Marcus Zimmermann

Labouseur Lab 7

Part One: Tycho CEO Fred Johnson has put together a spreadsheet of all the data he has so far, which he personally collected.

PackageID	TagNumber	InstallDate	SoftwareCostUSD
AC01	32808	09-13-2005	754.95
DB32	32808	12-03-2005	380.00
	37691	06-15-2005	380.00
DB33	57772	05-27-2005	412.77
WP08	32808	01-12-2006	185.00
	37691	06-15-2005	227.50
	57222	05-27-2005	170.24
WP09	59836	10-30-2005	35.00
	77740	05-27-2005	35.00

1. As he shows you the spreadsheet, having just signed your consulting agreement, he asks what you think of it. How do you reply?
   First of all, this spreadsheet contains important data and certainly provides me with a good sense of what the database should be capable of. Having this sort of foundational knowledge is crucial in database design. With that being said, it will definitely have to go through the process of normalization so that we can protect the data and ensure database flexibility. This spreadsheet won't work as a table in our database. The columns containing tag number, install date, and software cost all have multiple entries in some of the same rows. There is also redundancy throughout the table.
2. Put his data in 1NF and display it. (Show me the table; no SQL.)

PackageId	TagNumber	InstallDate	SoftwareCostUSD
AC01	32808	09-13-2005	754.95
DB32	32808	12-03-2005	380.00
DB32	37691	06-15-2005	380.00
DB33	57772	05-27-2005	412.77
WP08	32808	01-12-2006	185.00
WP08	37691	06-15-2005	227.50
WP08	57222	05-27-2005	170.24
WP09	59836	10-30-2005	35.00
WP09	77740	05-27-2005	35.00

3. What is the primary key?
   None of these columns serve as a unique identifier by themselves. However, a composite key that uses PackageId and TagNumber can be used as the primary key.

Part Two: Add two columns of new data: one column for software package name (e.g., Zork, Portal, etc.) and one for computer model (e.g., IBM, Apple, etc.). Be sure that your new data is consistent with the original data.

4. Display the new table.

PackageId	TagNumber	InstallDate	SoftwareCostUSD	SoftwarePackage	CompModel
AC01	32808	09-13-2005	754.95	Rosetta Stone	Lenovo
DB32	32808	12-03-2005	380.00	Microsoft Office	Lenovo
DB32	37691	06-15-2005	380.00	Microsoft Office	Lenovo
DB33	57772	05-27-2005	412.77	Photoshop	HP
WP08	32808	01-12-2006	185.00	IntelliJ	Lenovo
WP08	37691	06-15-2005	227.50	IntelliJ	Lenovo
WP08	57222	05-27-2005	170.24	IntelliJ	Dell
WP09	59836	10-30-2005	35.00	MacKeeper	Apple
WP09	77740	05-27-2005	35.00	MacKeeper	Apple

5. Identify and document all functional dependencies.
   The following functional dependencies can be found in the table above:
   TagNumber → CompModel
   PackageId → SoftwarePackage
   PackageId + TagNumber → InstallDate
   PackageId + TagNumber → SoftwareCostUSD
6. Explain why this new table is not in third normal form.
   For a table to be in third normal form, it must be in second normal form and have no multi-key dependencies. This table is not even in second normal form yet, so we can be sure that it is not in third normal form. If it were in second normal form, we wouldn't have any unwanted partial key dependencies. In other words, we wouldn't have dependencies that do not involve the primary key (or the entire composite key in our case). Because neither CompModel nor SoftwarePackage are dependent on both PackageId AND TagNumber, they should be placed in separate tables.

Part Three: Decompose your 1NF table into a set of tables that are in at least third normal form. (BCNF would be better.) Remember that it's wrong to add artificial keys to associative entities.

ComputerModels

TagNumber	CompModel
32808	Lenovo
37691	Lenovo
57772	HP
57222	Dell
59836	Apple
77740	Apple

Software

PackageId	SoftwarePackage
AC01	Rosetta Stone
DB32	Microsoft Office
DB33	Photoshop
WP08	IntelliJ
WP09	MacKeeper

Installations

PackageId	TagNumber	InstallDate	SoftwareCostUSD
AC01	32808	09-13-200	754.95
DB32	32808	12-03-2005	380.00
DB32	37691	06-15-2005	380.00
DB33	57772	5-27-2005	412.77
WP08	32808	01-12-2006	185.00
WP08	37691	06-15-2005	227.50
WP08	57222	5-27-2005	170.24
WP09	59836	10-30-2005	35.00
WP09	77740	05-27-2005	35.00

7. Identify all primary keys (determinants) for all tables.
   The primary key for ComputerModels is TagNumber
   The primary key for Software is PackageId
   The primary key for Installations is the composite key of PackageId AND TagNumber

8. Identify all functional dependencies for all tables.
   ComputerModels
   TagNumber → CompModel
   Software
   PackageId → SoftwarePackage
   Installations
   PackageId + TagNumber → InstallDate
   PackageId + TagNumber → SoftwareCostUSD

9. Explain why the new tables are in third normal form.
   The tables are in third normal form for two reasons. First, it is in second normal form. We removed the unwanted partial key dependencies. Second, there are no multi key dependencies. In other words, there are no non-key entities that are more dependent on another non-key entity than the primary key. Instead, each non-key entity is dependent on the primary key and nothing else.

10. Draw a beautiful E/R diagram.