ISE 211 - Industrial Information Systems

Databases and Analysis

Lecture 7 - Chapter 3.4

Data Modelling

Normalization

İzmir University of Economics

Halil POSACI

2011, İzmir

Agenda

- Avoidance Of Redundancy
  - Insertion Anomalies
  - Deletion Anomalies
  - Update Anomalies
  - Normal Forms
    - First Normal form
    - Second Normal Form
    - Third normal form
    - Denormalization
- HOMEWORK
- Lab

Normalization
Avoidance Of Redundancy

Table: Purchase_Order

PO NUMBER	RELEASE DATE	PO STATUS	PO AMT	VENDOR ID
2591	2/10/06	CLOSED	$4,300.00	V110
2592	2/10/06	OPEN	$505.50	V25
2593	2/11/06	OPEN	$4,000.00	V110
2594	2/12/06	OPEN	$3,280.00	V250
2595	2/15/06	OPEN	$500.00	V250
2596		HOLD	$1,000.00	V75

Table: Vendor

VENDOR ID	V NAME	V STREET	V CITY	V STATE	V ZIP
V110	Jersey Vegetable Co.	2 Main St.	Patterson	NJ	07055
V25	General Provisions	125 Common St.	Boise	ID	44830
V250	Spices Unlimited	25 Salty Lane	East Hampton	NY	10027
V75	Pasta Supply, Inc.	34 Henry St.	Philadelphia	PA	09098

a) 1:M Relationship

- What is Normalization?
  - The process of evaluating a data model to avoid data redundancy.
- What problems are caused by data redundancy?
  - Insertion Anomalies
  - Deletion Anomalies
  - Update Anomalies
- What is the solution?
  - Normalize the data model so that it is in 3rd Normal Form.

Table: PO_Vendor

PO NUM	RELEASE DAT	PO STAT	PO AMT	V NAME	V STREET	V CITY	V STATE	V ZIP
2591	2/10/2006	CLOSED	$4,300.00	Jersey Vegetable Co.	2 Main St.	Patterson	NJ	07055
2592	2/10/2006	OPEN	$505.50	General Provisions	125 Common St.	Boise	ID	44830
2593	2/11/2006	OPEN	$4,000.00	Jersey Vegetable Co.	2 Main St.	Patterson	NJ	07055
2594	2/12/2006	OPEN	$3,280.00	Spices Unlimited	25 Salty Lane	East Hampton	NY	10027
2595	2/15/2006	OPEN	$500.00	Spices Unlimited	25 Salty Lane	East Hampton	NY	10027
2596		HOLD	$1,000.00	Pasta Supply, Inc.	34 Henry St.	Philadelphia	PA	09098

b) Redundant data entries

Waste

Normalization
Insertion Anomalies

New Vendor CAUSES
Lots of empty cells

New Purchase Order may cause
Non unique Vendor info

- What is Normalization?
  - The process of evaluating a data model to avoid data redundancy.
- What problems are caused by data redundancy?
  - Insertion Anomalies
  - Deletion Anomalies
  - Update Anomalies
- What is the solution?
  - Normalize the data model so that it is in 3rd Normal Form.

PO NUM	RELEASE DAT	PO STAT	PO AMT	V NAME	V STREET	V CITY	V STATE	V ZIP
2591	2/10/2006	CLOSED	$4,300.00	Jersey Vegetable Co.	2 Main St.	Patterson	NJ	07055
2592	2/10/2006	OPEN	$505.50	General Provisions	125 Common St.	Boise	ID	44830
2593	2/11/2006	OPEN	$4,000.00	Jersey Vegetable Co.	2 Main St.	Patterson	NJ	07055
2594	2/12/2006	OPEN	$3,280.00	Spices Unlimited	25 Salty Lane	East Hampton	NY	10027
2595	2/15/2006	OPEN	$500.00	Spices Unlimited	25 Salty Lane	East Hampton	NY	10027
2596		HOLD	$1,000.00	Pasta Supply, Inc.	34 Henry St.	Philadelphia	PA	09098

b) Redundant data entries

Normalization
Deletion Anomalies

By deleting a PO you may remove a vendor completely

- What is Normalization?
  - The process of evaluating a data model to avoid data redundancy.
- What problems are caused by data redundancy?
  - ~~Insertion Anomalies~~
  - Deletion Anomalies
  - ~~Update Anomalies~~
- What is the solution?
  - Normalize the data model so that it is in 3rd Normal Form.

