Anomalies

Business Data Management and Analytics

Contributions by Arthur Adamopoulos and Vince Bruno

Anomalies & Dependencies

- Data Base Design
- Anomalies
- Dependency
- Database Design Process

MOBILE

	Name	Datatype	Size	Scale	Required
🔑	MOBILEID	INT	10	0	☑
	PHONENUMBER	VARCHAR	20	0	☐
	CUSTOMERID	INT	10	0	☑
	CUSTOMERNAME	VARCHAR	81	0	☐
	JOINED	DATE	10	0	☐
	PLANNAME	VARCHAR	20	0	☑
	CONNECTFEE	DECIMAL	8	2	☐
	PEAKFEE	DECIMAL	8	2	☐
	OFFPEAKFEE	DECIMAL	8	2	☐

- Is this database structure OK?

Sample Data

mobileid	phonenumber	customerID	customername	joined	planname	connectFee	PeakFee	OffPeakFee
4851	412122169	20698	GREGORY WARD HA...	1999-07-19	Yes10	1.00	1.05	0.90
5427	413668748	21808	MARY NOLA FEKONJA	1999-07-19	FreeStyle	3.95	0.50	0.50
5498	412070692	21934	KATHRYN JANE BRIG...	1999-07-18	Yes10	1.00	1.05	0.90
5281	413011201	21514	INTHIVARA ROWLAN...	1999-07-17	Yes10	1.00	1.05	0.90
4672	410390902	20350	CASSANDRA LILIAN ...	1999-07-16	Yes10	1.00	1.05	0.90
4765	412149725	20530	PETER FLATHER	1999-07-16	Yes40	1.75	0.45	0.42
4951	411544633	20900	MARK ALOYSIUS ST...	1999-07-15	Yes20	4.55	0.75	0.65
5625	411791395	22190	ANTHONY TAN	1999-07-15	FreeStyle	3.95	0.50	0.50
4537	410502306	20074	ROBERT JOHN GAZZ...	1999-07-13	FreeStyle	3.95	0.50	0.50
4669	413349000	20344	THAI DUY HONG CHIE...	1999-07-13	FreeStyle	3.95	0.50	0.50
5007	411579679	21012	JOSEPH WALKER	1999-07-11	Yes30	1.50	0.55	0.50
4878	410389897	20750	MATTHEW ALEXAND...	1999-07-10	Yes20	1.20	0.75	0.65
4995	410182045	20988	JOHN WILLIAM LIEW	1999-07-10	Yes20	1.20	0.75	0.65

Problems?

- Anything wrong with the previous database structure ?
- All I want to do is keep track of mobile details, that's all. The mobile phone number, the customers name, and the plan details.

Problems?

- Nothing very complex, anyone should be able to set up such a simple little database. Takes half an hour in MS-Access.
- After all, there is only one table !

Problems?

- What happens if I have a new mobile phone plan name to insert into my database ?
  Eg. Yes50, connect: $2, peak $0.21, offpeak $0.14

Problems?

- What happens if I want to alter a Connection Fee for a particular plan ? Eg. Change connection Fee for 'Yes10' plan changed from $1.00 to $1.05

Problems ?

- What happens if all the customers in a plan drop out ?
  Eg. Peter Flather (mobileID 4765) is deleted.

Problems ?

- Are all copies of the same plan details the same ?
  Eg. Yes20 connection Fee $1.20 or $4.20?

ANOMALIES

- An Inconsistency
- A database structure that will be prone to errors in the data.
- Not necessarily wrong, but will promote erroneous data rather than prevent it.
- Will usually also involve more effort in programs/users that maintain the file.

ANOMALIES

- INSERTION ANOMALY
  - Can not insert a value when we want
  - must wait for a un-related event to occur first
  - need to insert new value more than once

ANOMALIES

- UPDATE ANOMALY
  - Change of value must be done multiple times to each copy of the value
  - Multiple copies of the same value mean we do not know which is correct
  - Duplication wastes resources

