Relational Database Management System

Education, Training and Assessment

We enable you to leverage knowledge anytime, anywhere!

Infosys® | Building Tomorrow's Enterprise

Copyright Guideline

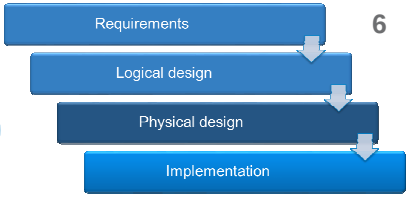
Confidential Information

- This Document is confidential to Infosys Limited. This document contains information and data that Infosys considers confidential and proprietary (“Confidential Information”).
- Confidential Information includes, but is not limited to, the following:
  - Corporate and Infrastructure information about Infosys
  - Infosys’ project management and quality processes
  - Project experiences provided included as illustrative case studies
- Any disclosure of Confidential Information to, or use of it by a third party, will be damaging to Infosys.
- Ownership of all Infosys Confidential Information, no matter in what media it resides, remains with Infosys.
- Confidential information in this document shall not be disclosed, duplicated or used – in whole or in part – for any purpose other than reading without specific written permission of an authorized representative of Infosys.
- This document also contains third party confidential and proprietary information. Such third party information has been included by Infosys after receiving due written permissions and authorizations from the party/ies. Such third party confidential and proprietary information shall not be disclosed, duplicated or used – in whole or in part – for any purpose other than reading without specific written permission of an authorized representative of Infosys.

Course Information

- Course Code: CCFP4.0-RDBMS
- Course Name: Relational Database Management System
- Document Number: RDBMS-02
- Version Number: 4.0

Conversion of entity relationship diagram to relational schema



Data representation in RDBMS(Revisit)

Attributes/Columns/Fields

NULL

Rows/Records/Tuples

custid	custname	custtype	cardtype
C1001	Jeremy	Regular	
C1002	Larry	Privileged	Silver
C1003	Henry	Privileged	Gold
C1004	Liza	Privileged	Platinum
C1005	Allen	Regular	

No. of Records/ Rows/ Tuples: Cardinality of the Relation

Relation/ Table

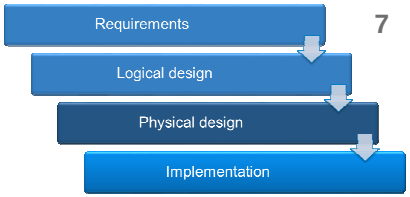
Name of the relation

No. of Attributes/Columns/Fields: Degree of the Relation

Attributes of the relation

Relation is usually represented as:

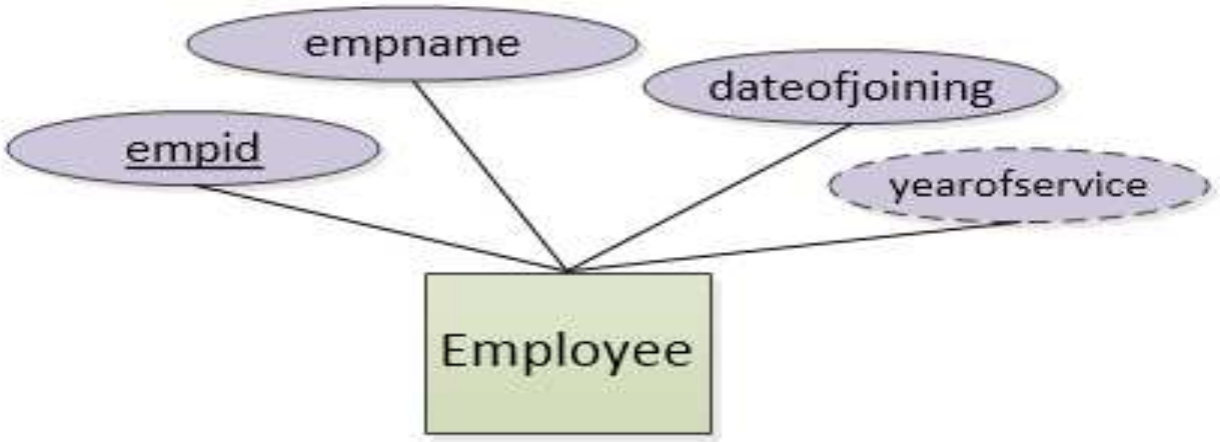
customer (custid, custname, custtype, cardtype)



Conversion of ER model to relational schema

- **Schema**
  - A description of a database
  - Specifies the relations, their attributes and the domains of the attributes
- **Conversion guidelines**
  - Each entity in ER diagram becomes a table in relational schema
  - Each single-valued attribute in ER diagram becomes a column of the table
  - Derived attributes of entities are ignored
  - Composite attributes of an entity are represented by its equivalent parts
  - Multi-valued attributes are kept in a separate table
  - The key attribute of an entity is chosen as the primary key of the table
  - Converting relationships is based on degree and cardinality of relationship

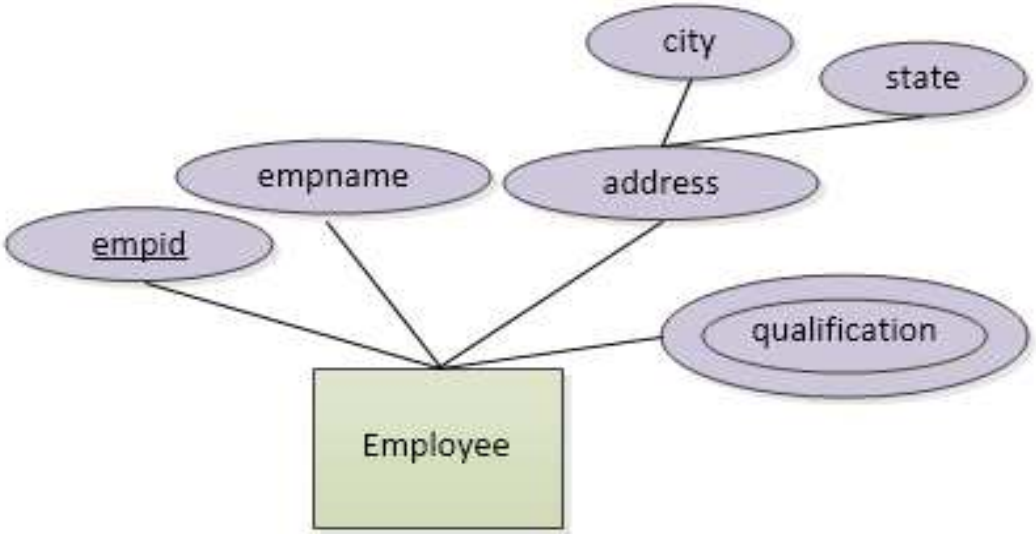
Example: Strong entity (1 of 2)



Relational Schema:

employee (empid, empname, dateofjoining)

Example: Strong entity (2 of 2)

