NYU TANDON SCHOOL OF ENGINEERING

Principles of Database Systems

Course Number: CSGY-6083
Section Number: B

Module 1

Instructor : Amit Patel
MS, PMP®, OCP®
Email: asp13@nyu.edu
patelamitnyu@gmail.com

Course Overview

- Introduction
- Basic Class Rules
- Syllabus
- Class Participation
- Individual Homework Assignments
- Group Project Work
- Mid-Term and Final Exams
- Software Installation

A) Must to have software

- Oracle SQL Data Modeler (for Database Design)
- Oracle SQL Live (Web based Oracle Database: No installation)

B) Need for project work

- Recommended: XAMPP (MySQL database, Apache, and PHP development environment)
- MySQL Workbench (For MySQL database interface)

[Student may choose any other RDBMS and programming language/framework that can be interfaced with database, e.g., Java, Python, Django]

C) Optional physical database

- Oracle Virtual Box and Oracle Database (Physical Local Database)
- Oracle SQL Developer (Client tool to interface with Database)

Assessment

Assessment Type	Description	Weight%
Participation	Ungraded assignments, forum discussions, meeting deadlines of assignment submission, class interactions	10%
Individual Assignment	Four problem sets, each with 5% weight	20%
Project Part I	For a given business case, Database design, development, and implementation	15%
Project Part II	Web application, interfaced with database	15%
Midterm Exam	Online Exam	20%
Final Exam	Online Exam (Cumulative)	20%

- Project group: Three students in team)
- Late submission: Maximum 1 day delay with 10% penalty.
- Any kind of copy/plagiarism will result into Zero points for entire assignment/exam/ project for all involved student and may be more severe penalty

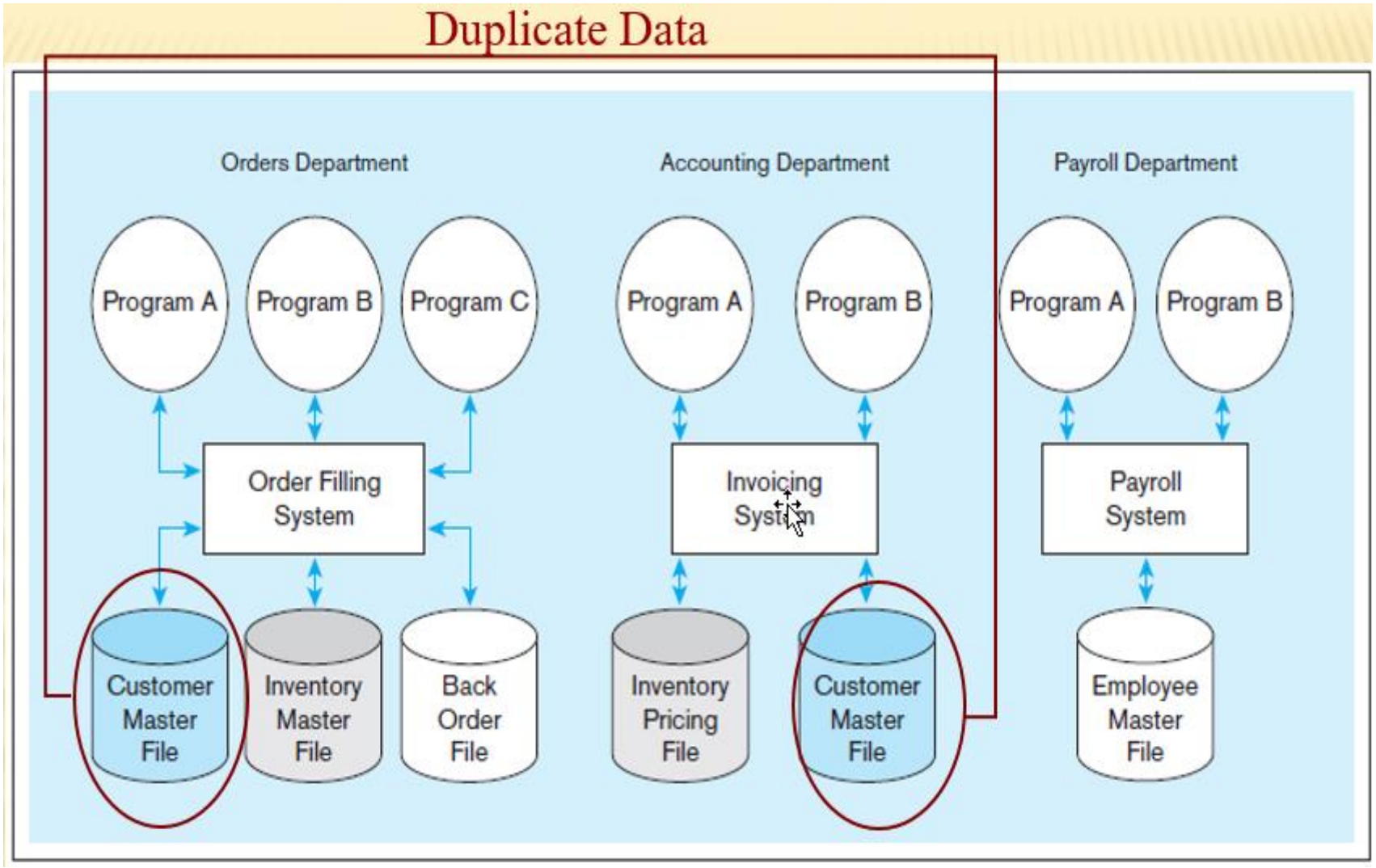
Module 1: Learning Objectives

- Need for Database Systems
- Advantage of Database Systems
- Cost of Risk of Database Systems
- Basic Database Terms
- Components of Database Environment
- Database System Development Life Cycle
- Roles of people involved
- Three Tier Schema Architecture of Database Systems
- Relational Database
- Relational Model
- Relational Algebra and Relational Calculus

File System Approach

NYU-TANDON									
Building_Num	School_Name	Address	City	Zip	Phone_No	Room_No	Type	Floor	Attribte/Field/ Column
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	221	Lecture	2	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	222	Lab	2	
Jabs	School of Engineering	6 METRO TECH Center	New York	11201	212-888-8888	223	Lecture	2	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	224	LAB	2	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	225	Lecture	2	
JABS	School of Engineering	6 Metro Tech Center	NYC	11201	212-888-8888	226	Lecture	2	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	226	lab	2	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	228	Lecture	2	
jAB	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	229	Lab	2	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	(212)-888-8888	230	Lecture	2	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	231	Meeting	2	
JABS	School of Engineering	6, METRO TECH Center	New York	11201	212-888-8888	232	Lecture	2	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	233	Facility	2	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	331	Lecture	3	
JABS	School of Engineering	6 Metro Tech Center	NEW YORK	11201	212-888-8888	332	Lecture	3	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	333	IT-Support	3	
JABS	School of Engg.	6 Metro Tech Center	New York	11201	212-888-8888	334	Lecture	3	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	2128888888	335	Lecture	3	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	336	Lab	3	
JABS	School of Engineering	6 Metro Tech Center	NY	11201	212-888-8888	337	Lab	3	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	338	Lab	3	
JABS	SCHOOL OF ENGINEERING	6 Metro Tech Center	New York	11201	212-888-8888	339	Lab	3	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	400	Lab	4	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	401	Meeting	4	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	402	Lecture	4	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	403	Lecture	4	
JABS	School of Engineering	6 Metro Tech Center	New York	11201	212-888-8888	475	Lecture	4	
Tuple/Record/Row									

Duplicate Data



Drawbacks of using file systems to store data

- Data redundancy and inconsistency
  - Multiple file formats, duplication of data in different files
- Difficulty in accessing data
  - No concurrent access to write
- Data isolation, Limitation of Data Sharing
  - Multiple files and formats, no easy way to share data
- Integrity problems
  - Integrity constraints of relevant data

