

# Laboratory 1

## Database Concepts

### References

#### Lecture Notes: Topic 1

#### Elmasri and Navathe, 2017: Chapters 1, 2, 3 and 5

In this session you will get to know the importance of databases and its design and implementation.

For this exercise the class will be divided into groups. Each group will be allocated questions from the exercises below. One member from each group will then provide the group's answers to the rest of the class for general discussion.

### Exercise 1

1. Given the Employee database below, answer the following questions:

EmpLastName	EmpFirstName	EmpTitle	CustomerName	CustomerAddress	TotalSales
Davolio	Nancy	Sales Representative	Andrew Tan	2A/507 James Ave. Victoria 3000	\$3,368
Fuller	Andrew	Vice President, Sales	David Ewing	780 Capital Way NSW 2890	\$298
Fuller	Andrew	Vice President, Sales	Donald Davies	908 Capital Way NSW 2890	\$2,599
Leverling	Janet	Sales Representative	Nadine Forward	722 Moss Bay Blvd. Victoria 3212	\$3,468
Peacock	Margaret	Sales Representative	Andy Tansil	4110 Redmond Rd. Victoria 3100	\$349
Buchanan	Steven	Sales Manager	Graham Ewing	14 Garrett Hill NSW 2300	\$89
Buchanan	Steven	Sales Manager	Mark Anderson	23 Garrett Rd. NSW 2300	\$349
Buchanan	Steven	Sales Manager	Tran Tan	30 Punt Hill Drive NSW 2300	\$149
Suyama	Michael	Sales Representative	Robert Chan	12 Coventry Rd. NSW 2000	\$2,799
King	Robert	Sales Representative	Dianne Rendell	10 Winchester Way Victoria 3122	\$227
Callahan	Laura	Marketing Manager	Thomas Emery	4726 Elliot St. Victoria 3000	\$2,799
Dodsworth	Amelia	Sales Representative	Daniel Smith	7 Houndstooth Rd. NSW 2100	\$2,799

- a) What problem would you encounter if you wanted to produce a listing by customer post-code?

Answer:

*Customer address includes: street name and number, suburb and post-code – all combined into one attribute. We cannot produce a listing by customer post-code only.*

- b) How would you solve this problem by altering the database structure?

Answer:

*To solve this problem we need to split the address attribute into three as follows:*

*STREET          STATE          POSTCODE*

- c) What data redundancies can you detect and how could these redundancies lead to data anomalies?

Answer:

*Note that data anomalies exist because any change in any field value must be made correctly in many places to maintain data consistency.*

*The details of employees who have more than one customer are replicated. Each time an existing employee deals with a customer all the details of that employee are re-entered into the database. This could lead to the following anomalies:*

- 1. Modification anomalies – if we need to change Steven Buchanan's employee title, we need to change it in multiple places, so when updating employee details records could be missed*
- 2. Insertion anomalies – if we need to add a new customer, an employee must also be added. This creates a high risk of data inconsistencies occurring as data may be entered in some records incorrectly.*
- 3. Deletion anomalies – if Steven Buchanan leaves, then all his customers are linked to a non-existent employee, so all related customer files need to be modified. Also note, if the customers Graham Ewing, Mark Anderson and Tran Tan are deleted, Steven Buchanan is also deleted.*

- d) Give some additional views that may be needed by other user groups for the employee database shown above.

Answer:

- 1. CUSTOMER view – needed by the Accounts department*

*CustomerName CustomerAddress TotalSales*

- 2. EMPLOYEE view – needed by Personnel*

*EmpLastName EmpFirstName EmpTitle*

- 3. PRODUCTIVE\_EMPLOYEE view – (employee whose sales > \$4000, needed by the company manager)*

*EmpLastName, EmpFirstName EmpTitle sum(TotalSales)*

## Exercise 2

Consider the following database tables:

### EmployeeTable

EmpNumber	EmpLastName	EmpFirstName	EmpAddress	EmpTitle
E0001	Davolio	Nancy	540 Capital Way NSW 2890	Sales Representative
E0002	Fuller	Andrew	14 Houndstooth Rd NSW 2100	Vice President, Sales
E0003	Leverling	Janet	12 James Ave NSW 2309	Sales Representative
E0004	Evans	Melissa	231 Capital Way NSW 2890	Sales Representative
E0005	Buchanan	Steven	15 Melton Place NSW 2890	Sales Manager
E0006	Suyama	Michael	22 Maple Dr NSW 2890	Sales Representative
E0007	King	Robert	2 Elm Parade NSW 2890	Sales Representative
E0008	Williams	Linda	26 Millers St NSW 2890	Marketing Manager
E0009	Dodsworth	Amelia	16 Garrett Rd NSW 2300	Sales Representative

### CustomerTable

CustNumber	CustLastName	CustFirstName	CustAddress
C0001	Tan	Andrew	2A/507 James Ave Victoria 3000
C0002	Ewing	David	780 Capital Way NSW 2890
C0003	Davies	Donald	908 Capital Way NSW 2890
C0004	Forward	Nadine	722 Moss Bay Blvd Victoria 3212
C0005	Tansil	Andy	4110 Redmond Rd Victoria 3100
C0006	Ewing	Graham	14 Garrett Hill NSW 2300
C0007	Anderson	Mark	23 Garrett Rd NSW 2300
C0008	Tan	Tran	30 Punt Hill Drive NSW 2300
C0009	Chan	Robert	12 Coventry Rd NSW 2000
C0010	Rendell	Dianne	10 Winchester Way Victoria 3122
C0011	Emery	Thomas	4726 Elliot St Victoria 3000
C0012	Smith	Daniel	7 Houndstooth Rd NSW 2100