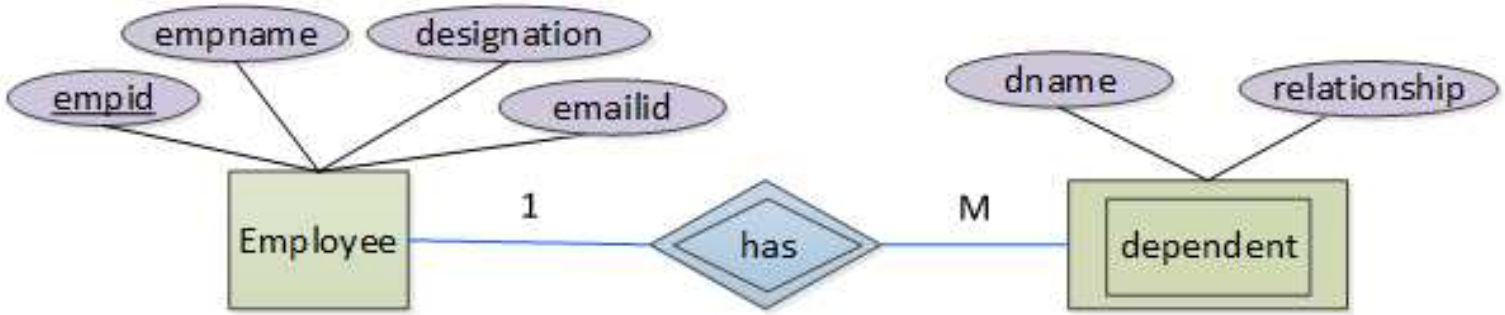


Relational Schema:

employee(empid, empname, state, city)

employeequalification (empid, qualification)

Example: Weak entity

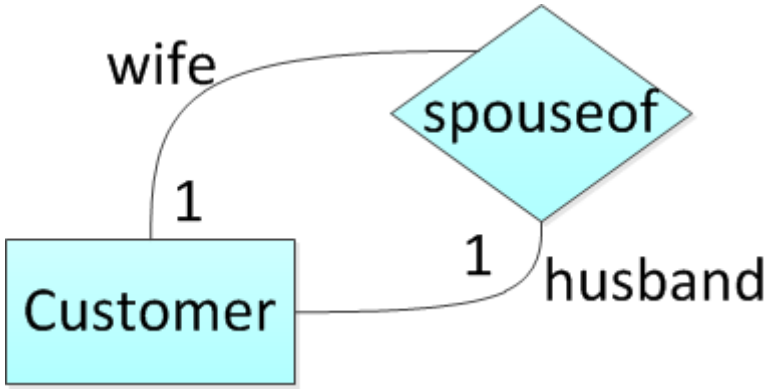


Relational Schema:

employee(empid,empname, designation, emailid)

dependent (empid, dname, relationship)

Example: Unary 1:1

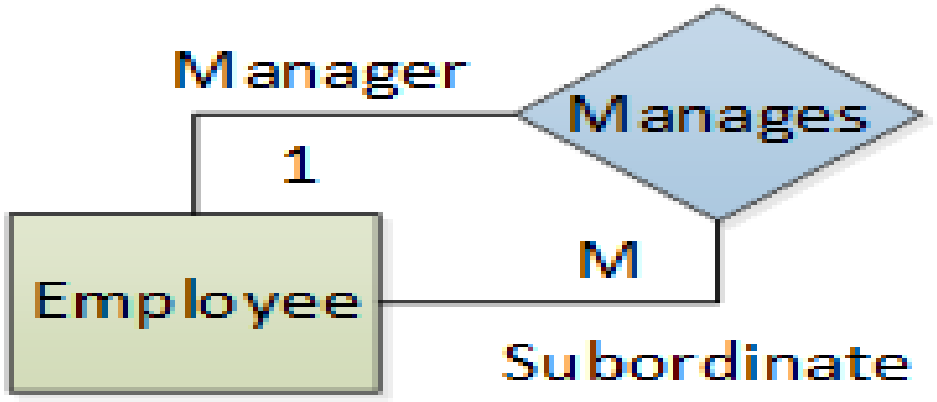


Relational Schema:

customer(customerid, customername,…spouse)

Example: Unary 1:M

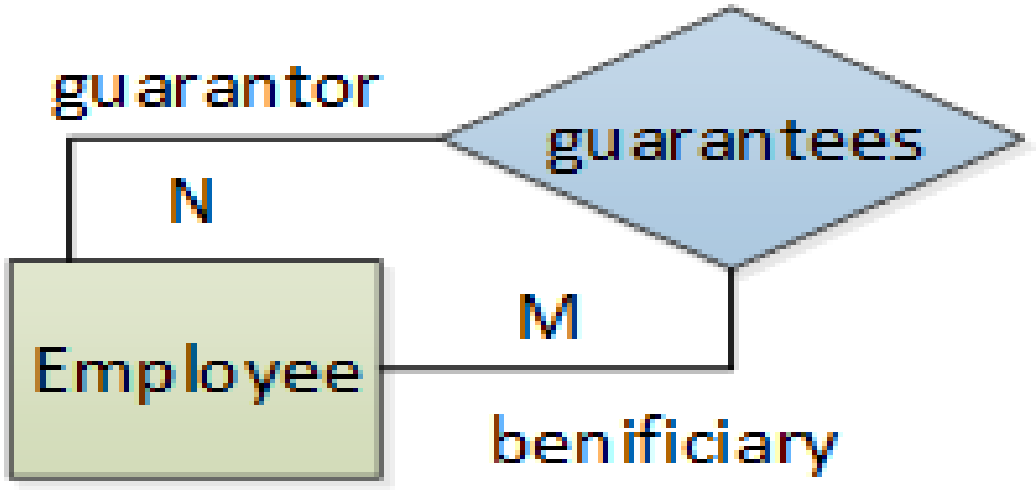
- The primary key of the table will itself become foreign key of the same table



Relational Schema:

employee(empid, empname, designation, ..., managerid)

Example: Unary M:N



Relational Schema:

employee(empid, empname, designation,….)

guaranty (guarantorid, beneficiaryid)

Example: Binary 1:1

- The key attribute of any of the participating entities in a relationship can become a foreign key in the other participating entity

Relational Schema:

employee(empid, empname, designation, ……., salary)

retailoutlet (retailoutletid, retailoutletlocation, retailouletmanagerid)

OR

employee(empid, empname, designation,……,salary, retailoutletid)

retailoutlet (retailoutletid, retailoutletlocation)

Example: Binary 1:M

- The key attribute of entity on the “1” side of the relationship becomes a foreign key of entity towards the “M” side

Relational Schema:

supplier (supplierid, suppliername, suppliercontactno, supplieremailid)

quotation (quotationid, itemcode, quotedprice, supplierid)

