Hyperiondev

TASK

Introduction to Databases

Visit our website

Introduction

WELCOME TO THE INTRODUCTION TO THE DATABASES TASK!

In this task, we discuss databases, the different types of databases and the DBMS. We also discuss the specialised terminologies used to describe computer files.

Get in touch

Connect for support

Remember that with our courses, you're not alone! You can contact your mentor to get support on any aspect of your course.

The best way to get help is to login to **www.hyperiondev.com/portal** to start a chat with your mentor. You can also schedule a call or get support via email.

Your mentor is happy to offer you support that is tailored to your individual career or education needs. Do not hesitate to ask a question or for additional support!

DATA VS INFORMATION

Imagine having a text file that looks like this:

```
Tripoli34Belgrade36Peshawar42Cordoba17Sydney22Pyongyang25Houston28Xi’an1
9Tehran23
```

This is what data looks like before it becomes information. You may recall that data and information are different. Data describes raw pieces of input with no context or meaning. Information, on the other hand, is processed and formatted data that can now be used to find patterns in the data, or even be able to make predictions based on picked-up trends. In order to do this, though, the data itself needs to be accurate and formatted in a way that makes it possible to process correctly. For example, if we were to format the data above, it would be much easier to work with:

```
Tripoli 34
Belgrade 36
Peshawar 42
Cordoba 17
Sydney 22
Pyongyang 25
Houston 28
Xi’an 19
Tehran 23
```

This format makes it much easier to create information. We can now see that we are looking at cities. If we knew the numbers were temperature highs in Celsius, we could process the data to make a table of information:

City	Temperature high (Celsius)
Tripoli	34
Belgrade	36
Peshawar	42
Cordoba	17
Sydney	22
Pyongyang	25
Houston	28

Xi'an	19
Tehran	23

Now that the information is easily understood, we could start to pick up patterns about which cities around the world are hottest on that particular recorded day. We could also expand this table with more information, such as high temperatures of each city every day for three years. That information could then be used to predict if the peak temperatures are rising over the years, which could impact long term crop growth, for example. This could lead to decisions being made about budget and resource distributions in the area. The aim is to be able to make strategic decisions based on extrapolated information acquired from the data.

THE DATABASE

A database is what is used to turn data into information. In the above example, we used a table to do this, which made the information easy to access and process. Databases store both the processed information and metadata. Metadata is data about the data itself. For example, the table above will store information like: only a number can be input as a temperature, no element can be left empty, but a temperature can be zero, etc.

There are various types of databases, which are divided by category. We'll look at each category below and which databases fit with them:

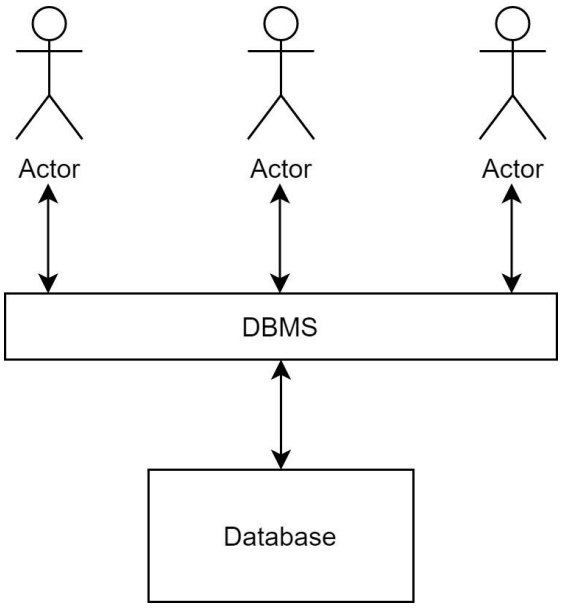
- **Number of users:** This is based on how many people can use the database at a time.
    - **Single-user:** Only one user at a time can access the database.
    - **Multi-user:** Multiple users at a time can access the database.
- **Database locations:** This is based on where the database is situated.
    - **Centralised:** The database only supports data from a single sight.
    - **Distributed:** The database supports data from multiple sites.
- **How the database is used:** This is based on the role of the database in the business.
    - **Operational:** The database supports the basic day-to-day operations of a business.

    - **Data warehouse:** The database gathers information that can be extracted and used to make strategic decisions for the business.

- **Structure:** This is based on how structured the data in the database is.
    - **Unstructured:** This is a database of raw data.
    - **Semistructured:** This is a database of data that has been somewhat processed, but is not yet completely structured.
    - **Structured:** The data are well-formatted and easy to process into information.

