**Relational Database Concepts**

**Learning Objectives**

*After completing this topic, you should be able to*

* *describe the features and recognize the types of relational databases*
* *recognize how to model the database*

**1. Features and types of RDBMSs**

A relational database is a collection of tables made up of data that are related to each other by common characteristics. This data can be reorganized and grouped in many different ways without modifying the database tables.

For example, a business order entry database could include two tables – Customers and Orders. The Customers table would include information about the names, addresses, and phone numbers of customers. The Orders table would include information about products ordered, the customers, order dates, and sales prices.

A sales manager can reorganize the table to view only the number of orders placed by each customer on a given date. A finance manager in the same company can group the table entries to view only the names and addresses of customers with pending payments.

Such a grouping of data is based on the relational database model. This model was conceived by Dr. Edgar Frank "Ted" Codd in the 1970s in a paper titled "A Relational Model of Data for Large Shared Data Banks," which he published while working as a researcher at IBM. The model is based on set theory and predicate logic, branches of mathematics.

The relational model was designed to simplify the database structure. It replaced the hierarchical model with simple tables containing rows and columns.

In the 1980s, Dr. Codd published 13 rules, zero to 12, for determining whether a database is relational. This was done to prevent DBMS vendors from labeling their products as relational, even though they were not fully relational.

**Supplement**

*Selecting the link title opens the resource in a new browser window.*

**Job Aid**

Use the job aid [**Codd's 13 Rules**](javascript:doWindow('../../html/jadf_dbfn_a01_it_enust1201_frame.html')) to review the rules that help you determine the characteristics of a relational database.

Dr. Codd identified three main components of the relational model – manipulation, structure, and integrity.

The relational model consists of manipulative components, which are high level operators that you can use to convert one set of relations to another. Some of the relational operators used for data manipulation are union, intersection, difference, and division. These relate to mathematical relational algebra and relational calculus, which Codd originally specified.

The structural component defines how data is represented. In a relational database, all the information is presented as relations, also known as tables. These relations are composed of sets of tuples that have the same attributes. A tuple can be thought of as a row. Whereas an attribute refers to a column and for each attribute there is a set of values. These relations are based on relations in set theory, a mathematical concept.

The integrity component defines methods of checking the validity and consistency of stored data. These methods are called constraints and are designed to enforce rules within and between relations in a database. A database provides various integrity rules, such as entity integrity, domain integrity, and referential integrity.  
  
Entity integrity states that a primary key column cannot hold a null value whereas domain integrity validates entries for a specific column. Referential integrity, in contrast, governs how tables relate to one another.

**Question**

Match each component of the relational model with its correct description.

**Options:**

1. Structural
2. Integrity
3. Manipulative

**Targets:**

1. Components that define how information is represented in a database
2. Rules that protect the consistency of a database
3. Relational operators that can be used to convert one set of relations into another

**Answer**

*The structural component defines how information is represented or structured in a database. In the relational model, data is represented in relations, which are composed of tuples and attributes.*

*The integrity component defines constraints or rules that validate and protect the consistency of data stored in a database. Some key integrity rules are entity integrity, referential integrity, and domain integrity.*

*The manipulative components are a set of relational operators, such as union, intersection, difference, and division, that convert one relation to another.*

**Correct answer(s):**

Target 1 = Option A

Target 2 = Option B

Target 3 = Option C

The data structure of a relational database mainly consists of four components:

**relation**

A relational database stores all its data inside relations, commonly known as tables. A relation is a combination of tuples and attributes. In the example displayed, the Sales table represents a relation.  
  
You can perform a number of operations on a relation. For example, you can retrieve a subset of tuples or a subset of attributes from a relation. Alternatively, you can combine items from two relations into one.   
*The Sales table contains five columns – Order ID, Order Date, Customer ID, Customer Name, and Sales Amount ($). The table contains four tuples.*

**tuple**

In the relational model, a tuple refers to a row. A tuple consists of a set of attributes, and each tuple is linked to an entity.  
  
If the values of a tuple are unknown or there are values that don't apply to that tuple, you use a NULL value.   
*The Sales table contains four tuples. The first tuple contains 001 as the Order ID, 09/09/2011 as the Order Date, 101 as the Customer ID, Portage Airlines as the Customer Name, and 244,500 as the Sales Amount ($).*

**attribute, and**

The attributes of a relation refer to its columns. An attribute is the smallest data structure in a relational database, and gives its name to the column header in a relation. An attribute takes its value from an appropriate domain.   
*There are five attributes in the Sales table – Order ID, Order Date, Customer ID, Customer Name, and Sales Amount ($). The Order ID attribute contains values, such as 001, 002, 003, and 004.*

**domain**

A domain is a set of values that are assigned to an attribute and can be specified by assigning a data type. The data type describes the kind of values an attribute can contain. For example, in the Sales table, the data type for the Order Date domain is DATE specified as DD/MM/YYYY. This ensures that only values in the appropriate date format can be entered in this domain.  
  
You can specify a name and format for a domain or other information that assists with interpreting the values in a domain.