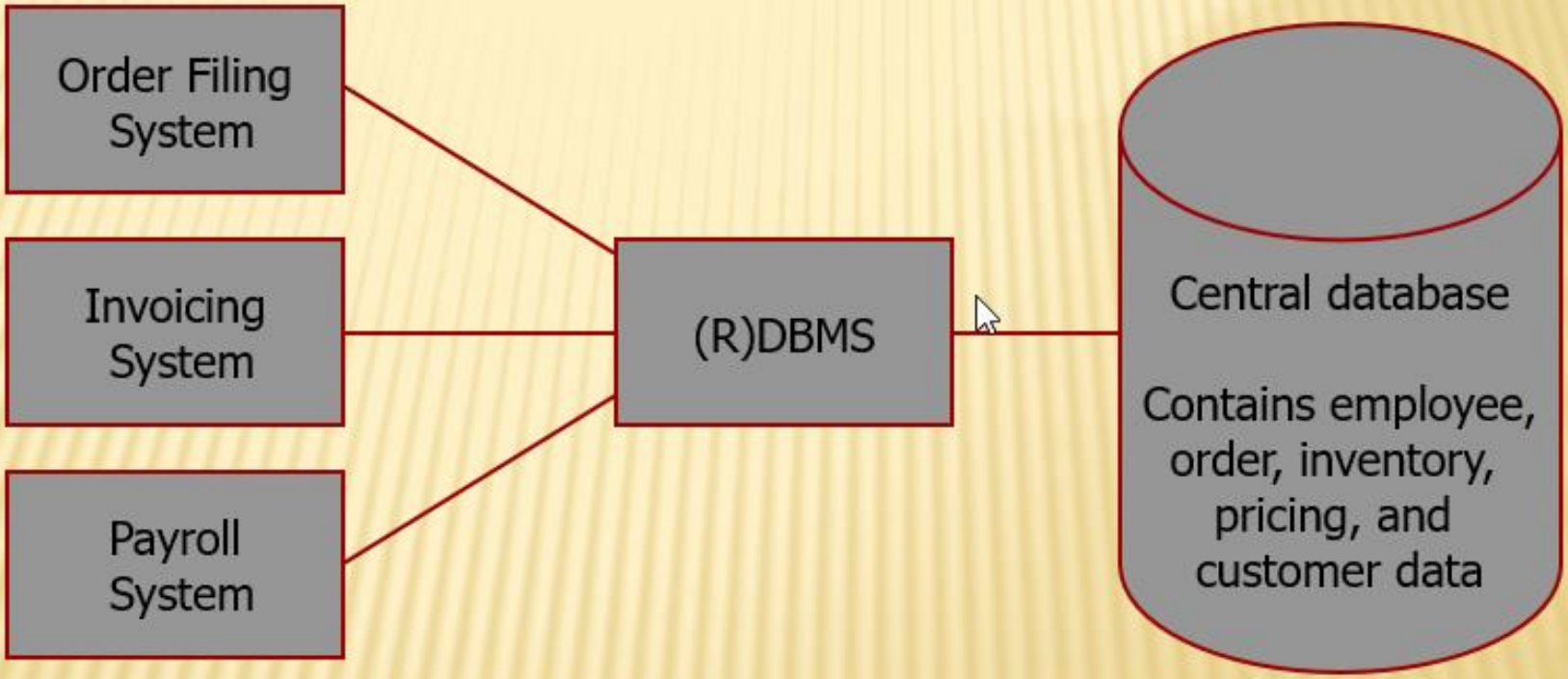
...Drawbacks of using file systems to store data

- Constraint on Atomicity of updates
  - Failures may leave database in an inconsistent state with partial updates carried out
  - Example: Transfer of funds from one account to another should either complete or not happen at all
- Constraint on Concurrent Access by multiple users
  - Concurrent access needed for performance
  - Uncontrolled concurrent accesses can lead to inconsistencies
    - Example: Two people reading a balance (say 100) and updating it by withdrawing money (say 75 each) at the same time
- Security problems
  - Hard to provide user access to some, but not all, data

Database systems offer solutions to all the above problems.

DATABASE MANAGEMENT SYSTEM

- A software system that is used to create, maintain, and provide controlled access to user databases



DATABASE APPROCH TO BUSSINESS SOLUTIONS

Customers Facing

Web Server

Multi User, Online, Concurrent Transitions

Read Write

OLTP DATABASE

OLTP (Online Transaction Processing)

High Quality Data - Consistent, Integrated, Centralized, Sharable, Securable, Scalable, Auditable, Recoverable
(CORPORATE ASSET)

Business users Facing

Data Driven Business Decisions

Data Analytic and Presentation Tools

Read Only

DW DATABASE

DW (Data Warehouse),
DSS(Decision Support System)
INFORMATION FACTORY

Information (processed Data) – Meaningful Insight

TABLE 1.1

TYPES OF DATABASES

PRODUCT	NUMBER OF USERS			DATA LOCATION		DATA USAGE		XML
	SINGLE USER	MULTIUSER						
		WORKGROUP	ENTERPRISE	CENTRALIZED	DISTRIBUTED	OPERATIONAL	ANALYTICAL	
MS Access	X	X		X		X		
MS SQL Server	X*	X	X	X	X	X	X	X
IBM DB2	X*	X	X	X	X	X	X	X
MySQL	X	X	X	X	X	X	X	X
Oracle RDBMS	X*	X	X	X	X	X	X	X

*Vendor offers single-user/personal or Express DBMS versions

Database Applications

- **Banking:** clients, financial transactions
- **Airlines:** reservations, schedules
- **Universities:** courses , faculties, students, grades
- **Sales:** customers, products, purchases
- **Online retailers:** order tracking, delivery, feedback
- **Manufacturing:** production, inventory, orders
- **Human resources:** employees, benefits, deductions

Databases touch all aspects of our lives!

single-user database
A database that supports only one user at a time.

desktop database
A single-user database that runs on a personal computer.

multiuser database
A database that supports multiple concurrent users.

workgroup database
A multiuser database that usually supports fewer than 50 users or is used for a specific department in an organization.

enterprise database
The overall company data representation, which provides support for present and expected future needs.

centralized database
A database located at a single site.

distributed database
A logically related database that is stored in two or more physically independent sites.

cloud database
A database that is created and maintained using cloud services, such as Microsoft Azure or Amazon AWS.

operational database
A database designed primarily to support a company's day-to-day operations. Also known as a *transactional database*, *OLTP database*, or

Extensible Markup Language (XML)
A metalanguage used to represent and manipulate data elements. Unlike other markup languages, XML permits the manipulation of a document's data elements.

analytical database
A database focused primarily on storing historical data and business metrics used for tactical or strategic decision making.

data warehouse
A specialized database that stores historical and aggregated data in a format optimized for decision support.

online analytical processing (OLAP)
A set of tools that provide advanced data analysis for retrieving, processing, and modeling data from the data warehouse.

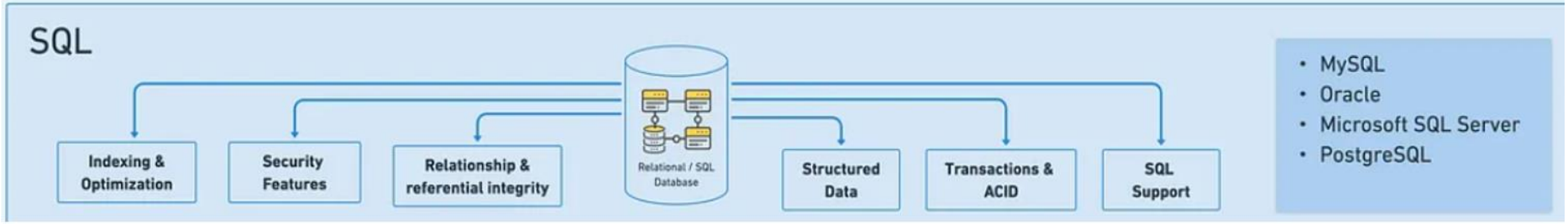
Costs and Risks of the Database Approach

- New, specialized personnel
- Installation and management cost and complexity
- Conversion costs
- Need for explicit backup and recovery
- Organizational conflict

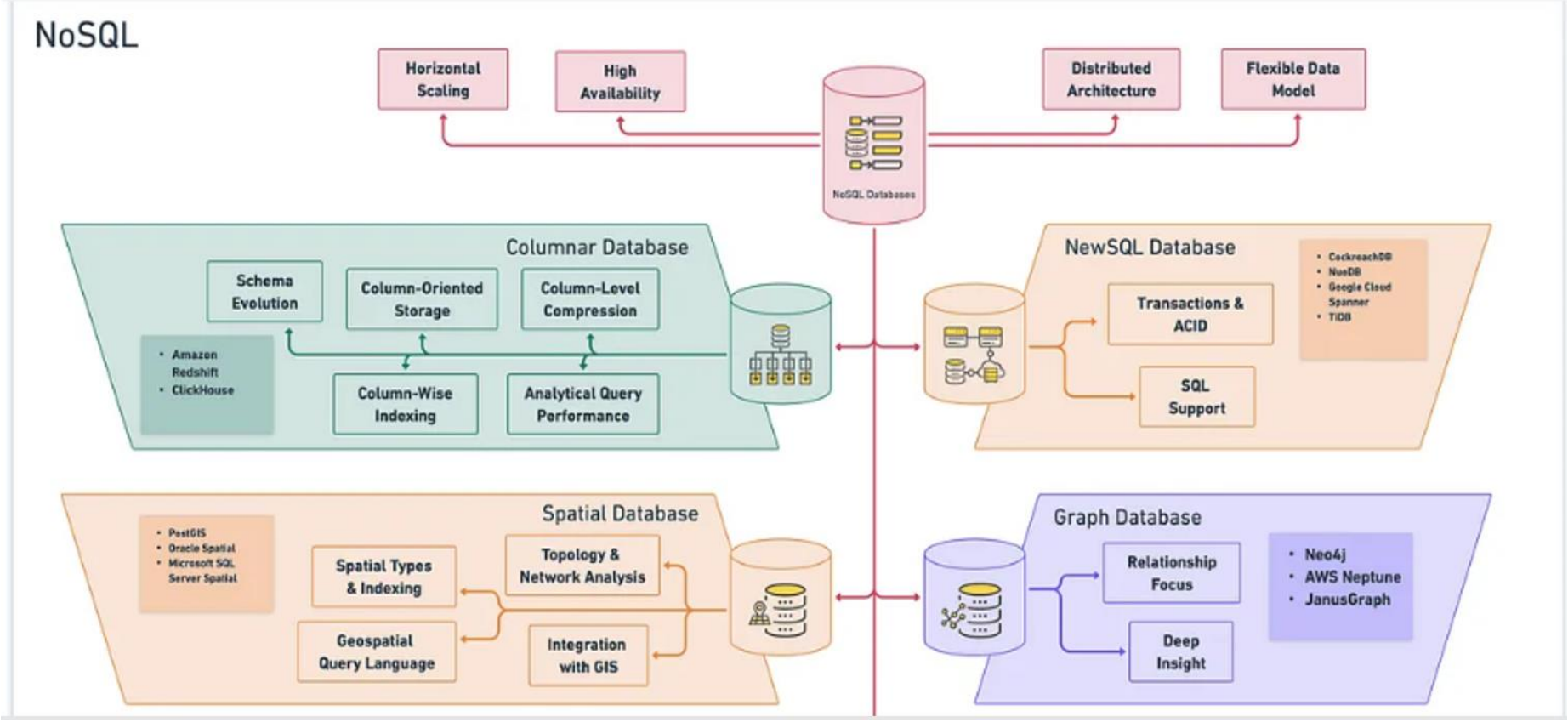
Most popular RDBMS are:

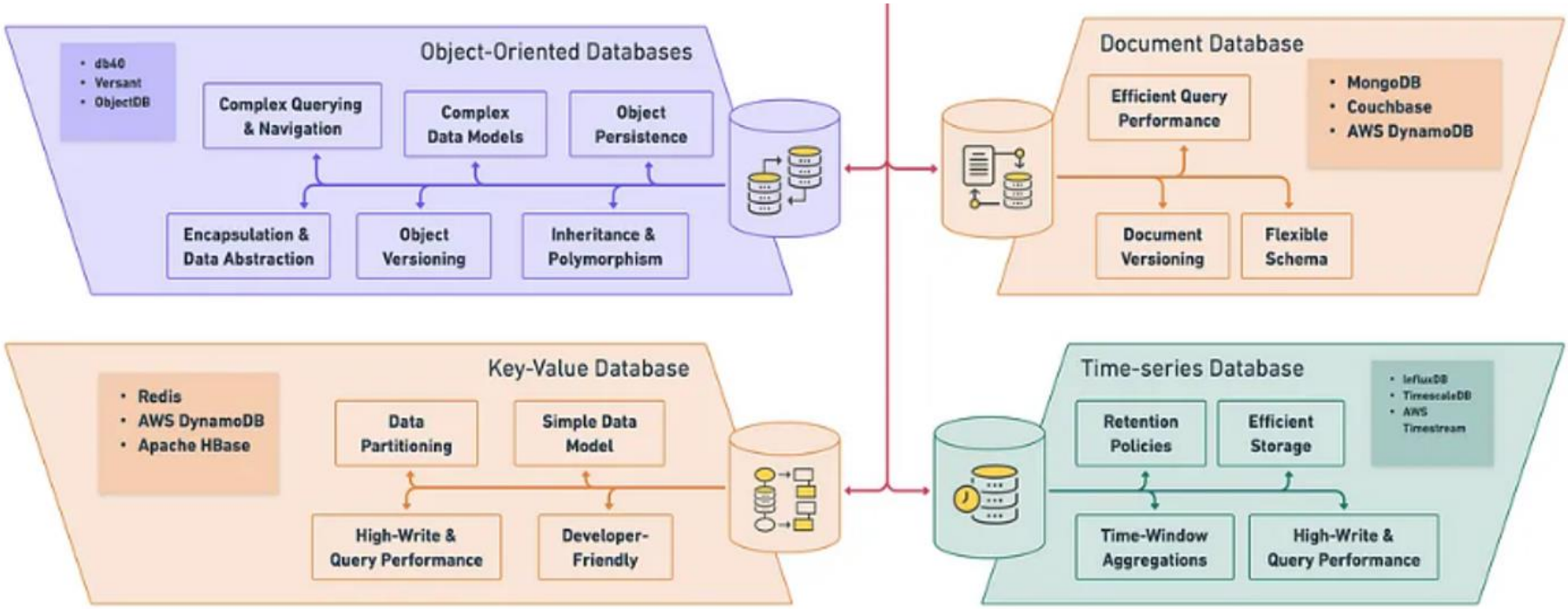
- Oracle
- MySQL
- SQL Server
- PostgreSQL
- IBM DB2

TYPES OF DATABASES



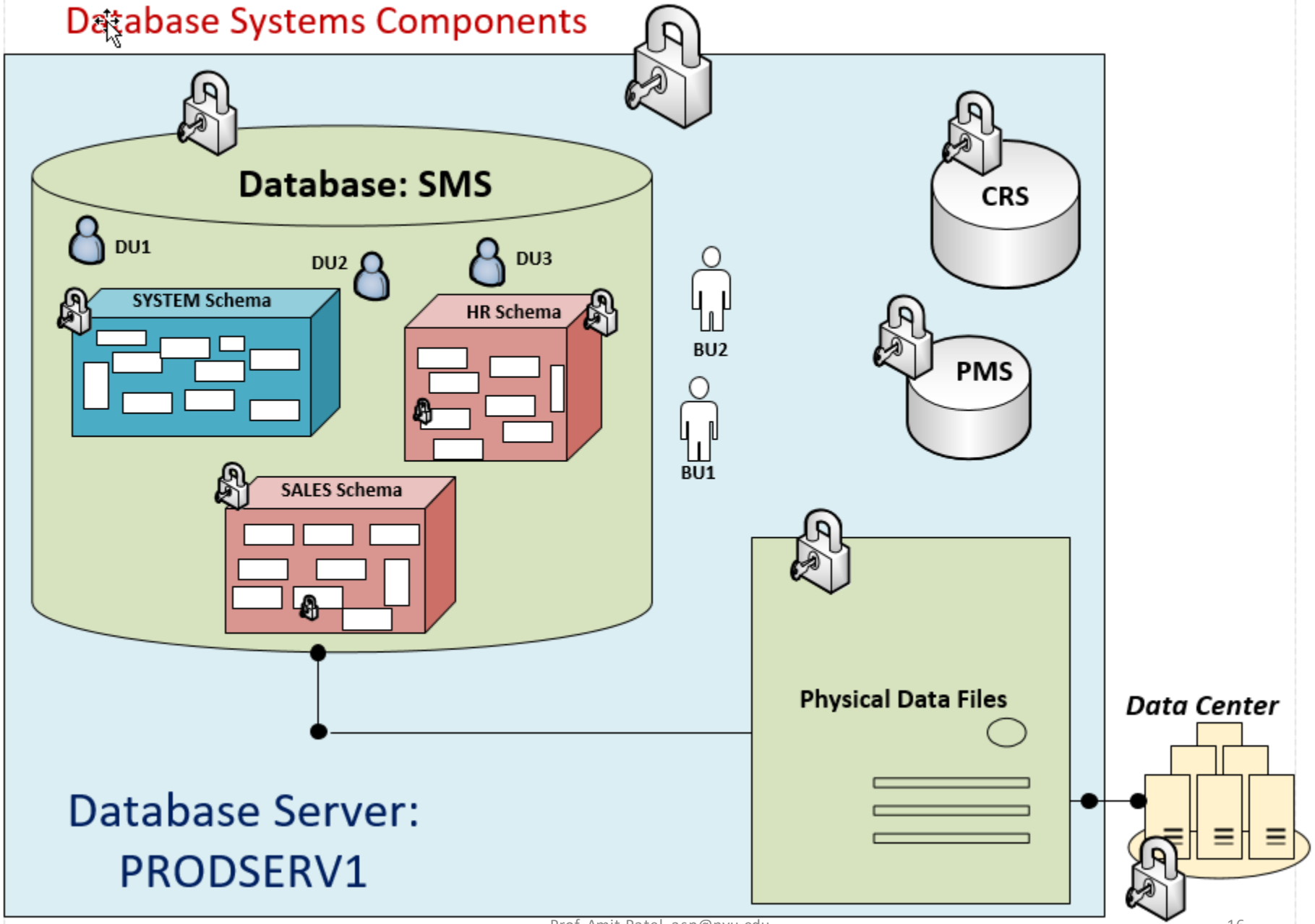
https://blog.devgenius.io/mastering-the-database-duality-exploring-the-realm-of-sql-and-nosql-with-cheatsheet-33a73f752460