DBMS

A database management system (DBMS), as the name suggests, is software that is used to manage the database. For example, the DBMS will automatically put processed data into the table above so that a user can interpret it easily. In this way, it is the middleman between the user and the database:



Coronel, Morris and Rob (2011, p. 8) describe the following advantages of a DBMS:

- **Improved data sharing:** The DBMS helps create an environment in which end-users have better access to more and better-managed data.
- **Improved data security:** The DBMS makes it easy for companies to enforce data privacy and security policies.

- **Better data integration:** An integrated view of the organisation's operations and a clearer view of the big picture is promoted by wider access to well-managed data.
- **Minimised data inconsistency:** Data inconsistency occurs when different versions of the same data appear in different places. A properly designed database greatly reduces the probability of data inconsistency.
- **Improved data access:** A query is a specific request for data manipulation (e.g. to read or update the data) sent to the DBMS. The DBMS makes it possible to produce quick answers to spur-of-the-moment queries.
- **Improved decision making:** Better quality information (on which decisions are made) is generated due to better-managed data and improved data access.
- **Increased end-user productivity:** The availability of data and the tools that transform data into usable information encourages end-users to make quick, informed decisions.

Historically, DBMSes have been relational, such as SQL (Structured Query Language). This means that they stored data in tables that relate to one another. For example, you could have a table of people's *names* and another table of *job title*. The tables will relate to one another because each person in the first table will have a job title listed in the second table.

In contrast, the NoSQL (Not only SQL) database is a new generation of DBMS that is not based on the traditional database model. NoSQL databases generally have the following characteristics (Coronel, & Morris, 2016):

- They are not based on the relational model
- They support distributed database architectures
- They provide high scalability, high availability and fault tolerance
- They support very large amounts of sparse data
- They are geared toward performance rather than transaction consistency

BASIC FILE TERMINOLOGY

Here is some jargon that is used when describing the details of a computer file:

Data: As you know by now, data are raw facts with no meaning or context yet, such as a random string of numbers. It requires only 1 byte of computer storage.

Field: This is a character or group of characters that have a specific meaning, like the alphabet. It is used to define and store data.

Record: This is a connected set of fields that describe a particular element.

File: This is a collection of related records.

Have a look at the ANIMAL file below:

A_SPECIES	A_SERIAL	A_REGION	A_NAME	A_GENUS	A_DOB	H_NAME
Lion	0002002	Namibia	Leah Laved	Panthera	16 April 2004	John Mack
Tiger	102030421	Serbia	Ted Tix	Panthera	23 March 1996	Lindy Queue
Jellyfish	4365	Atlantic Ocean	Jill Jam	Chrysaora	8 September 2017	Sam Kins

A_SPECIES = animal species
A_SERIAL = animal serial number
A_REGION = animal region
A_NAME = animal name
A_GENUS = animal genus
A_DOB = animal date of birth
H_NAME = human's name who spotted the animal

Using the proper file terminology you can identify the file components shown in the table above. The ANIMAL file contains 3 records. Each record is composed of 7 fields: A_SPECIES, A_SERIAL, A_REGION, A_NAME, A_GENUS, A_DOB and H_NAME. The 3 records are stored in a file named ANIMAL since it contains animal data.

Compulsory Task

Answer the following questions:

- Create a text file called **databases.txt** where you will answer the following questions.
- Explain the difference between data and information.
- Research “sparse data”. What is the alternative? What are the differences between the two?
- Research and list 3 relational DBMSes and 3 NoSQL DBMSes.
- Look up “transaction consistency”. Describe it in your own words.
- Given the file below, answer the following questions:
    - How many records does the file contain?
    - How many fields are there per record?

PROJECT_CODE	PROJECT_MANAGER	MANAGER_PHONE	MANAGER_ADDRESS	PROJECT_BID_PRICE
21-5U	Holly Parker	33-5-592000506	180 Boulevard du General, Durban, 64700	R13179975.00
21-7Y	Jane Grant	0181-898-9909	218 Clark Blvd, Cape Town, WC, TRY	R45423415.00
25-9T	George Dorts	0181-227-1245	124 River Dr, Cape Town, WC, 7YU	R78287312.00
29-7P	Holly Parker	33-5-592000506	180 Boulevard du General, Durban, 64700	R20883467.00

Rate us
Share your thoughts

HyperionDev strives to provide internationally-excellent course content that helps you achieve your learning outcomes.

Think that the content of this task, or this course as a whole, can be improved or think we've done a good job?

Click here to share your thoughts anonymously.

References:

Coronel, C., & Morris, S. (2016). *Database Systems: Design, Implementation and Management* (12th ed., pp. 1-70). Mason, OH: Cengage Learning.

Coronel, C., Morris, S., & Rob, P. (2011). *Database systems* (9th ed., pp. 4-26). Boston: Cengage Learning.