CO527: Advanced Database Systems

Sampath Deegalla

dsdeegalla@pdn.ac.lk, sampath@eng.pdn.ac.lk

References

- Ramez Elmasri and Shamkant B. Navathe, Fundamentals of Database Systems

- Raghu Ramakrishnan and Johannes Gehrke, Database Management Systems
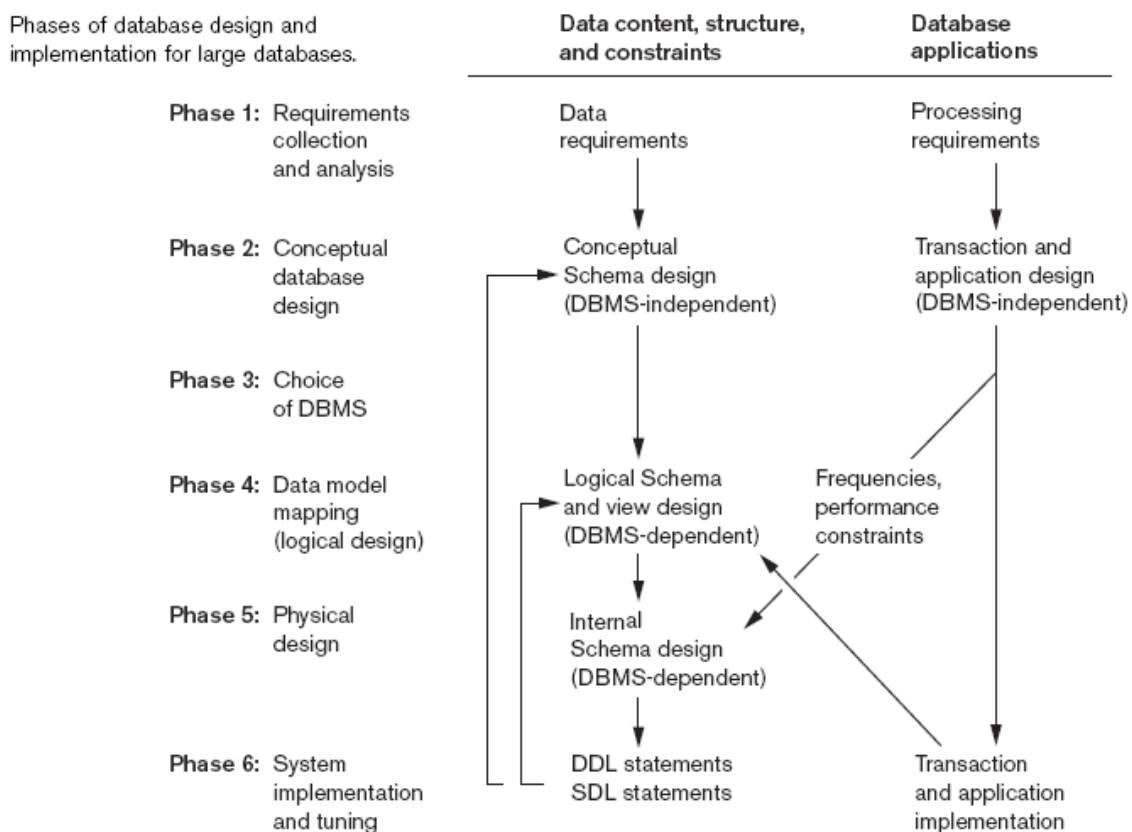
Topics

- Data storage and indexing structures

  - Storage and file structures, indexed, hashed and signature files, Btrees, sparse and dense indexes, variable length records

- Query Optimization and database tuning

  - Translating SQL to relational algebra, Algorithms for relational algebra operations, Query optimization: rule based and cost based approaches, Database design and tuning in relational databases

- Transaction Processing

  - Transactions, ACID properties, Concurrency control, Serialization, Failure and recovery, SQL support for transactions: TCL

- Database security

  - Security issues, Access control: granting and revoking privileges, Multilevel security, Encryption

- Distributed database concepts

  - data fragmentation, replication and allocation; distributed query processing, distributed transaction model, concurrency control, homogeneous and heterogeneous environments

- Object databases and Object-relational databases

  - Current trends of database technology; Object databases : Object identity, Object structure, Type constructors, Encapsulation, Type and Class Hierarchies and Inheritance; Object-relational databases

- Other data models

  - Different types of data: structured, semi-structured and unstructured data; XML; JASON; Comparison of other data models with relational model