PO NUM	RELEASE DAT	PO STAT	PO AMT	V NAME	V STREET	V CITY	V STATE	V ZIP
2591	2/10/2006	CLOSED	$4,300.00	Jersey Vegetable Co.	2 Main St.	Patterson	NJ	07055
2592	2/10/2006	OPEN	$505.50	General Provisions	125 Common St.	Boise	ID	44830
2593	2/11/2006	OPEN	$4,000.00	Jersey Vegetable Co.	2 Main St.	Patterson	NJ	07055
2594	2/12/2006	OPEN	$3,280.00	Spices Unlimited	25 Salty Lane	East Hampton	NY	10027
2595	2/15/2006	OPEN	$500.00	Spices Unlimited	25 Salty Lane	East Hampton	NY	10027
2596		HOLD	$1,000.00	Pasta Supply, Inc.	34 Henry St.	Philadelphia	PA	09098

b) Redundant data entries

Normalization
Update Anomalies

You may need to update lots of lines for one vendor

- What is Normalization?
  - The process of evaluating a data model to avoid data redundancy.
- What problems are caused by data redundancy?
  - Insertion Anomalies
  - Deletion Anomalies
  - Update Anomalies
- What is the solution?
  - Normalize the data model so that it is in 3rd Normal Form.

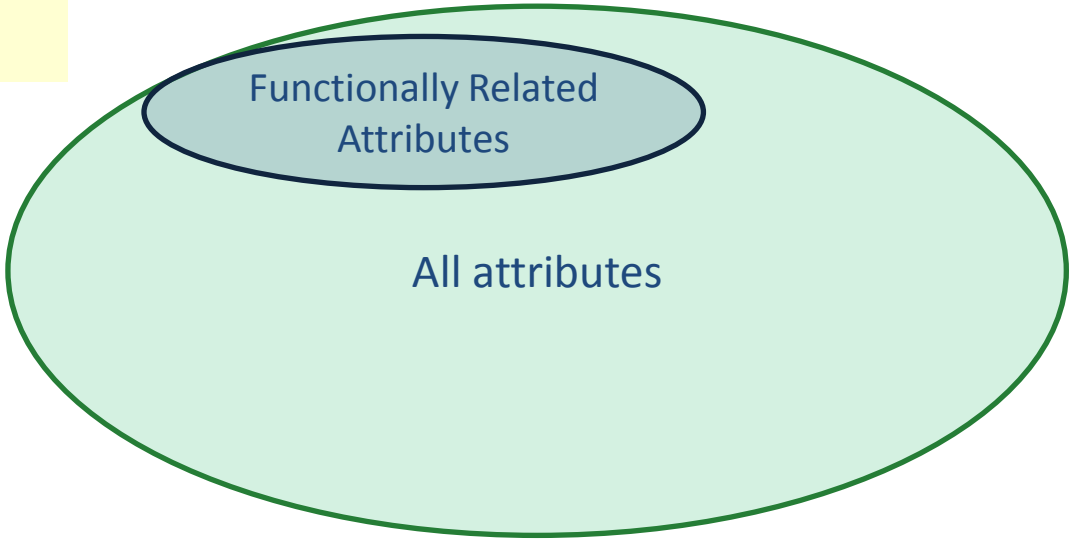
PO NUM	RELEASE DAT	PO STAT	PO AMT	V NAME	V STREET	V CITY	V STATE	V ZIP
2591	2/10/2006	CLOSED	$4,300.00	Jersey Vegetable Co.	2 Main St.	Patterson	NJ	07055
2592	2/10/2006	OPEN	$505.50	General Provisions	125 Common St.	Boise	ID	44830
2593	2/11/2006	OPEN	$4,000.00	Jersey Vegetable Co.	2 Main St.	Patterson	NJ	07055
2594	2/12/2006	OPEN	$3,280.00	Spices Unlimited	25 Salty Lane	East Hampton	NY	10027
2595	2/15/2006	OPEN	$500.00	Spices Unlimited	25 Salty Lane	East Hampton	NY	10027
2596		HOLD	$1,000.00	Pasta Supply, Inc.	34 Henry St.	Philadelphia	PA	09098

b) Redundant data entries

Normalization
Normal Forms

Subset of attributes with functional dependencies
should be isolated as a separate Entity (Table)
Eg: Vendor(VendId, VendStreet, VendCity, Vend State, VendZip)

- Design a schema that does not suffer from the insertion, deletion and update anomalies.
- If there are any anomalies present, then note them so that applications can be made to take them into account.



