SQL program and you might have to be a good program and in C or C plus plus or python of these, but that is it that is a very very emerging area.

So, if you can acquire a little bit of besides database you know he said that the basics of the database along with that if you pick up few basics or from here, you will be able to enter into the space and that will give you a very very bright future in my view otherwise you can focus on the application programming stat as I have mentioned. So, this is the basically skill profile matrix that you have mapping that you have that you can focus on.

(Refer Slide Time: 33:49)



Companies in RDBMS & Big Data Space

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018

All consultancies and product development companies use database management systems as backend:

1. TCS
2. Cognizant Ltd
3. IBM
4. Lexmark
5. Accenture
6. Intel
7. Amazon
8. NetApp
9. Ericsson
10. PWC
11. Deloitte
12. Tech mahindra
13. Wipro
14. HCL
15. Mindtree
16. Mphasis
17. Capegemini
18. Microsoft
19. Abobe
20. and so on.

DB Application Development

Companies working to develop new DBMS products and provide services:

1. Oracle - Oracle
2. Teradata - Teradata
3. IBM - DB2, Informix
4. Microsoft - SQL Server, My SQL
5. SAP - Sybase
6. Amazon - SimpleDB

DB System Development

Companies working to develop new DBMS products in NoSQL area and providing services:

1. KPMG, Google, SAP, Nokia, Adobe, Facebook - MangoDB (Document)
2. Microsoft, Ebay, Walmart, Nvidia, Sony - Cassandra (Column)
3. Uber, Twitter, Instagram, Airbnb - Redis (key value)
4. Blockchain, IBM, Walmart, ebay, Cisco, Microsoft - Neo4j (graph stores)

Big Data Application or System Development

Database System Concepts - 6th Edition

40.46

©Silberschatz, Korth and Sudarshan

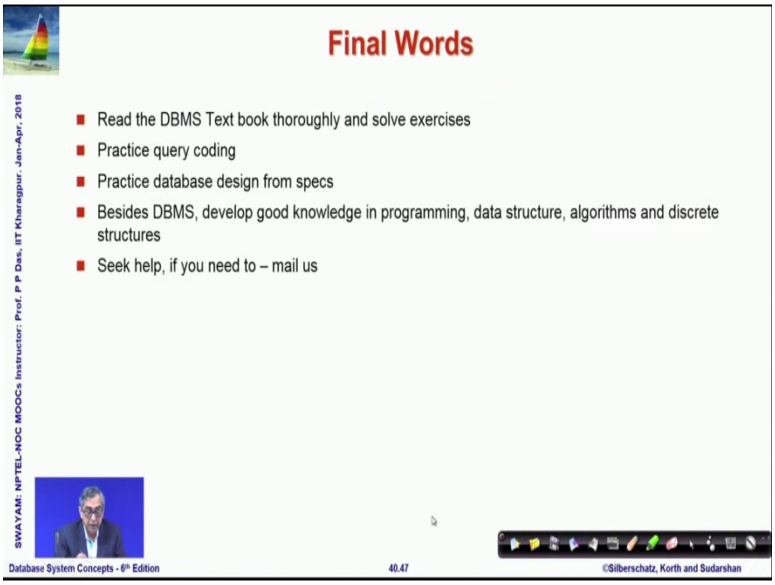
So, finally, before I close here a glimpses of companies that are in the very active in the RDMS space really really any big organization you talk about, they have consultancy projects, product development different database management back end services and so on. So, DB application development, I have listed some around 20 companies, but there are really 100s of them almost. Any big organization in any area you think of, they require databases. So, in terms of beta based application programmer and senior programmer and to some extent architect, you have a wide range of jobs available which you may just grab; if you have been able to study write the basics of the database. The second group of companies which I show here, these are system development companies who are actually working on the new DBMS products and services around that.

So, these are companies like Oracle, Teradata or Microsoft and so on, naturally these are big companies and you need more lot of more skills besides the database like I said algorithms programming and all that to crack a job here and here are some of some companies which I have mentioned, but there are many others who are focusing on the big data space.

So, I have tried to you know these may not be absolutely accurate because you know these are all collected from different sources, but these are the different companies and the kind of non relational database that they are focusing with working with. So, if you pick up certain skills in those in a certain non-relational no SQL database, then you can target the corresponding companies better or other companies and you can see that all. You know new generation companies, the companies were working for products for the next 10, 10, 15 years are in this space.

So, there are whole lot of opportunities for you all if you if you prepare a little hard, then you will I mean job will run after, you will not have to run after the job.

(Refer Slide Time: 35:56)



Final Words

- Read the DBMS Text book thoroughly and solve exercises
- Practice query coding
- Practice database design from specs
- Besides DBMS, develop good knowledge in programming, data structure, algorithms and discrete structures
- Seek help, if you need to – mail us

SWAYAM: NPTEL-NOC MOOCs Instructor: Prof. P P Das, IIT Kharagpur. Jan-Apr, 2018
Database System Concepts - 6th Edition
40.47
©Silberschatz, Korth and Sudarshan

So, with that I conclude this course a couple of final words the hygiene words. Read the DBMS textbook thoroughly and solve exercises. There is no shortcut to that, there is no other way to master the horse other than this you must practice query coding as much as you can, practice database design from specification. We are releasing a tutorial on this where for a hospital management system we are showing from the specification how you can do the initial schema and the refinements and finally, how can you implement it using my SQL.

So, do similar practices very heavily. Keep in mind the database the knowledge of database system alone will not be good enough to get a good job, get a good placement. So, develop good knowledge in programming data structure, algorithms and discrete structures; these are the minimum required around the database systems which will really make you powerful and if you need we are there to help you.

As long as the course is on, the forum would be on. You can post in the forum beyond that also if you need help, please ask for it mail us and wish you all the very best with your course in your examination and the future course of your profession in life, all the very best.