- Data models for advanced applications

  - Active, temporal, multimedia and deductive databases

1 Review

Typical DBMS Functionality

- *Define* a database

  - in terms of data types, structures and constraints

- *Construct* or *Load* the Database on a secondary storage medium

- *Manipulating* the database

  - querying, generating reports, insertions, deletions and modifications to its content

- *Concurrent Processing* and *Sharing* by a set of users and programs - yet, keeping all data valid and consistent

- Other features:

  - *Protection* or *Security measures* to prevent unauthorized access
  - *Active processing* to take internal actions on data
  - *Presentation* and *Visualization* of data

Phases of Database Design and Implementation Process

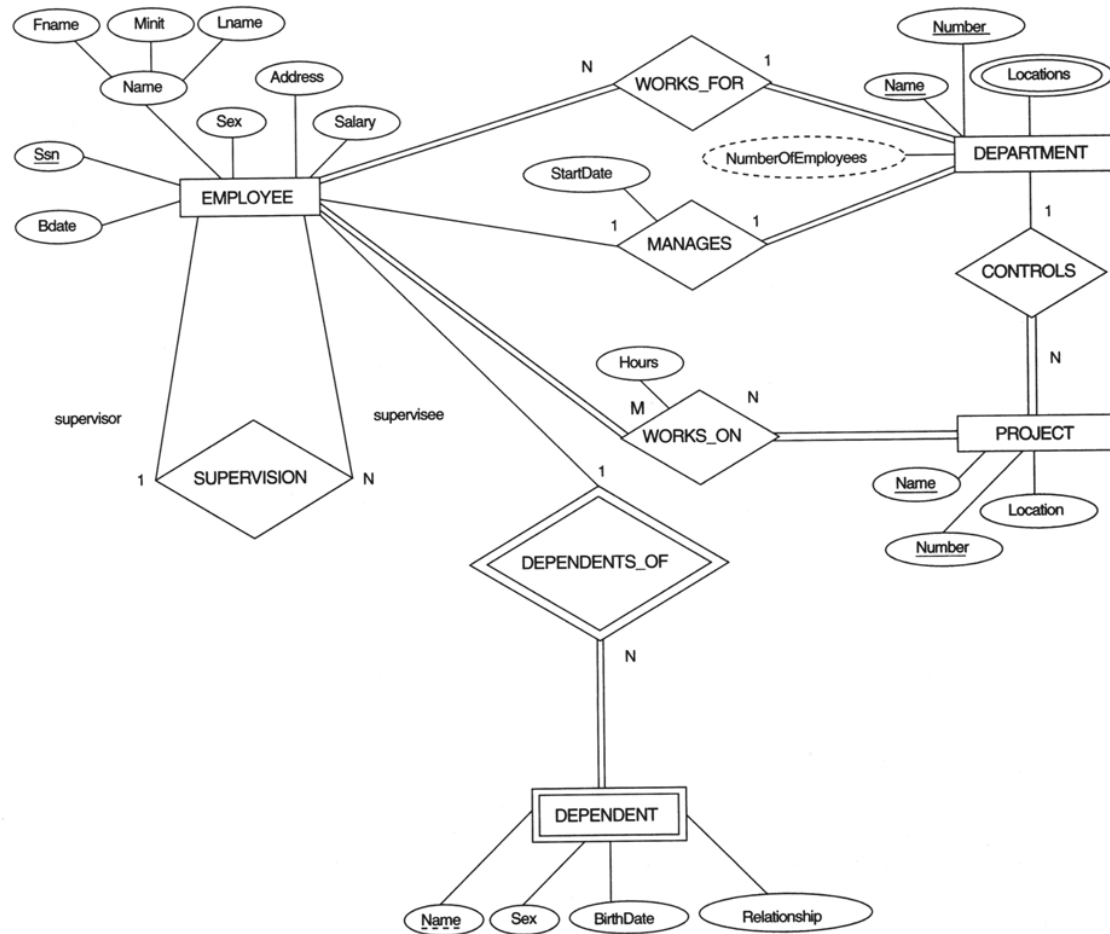| Phases of database design and implementation for large databases.	Data content, structure, and constraints	Database applications
Phase 1: Requirements collection and analysis	Data requirements	Processing requirements
Phase 2: Conceptual database design	Conceptual Schema design (DBMS-independent)	Transaction and application design (DBMS-independent)
Phase 3: Choice of DBMS		
Phase 4: Data model mapping (logical design)	Logical Schema and view design (DBMS-dependent)	Frequencies, performance constraints
Phase 5: Physical design	Internal Schema design (DBMS-dependent)	
Phase 6: System implementation and tuning	DDL statements SDL statements	Transaction and application implementation