Normalization

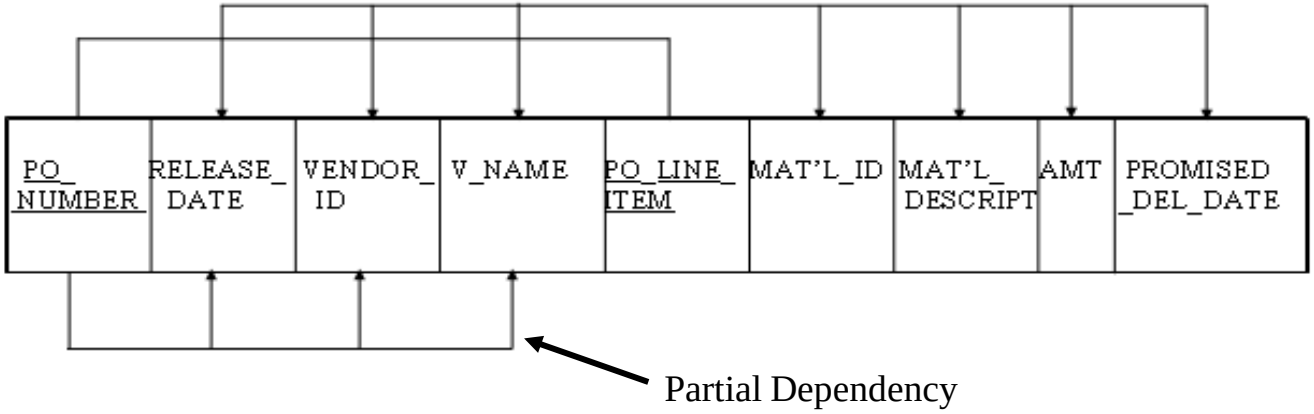
First Normal Form

Po_no	Date	Ven ID	Vendor	Line item	Mat'l ID	Description	Amt	Del dat
3502	5/28/06	V25	Gen. Provisions	1	RM805	tomato paste	4760	6/25/
3503	5/29/06	V110	Jersey Vegetables	1	RM201	carrots, whole	1000	6/20/0
				2	RM202	carrots, diced	1400	6/20/0
				3	RM201	carrots, whole	1300	7/25/0
3504	5/29/06	V250	Spices Unlimited	1	RM305	paprika	800	6/30/0
				2	RM340	sugar, bulk	1000	6/30/06

Original purchase order form From Gemba

PO NUMBER	RELEASE	VENDOR	V NAME	PO LINE I	MAT'L I	MAT'I DESC	AMT	DEL DAT
3502	5/28/06	V25	Gen.	1	RM805	Tomato paste	4760	6/25/06
3503	5/29/06	V110	Jersey	1	RM201	Carrots,	1000	6/20/06
3503	5/29/06	V110	Jersey	2	RM202	Carrots, diced	1400	6/20/06
3503	5/29/06	V110	Jersey	3	RM201	Carrots,	1300	7/25/06
3504	5/29/06	V250	Spices	1	RM305	Paprika	800	6/30/06
3504	5/29/06	V250	Spices	2	RM340	Sugar, bulk	1000	6/30/06

Table in First Normal form



Dependency Diagram

Normalization

First Normal Form

Table in First Normal form

- Disallows
  - composite attributes
  - Multi-valued attributes
  - **nested relations**

PO NUMBER	RELEASE	VENDOR	V NAME	PO LINE I	MAT'L I	MAT'I DESC	AMT	DEL DAT
3502	5/28/06	V25	Gen.	1	RM805	Tomato paste	4760	6/25/06
3503	5/29/06	V110	Jersey	1	RM201	Carrots,	1000	6/20/06
3503	5/29/06	V110	Jersey	2	RM202	Carrots, diced	1400	6/20/06
3503	5/29/06	V110	Jersey	3	RM201	Carrots,	1300	7/25/06
3504	5/29/06	V250	Spices	1	RM305	Paprika	800	6/30/06
3504	5/29/06	V250	Spices	2	RM340	Sugar, bulk	1000	6/30/06

Normalization
Conversion to Second Normal Form

- Process of going from 1NF to 2NF
  1. Write each key component that has partial dependency on a separate line and the original key on the last line.

PO_NUMBER

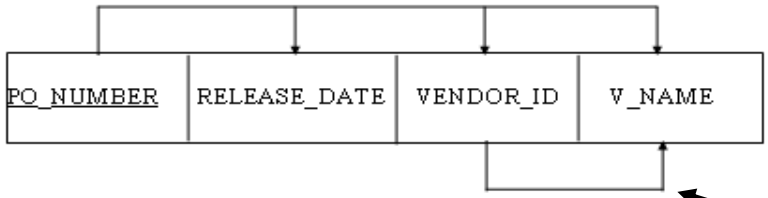
PO_NUMBER, PO_LINE_ITEM

  2. Define an entity name for each primary key and write the primary key attribute and dependent attributes of the entity.