There are several types of relational databases available in the market. Some commonly used databases are

* HSQLDB and OpenOffice.org Base
* Microsoft Office Access
* Microsoft SQL Server
* Mysql
* PostgreSQL
* Oracle, and
* IBM DB2

Short for Hyper Structured Query Language Database, HSQLDB is an open source RDBMS. HSQLDB is written in Java and is available for free under a Berkeley Software Distribution license. It conforms to almost all SQL:1992 Advanced Level and SQL:2008 standards.

**Note**

*Berkeley Software Distribution or BSD is a modified Unix operating system developed and distributed by the University of California, Berkeley, California.*

HSQLDB is the default database engine used in many open source software programs, including OpenOffice.org Base. Using HSQLDB, Base implements a one-file database, that is, it creates databases that contain reports, forms, queries, and tables in one file. This helps to provide end users with easy access to data.

Base is meant to be used as a desktop database. You can use Base to create new tables for your data, perform simple or complex sorts, create complex queries, and connect to several other RDBMSs. It also runs on multiple platforms, including Unix, Linux, and Microsoft Windows.

Like OpenOffice.org Base, Microsoft Access is also a desktop file-based database system that is written in Microsoft Visual Basic language, and is generally included with Microsoft Office suite applications.

Like Base, Access stores all database tables, forms, queries, and reports as a single file. This file is stored in the Microsoft Jet Database engine that works in combination with a graphical user interface and software development tools. In addition to its own Jet Database engine, Access can be used with other databases and applications.

In addition to Access, which is a desktop database, Microsoft has developed an enterprise-level relational database server called SQL Server. As the name suggests, SQL Server's primary query language is SQL, which uses Microsoft implementation called, Transact-SQL.

SQL Server stores and retrieves data queried by applications and users. It can be used for business administration, application development, business intelligence, reporting, and integration.

MySQL is another RDBMS that runs as a server and provides access to multiple databases. It is written in C and C++ programming languages, and is open source.

MySQL runs on several platforms, including Microsoft Windows, Apple iOS, Solaris, and Linux. It is considered to be simple, agile, and easy to use, and is also supported by several open source applications, including OpenOffice.org Base.

Unlike some relational database types, MySQL offers a distinguishing feature by allowing users to query and retrieve data from multiple storage engines. A database table can be created with different storage engines, each with unique capabilities. This gives the users freedom to choose the storage engine that best meets their application needs.

Another popular relational database is PostgreSQL, sometimes referred to as Postgres. It is an object-oriented DBMS, or OODBMS, and is written in the SQL query language. It can run on several platforms, including Windows, Linux, Solaris, and Apple OS X. Like MySQL and HSQLDB, PostgreSQL is a free and open source software program.

Just like MySQL, OpenOffice.org Base can be used as the graphic user interface for PostgreSQL.

Whereas PostgreSQL is an object-oriented DBMS, Oracle database is an object-relational DBMS, or ORDBMS, developed and marketed by Oracle Corporation.

Oracle database complies with SQL:2008 Standard and supports most of its features.

Oracle database uses three key programming languages – SQL, Java, and PL/SQL, short for Procedural Language/Structural Query Language.

**Note**

*PL/SQL is Oracle Corporation's built-in procedural language for its database.*

Like Oracle database, IBM DB2 is an enterprise-level RDBMS, and is developed and marketed by IBM Corporation. IBM DB2 runs on several operating systems, including Windows, Linux, and Unix.

IBM DB2 is based on its own built-in procedural extension language SQL PL, but it provides support for other languages, which include Java, Oracle's PL/SQL, and Microsoft's .NET languages.

**Question**

What are the key features of a relational database?

**Options:**

1. A relational database stores all its data inside relations
2. In a relational database, an attribute consists of a collection of tuples
3. In a relational database, the tuples are assigned unique data types
4. In a relational database, the attributes of a relation refer to its columns

**Answer**

***Option 1:*** *This option is correct. In a relational database, data is stored inside relations, which are tables. They consist of tuples and attributes.*

***Option 2:*** *This option is incorrect. In a relational database, a tuple consists of a set of attributes, and each tuple is linked to an entity.*

***Option 3:*** *This option is incorrect. Domains in a relational database are assigned a data type. The data type specifies the kind of value an attribute can contain.*

***Option 4:*** *This option is correct. The attributes of a relation are commonly referred to as the columns of a table. Attributes are the smallest data structure in a relational database.*

**Correct answer(s):**

1. A relational database stores all its data inside relations  
4. In a relational database, the attributes of a relation refer to its columns

**Question**

Match each type of RDBMS with its description.

**Options:**

1. OpenOffice.org Base
2. Microsoft Access
3. SQL Server
4. MySQL

**Targets:**

1. An open source RDBMS that uses HSQLDB as the default database engine
2. A file-based database system that is written in the Visual Basic language
3. An enterprise-level RDBMS developed and marketed by Microsoft
4. An open source RDBMS that is written in C and C++

**Answer**

*OpenOffice.org Base is an open source RDBMS that is available under a BDS license, and it uses HSQLDB as its default database engine.*

*Microsoft Access is written in Microsoft Visual Basic programming language and stores data in the Microsoft Jet Database engine.*

*Microsoft SQL Server is an enterprise-level RDBMS, and it uses SQL as its query language.*

*MySQL is an open source RDBMS that is written in C and C++ and provides access to multiple databases.*

**Correct answer(s):**

Target 1 = Option A

Target 2 = Option B

Target 3 = Option C

Target 4 = Option D

**Question**

Match each type of RDBMS with its description.