Example: Binary M:N

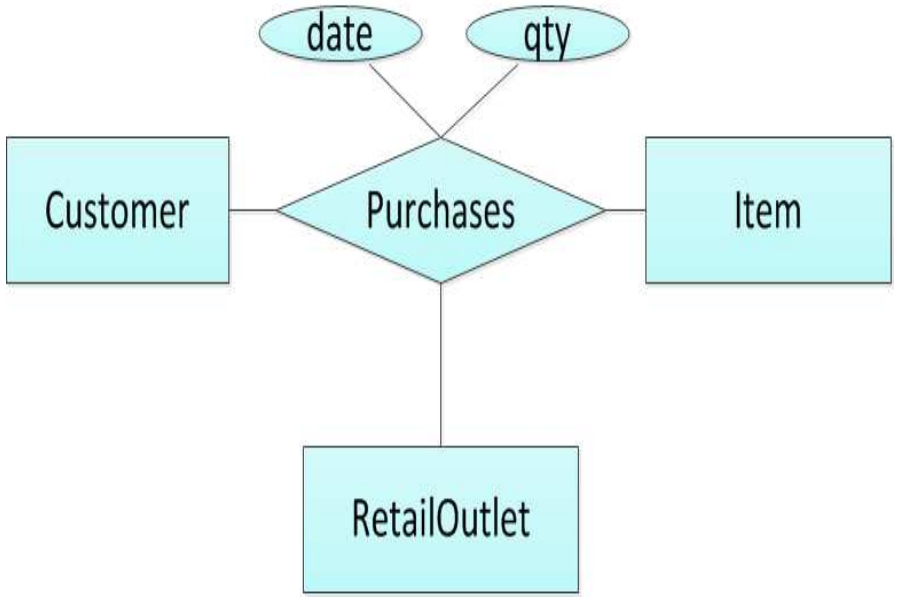
Relational Schema:

customer (customerid,customertype,customername,emailid,contactno,address)

retailoutlet (retailoutletid, retailoutletlocation)

purchasesfrom (customerid, retailoutletid)

Example: Ternary relationship



Relational Schema:

customer (customerid,customertype,customername,emailid,contactno,address)

retailoutlet (retailoutletid, retailoutletlocation)

item (itemid, description, reorderlevel, itemclass)

purchases(customerid, retailoutletid, itemcode, date, qty)

Guided Activity: CC-FP4.0-RDBMSAssignments, Database Basics, Assignment 5 (Estimated Time: 45 mins.)

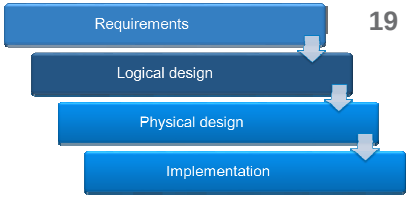
Summary:

ER model

Conversion of ER model to relational schema

Normalization

Database life cycle – Logical design



Top down approach (Entity – Relationship (ER) model)

- This approach is used when application **requirements are clear**
- Represents the application requirements in a pictorial form
- The real world objects and their corresponding attributes are identified from the requirements – hence it is top down
- This model helps in
  - analysis and design
  - re-validating the requirements

Bottom up approach (Normalization)

- This approach is used when application **requirements are not very clear**
- First define the required data items and then group the related data items
- Further refinement may be carried out depending on the application need

Normalization - A Bottom Up Approach

Functional dependency

Guided Activity: CC-FP4.0-RDBMSAssignments - Database basics, Assignment 6 (Estimated Time: 30 mins.)

Solution to Guided activity:

- The identified **Relation** with respective attributes as per the requirement is:
  - Name of the relation is : **retailoutletstock**
  - Attributes of the relation are**:**

retailoutletid, itemcode, description, retailoutletlocation (street name, city, pincode), qtyavailable, retailunitprice, itemclass

The above relation is represented as :

retailoutletstock (retailoutletid, itemcode, description, retailoutletlocation(streetname, city, pincode), qtyavailable, retailunitprice, itemclass)

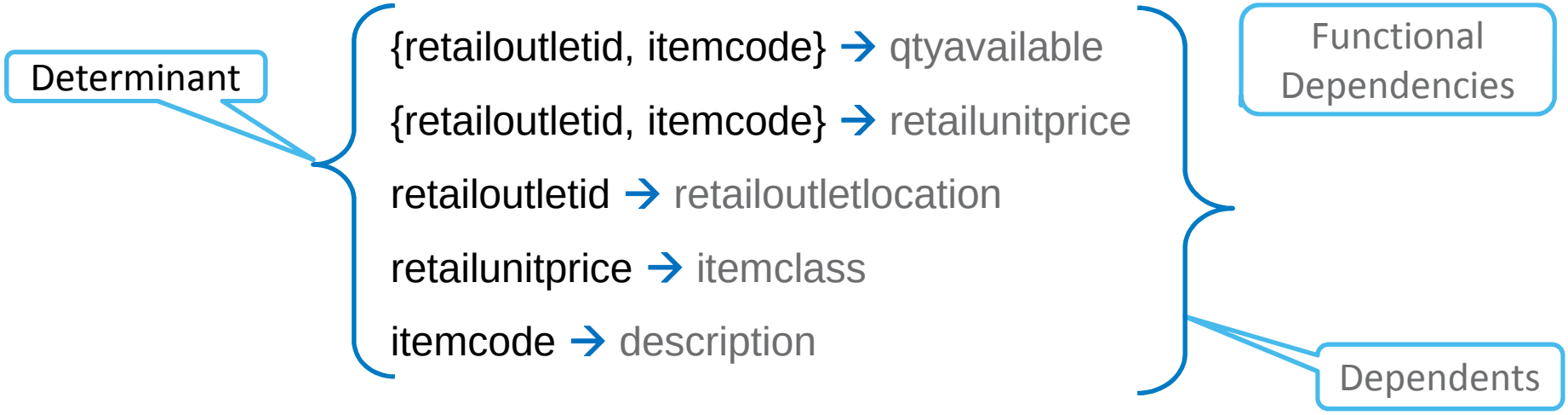
Functional dependency (contd.)

Dependency in a Relation:

- An attribute of a relation can be determined by knowing one/more attributes of the same relation
- The **attribute** which determines the value of other attributes is known as "Determinant"

Guided Activity: Identify all the possible dependencies in the relation retailoutletstock
(Estimated Time: 10 mins.)

retailoutletstock (retailoutletid, itemcode, description, retailoutletlocation(streetname, city, pincode), qtyavailable, retailunitprice, itemclass)

