Hyperiondev

TASK

Introduction to Databases

Visit our website

Introduction

Welcome to the Introduction to 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

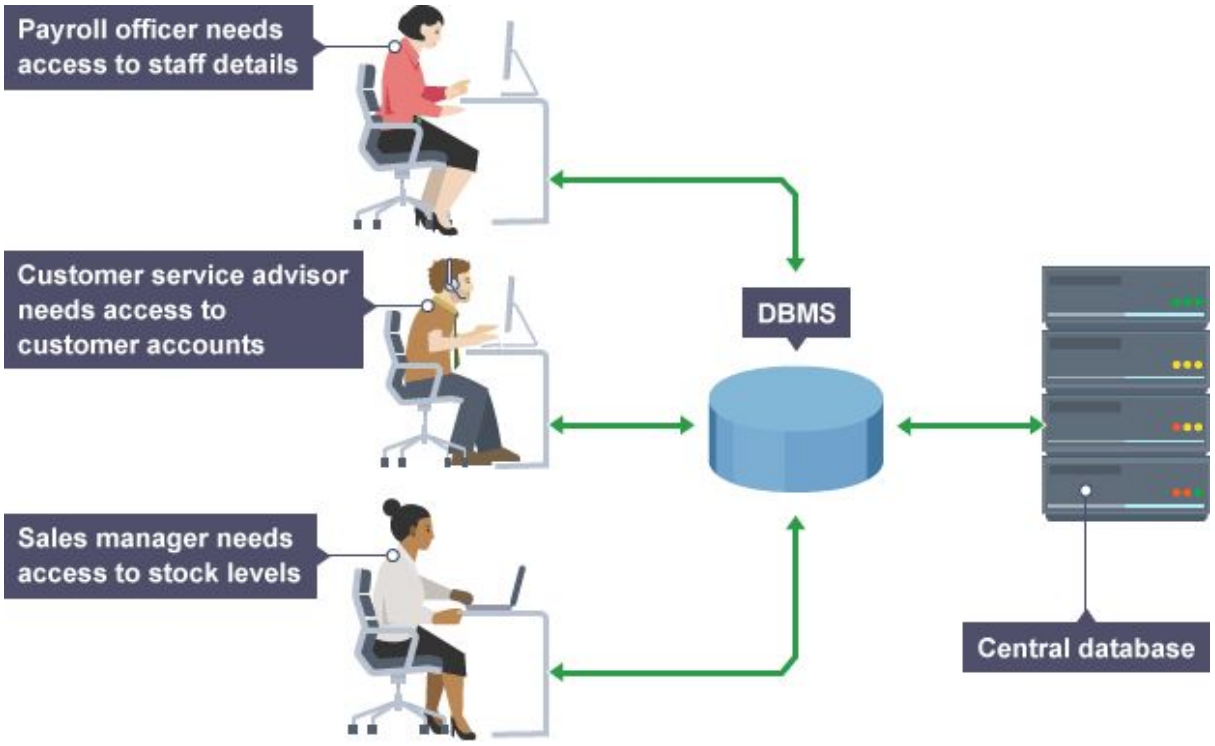
In order to properly understand databases, you must first understand the difference between data and information. Simply put, data are raw facts. The word raw indicates that the facts have not been processed to reveal their meaning yet. Information, on the other hand, is the result of processing raw data to reveal its meaning. Data processing can be as simple as organising data to reveal patterns or as complex as making forecasts or drawing inferences using statistical modelling.

Accurate, relevant and timely information is the foundation of decision making. Timely and useful information requires accurate data which must be generated properly and stored in a format that is easy to access and process. The data environment should also be carefully managed just like any other basic resource.

The Database and DBMS

A computer database is used to efficiently manage data. A database is a shared, integrated computer structure that stores a collection of end-user data, which are raw facts that are of interest to the end user, and metadata, which are data about data through which the end user data are integrated and managed. Metadata provides a description of the data characteristics and the set of relationships that link the data found within the database.

A database can be thought of as a well organised electronic filing cabinet where powerful software, known as a database management system, helps manage the contents of the cabinet. A database management system (DBMS) is a collection of programs that manage the database structure and controls access to the data stored in the database.



The DBMS manages the interaction between the end user and the database (bbc.co.uk)

The illustration above shows how the DBMS serves as an intermediary between the user and database. The DBMS receives all application requests and translates them into the complex operations required to fulfil those requests.

Much of the database's internal complexity is hidden from the application programs and end users by the DBMS. There are some very important advantages to having a DBMS between the end users application and the database. Firstly, the DBMS allows the data in the database to be shared among multiple applications or users. Secondly, the DBMS integrates many different users' views of the data into a single data repository.

The DBMS helps make data management much more efficient and effective and provides advantages such as:

- **Improved data sharing:** The DBMS helps create an environment in which end-users have better access to more and better-managed data.

- **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.

Types of Databases

There are many different types of databases. These databases can be classified according to the number of users supported, where the data are located, the type of data stored, the intended data usage and the degree to which the data are structured.