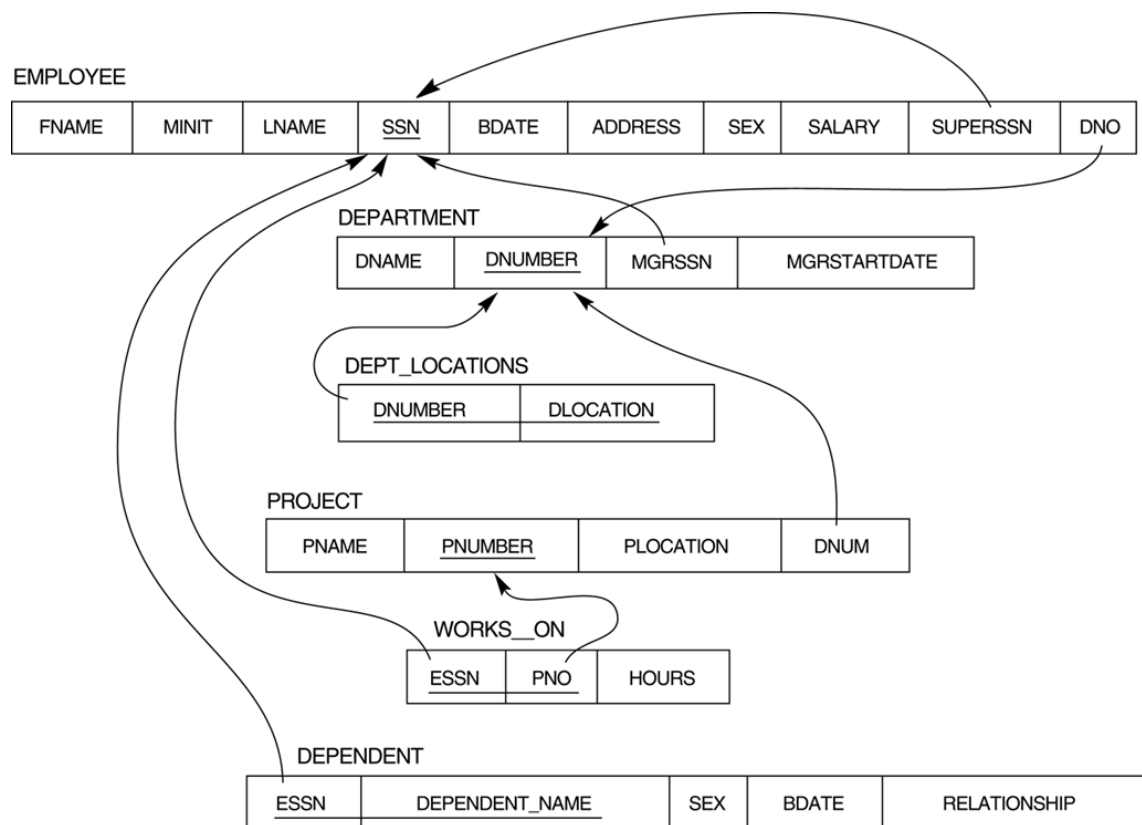Figure 1: An ER schema diagram for the COMPANY database

Figure 2: Referential integrity constraints displayed on the COMPANY relational database schema

EMPLOYEE

FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-07-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	null	1

DEPT_LOCATIONS

DNUMBER	DLOCATION
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

DEPARTMENT

DNAME	DNUMBER	MGRSSN	MGRSTARTDATE
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

WORKS_ON

ESSN	PNO	HOURS
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	null

PROJECT

PNAME	PNUMBER	PLOCATION	DNUM
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

DEPENDENT

ESSN	DEPENDENT_NAME	SEX	BDATE	RELATIONSHIP
333445555	Alice	F	1986-04-05	DAUGHTER
333445555	Theodore	M	1983-10-25	SON
333445555	Joy	F	1958-05-03	SPOUSE
987654321	Abner	M	1942-02-28	SPOUSE
123456789	Michael	M	1988-01-04	SON
123456789	Alice	F	1988-12-30	DAUGHTER
123456789	Elizabeth	F	1967-05-05	SPOUSE