NoSQL
Horizontal Scaling
High Availability
NoSQL Databases
Distributed Architecture
Flexible Data Model
Columnar Database
Schema Evolution
Column-Oriented Storage
Column-Level Compression
Amazon Redshift
ClickHouse
Column-Wise Indexing
Analytical Query Performance
NewSQL Database
CockroachDB
NuoDB
Google Cloud Spanner
TiDB
Transactions & ACID
SQL Support
Spatial Database
PostGIS
Oracle Spatial
Microsoft SQL Server Spatial
Spatial Types & Indexing
Topology & Network Analysis
Geospatial Query Language
Integration with GIS
Graph Database
Relationship Focus
Neo4j
AWS Neptune
JanusGraph
Deep Insight



Please refer to the document about No-SQL database, and MongoDB posted on course website: Additional Material→ No-SQL Databases: MongoDB

Database Systems Components



Basic Database Terms

- **Entity (Table/Relation) :**

  Noun describing people / place / object /event /concept

People: e.g. Employee, Customer, Student, Faculties, Doctor, Patient
Place: e.g. Department, Country, City, Warehouse, Hospital, College
Object: e.g. Car, Motorcycle, Part, Laptop, Television, Phone
Event: e.g. Course, Project, Invoice, Payment, Purchase, Sale

- **Attribute (Column/Field/Domain) : Characteristics of an entity**

Table: EMPLOYEE

EMPNO	ELNAME	EFNAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO	EMAIL	PHONE_NUMBER	SSN
7369	SMITH	CHARLES	CLERK	7902	8-Oct-01	2400		20	csmith@abc.com	212-212-2100	888-88-8800
7499	ALLEN	GRAYSON	SALESMAN	7698	12-Dec-01	4800	600	30	gallen@abc.com	212-212-2101	888-88-8801
7521	WARD	MATTHEW	SALESMAN	7698	14-Dec-01	3750	1000	30	mward@abc.com	212-212-2102	888-88-8802
7566	JONES	NICHOLAS	MANAGER	7839	22-Jan-02	8925		30	njones@abc.com	212-212-2103	888-88-8803
7654	MARTIN	CHRIS	SALESMAN	7698	20-Jul-02	3750	2800	30	cmartin@abc.com	212-212-2104	888-88-8804
7698	GRIFFIN	BLAKE	MANAGER	7839	20-Feb-02	8550		30	bgriffin@abc.com	212-212-2105	888-88-8805
7782	CLARK	KENT	MANAGER	7839	31-Mar-02	7350		10	kclark@abc.com	212-212-2106	888-88-8806
7788	BOOKER	DEVIN	ANALYST	7566	8-Feb-08	9000		20	dbooker@abc.com	212-212-2107	888-88-8807
7839	KING	MARTIN	PRESIDENT		8-Sep-02	15000		10	mking@abc.com	212-212-2108	888-88-8808
7844	TURNER	WILLIAM	SALESMAN	7698	30-Jun-02	4500	0	30	wturner@abc.com	212-212-2109	888-88-8809
7876	ADAMS	JOHN	CLERK	7788	13-Mar-08	3300		20	jadam@abc.com	212-212-2110	888-88-8810
7900	JAMES	LEBRON	CLERK	7698	24-Sep-02	2850		30	ljames@abc.com	212-212-2111	888-88-8811
7902	FORD	CHRISTIAN	ANALYST	7566	24-Sep-02	9000		20	cford@abc.com	212-212-2112	888-88-8812
7934	MILLER	MIKE	CLERK	7782	14-Nov-02	3900		10	mmiller@abc.com	212-212-2113	888-88-8813

Basic database Terms (Cont'd..)

Data: Essentially the plain facts collected during the business operation

- Records wide range of business activities
- Can be external or internal to business
- Basis for all meaningful insights that helps making crucial business decision
- Element of constructed information

Data can be structured or unstructured or semi-structured

- Structured Data: Number, Text, Date
- Unstructured Data: Images, Video, Sound, Document
- Semi-Structured Data: Email, Music Album

▪**Information**: processed data → to derive meaningful and actionable insight → that increases knowledge of person/org. using it → to help making decisions/conclusion (information can be descriptive/predictive/ prescriptive

▪**Analysis**: systemic process of turning row data into information

Basic database Terms (Cont'd..)

- **Schema:** Logical grouping of related entities and associated database programs

- **Database:** organized collection of logically related schema

- **DBMS** (Database Management System): Software that enables to create, maintain, secure, store, retrieve, manipulate entities and data

- **RDBMS** (Relational Database Management System): DBMS that establishes relationship among entities based upon defined relationship via common attribute(s)

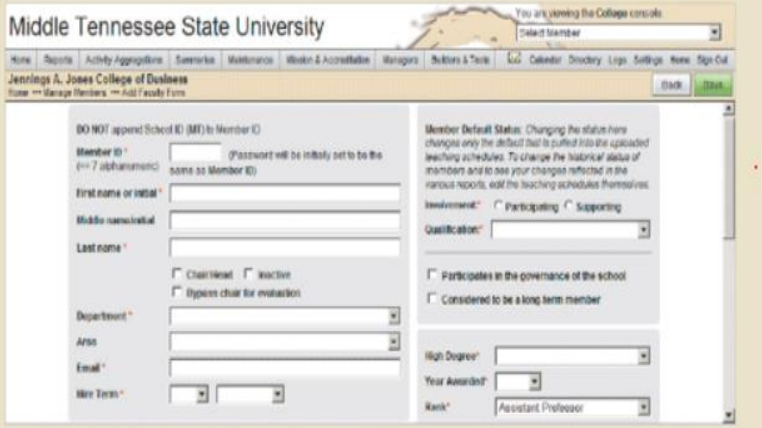
- **Metadata**: data that describes the properties and context of user data (data about data) [data type, size, optional/mandatory, constraint, description]

Table Name EMP (EMPLOYEE TABLE)

Column Name	Datatype	Size	Optional/Mandatory	Comment
EMPNO	NUMERIC	4	Mandatory	Employee ID Number
EFNAME	VARCHAR	30	Mandatory	Employee Frist Name
ELNAME	VARCHAR	30	Mandatory	Employee Last Name
JOB	VARCHAR	30	Mandatory	Employee Functional Role
MGR	NUMERIC	4	Optional	Employee Manager ID
HIREDATE	DATE		Mandatory	Employee Join Date
SAL	NUMERIC	(7,2)	Mandatory	Employee Monthly Salary in USD
COMM	NUMERIC	(7,2)	Optional	Employee commission
EMAIL	VARCHAR	30	Mandatory	EMPLOYEE's EMAIL ADDRESS
PHONE_NUMBER	VARCHAR	12	Mandatory	EMPLOYEE's PHONE NUMBER
SSN	VARCHAR	11	Mandatory	SOCIAL SECURITY NUMBER