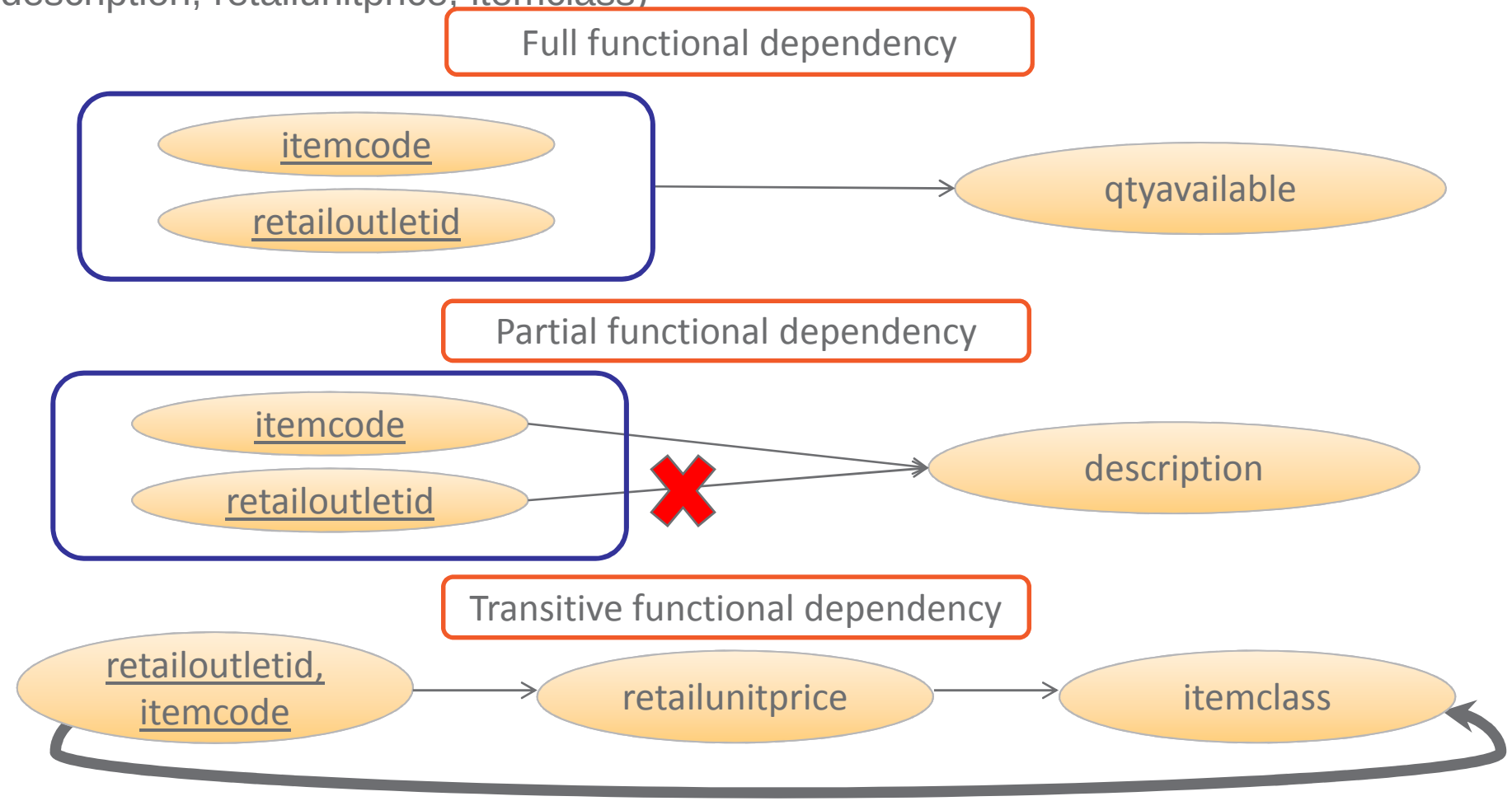


Functional dependency (contd.)

- In a given relation R, A and B are attributes.
- Attribute B is **functionally dependent** on attribute A if each value of A determines **EXACTLY ONE** value of B, which is represented as A -> B (A can be composite in nature).
- A is called **determinant** and B is called **dependent**

Functional dependency: types

retailoutletstock (retailoutletid, itemcode, retailoutletlocation, qtyavailable, description, retailunitprice, itemclass)



Modification anomalies

retailoutletstock

retailoutletid	retailoutletlocation	itemcode	description	qtyavailable	retailunitprice	itemclass
R1001	King Street, California, 123456	I1001	Britannia Marie Gold Cookies	25	1600	B
R1001	King Street, California, 123456	I1002	Brooke Bond Tea	30	850	C
R1002	Victoria Street, New York, 654897	I1001	Britannia Marie Gold Cookies	25	1650	B
R1002	Victoria Street, New York, 654897	I1003	Best Rice	50	6600	A
R1002	Victoria Street, New York, 654897	I1004	Modern Bread	20	5200	A
R1003	Saint James Avenue, Dallas, 426987	I1005	Bourbon Cookies	20	5500	A

- What will happen if we try to insert(add) the details of a new retail outlet that currently has no items in its stock?

NULL values would be inserted into the itemdetails column, which is not preferable

Modification anomalies

retailoutletstock

retailoutletid	retailoutletlocation	itemcode	description	qtyavailable	retailunitprice	itemclass
R1001	King Street, California, 123456	I1001	Britannia Marie Gold Cookies	25	1600	B
R1001	King Street, California, 123456	I1002	Brooke Bond Tea	30	850	C
R1002	Victoria Street, New York, 654897	I1001	Britannia Marie Gold Cookies	25	1650	B
R1002	Victoria Street, New York, 654897	I1003	Best Rice	50	6600	A
R1002	Victoria Street, New York, 654897	I1004	Modern Bread	20	5200	A
R1003	Saint James Avenue, Dallas, 426987	I1005	Bourbon Cookies	20	5500	A

- What will happen if we try to delete the item details for an Item I1005?

The details of the retail outlet R1003 will also be lost from the database

Modification anomalies

retailoutletstock

retailoutletid	retailoutletlocation	itemcode	description	qtyavailable	retailunitprice	itemclass
R1001	King Street, California, 123456	I1001	Britannia Marie Gold Cookies	25	1600	B
R1001	King Street, California, 123456	I1002	Brooke Bond Tea	30	850	C
R1002	Victoria Street, New York, 654897	I1001	Britannia Marie Gold Cookies	25	1650	B
R1002	Victoria Street, New York, 654897	I1003	Best Rice	50	6600	A
R1002	Victoria Street, New York, 654897	I1004	Modern Bread	20	5200	A
R1003	Saint James Avenue, Dallas, 426987	I1005	Bourbon Cookies	20	5500	A

- How many rows will be updated if the retail outlet location of R1002 is changed from Victoria Street to Saint John Street?

3 rows will be updated.

Modification anomalies

retailoutletstock

retailoutletid	retailoutletlocation	itemcode	description	qtyavailable	retailunitprice	itemclass
R1001	King Street, California, 123456	I1001	Britannia Marie Gold Cookies	25	1600	B
R1001	King Street, California, 123456	I1002	Brooke Bond Tea	30	850	C
R1002	Victoria Street, New York, 654897	I1001	Britannia Marie Gold Cookies	25	1650	B
R1002	Victoria Street, New York, 654897	I1003	Best Rice	50	6600	A
R1002	Victoria Street, New York, 654897	I1004	Modern Bread	20	5200	A
R1003	Saint James Avenue, Dallas, 426987	I1005	Bourbon Cookies	20	5500	A

- What are the details we need to insert when new items are supplied to a retail outlet?