Figure 3: One possible database state for the COMPANY relational database schema

Review Assignment: Create a sql script to populate a MySQL database using the Figure3. Submit your script as exxxxxco525ra.sql where xxxxx stands for your registration number. (e.g. If you number is E/15/450 then the file name should be e15450co525ra.sql). The deadline is 12.02.2020.
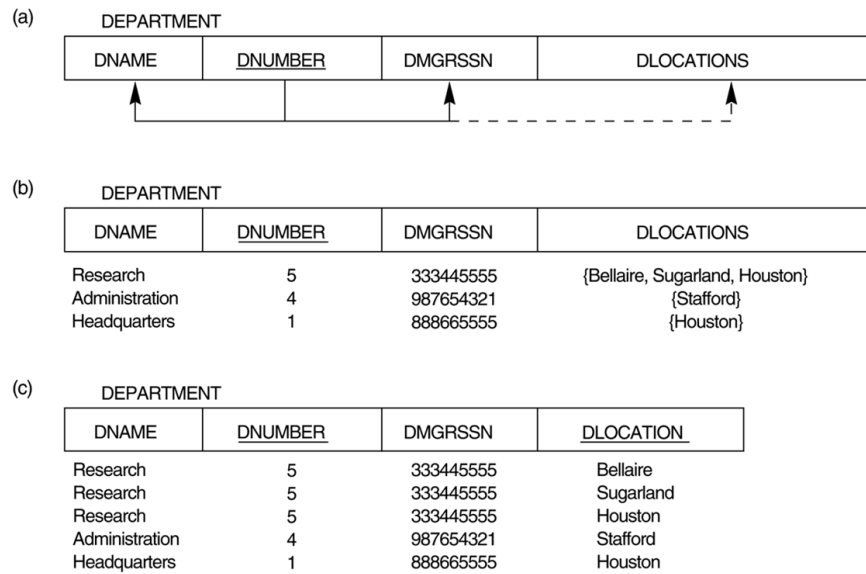
(a) DEPARTMENT

DNAME	DNUMBER	DMGRSSN	DLOCATIONS

(b) DEPARTMENT

DNAME	DNUMBER	DMGRSSN	DLOCATIONS
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

(c) DEPARTMENT

DNAME	DNUMBER	DMGRSSN	DLOCATION
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

Figure 4: Normalization into 1NF

(a) EMP_PROJ

SSN	ENAME	PROJS	
		PNUMBER	HOURS

(b) EMP_PROJ

SSN	ENAME	PNUMBER	HOURS
123456789	Smith,John B.	1	32.5
		2	7.5
666884444	Narayan,Ramesh K.	3	40.0
453453453	English,Joyce A.	1	20.0
		2	20.0
333445555	Wong,Franklin T.	2	10.0
		3	10.0
		10	10.0
		20	10.0
999887777	Zelaya,Alicia J.	30	30.0
		10	10.0
987987987	Jabbar,Ahmad V.	10	35.0
		30	5.0
987654321	Wallace,Jennifer S.	30	20.0
		20	15.0
888665555	Borg,James E.	20	null