FIGURE 1.2 TRANSFORMING RAW DATA INTO INFORMATION

a) Data entry screen



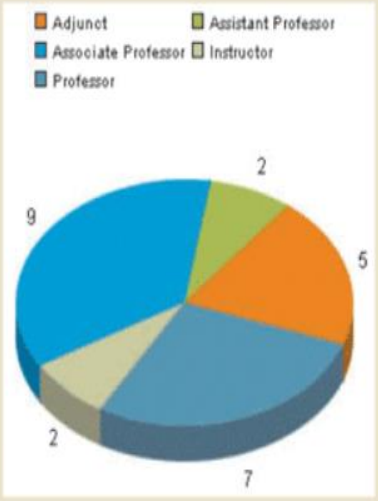
b) Raw data

Id	LastName	MidName	FirstName	DeptCode	Office	Email	Rank	HireYear	Degree
1	Washington	A	George	MGMT	N135	gwashington@mtsu.edu	Professor		2001 Ph.D.
2	Adams		John	FIN	N213	jadams@mtsu.edu	Professor		1994 Ph.D.
3	Jefferson	L	Thomas	ECON		tjefferson@mtsu.edu	Instructor		2002 M.B.A.
4	Madison	D	James	FIN	N236	jmadison@mtsu.edu	Associate Professor		1994 Ph.D.
5	Monroe	N	James	ACCT	N211	jmonroe@mtsu.edu	Assistant Professor		1995 Ph.D.
6	Adams	G	John	ACCT	N218	jqadams@mtsu.edu	Associate Professor		1999 Ph.D.
7	Jackson	C	Andrew	ECON	N233	ajackson@mtsu.edu	Associate Professor		1999 Ph.D.
8	Van Buren	T	Martin	FIN	N208	mvanburen@mtsu.edu	Professor		1988 Ph.D.
9	Harrison	R	William	MKTG	N218	wharrison@mtsu.edu	Professor		1994 Ph.D.
10	Tyler	M	John	MGMT		JTyler@mtsu.edu	Assistant Professor		2000 Ed.D.
11	Polk		Cheryl	MKTG	N240	cpolk@mtsu.edu	Associate Professor		2002 Ph.D.
12	Taylor	G	Zachary	ACCT	N215	ztaylor@mtsu.edu	Associate Professor		1986 Ph.D.
13	Fillmore		Millard	JCB	N219	mfillmore@mtsu.edu	Professor		1982 Ph.D.
14	Pierce	A	Franklin	MKTG	N259	pfranklin@mtsu.edu	Instructor		2005 M.B.A.
15	Buchanan	T	James	MGMT	N246	jbuchanan@mtsu.edu	Associate Professor		1996 D.B.A.
16	Lincoln	W	Leroy	MGMT	N258	llincoln@mtsu.edu	Associate Professor		1996 Ph.D.
17	Johnson		Andrew	ISYS	N268	ajohnson@mtsu.edu	Professor		1987 Ph.D.
18	Grant		Kelly	MKTG	N228	kgrant@mtsu.edu	Assistant Professor		1995 D.B.A.
19	Rutherford		Hayes	ACCT	N408	hrutherford@mtsu.edu	Professor		1982 Ph.D.
20	Garfield	T	Denise	ACCT		dgarfield@mtsu.edu	Assistant Professor		2010 Ph.D.
21	Arthur		Emily	ACCT	N413	earthur@mtsu.edu	Associate Professor		2001 J.D.
22	Cleveland	G	Robert	ACCT	N401	rcleveland@mtsu.edu	Associate Professor		1997 Ph.D.
23	Harrison	X	Pamela	BULA	N406	pharrison@mtsu.edu	Associate Professor		2001 J.D.
24	McKinley	D	Priscilla	ISYS	N263	pmckinley@mtsu.edu	Adjunct		1994 M.S.
25	Roosevelt	F.	Hillary	MGMT	N234	hroosevelt@mtsu.edu	Associate Professor		2002 Ph.D.
26	Wilson		Laura	BCEN	N448	lwilson@mtsu.edu	Professor		1992 Ph.D.
27	Harding		Warren	MKTG	N214	wharding@mtsu.edu	Professor		1984 Ed.D.
28	Coolidge		Calvin	ECON	N216	ccoolidge@mtsu.edu	Professor		1975 Ph.D.
29	Hoover		Lisa	MGMT		lhoover@mtsu.edu	Adjunct		1978 M.B.A.
30	Truman		Betty	ACCT	N416	btruman@mtsu.edu	Professor		1971 Ed.D.
31	Johnson		Robert	BCEN	N248	rjohnson@mtsu.edu	Professor		2001 Ph.D.

c) Information in summary format

Rank	COUNT	%/INFS	TOT/COL	%/COL. TOT.	%/COL. FAC.
Adjunct	5	20.00%	23	21.74%	3.27%
Assistant Professor	2	8.00%	28	7.14%	1.31%
Associate Professor	9	36.00%	37	24.32%	5.88%
Instructor	2	8.00%	18	11.11%	1.31%
Professor	7	28.00%	47	14.89%	4.58%

d) Information in graphical format



Raw data without context has no value addition. Context in Data helps users understand data

Graphical displays turn raw data into useful information that managers and business can use for decision making and interpretation