### Sales Table

EmpNumber	CustNumber	ItemCode	Quantity
E0001	C0001	I0001	1
E0001	C0001	I0009	1
E0002	C0002	I0006	2
E0002	C0003	I0002	1
E0003	C0004	I0007	2
E0003	C0004	I0002	1
E0003	C0004	I0009	1
E0003	C0005	I0004	1
E0005	C0006	I0005	1
E0005	C0007	I0004	1
E0005	C0008	I0006	1
E0006	C0009	I0001	1
E0007	C0010	I0008	3
E0009	C0011	I0001	1
E0009	C0012	I0001	1

## Item Table

ItemCode	ItemDescription	UnitPrice
I0001	Plasma HD TV 127cm	\$2,799
I0002	LCD HD TV 106cm	\$2,599
I0003	LCD Widescreen TV 57cm	\$749
I0004	Mini HiFi System	\$349
I0005	Micro HiFi System	\$89
I0006	MP3 Player 2GB	\$149
I0007	MP3 Player 4GB	\$175
I0008	Multimedia Player 2GB	\$79
I0009	Home Theatre System	\$569

1. Which attribute is the primary key in the Employee table?

Answer: *EmpNumber*

2. What is the primary key of the Sales table?

Answer: *EmpNumber, CustNumber, ItemCode*

3. How are the Employee and Customer tables related? What constraint ensures the integrity of this relationship?

Answer:

*The Employee and Customer tables are related via the Sales Table. EmpNumber in the Sales table references EmpNumber in the Employee table and CustNumber in the Sales table references CustNumber in the Customer table.*

*The referential integrity constraint ensures the integrity of this relationship. This implies that for every EmpNumber in Sales there must be a matching EmpNumber in Employee and for every CustNumber in Sales there must be a matching CustNumber in Customer.*

4. What other relationships are there between these tables?

Answer:





*There is a relationship between the Sales table and the Item table. For every ItemCode in the Sales table there must be a matching ItemCode in the Item table. As in 3 above, this is ensured by the referential integrity constraint.*

5. What information can be derived from the database above that is not explicitly stored?

Answer:

- *The total number of each item sold.*
  - *The most popular item and the least popular item*
- *The item combinations that are popular.*
- *The total amount of sales for each employee.*
- *The total amount spent by each customer.*

## Helpful Tips: Simple ER-Model Concept

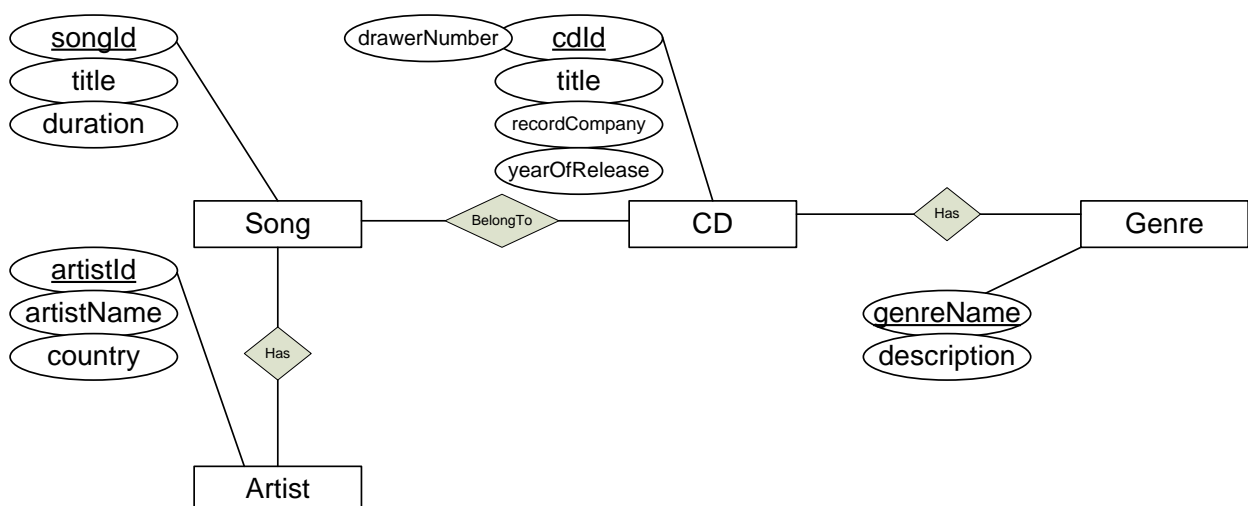
Step	ER-Model	Relational Model
1	Entity 	1 table A
2	Attribute 	1 attribute (name: b)
3	Primary key (PK) 	1 primary key (name: a)
4	Relationship 	1 relationship( name : r)

### Exercise 3

Consider the following problem description:

Arthur would like to create a database to catalogue his entire CD collection. He stores his CDs in individually numbered CD drawers. Each draw contains CDs of a specific genre. He wants to be able to search his database by either the CD title or artist and have the database tell him which draw of his collection the CD is in. Being the music aficionado that he is, Arthur would also like to store information on the artist, CD and genre to which the CD belongs. Artist information includes: individual/group name and their country of origin. If the CD is a compilation Arthur doesn't worry about keeping the artist information. If the CD is a classical CD, Arthur keeps the name of the composer, not performer. CD information includes: the title of the CD, the record company that published the CD, the year of release and a track listing, including the duration of each song. Genre information includes: the name of the genre and a description of the genre.

Construct a simple ER-Model for the above problem description. (only include the entity, attribute, primary key and the relationship)



Please note, if we establish a relationship between Artist and CD, and remove the relationship between Song and Artist, the model will still work. Also, if you create an entity for drawer, that is also not conceptually incorrect. It is just a bit inefficient.