Apart from all necessary details, retailoutletlocation will also be inserted, which is redundant

Modification anomalies

retailoutletstock

retailoutletid	retailoutletlocation	itemcode	description	qtyavailable	retailunitprice	itemclass
R1001	King Street, California, 123456	I1001	Britannia Marie Gold Cookies	25	1600	B
R1001	King Street, California, 123456	I1002	Brooke Bond Tea	30	850	C
R1002	Victoria Street, New York, 654897	I1001	Britannia Marie Gold Cookies	25	1650	B
R1002	Victoria Street, New York, 654897	I1003	Best Rice	50	6600	A
R1002	Victoria Street, New York, 654897	I1004	Modern Bread	20	5200	A
R1003	Saint James Avenue, Dallas, 426987	I1005	Bourbon Cookies	20	5500	A

Following challenges are observed in the relation:
insert , delete, update anomalies and **data redundancy**

- Functional dependencies may lead to anomalies.
- To minimize anomalies there is need to refine functional dependencies.
- This process is called **Normalization**

Definition: Normalization

- It is a step-by-step process where in a complex relation is decomposed into simple relations
- It is a formal process of achieving a good database design
- "**Normal Forms**" (NF) are the different stages of normalization
  - 1NF
  - 2NF
  - 3NF

First Normal Form: 1NF

- A relation R is in 1NF if and only if :
  - All the attributes of R are atomic in nature
  - There should not be any multi valued attribute

retailoutletstock

retailoutletid	street	city	zipcode	itemcode	description	qtyavailable	retailunitprice	itemclass
R1001	King Street	California	123456	I1001	Britannia Marie Gold Cookies	25	1600	B
R1001	King Street	California	123456	I1002	Brooke Bond Tea	30	850	C
R1002	Victoria Street	New York	654897	I1001	Britannia Marie Gold Cookies	25	1650	B
R1002	Victoria Street	New York	654897	I1003	Best Rice	50	6600	A
R1002	Victoria Street	New York	654897	I1004	Modern Bread	20	5200	A
R1003	Saint James Avenue	Dallas	426987	I1005	Bourbon Cookies	20	5500	A

Second Normal Form: 2NF

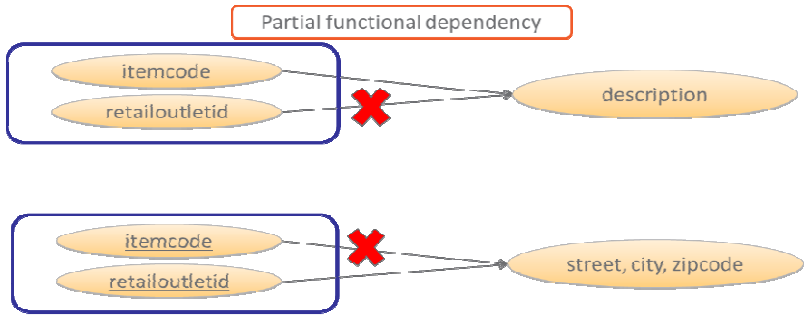
- A relation R is in second normal form if and only if :
  - R is already in 1NF, and
  - There is no partial dependency in R which exists between non-key attributes and key attributes
- To make a table 2NF compliant, remove all such partial dependencies and decompose the relation

item

itemcode	description
I1001	Britannia Marie Gold Cookies
I1002	Brooke Bond Tea
I1003	Best Rice
I1004	Modern Bread
I1005	Bourbon Cookies

retailoutlet

retailoutletid	street	city	zipcode
R1001	King Street	California	123456
R1002	Victoria Street	New York	654897
R1003	Saint James Avenue	Dallas	426987



retailstockdetails

retailoutletid	itemcode	qtyavailable	retailunitprice	itemclass
R1001	I1001	25	1600	B
R1001	I1002	30	850	C
R1002	I1001	25	1650	B
R1002	I1003	50	6600	A
R1002	I1004	20	5200	A
R1003	I1005	20	5500	A

Normalization (contd.)

retailstockdetails

retailoutletid	itemcode	qtyavailable	retailunitprice	itemclass
R1001	I1001	25	1600	B
R1001	I1002	30	850	C
R1002	I1001	25	1650	B
R1002	I1003	50	6600	A
R1002	I1004	20	5200	A
R1003	I1005	20	5500	A

Are all anomalies eliminated?

Normalization (contd.)

retailstockdetails

retailoutletid	itemcode	qtyavailable	retailunitprice	itemclass
R1001	I1001	25	1600	B
R1001	I1002	30	850	C
R1002	I1001	25	1650	B
R1002	I1003	50	6600	A
R1002	I1004	20	5200	A
R1003	I1005	20	5500	A

- What will happen if we try to delete the record containing item code I1002?

The definition of item class C will also be lost from the database

Normalization (contd.)

retailstockdetails

retailoutletid	itemcode	qtyavailable	retailunitprice	itemclass
R1001	I1001	25	1600	B
R1001	I1002	30	850	C
R1002	I1001	25	1650	B
R1002	I1003	50	6600	A
R1002	I1004	20	5200	A
R1003	I1005	20	5500	A

- What if there is a change in the business rule for the item class , say the retail unit price range has been increased by 1000 for every item class?

retailunitprice	itemclass
<1000	C
>=1000 and <5000	B
>=5000 and <100000	A

Multiple rows will be updated

Normalization (contd.)

retailstockdetails

retailoutletid	itemcode	qtyavailable	retailunitprice	itemclass
R1001	I1001	25	1600	B
R1001	I1002	30	850	C
R1002	I1001	25	1650	B
R1002	I1003	50	6600	A
R1002	I1004	20	5200	A
R1003	I1005	20	5500	A

insert , delete, update anomalies still exist which needs to be eliminated. This necessitates for further normalization

The root cause of these anomalies is transitive dependency between retailoutletid, itemcode, retailunitprice and itemclass. This needs to be eliminated

Third Normal Form: 3NF

- A relation R is said to be in the Third Normal Form (3NF) if and only if :
  - It is in 2NF and
  - There is no transitive dependency which exists between key attributes and non-key attributes through another non-key attribute
- To make a table 3NF compliant, we have to remove all such transitive dependencies by decomposing the relation

retailstock

retailoutletid	itemcode	qtyavailable	retailunitprice
R1001	I1001	25	1600
R1001	I1002	30	850
R1002	I1001	25	1650
R1002	I1003	50	6600
R1002	I1004	20	5200
R1003	I1005	20	5500

itemclass

retailunitprice	itemclass
1600	B
850	C
1650	B
6600	A
5200	A
5500	A

Improved database design

- Storing itemclass for every retailunitprice is not the efficient way, so it is more appropriate to store it in the following way:

itemclassrange

minretailunitprice	maxretailunitprice	itemclass
1	999	C
1000	4999	B
5000	99999	A

Normalization guidelines

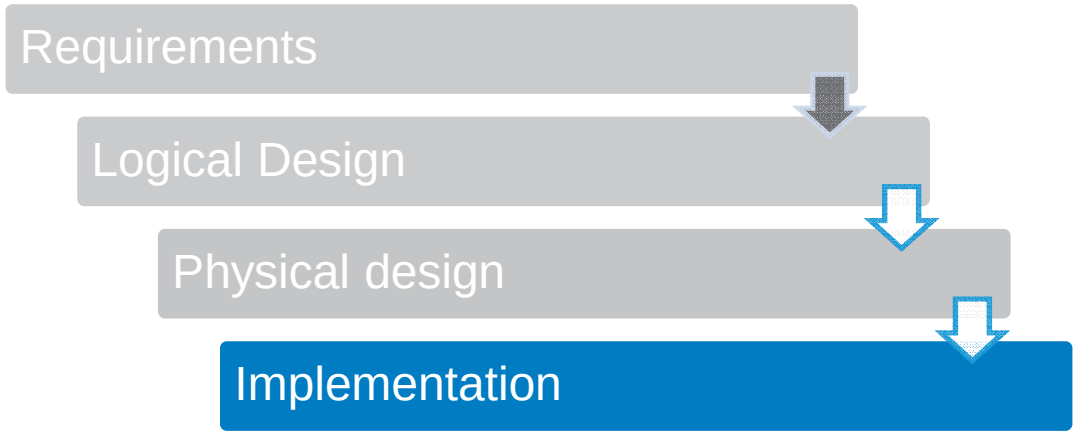
- Depending on the business requirements, the tables can be normalized up to 2nd normal form or 3rd normal form
- Tables in 3 NF are preferred in applications with extensive data modifications,
- Tables in 2 NF are preferred in applications with extensive data retrieval
  - Reason: retrieving data from multiple tables is a costly operation
- Converting the tables from higher normal form to lower normal form is called "Denormalization"

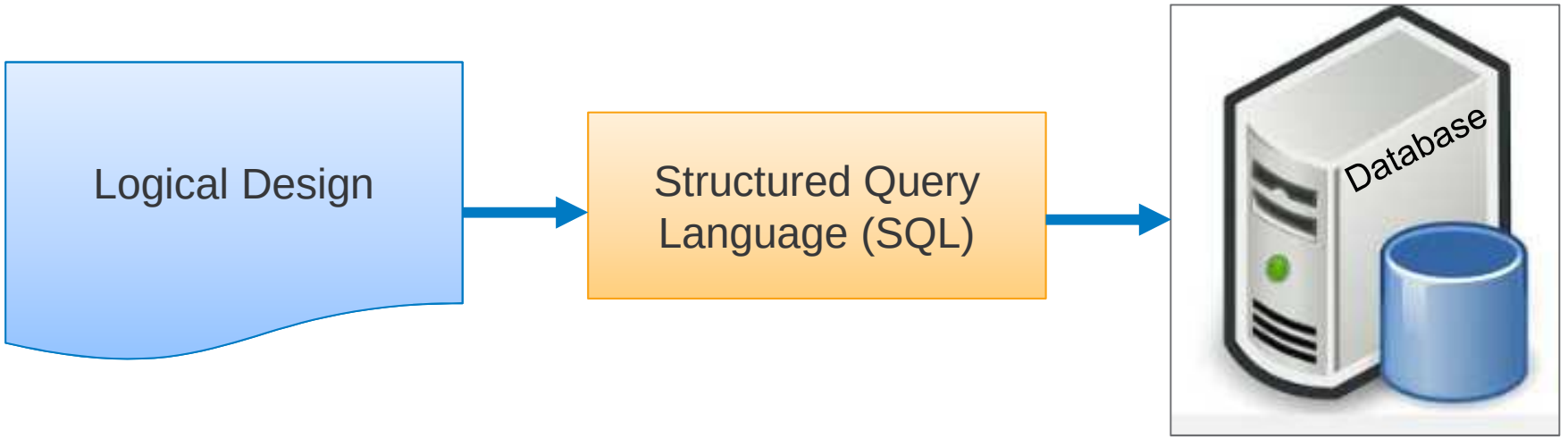
Normalization

Guided Activity: CC-FP4.0-RDBMSAssignments - Database basics, Assignment 7 (Estimated Time: 50 mins.)

- To summarize:
  - Need for normalization
  - Functional dependencies
  - Types of functional dependencies
  - Normal forms