**Options:**

1. PostgreSQL
2. Oracle
3. IBM DB2

**Targets:**

1. An OODBMS that is written in SQL query language
2. An ORDBMS that complies with SQL:2008 Standard
3. An enterprise-level RDBMS that uses SQL, Java, Microsoft .NET, and PL/SQL as its programming languages

**Answer**

*PostgreSQL is an open source, OODBMS and is written in the SQL query language.*

*Oracle is an ORDBMS developed and marketed by Oracle Corporation. It complies with SQL:2008 Standard.*

*IBM DB2 is an enterprise-level RDBMS that uses SQL, Java, Microsoft .NET, and PL/SQL as its programming languages.*

**Correct answer(s):**

Target 1 = Option A

Target 2 = Option B

Target 3 = Option C

**2. Database modeling**

When designing a relational database system, you use entity-relationship modeling. This is a database modeling technique that is used to represent the data requirements of a business into a logical structure. These logical structures become the physical relations, attributes, and tuples of the database.

The business processes are represented in the form of entities, with defined relationships between them. The entity may be a place, person, object, or event for which data is collected. For example, in a Sales database, entities would include customers, products, and orders.

The relationship represents how one entity is linked with other entities. In the Sales database example, the order that a customer places for products represents how the customers entity is linked to the orders entity, and how the orders entity is linked to the products entity to retrieve accurate data.

Diagrams developed by using this model are referred to as Entity-Relationship Diagrams or ERDs. Creating an ERD helps you to understand an organization's data needs, and the ERD can be used as a schema diagram for identifying the required system's database.

An ERD consists of data flow diagrams or DFDs that organize the business requirements into categories and allow you to map the relationships between categories. DFDs help you to visualize the flow of data through a system – which entities the data will come from and go to, where the data will be stored, and what processes will be performed on the data.

Because an ERD represents the data requirements of a business in a graphical form, it is easy for users and developers to understand the database structure.

Data modeling is an iterative process, and you can easily adapt an ERD to accommodate any required changes to the model or data.

**Question**

Which statements about database modeling are correct?

**Options:**

1. Database modeling represents data in physical structures
2. Data modeling defines the relationships between entities
3. DFDs represent the flow of data through an information system
4. Entity-relationship modeling is a nonrepetitive process

**Answer**

***Option 1:*** *This option is incorrect. An ERD is used to transform the data requirements of a business into logical data structures.*

***Option 2:*** *This option is correct. Data modeling helps you to define the relationships between entities in a database.*

***Option 3:*** *This option is correct. A DFD is a graphic representation of the flow of data through an information system. It shows where data is stored, where it will originate from and go to, and what processes will be performed on the data.*

***Option 4:*** *This option is incorrect. Entity-relationship modeling is an iterative process, and you can make required changes to the data at any time.*

**Correct answer(s):**

2. Data modeling defines the relationships between entities  
3. DFDs represent the flow of data through an information system

**Summary**

A relational database is a collection of related data items that are organized in the form of tables. Dr. Edgar Codd developed the relational database model in a paper that he wrote in 1970. In the paper, Codd identified 13 rules that define the characteristics of a relational database. Additionally, the relational model explains the various parts of a table, which include attributes, domains, relations, and tuples.  
  
Database modeling is a technique that is used to represent the data requirements of a business into a logical structure.

**Table of Contents**

|  [**Top of page**](http://sumt.wipro.com/ITMS/data/20121227_161050_2388/df_dbfn_a01_it_enus_pif/Content/cca/df_dbfn_a01_it_enus/output/t12/misc/transcript.html#pagetop) |

|  [**Learning Objectives**](http://sumt.wipro.com/ITMS/data/20121227_161050_2388/df_dbfn_a01_it_enus_pif/Content/cca/df_dbfn_a01_it_enus/output/t12/misc/transcript.html#objective) |

| [**1.Features and types of RDBMSs**](http://sumt.wipro.com/ITMS/data/20121227_161050_2388/df_dbfn_a01_it_enus_pif/Content/cca/df_dbfn_a01_it_enus/output/t12/misc/transcript.html#section_1) |

| [**2.Database modeling**](http://sumt.wipro.com/ITMS/data/20121227_161050_2388/df_dbfn_a01_it_enus_pif/Content/cca/df_dbfn_a01_it_enus/output/t12/misc/transcript.html#section_2) |

| [**Summary**](http://sumt.wipro.com/ITMS/data/20121227_161050_2388/df_dbfn_a01_it_enus_pif/Content/cca/df_dbfn_a01_it_enus/output/t12/misc/transcript.html#section_3) |