ANOMALIES

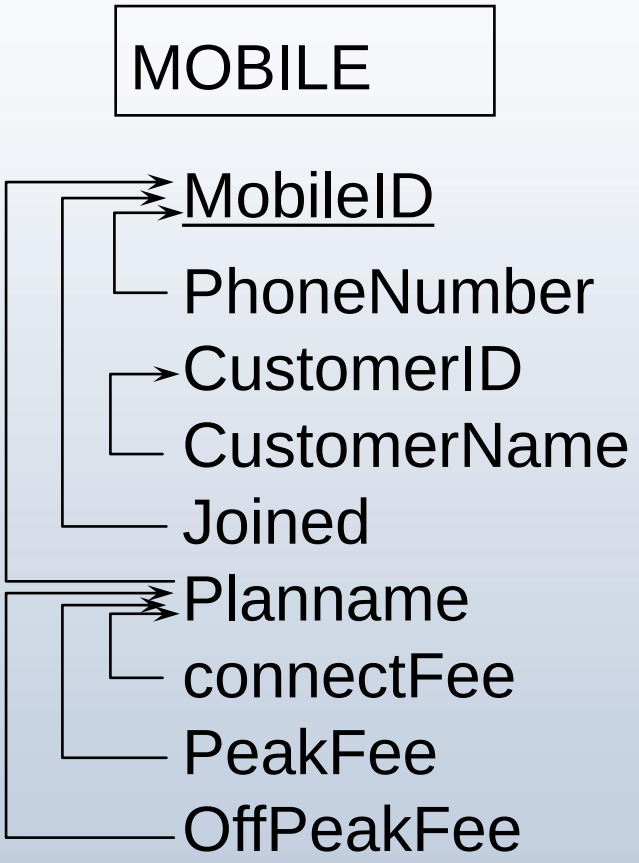
- DELETE ANOMALY
  - A value is deleted "accidentally" due to an unrelated event occurring
  - need to delete value more than once

Desired Database Structure

- INSERT
  - Insert when required, in one place, once.
- UPDATE
  - Update in one place, once.
  - Only one copy of each piece of data.
- DELETE
  - Delete in one place, once.
  - Deletions not caused by unrelated events.

DEPENDENCY

- If field A is dependent on field B, A cannot exist until B exists.
  - Phone number and Joined depend on mobileID.
  - What does connectFee depend on?
  - What does peakFee depend on?
  - What does customerName depend on?



DEPENDENCY

- The field phonenumber stores phone numbers.
- Every person in the street has a phone number. Would you put all those phone numbers in your database ?
- We say that phonenumber depends on mobileID, because a phonenumber value would not exist unless a corresponding mobileID existed within mobile phone company.

Dependency

"A FIELD MUST DEPEND ON THE KEY,
THE WHOLE KEY,
AND NOTHING BUT THE KEY,
SO HELP ME CODD"

- Every field in a table should be dependent on the whole primary key.
- If it is not, it should be in another table !
- If there is no other table to fit it, create a new table!!

Data Design Aims/Steps

- Identify Entities/Tables
- Put Fields where they belong
  - where they depend on the whole primary key
  - If cannot be placed, usually signifies a missing entity/table
- ELIMINATE ANOMALIES

Solution ?

MOBILE	PLAN	CUSTOMER
mobileID	planName	customerID
phoneNumber	connectFee	surname
Joined	peakFee	given
Planname	offPeakFee	
customerID		

Database Design Process

- Like most design activities, data design is not an exact science. Performing it involves experience and there is no single answer for any problem.
- For small tasks, I encourage you to use an "intuitive", bit by bit approach.
  - May not work for large tasks!

Database Design Process

- Go through the problem specification (if there is one), picking out any possible attribute/field etc. you think might be important. - Create a big list.
- Work on both an E-R model and a Relational model, at the same time! (Make sure they always correspond)

Database Design Process

- Start by identifying any "easy" entities.
  - People are usually easy (STAFF, STUDENT, CUSTOMER etc.)
  - Fill in the obvious attributes.
- Look for straightforward relationships, keeping both E-R model and relational model corresponding (relationships in E-R become foreign keys!)

Database Design Process

- Continue an iterative process until you have filled in as much as required.
- Part of the design process is deciding what will not be stored. You have to define the scope of the database.
- Continually ensure fields are dependent on their keys and there are no anomalies.

Database Design Process

- Ask yourself for each field you place in a table:
  - What happens if I insert ?
  - What happens if I update ?
  - What happens if I delete ?
  - Does this field belong here ?
  - Does this field depend on the WHOLE PRIMARY KEY?

Business Case Study: Newsagent (1 of 3)

- At my local newsagency, I got to know the owner. Over lunch we discussed the way in which the business was run, especially in relation to the newspaper ordering and delivery system.
- Currently a register of customers and their orders are kept in a book. The usual information is kept about a customer, the most important being the delivery address.
- Customers order a newspaper to be delivered on a specific day, multiple days or the whole weeks. They may request one paper or a number of different papers. The newsagent must keep a record of all this so that they can organise the paper rounds.

Business Case Study: Newsagent (2 of 3)

- The paper rounds are performed by boys on bikes. They arrive early each day (about 6:00am) and are given a paper round. A paper round consists of streets and specific houses where customers require a particular newspaper delivered.
- The newsagent receives stocks of a range of different papers for distribution. They consist of the standard Herald SUN, AGE, Australian, to the multi-cultural papers, such as Il Globo. The newsagency also has commenced delivery of magazines to customers.

Business Case Study: Newsagent (3 of 3)

- Each item is wrapped in glad wrap before delivery, in order to protect it from wet weather.
- Provide the following:
  - Entity Relationship Diagram.
  - Relational Model (showing fields, primary keys, foreign keys and concatenated keys).
  - Justify the choices you have made. Are the entities and relationships you have chosen free of problems? What assumptions have been made?