Database lifecycle revisited

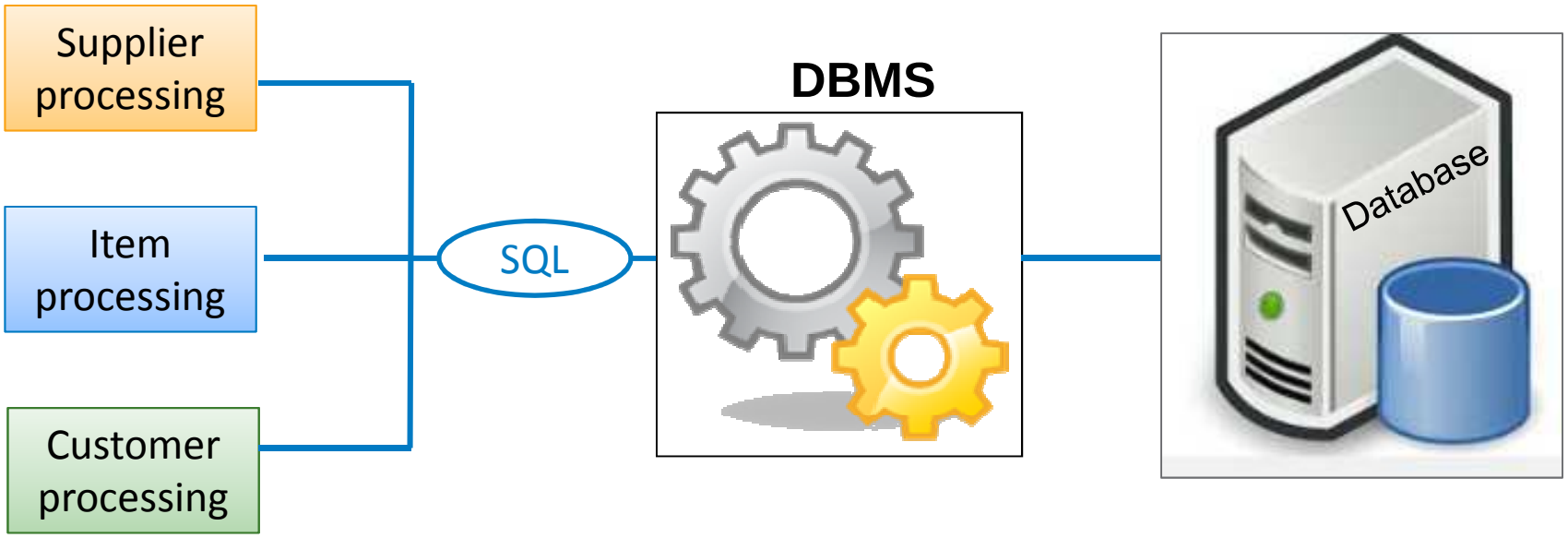




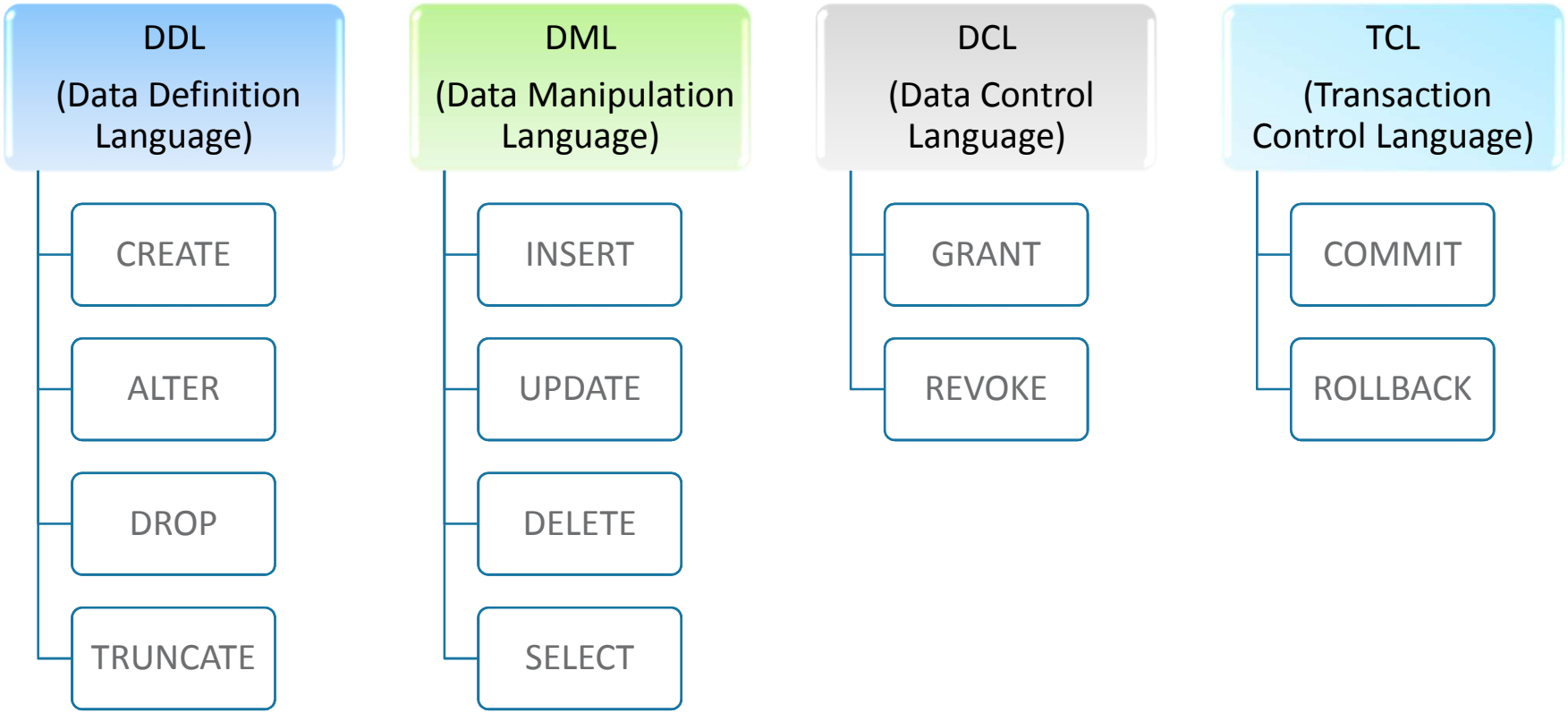
SQL

Demo: EasyShop application

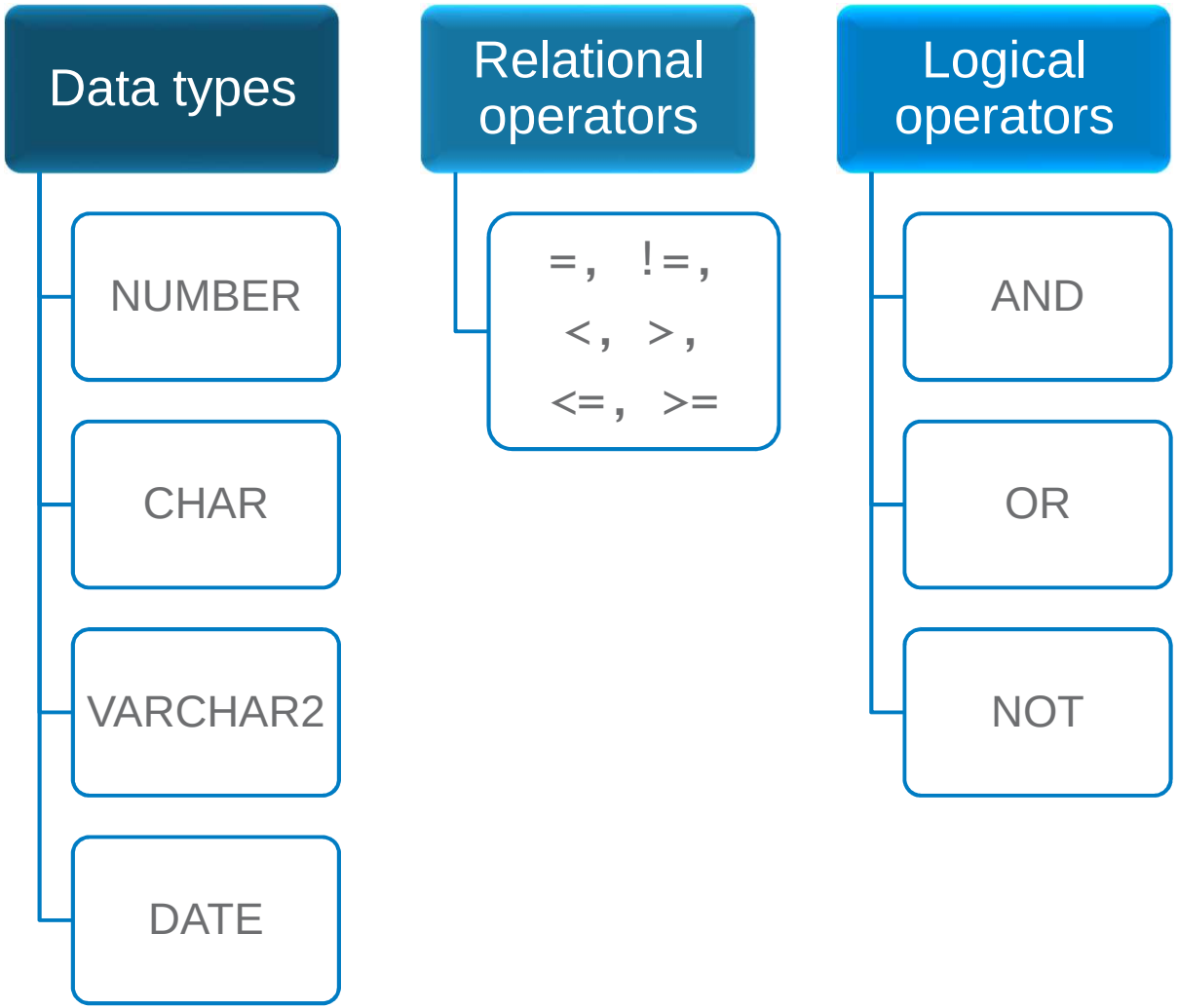
Modules in EasyShop



SQL statements



Data types and operators in SQL



DDL and DML

Constraints - Types:

- UNIQUE
- NOT NULL
- PRIMARY KEY
- CHECK
- FOREIGN KEY

Constraints - Levels:

- Column level
- Table level

Guided Activity / Demos: CC-FP4.0-RDBMSAssignments - SQL basics, Assignments 1, 2 and 3 (Estimated Time: 80 mins.)