PURCHASE_ORDER(PO_NUMBER, RELEASE_DATE, VENDOR_ID, V_NAME)

PO_DETAIL(PO_NUMBER, PO_LINE_ITEM, MAT'L_ID, MAT'L_DESCRIPT,AMT, PROMISED_DEL_DATE)

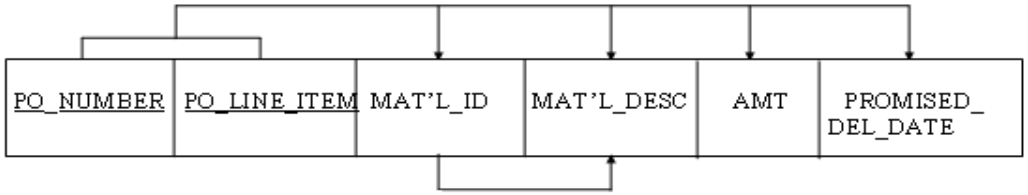
Table: PURCHASE_ORDER:



Transitive Dependency

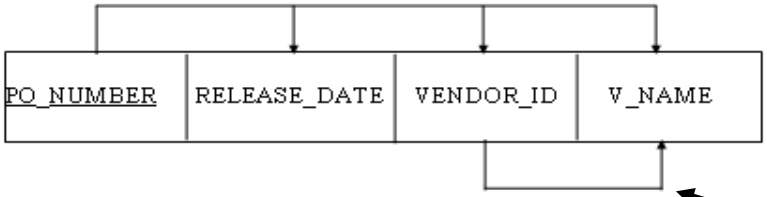
Dependency Diagram

Table: PO_DETAIL:



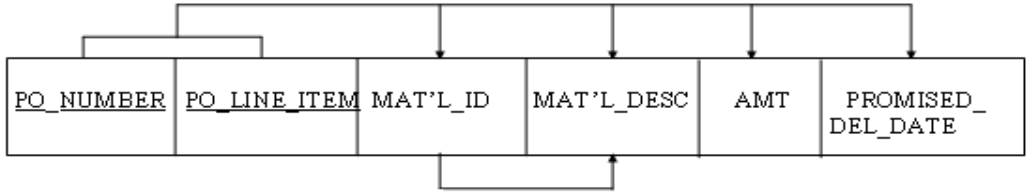
Normalization
Conversion to Third Normal Form

Table: PURCHASE_ORDER:



Transitive Dependency

Table: PO_DETAIL:



Remove Transitive Dependencies

3 NF

PURCHASE_ORDER(PO_NUMBER, RELEASE_DATE, VENDOR_ID)

VENDOR(VENDOR_ID, V_NAME)

PO_DETAIL(PO_NUMBER, PO_LINE_ITEM, MAT'L_ID, AMT, PROMISED_DEL_DATE)

MATERIAL(MAT'L_ID, MAT'L_DESCRIPTION)

DeNormalization

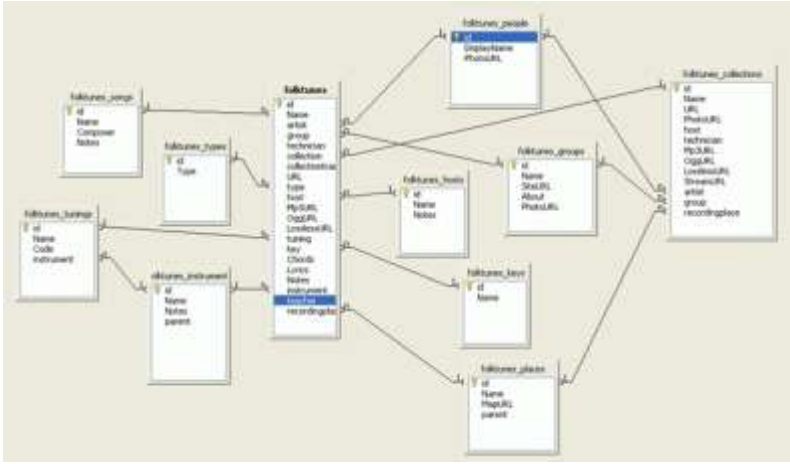
Too much normalization

NOT ALWAYS

De-normalization **may** supply high performance

Too many joins in SQL

Bad Performance – high CPU usage



Classwork - Labwork

- Try to write down your project draft analysis similar to as given in Case 3.10 of the book.

  If you don't have any idea about your project, work on with some others project. You can be a team of 6 at most. But get ready for next week to have your own project for your team of 2.
- If there are forms, reports, documents or proposed ones; try to sketch them by hand within Lab hour.

  If you don't have any idea, have them till next week.
- Try to draw a very draft E-R diagram for your project. - Try to draw E-R diagram for 3.10

  Compare two E-R diagrams

  a. Write down at least 3 similarities between two diagrams

  b. Write down differences between two diagrams. (10 of them at most)
- **Select one** of the E-R diagrams and try to create Data Base within access.

  Try to create Tables, DB diagram and show relations between table schemas.

Handle your paper works with 6 names at most.

Send by e-mail to hposaci@quiztechnology.com (Do not forget to write names)

Questions

Questions?

532 2877127

hposaci@quiztechnology.com