Three Schema Architecture of DBMS

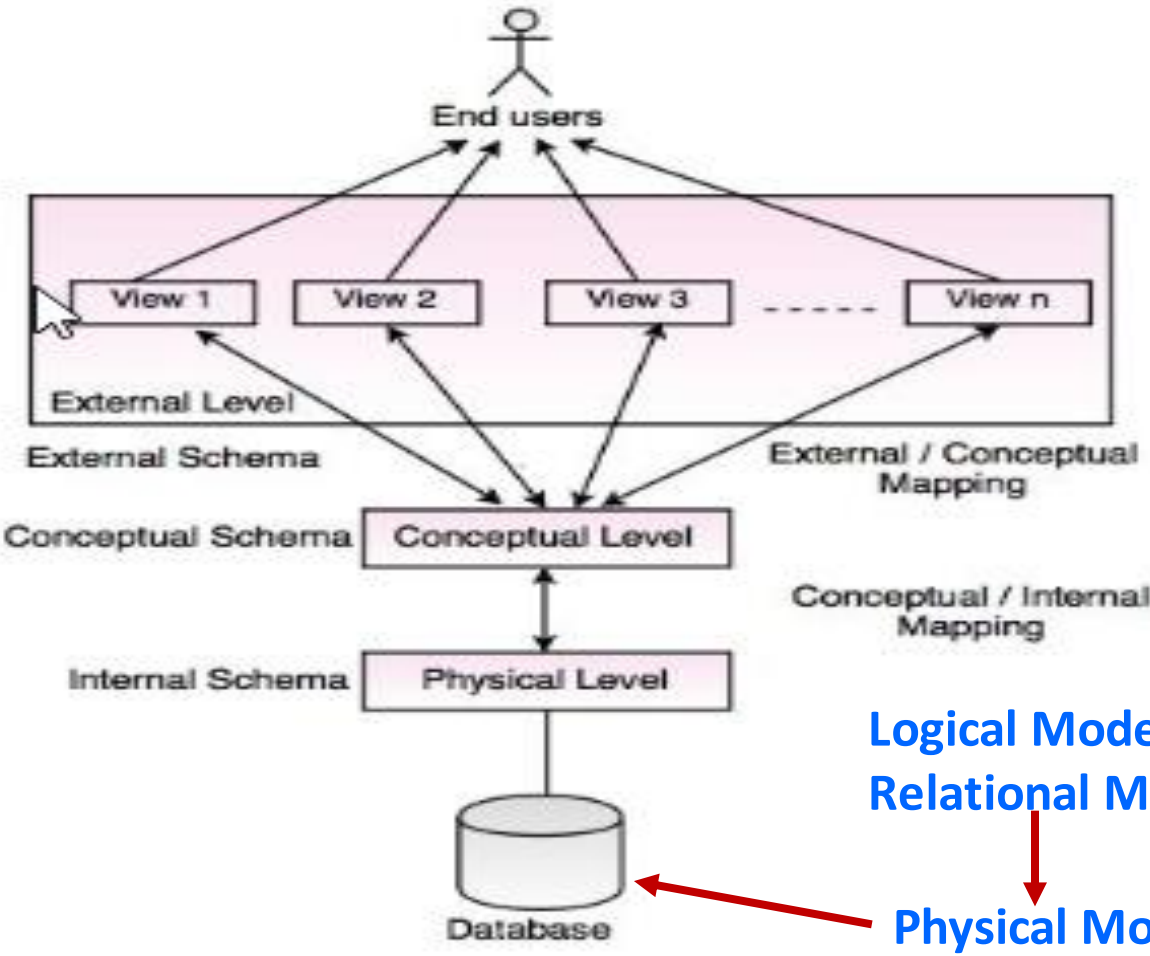


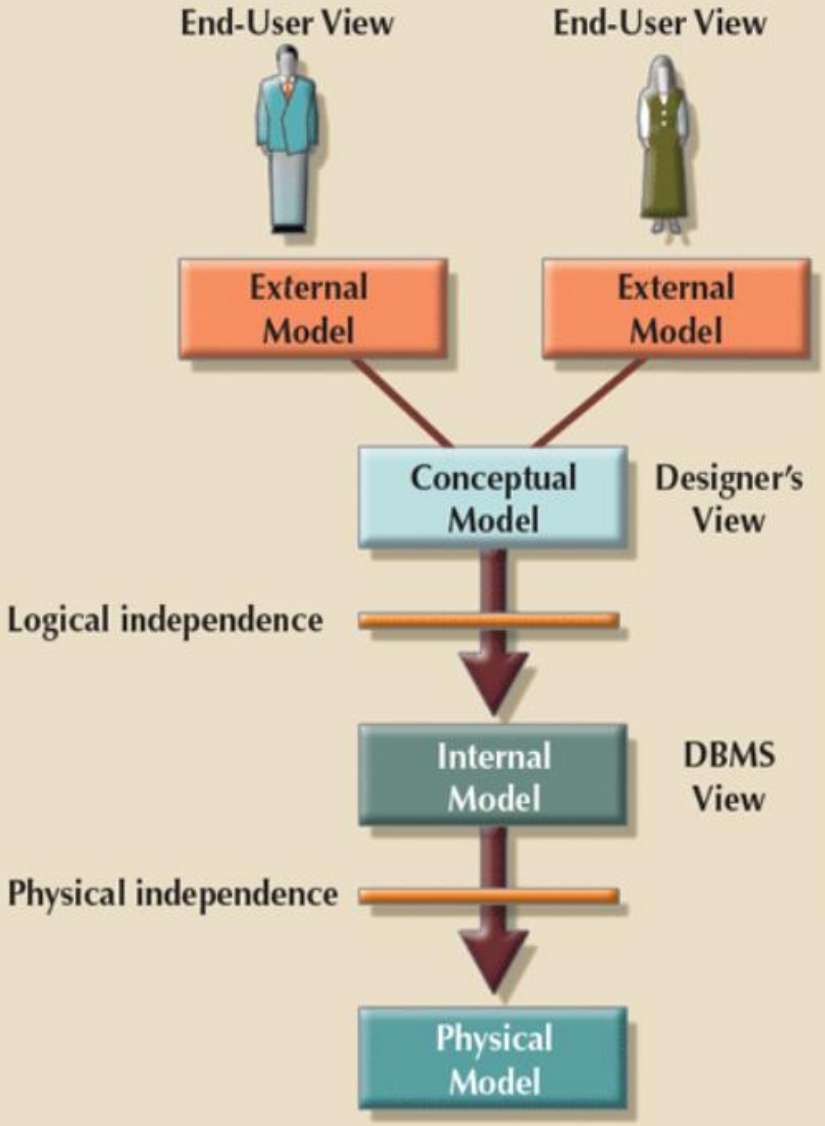
Fig. Three Level Architechture of DBMS

Enterprise Data Model – High Level

Project Data Model – Conceptual

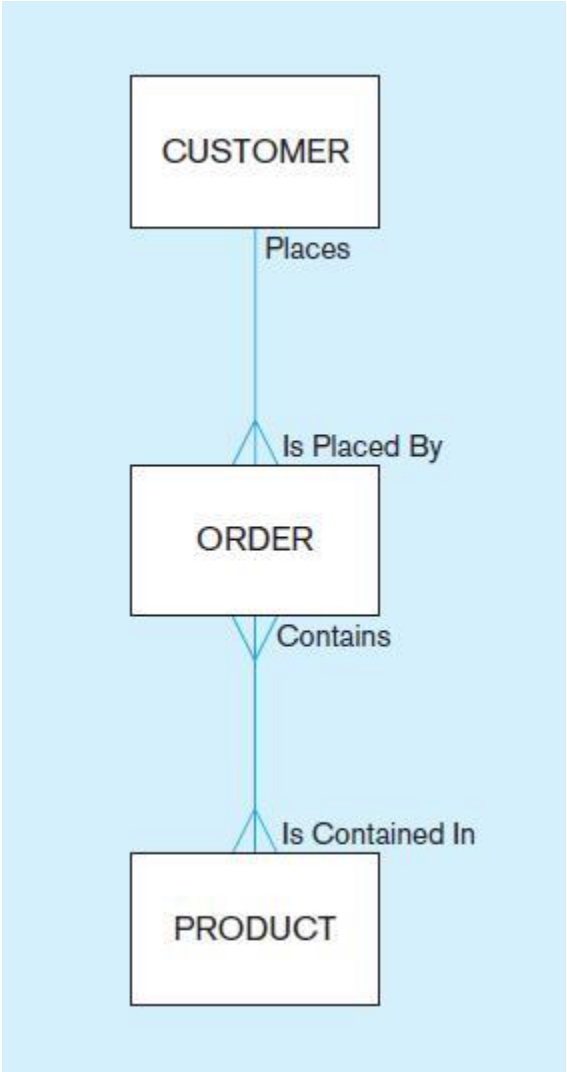
Logical Model and Relational Model

Physical Model



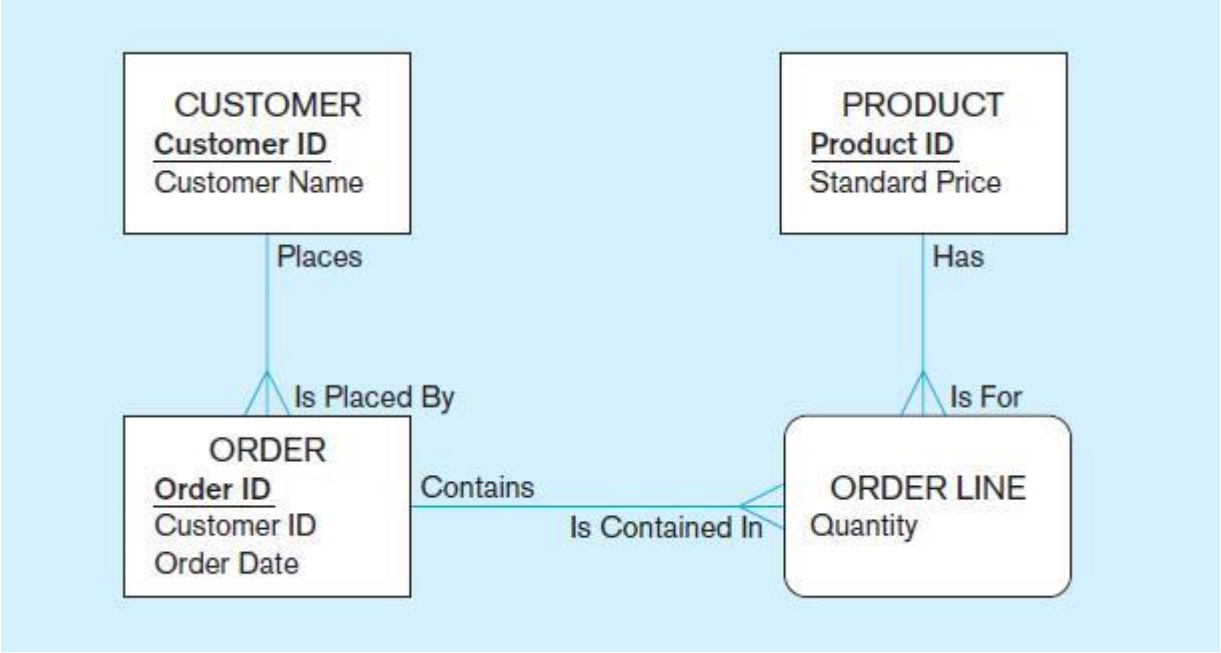
Degree of Abstraction		Characteristics
High	ER Object-Oriented	Hardware-independent Software-independent
Medium	Relational	Hardware-independent Software-dependent
Low	Network Hierarchical	Hardware-dependent Software-dependent

Comparison of enterprise and project level data models



Segment of an enterprise data model

Segment of a project-level data model



Two Approaches to Database and Information System Development

- SDLC
  - System Development Life Cycle
  - Detailed, well-planned development process
  - Time-consuming, but comprehensive
  - Long development cycle
- Prototyping
  - Rapid application development (RAD)
  - Cursory attempt at conceptual data modeling
  - Define database during development of initial prototype
  - Repeat implementation and maintenance activities with new prototype versions

DBLC (Database Life Cycle): The DBLC is composed of six phases:

PLANNING
Initial assessment
Feasibility study

ANALYSIS
User requirements
Study of existing systems
Logical system design

DETAILED SYSTEMS DESIGN
Detailed system specifications

IMPLEMEN
Coding, testing, debugging
Installation, fine-tuning, Documentation,
User training

MAINTENANCE
Changes, Upgrades, Bug fixes (patching)
Enhancements
Backups
Performance Tuning
Capacity Planning and Provisions

Managing database Projects: People Involved

- Business Analysts (BA)
- Systems Analysts (SA)
- Database Architects (DA)
- Database Modelers / Database Designers
- Database Developers/ Database Programmers
- Data Administrators (DBA)
- Data Steward / Chief Data Officer (CD0)
- Data Analyst
- Data Scientist
- Project Manager (PM)

DATABASE CAREER OPPORTUNITIES		
JOB TITLE	**DESCRIPTION**	**SAMPLE SKILLS REQUIRED**
Database Developer	Create and maintain database-based applications	Programming, database fundamentals, SQL
Database Designer	Design and maintain databases	Systems design, database design, SQL
Database Administrator	Manage and maintain DBMS and databases	Database fundamentals, SQL, vendor courses
Database Analyst	Develop databases for decision support reporting	SQL, query optimization, data warehouses
Database Architect	Design and implementation of database environments (conceptual, logical, and physical)	DBMS fundamentals, data modeling, SQL, hardware knowledge, etc.
Database Consultant	Help companies leverage database technologies to improve business processes and achieve specific goals	Database fundamentals, data modeling, database design, SQL, DBMS, hardware, vendor-specific technologies, etc.
Database Security Officer	Implement security policies for data administration	DBMS fundamentals, database administration, SQL, data security technologies, etc.
Cloud Computing Data Architect	Design and implement the infrastructure for next-generation cloud database systems	Internet technologies, cloud storage technologies, data security, performance tuning, large databases, etc.
Data Scientist	Analyze large amounts of varied data to generate insights, relationships, and predictable behaviors	Data analysis, statistics, advanced mathematics, SQL, programming, data mining, machine learning, data visualization

Table (Relation/Entity)

Columns / Attributes / Fields / Domains

The data is stored in various Tables (Relations / Entities)

Example of tabular data in relational model

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

Rows / Records / Tuples / Instances

(a) The *instructor* table

A Sample Relational Database

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

(a) The *instructor* table

dept_name	*building*	*budget*
Comp. Sci.	Taylor	100000
Biology	Watson	90000
Elec. Eng.	Taylor	85000
Music	Packard	80000
Finance	Painter	120000
History	Painter	50000
Physics	Watson	70000

(b) The *department* table

Relation (TABLE)

- A relation is a named, two-dimensional table of data.
- A table consists of rows (records) and columns (attribute or field).