A database can be classified as either single-user or multi-user. A database that only supports one user at a time is known as a single-user database. With a single user database, if user A is using the database, users B and C must wait until user A is done. A desktop database is a single-user database that runs on a personal computer. A multi-user database, on the other hand, supports multiple users at the same time. A workgroup database is a multi-user database that supports a relatively small number of users (usually less than 50) or a specific department within an organisation. When a multi-user database supports many users (more

than 50) and is used by the entire organisation, across many departments, it is known as an enterprise database.

A database can also be classified based on location. A centralised database is a database that supports data located at a single site, while a distributed database supports data distributed across several different sites.

Today, however, the most popular way of classifying databases is based on how they will be used and on the time sensitivity of the information gathered from them. An operational database is a database that is designed to primarily support a company's day-to-day operations. Operational databases are also known as online transaction processing (OLTP), transactional or production databases.

The degree to which data is structured is another way of classifying databases. Data that exist in their original, or raw, state are known as unstructured data. In other words, they are in the format in which they were collected. Unstructured data therefore exist in a format that does not lend itself to the processing that yields information. Structured data are the result of formatting unstructured data to facilitate storage, use and the generation of information. You apply structure based on the type of processing that you intend to perform on the data. For example, imagine that you have a stack of printed invoices. If you just want to store these invoices so that you are able to retrieve them display them later, you can scan them and save them in a graphical format. However, if you what to derive information from them, such as monthly totals or average sales, having the invoices in a graphical format will not be useful. You could instead store the invoice data in a structured spreadsheet format so that you can perform the desired computations.

Analytical databases focus on storing historical data and business metrics used exclusively for tactical or strategic decision making. They typically comprise of two components; a data warehouse and an online analytical processing (OLAP) front end. Analytical databases allow the end user to perform advanced data analysis of business data using sophisticated tools. A data warehouse, on the other hand, focuses on storing data used to generate information required to make tactical or strategic decisions.

The table below compares a number of database management systems:

Product	Number of Users			Data Location		Data Usage	
	Single-user	Multi-user		Centralised	Distributed	Operational	Analytical
		Workgroup	Enterprise				
MS Access	X	X		X		X	
MS SQL Server	X	X	X	X	X	X	X
IBM DB2	X	X	X	X	X	X	X
MySQL	X	X	X	X	X	X	X
ORACLE RDBMS	X	X	X	X	X	X	X

NoSQL (Not only SQL) database is a new generation of database management system that is not based on the traditional database model. Believe it or not, you are using a NoSQL database every time you search for a product on Amazon, watch a video on Youtube or send a message to a friend on Facebook. NoSQL databases generally have the following characteristics:

- 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

Specialised vocabulary is required for the description of computer file:

Data: Data are raw facts, such as a telephone number, customer name or birth date. Unless they have been organised in some logical manner, data have little meaning. A single character is the smallest piece of data that can be recognised by a computer. It requires only 1 byte of computer storage.

Field: A field is a character or group of characters that have a specific meaning. It is used to define and store data.

Record: A record is a logically connected set of one or more fields that describes a person, place or thing.

File: A file is a collection of related records.

Below is the CUSTOMER file:

C_NAME	C_PHONE	C_ADDRESS	C_POSTCODE	A_NAME	A_PHONE	TP	AMT	REN
Alfred Smith	082 345 2341	207 Willow St, Port Elizabeth	6390	Leah Hahn	084 259 2073	T1	R100.00	05-Apr-2021
Kathy Dunne	083 567 9012	556 Bad St, Cape Town	7100	Alex Alby	085 785 3938	S2	R250.00	16-Jun-2021
Paul Farris	076 782 1232	2148 High St,Benoni	1522	Leah Hahn	084 259 2073	T2	R850.00	22-sep-2021

C_NAME = customer name
C_PHONE = customer phone
C_ADDRESS = customer address
C_POSTCODE = customer postcode
A_NAME = agent name
A_PHONE = agent phone
TP = insurance type
AMT = insurance policy amount in thousands of R
REN = Insurance renewal date

Using the proper file terminology you can identify the file components shown in the table above. The CUSTOMER file contains 3 records. Each record is composed of 9 fields: C_NAME, C_PHONE, C_ADDRESS , C_POSTCODE, A_NAME , A_PHONE , TP , AMT and REN. The 3 records are stored in a file named CUSTOMER since it contains customer data.

Compulsory Task

Answer the following questions:

- Define each of the following terms:
  - Data
  - Field
  - Record
  - File
- What is a DBMS and what are its advantages?
- Explain the difference between data and information.
- What is metadata?
- 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, Paris, 64700	$13179975.00
21-7Y	Jane Grant	0181-898-9909	218 Clark Blvd, London, NW3, TRY	$45423415.00
25-9T	George Dorts	0181-227-1245	124 River Dr, London, N6, 7YU	$78287312.00
29-7P	Holly Parker	33-5-592000506	180 Boulevard du General, Paris, 64700	$20883467.00

Rate us

Share your thoughts

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

Think 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.