(c) EMP_PROJ1

SSN	ENAME

EMP_PROJ2

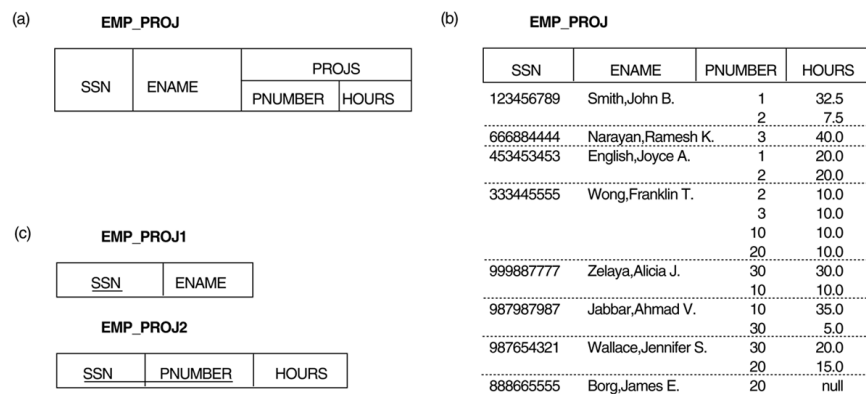| SSN	PNUMBER	HOURS

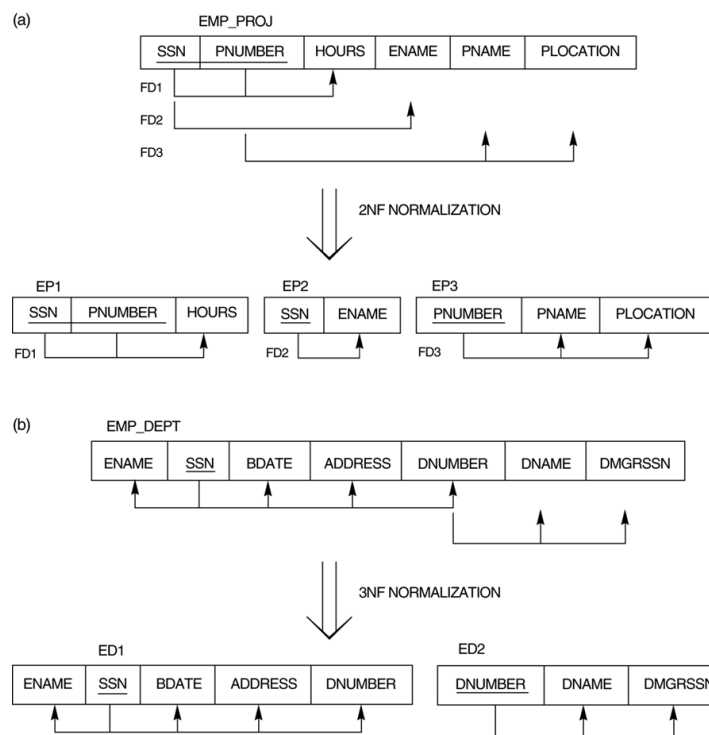Figure 5: Normalization nested relations into 1NF
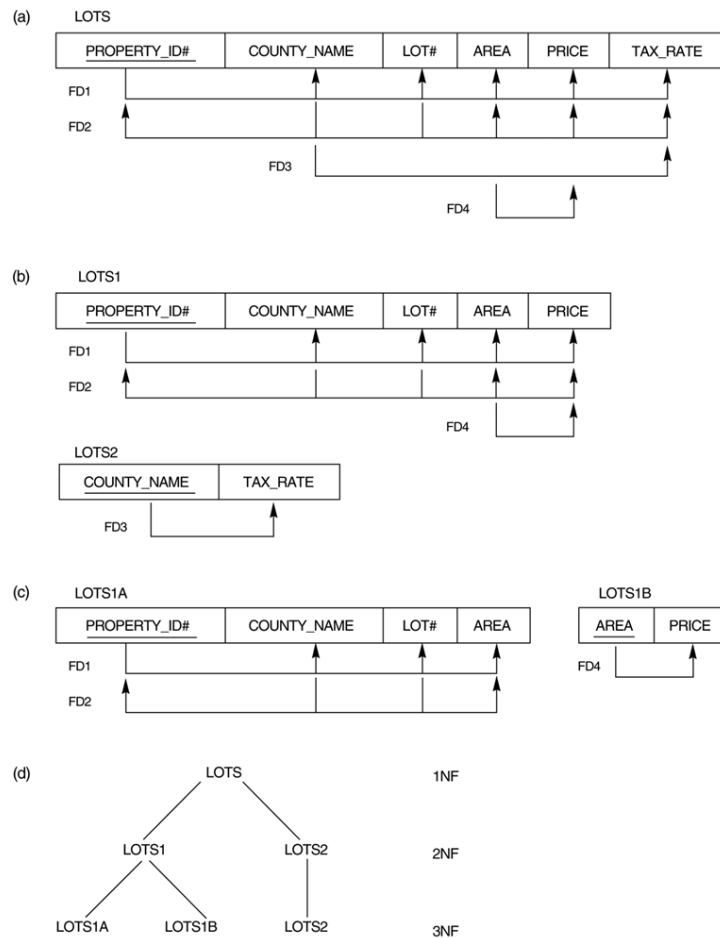
6

Figure 6: Normalizing into 2NF and 3NF

Figure 7: Normalizing into 2NF and 3NF

The relational algebra and calculus

- unary relational operators: SELECT σ, PROJECT π, RENAME

- set theoretic operations: UNION $\cup$, INTERSECTION $\cap$, SET DIFFERENCE $-$, CARTESIAN PRODUCT $\times$

- join operations: JOIN, OUTER JOIN