ALTER TABLE

- ALTER TABLE statement is used to modify the structure of an existing table
- Modification of structure includes
  - Adding/Dropping columns or constraints
  - Modifying the data type or size of columns

Guided Activity: CCFP4.0-RDBMSAssignments - SQL Basics, Assignment 4, 5 (Estimated Time: 50 mins.)

SELECT

Operators:

- IN
- LIKE

- IS NULL
- BETWEEN AND

Guided Activity: To familiarize with the tables of EasyShop retail application, read the documents:
CCFP4.0-RDBMS-EasyShop Retail Application_DB_Structure.docx and
CCFP4.0-RDBMS-EasyShop Retail Application_DB_Design.docx (Estimated Time: 20 mins.)

Guided Activity: CCFP4.0-RDBMSAssignments - SQL Basics, Assignment 7, 8
(Estimated Time: 90 mins.)

DISTINCT, ORDER BY

- DISTINCT is used to get the discrete values from column(s)
- ORDER BY is used to sort the records while retrieving
  - The data stored in the table will not be sorted
  - By default the order is ASCENDING

Guided Activity: CCFP4.0-RDBMSAssignments - SQL Basics, Assignment 9, 10 (Estimated Time: 60 mins.)

CASE

CASE

Requirement:

- Salary increments are to be given to all employees of EasyShop based on their role. The percentage increase is given below:

Designation	Increment in %
Security	25
Billing Staff	20
Administrator	10
Manager	5
Others	2

- Write a query to display the employee details along with their increased salary

Guided Activity : CCFP4.0-RDBMSAssignments - SQL Basics, Assignment 11, 12 (Estimated Time: 60 mins.)

SQL functions

Single row functions

- **Numeric**
  - ROUND
  - CEIL
  - FLOOR
  - ABS
- **Character**
  - LOWER
  - UPPER
  - SUBSTR
  - LENGTH
- **DATE**
  - ADD_MONTHS
  - MONTHS_BETWEEN
  - TO_CHAR
- **Conversion**
  - TO_CHAR
  - TO_NUMBER
  - TO_DATE
- **Others**
  - NVL

Multi row functions (Aggregate functions)

- MIN
- MAX
- SUM
- AVG
- COUNT

Guided activity

- Retrieve the retail outlet id and associated manager id to the retail outlet. If a retail outlet doesn't have a manger display "Manager not allocated".

```
SELECT retailoutletid, NVL(TO_CHAR(retailoutletmanagerid),
'Manager not allocated') retailoutletmanagerid
FROM retailoutlet;
```

Guided Activity: CCFP4.0-RDBMSAssignments - SQL Basics, Assignment 13, 14
(Estimated Time: 60 mins.)

GROUP BY

GROUP BY

Manager of EasyShop needs a report of total amount collected daily

Demo: EasyShop application → Bill Amount – All Dates report

purchasebill

billid	retailoutletid	customerid	billamount	billdate
10001	R1001	1001	10355	02-JUN-2011
10002	R1001	1002	9021	02-JUN-2011
10003	R1001	1002	4900	03-JUN-2011
10004	R1002	1003	2940	03-JUN-2011
10005	R1002	1005	7663	04-JUN-2011
10006	R1002	1004	8730	04-JUN-2011
10007	R1002	1001	4900	05-JUN-2011

Expected output:

billdate	SUM(billamount)
02-JUN-2011	19376
03-JUN-2011	7840
04-JUN-2011	16393
05-JUN-2011	4900

```
SELECT billdate, SUM(billamount) FROM purchasebill GROUP BY billdate;
```

Working of GROUP BY

```
SELECT billdate, SUM(billamount) FROM purchasebill GROUP BY
billdate;
```

purchasebill

billid	retailoutletid	customerid	billamount	billdate
10001	R1001	1001	10355	02-JUN-2011
10002	R1001	1002	9021	02-JUN-2011
10003	R1001	1002	4900	03-JUN-2011
10004	R1002	1003	2940	03-JUN-2011
10005	R1002	1005	7663	04-JUN-2011
10006	R1002	1004	8730	04-JUN-2011
10007	R1002	1001	4900	05-JUN-2011

billdate	SUM(billamount)
02-JUN-2011	19376
03-JUN-2011	7840
04-JUN-2011	16393
05-JUN-2011	4900

HAVING

- Manager of EasyShop needs a report, that has daily total collection amount greater than 5000, billed everyday

  Expected output:

Write a SQL query for this requirement

billdate	SUM(billamount)
02-JUN-2011	19376
03-JUN-2011	7840
04-JUN-2011	16393

Demo: EasyShop application → Bill Amount – Purchase Based report

```
SELECT billdate, SUM(billamount) FROM purchasebill GROUP
BY billdate WHERE SUM(billamount) > 5000;
```

Solution:

```
SELECT billdate, SUM(billamount) FROM purchasebill GROUP
BY billdate HAVING SUM(billamount) > 5000;
```

Guided Activity: CCFP4.0-RDBMSAssignments – Assignments on GROUP BY and HAVING, Assignment 1a, 1b (Estimated Time: 40 mins.)

Summary

- Conversion of ER model to relational schema
- Normalization
- Introduction to SQL
- SQL data types and operators
- DDL statements
- DML statements
- Operators in SQL
- Usage of SELECT Statements with
  - Operators
  - DISTINCT
  - ORDER BY
  - CASE

Summary

- SQL Functions
  - Single row
  - Multi row (Aggregate)
- GROUP BY
- HAVING

Self-Study

Refer to NPTEL course: http://nptel.ac.in/courses.php

Course : Course : NPTEL >> Computer Science and Engineering >> **Database Design**

Videos:

- ER Model to Relational Mapping
- Functional Dependencies and Normal Form
- ER Model to Relational Model Mapping
- Structured Query Language
- Structured Query Language II

Refer to:

https://class.stanford.edu/courses/Home/Databases/Engineering/about

References

- Abraham Silberschatz, Henry Korth and S. Sudarshan, Database System Concepts Jan 27, 2010
- C.J. Date, Database Design and Relational Theory: Normal Forms and All That Jazz (Theory in Practice) Apr 24, 2012
- Kevin Loney, George Koch “Oracle 9i, The Complete reference” Oracle Press
- http://en.wikipedia.org/wiki/Entity%E2%80%93relationship_model
- http://www.w3schools.com/sql/default.asp
- http://docs.oracle.com/cd/E11882_01/server.112/e41084/toc.htm
- http://online.stanford.edu/course/intro-to-databases-winter-2014
- http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-830-database-systems-fall-2010/lecture-notes/
- http://www.techopedia.com/definition/1245/structured-query-language-sql

Thank You

Infosys® | Building Tomorrow's Enterprise