**CCT College Dublin**

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | Databases |
| **Assessment Title:** | Databases CA1 |
| **Lecturer Name:** | Marina Iantorno |
| **Student Full Name:** | Michael Gelezov |
| **Student Number:** | sba23143 |
| **Assessment Due Date:** | 18/02/2024 |
| **Date of Submission:** | 17/02/2024 |

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |

**Overview of MYSQL**

MYSQL is known as a relational database management system. More specifically, data is stored in what is called a relational database. What it does is store data and organize it regarding its relational model which is made up of tables, columns, rows and relationships between tables.

MYSQL was developed by a company called Oracle and is compatible with many operating systems such as Windows, Linux and Mac. It is also open-sourced so any user or company can use it, modify or publish it anywhere without issues from the company that developed it.

**Overview of Databases**

There are a variety of different databases that are used in MYSQL;

Relational databases were first created around 1970, which continue to be used today.

Relational databases store data in several related tables, in those tables are columns and rows which data is stored in.

Relational databases give you allowance of creating, updating and administering of a relational database.

NoSQL databases are databases that do not have MYSQL as the primary access language.

These are often labelled as non-relational databases meaning they do not have to adhere to a schema that has already being organized for the data to follow.

NoSQL is used for storing unstructured data or somewhat structured data and is useful when you’re making changes to the data without using applications that are connected to the data.

Graph databases are a type of database used in NoSQL and is based on graph theory, Graph databases are used to identify and work with data points meaning they are used to analyze relationships in between data points which prevents fraud or can be used to data mine customer details.

Wide column databases are column stores for data storage, rather than storing in rows or tables wide column databases store their data in columns.

Wide column databases store petabytes as their data, meaning it’s easier to handle large amounts of data on large applications.

Columnar databases are used often in data warehousing and data is stored in columns instead of rows, reason why they are stored in such conditions is that they are easy to use in analytics sense regarding queries any data that does not apply to the query is ignored by default.

**Concept of data warehousing**

Data warehousing is a repository for information that is able to be analyzed in order to make decisions in a more open-minded way without bumping into problems.

Data goes through a data warehouse via transactional systems, relational databases and a variety of other ways.

Data warehouses can be accessed through BI Tools (Business intelligence) which are used by a lot of people like business analysts, decision makers, data scientists and data engineers.

In business data warehousing is relied on heavily, as it can store reports, dashboards and analytics tools which can lessen the input and output of data more efficiently and give results at a quicker pace to users using the warehouse.

In MYSQL data warehousing is convenient as you can store data in databases you create which you can use elsewhere through connecting MYSQL to Java or connecting MYSQL to HTML, however it cannot maintain large amounts of data therefore some people rely on other methods of storing large quantity of data.

**Concept of data mining**

Data mining is a process that searches and analyzes a huge amount of data to identify patterns and extract data.

Data mining is used by a lot of companies in order to obtain information about their customers which helps increase sales of their products, place marketing strategies to maximize the efficiency of sales and decrease any costs that can occur.

What data mining involves is a lot of exploring and going through large amounts of data to observe any useful patterns and trends. This is often used in detecting fraud or filtering spam or credit risk management but can also be used in marketing to observe opinions of a targeted group of people.

What makes data mining for SQL convenient is you can make different rules to suit your needs, identify patterns in your data and analyze more complex data. In SQL data mining is perfect for analysis and reports about what is going on in your data.

**Big data storage**

When it comes to storage of big data, we must ask if there is more than one way to store big data, one such way is MYSQL but it’s not the only way to store large quantity of data.

Cloud storage can store big data, you use cloud storage when you are using iCloud or Google Drive to store your data meaning anything you store into iCloud or Google Drive you store it electronically meaning you can use it or access it anywhere and there is no limit as to how much data you can store.

Cloud storage gives users easier usability and accessibility so if a user wanted to access something from Google Drive all they have to do is enter their credential details and the data is available for them to access

Object storage acknowledges data as objects, the data is stored in a large depository which can be accessed anywhere from physical storage devices.

Object storage has data that are considered files or objects in this case which includes metadata, more meta data is added to allow data accessibility excluding hierarchy in the process all the objects in object storage are given a unique address so that the user is able to specify which data they wish to access or use.

Data lake is a central repository for mega data and can store structured, unstructured or semi-structured data meaning that your data is more flexible to use.

When the data is stored, it is given a unique identifier for easier usage and accessibility for the user. While they are able to store big data through data warehousing both data lake and warehousing won’t be considered the same, as to why it is not is simply because that the data in data lake storage actually has no function therefore has no fulfilment whereas data warehousing stores data that already has a function it can perform.

**Harvard References**

Matillion. (n.d.). *The Types of Databases (with Examples)*. [online] Available at: <https://www.matillion.com/blog/the-types-of-databases-with-examples>.

‌

Juillet, R. (2022). *What Are The Storage Methods For Big Data?* [online] Bocasay. Available at: <https://www.bocasay.com/how-to-store-big-data/>.

‌

Amazon Web Services, Inc. (n.d.). *What is a Data Warehouse? - Data Warehouse Explained - AWS*. [online] Available at: [https://aws.amazon.com/what-is/data-warehouse/#:~:text=A%20data%20warehouse%20is%20a](https://aws.amazon.com/what-is/data-warehouse/" \l ":~:text=A data warehouse is a).

Twin, A. (2023). *Data Mining Definition*. [online] Investopedia. Available at: [https://www.investopedia.com/terms/d/datamining.asp#:~:text=Data%20mining%20is%20the%20process](https://www.investopedia.com/terms/d/datamining.asp" \l ":~:text=Data mining is the process).

‌

Cybiant. (n.d.). *Big Data Storage Methods - Cybiant Knowledge Centre*. [online] Available at: <https://www.cybiant.com/knowledge/big-data-storage-methods/>.

Talend - A Leader in Data Integration & Data Integrity. (n.d.). *What is MySQL? Everything You Need to Know*. [online] Available at: [https://www.talend.com/resources/what-is-mysql/#:~:text=MySQL%20is%20a%20relational%20database](https://www.talend.com/resources/what-is-mysql/" \l ":~:text=MySQL is a relational database).