Requirements for a table to qualify as a relation:

- Every attribute value must be atomic (not multivalued, not composite).
- Every row must be unique (can't have two rows with exactly the same values for all their fields).
- Table and Attributes (columns) in tables must have unique names, single world, can not be reserved words, and of no more than 30 characters.
- The order of the columns must be irrelevant.
- The order of the rows must be irrelevant.

Oracle Reserved Words:

https://docs.oracle.com/database/121/SQLRF/ap_keywd001.htm#SQLRF55621

A Relation(Entity/Table), Not Well Structure

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO	DNAME	LOC
7782	CLARK	MANAGER	7839	6/9/1981	2450		10	ACCOUNTING	NEW YORK
7839	KING	PRESIDENT		11/17/1981	5000		10	ACCOUNTING	NEW YORK
7934	MILLER	CLERK	7782	1/23/1982	1300		10	ACCOUNTING	NEW YORK
7566	JONES	MANAGER	7839	4/2/1981	2975		20	RESEARCH	DALLAS
7902	FORD	ANALYST	7566	12/3/1981	3000		20	RESEARCH	DALLAS
7876	ADAMS	CLERK	7788	5/23/1987	1100		20	RESEARCH	DALLAS
7369	SMITH	CLERK	7902	12/17/1980	800		20	RESEARCH	DALLAS
7788	SCOTT	ANALYST	7566	4/19/1987	3000		20	RESEARCH	DALLAS
7521	WARD	SALESMAN	7698	2/22/1981	1250	500	30	SALES	CHICAGO
7844	TURNER	SALESMAN	7698	9/8/1981	1500	0	30	SALES	CHICAGO
7499	ALLEN	SALESMAN	7698	2/20/1981	1600	300	30	SALES	CHICAGO
7900	JAMES	CLERK	7698	12/3/1981	950		30	SALES	CHICAGO
7698	BLAKE	MANAGER	7839	5/1/1981	2850		30	SALES	CHICAGO
7654	MARTIN	SALESMAN	7698	9/28/1981	1250	1400	30	SALES	CHICAGO
								OPERATIONS	BOSTON

Well Structured Relations with Relationships

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
7369	SMITH	CLERK	7902	17-Dec-80	800		20
7499	ALLEN	SALESMAN	7698	20-Feb-81	1,600	300	30
7521	WARD	SALESMAN	7698	22-Feb-81	1,250	500	30
7566	JONES	MANAGER	7839	2-Apr-81	2,975		20
7654	MARTIN	SALESMAN	7698	28-Sep-81	1,250	1,400	30
7698	BLAKE	MANAGER	7839	1-May-81	2,850		30
7782	CLARK	MANAGER	7839	9-Jun-81	2,450		10
7788	SCOTT	ANALYST	7566	19-Apr-87	3,000		20
7839	KING	PRESIDENT		17-Nov-81	5,000		10
7844	TURNER	SALESMAN	7698	8-Sep-81	1,500		30
7876	ADAMS	CLERK	7788	23-May-87	1,100		20
7900	JAMES	CLERK	7698	3-Dec-81	950		30
7902	FORD	ANALYST	7566	3-Dec-81	3,000		20
7934	MILLER	CLERK	7782	23-Jan-82	1,300		10
7999	ROBERT	ANALYST	7782	12-May-92	3,200		60

Primary Key (unique, Not Null)

Foreign Key (Referencing to PK of Other Table)

Null

Null: Value not assigned, not eligible, Unknown, absent. Null IS NOT blank or zero

DEPTNO	DNAME	LOC
10	ACCOUNTING	NEW YORK
20	RESEARCH	DALLAS
30	SALES	CHICAGO
40	OPERATIONS	BOSTON

Primary Key (unique, Not Null)

Relational Query Languages

- Procedural vs .non-procedural, or declarative
- "Pure" languages:
  - Relational algebra
  - Tuple relational calculus (TRC)
  - Domain relational calculus (DRC)

- The above 3 pure languages are equivalent in computing power

Relational Algebra

- Procedural language
- Six basic operators
  - select: σ (Unary operator)
  - project: $\prod$ (Unary operator)
  - union: $\cup$ (Binary operator)
  - set difference: – (Binary operator)
  - Cartesian product: x (Binary operator)
  - rename: ρ *(Unary operator)*
- The operators take one or two relations as inputs and produce a new relation as a result.

- Unary operator: takes one relation as input

- Binary Operator: takes pair of relations as input

Relational Algebra Operators

Symbol (Name)	Example of Use
σ (Selection)	$\sigma_{salary >= 85000}(instructor)$
	Return rows of the input relation that satisfy the predicate.
Π (Projection)	$\Pi_{ID, salary}(instructor)$
	Output specified attributes from all rows of the input relation. Remove duplicate tuples from the output.
x (Cartesian Product)	*instructor* x *department*
	Output pairs of rows from the two input relations that have the same value on all attributes that have the same name.
∪ (Union)	$\Pi_{name}(instructor) \cup \Pi_{name}(student)$
	Output the union of tuples from the *two* input relations.
– (Set Difference)	$\Pi_{name}(instructor) - \Pi_{name}(student)$
	Output the set difference of tuples from the two input relations.
⋈ (Natural Join)	*instructor* ⋈ *department*
	Output pairs of rows from the two input relations that have the same value on all attributes that have the same name.

Relational Algebra Operators

Relational Algebra uses:

Comparative operators: >, <, <= , >=, =, #

Logical operators: AND (^) , OR (v), NOT (-l)

Arithmetic operators: + , - , * , /

Aggregate functions: COUNT, SUM, MIN, MAX, AVG

Example of Relational Tables

SQL Worksheet

```
1  SELECT *
2  FROM AP_EMP;
```

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
7369	SMITH	CLERK	7902	08-OCT-01	2400	-	20
7499	ALLEN	SALESMAN	7698	12-DEC-01	4800	600	30
7521	WARD	SALESMAN	7698	14-DEC-01	3750	1000	30
7566	JONES	MANAGER	7839	22-JAN-02	8925	-	20
7654	MARTIN	SALESMAN	7698	20-JUL-02	3750	2800	30
7698	BLAKE	MANAGER	7839	20-FEB-02	8550	-	30
7782	CLARK	MANAGER	7839	31-MAR-02	7350	-	10
7788	STUDENT	ANALYST	7566	08-FEB-08	9000	-	20
7839	KING	PRESIDENT	-	08-SEP-02	15000	-	10
7844	TURNER	SALESMAN	7698	30-JUN-02	4500	0	30
7876	ADAMS	CLERK	7788	13-MAR-08	3300	-	20
7900	JAMES	CLERK	7698	24-SEP-02	2850	-	30
7902	FORD	ANALYST	7566	24-SEP-02	9000	-	20
7934	MILLER	CLERK	7782	14-NOV-02	3900	-	10

Download CSV

14 rows selected.

SQL Worksheet

```
1  SELECT *
2  FROM AP_DEPT;
```

DEPTNO	DNAME	LOC
10	ACCOUNTING	NEW YORK
20	RESEARCH	DALLAS
30	SALES	CHICAGO
40	OPERATIONS	BOSTON

Download CSV

4 rows selected.

DEPTNO is the Primary Key in AP_DEPT table

EMPNO is the Primary Key in AP_EMP table

DEPTNO Is the Foreign Key for AP_EMP table, referencing to Primary Key of AP_DEPT table

Relational Algebra Example: SELECTION

SQL Worksheet

```
1  SELECT *
2  FROM AP_EMP
3  WHERE DEPTNO=10;
```

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
7782	CLARK	MANAGER	7839	31-MAR-02	7350	-	10
7839	KING	PRESIDENT	-	08-SEP-02	15000	-	10
7934	MILLER	CLERK	7782	14-NOV-02	3900	-	10

Download CSV

3 rows selected.

Relational Algebra Example: PROJECTION

SQL Worksheet

```
1  SELECT EMPNO, ENAME,DEPTNO, SAL
2  FROM AP_EMP;
3
```

EMPNO	ENAME	DEPTNO	SAL
7369	SMITH	20	2400
7499	ALLEN	30	4800
7521	WARD	30	3750
7566	JONES	20	8925
7654	MARTIN	30	3750
7698	BLAKE	30	8550
7782	CLARK	10	7350
7788	STUDENT	20	9000
7839	KING	10	15000
7844	TURNER	30	4500
7876	ADAMS	20	3300
7900	JAMES	30	2850
7902	FORD	20	9000
7934	MILLER	10	3900

Download CSV

14 rows selected.

Relational Algebra Example: SELECTION and PROJECTION

SQL Worksheet

```
1  SELECT EMPNO, ENAME,DEPTNO, SAL
2  FROM AP_EMP
3  WHERE DEPTNO=10;
4
```

EMPNO	ENAME	DEPTNO	SAL
7782	CLARK	10	7350
7839	KING	10	15000
7934	MILLER	10	3900

Download CSV

3 rows selected.

Relational Algebra Example: UNION

SQL Worksheet

```
1  SELECT EMPNO, ENAME,DEPTNO,JOB
2  FROM AP_EMP
3  WHERE DEPTNO=10
4  union
5  SELECT EMPNO, ENAME,DEPTNO, JOB
6  FROM AP_EMP
7  WHERE JOB='CLERK'
8
9
```

EMPNO	ENAME	DEPTNO	JOB
7369	SMITH	20	CLERK
7782	CLARK	10	MANAGER
7839	KING	10	PRESIDENT
7876	ADAMS	20	CLERK
7900	JAMES	30	CLERK
7934	MILLER	10	CLERK

Download CSV

6 rows selected.

Relational Algebra Example: SET DIFFERENCE

SQL Worksheet

```
1  SELECT EMPNO, ENAME,DEPTNO,JOB
2  FROM AP_EMP
3  WHERE DEPTNO=10
4  minus
5  SELECT EMPNO, ENAME,DEPTNO, JOB
6  FROM AP_EMP
7  WHERE JOB='CLERK'
8
9
```

EMPNO	ENAME	DEPTNO	JOB
7782	CLARK	10	MANAGER
7839	KING	10	PRESIDENT

Download CSV

2 rows selected.

Relational Algebra Example: CARTESIAN JOIN

SQL Worksheet

```
1  SELECT *
2  FROM AP_EMP, AP_DEPT
3
4
```

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO	DEPTNO	DNAME	LOC
7369	SMITH	CLERK	7902	08-OCT-01	2400	-	20	10	ACCOUNTING	NEW YORK
7499	ALLEN	SALESMAN	7698	12-DEC-01	4800	600	30	10	ACCOUNTING	NEW YORK
7521	WARD	SALESMAN	7698	14-DEC-01	3750	1000	30	10	ACCOUNTING	NEW YORK
7566	JONES	MANAGER	7839	22-JAN-02	8925	-	20	10	ACCOUNTING	NEW YORK
7654	MARTIN	SALESMAN	7698	20-JUL-02	3750	2800	30	10	ACCOUNTING	NEW YORK
7698	BLAKE	MANAGER	7839	20-FEB-02	8550	-	30	10	ACCOUNTING	NEW YORK
7782	CLARK	MANAGER	7839	31-MAR-02	7350	-	10	10	ACCOUNTING	NEW YORK
7788	STUDENT	ANALYST	7566	08-FEB-08	9000	-	20	10	ACCOUNTING	NEW YORK
7839	KING	PRESIDENT	-	08-SEP-02	15000	-	10	10	ACCOUNTING	NEW YORK
7844	TURNER	SALESMAN	7698	30-JUN-02	4500	0	30	10	ACCOUNTING	NEW YORK
7876	ADAMS	CLERK	7788	13-MAR-08	3300	-	20	10	ACCOUNTING	NEW YORK
7900	JAMES	CLERK	7698	24-SEP-02	2850	-	30	10	ACCOUNTING	NEW YORK
7902	FORD	ANALYST	7566	24-SEP-02	9000	-	20	10	ACCOUNTING	NEW YORK
7934	MILLER	CLERK	7782	14-NOV-02	3900	-	10	10	ACCOUNTING	NEW YORK
7369	SMITH	CLERK	7902	08-OCT-01	2400	-	20	20	RESEARCH	DALLAS
7499	ALLEN	SALESMAN	7698	12-DEC-01	4800	600	30	20	RESEARCH	DALLAS
7521	WARD	SALESMAN	7698	14-DEC-01	3750	1000	30	20	RESEARCH	DALLAS
7566	JONES	MANAGER	7839	22-JAN-02	8925	-	20	20	RESEARCH	DALLAS
7654	MARTIN	SALESMAN	7698	20-JUL-02	3750	2800	30	20	RESEARCH	DALLAS
7698	BLAKE	MANAGER	7839	20-FEB-02	8550	-	30	20	RESEARCH	DALLAS
7782	CLARK	MANAGER	7839	31-MAR-02	7350	-	10	20	RESEARCH	DALLAS
7788	STUDENT	ANALYST	7566	08-FEB-08	9000	-	20	20	RESEARCH	DALLAS

Relational Algebra Example: NATURAL JOIN

SQL Worksheet

```
1  SELECT *
2  FROM AP_EMP NATURAL JOIN AP_DEPT
3
4
```

DEPTNO	EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DNAME	LOC
20	7369	SMITH	CLERK	7902	08-OCT-01	2400	-	RESEARCH	DALLAS
30	7499	ALLEN	SALESMAN	7698	12-DEC-01	4800	600	SALES	CHICAGO
30	7521	WARD	SALESMAN	7698	14-DEC-01	3750	1000	SALES	CHICAGO
20	7566	JONES	MANAGER	7839	22-JAN-02	8925	-	RESEARCH	DALLAS
30	7654	MARTIN	SALESMAN	7698	20-JUL-02	3750	2800	SALES	CHICAGO
30	7698	BLAKE	MANAGER	7839	20-FEB-02	8550	-	SALES	CHICAGO
10	7782	CLARK	MANAGER	7839	31-MAR-02	7350	-	ACCOUNTING	NEW YORK
20	7788	STUDENT	ANALYST	7566	08-FEB-08	9000	-	RESEARCH	DALLAS
10	7839	KING	PRESIDENT	-	08-SEP-02	15000	-	ACCOUNTING	NEW YORK
30	7844	TURNER	SALESMAN	7698	30-JUN-02	4500	0	SALES	CHICAGO
20	7876	ADAMS	CLERK	7788	13-MAR-08	3300	-	RESEARCH	DALLAS
30	7900	JAMES	CLERK	7698	24-SEP-02	2850	-	SALES	CHICAGO
20	7902	FORD	ANALYST	7566	24-SEP-02	9000	-	RESEARCH	DALLAS
10	7934	MILLER	CLERK	7782	14-NOV-02	3900	-	ACCOUNTING	NEW YORK

Download CSV

14 rows selected.

Relational Algebra Example: RENAME

```
1  select ename AS "Employee Name", sal "Monthly Salary in USD"
2  from AP_EMP;
```

Employee Name	Monthly Salary in USD
SMITH	2400
ALLEN	4800
WARD	3750
JONES	8925
MARTIN	3750
BLAKE	8550
CLARK	7350
STUDENT	9000
KING	15000
TURNER	4500
ADAMS	3300
JAMES	2850
FORD	9000
MILLER	3900

Relational Algebra vs. Relational Calculus

- **Relational-Algebra:** is procedural query language that provides a sequence of procedures that generates the answer of to the query

- **Relational Calculus :** is non-procedural query language that describes desired information without giving a specific procedure for obtaining the information,. explains what to do but not how to do

Two forms: closely related

✓ **Tuple Relational Calculus (TRC) :** takes values from an entire tuple

✓ **Domain Relational Calculus (DRC):** takes values from attributes

Tuple Relational Calculus

- A nonprocedural query language, where each query is of the form

$$\{t \mid P(t)\}$$

- It is the set of all tuples *t* such that predicate *P* is true for *t*
- *P* is a *formula* similar to that of the predicate calculus
- P(t) = known as Predicate and these are the conditions that are used to fetch t

TRC Examples

Table Loan

LOAN NUMBER	BRANCH NAME	AMOUNT
L33	ABC	10000
L35	DEF	15000
L49	GHI	9000
L98	DEF	65000

Queries-1: Find the loan number, branch name, amount of loans of greater than or equal to 10000 amount.

Queries-2: Find the loan number for each loan of an amount greater or equal to 10000.

Question: 1

{t| t ∈ loan ∧ t[amount]>=10000}

t[amount] is known as tuple variable

Resulting relation:

LOAN NUMBER	BRANCH NAME	AMOUNT
L33	ABC	10000
L35	DEF	15000
L98	DEF	65000

Question: 2

{t| ∃ s ∈ loan(t[loan number] = s[loan number] ∧ s[amount]>=10000)}

Resulting relation:

LOAN NUMBER
L33
L35
L98

Domain Relational Calculus

- A nonprocedural query language equivalent in power to the tuple relational calculus
- Each query is an expression of the form:

$$\{ < x_1, x_2, \ldots, x_n > \mid P(x_1, x_2, \ldots, x_n) \}$$

  - $x_1, x_2, \ldots, x_n$ represent domain variables
  - P represents a formula similar to that of the predicate calculus

DRC Examples

Table Loan

LOAN NUMBER	BRANCH NAME	AMOUNT
L01	Main	200
L03	Main	150
L10	Sub	90
L08	Main	60

Query-1: Find the loan number, branch name, amount of loans of greater than or equal to 100 amount.

Query-2: Find the loan number for each loan of an amount greater or equal to 150.

Query 1:

$$\{\langle l, b, a\rangle \mid \langle l, b, a\rangle \in loan \wedge (a \geq 100)\}$$

Resulting relation:

LOAN NUMBER	BRANCH NAME	AMOUNT
L01	Main	200
L03	Main	150
L10	Sub	90

Query 2:

$$\{\langle l\rangle \mid \exists\, b, a\ (\langle l, b, a\rangle \in loan \wedge (a \geq 150)\}$$

Resulting relation:

LOAN NUMBER
L01
L03