9
```
SELECT Sum(amount) AS total amount
FROM (
        SELECT amount ~~FA~~
        FROM online_sales
        union all
        SELECT ~~Ga~~ amount
        FROM store_sales
        ) AS combine_sales
```

total - amount
1400

Research Assignment

1 The main types of Databases are relational databases and non-relational databases

2 An RDBMS is a Software System used to create, manage and administer relational databases

3 A Primary key is a column in a table that uniquely identifies each row in that table, ~~as opposed~~ ~~a~~ whereas a foreign key is a column in one table that refers to ~~an~~ a primary key in another table

4 Database normalization is the systematic process of structuring a relational database to minimize data redundancy and undesirable anomalies

5 A database Schema is the logical design or Structure of the entire database

6 Structured data - Highly organized, fixed format & fits neatly in a table

Semi-Structured data - Has some organisational properties but no rigid structure

unstructured Has no predefined format or organization Very difficult to categorize

7 A fact table normally stores quantitative metrics or we can say numeric value and it is usually very large

whereas a dimension table normally store descriptive attributes or we can say textual data and it is usually small

8 A data model is a conceptual representation of the data Structure used by a database. It defines the data elements and relationship between them. a data ~~model~~ model is important because it provides a blueprint for the physical database design, ensuring all necessary data is captured, relationships are ~~diff~~ defined collectly and the business rules are enforced.

9 ~~to~~ Database - optimized for real-time operational ~~transaction~~ transactions

Data warehouse - optimized for historical analysis and reporting

Data Lake - Stores raw, ~~unstructured~~ unorganized data ~~as~~ at a massive scale, including Structured, Semi-Structured and unstructured data, for future processing analysis and Machine learning.

10 ~~a dart~~
A data mart is a subset of a data warehouse that is specifically designed for and focused on a single line of business team, or functional area

A ~~datawarehouse~~ ~~A~~ data warehouse is enterprise-wide covering many subject areas, A data mart is smaller, more focused and serves the analytical needs of a specific group of users.

11 A query language is a specialized programming language designed to retrieve and manage data from database
SQL is the most common because it is the standard language for relational database management systems and it is english-like syntax which makes easy to read

12 Indexes are special lookup tables that a database search engine can use to speed up data retrieval. They reduce the need ~~to~~ for the database to perform a full table scan.

13 A transaction is a single logical unit of work which may contain one ~~are~~ or more SQL statements

~~ACID~~ ACID properties are a set of properties that gurantee that database transactions are process reliably

14 a database View is a virtual ~~tabe~~ table whose contents are defined by a ~~q~~ query. I doesn't Store data itself but rather provides a dynamic window into the data Stored in the base tables

It can have negative and positive impact.

Negative: views can Sometimes decrease performance especially if they are based on complex Joins

Positive: views can improve performance if they pre-filter or pre-aggregate data

15 Views - A ~~ver~~ virtual table based on a SELECT Statement. used for ~~sp~~ Simplification and Security

Stored procedures - a Set of precompiled SQL statement Stored in the database. They accept input parameters and return output values.

Triggers - A Special type of Stored procedure that executes automatically. ~~whe~~ when a Specific event occurs on a specific table or view.

16 The ~~differen~~ ~~difer~~ difference between ETL and ELT is that with ETL, tranformation is done before loading and only tranformed, cleaned data is loaded as opposed to an ELT where ~~traf~~ transformation is done after ~~loading~~ loading and raw data is loaded first, then transformed

17 ~~with Batch processing data is input a large finite chunks of historical~~.

with Batch processing, it is finite, large Chucks of historical data. ~~where~~ whereas with Stream processing, it is infinite, continuous flow of Small data records

18 A Join is a clause in SQL used to combine rows from two ~~are~~ or more table based on a related column between them.

Types of Joins

Left Join

INNer Join

Right Join

Full (outer) Join

19 Referential ~~H~~ integrity is a concept that ensures that relationships between tables remain Consistent.

It is important because it prevents the creation of orphan records - rows that reference a non-existent value in another table

20 It decreases performance because any update, Insertion, deletion of data must be done in multiple places which increase transaction processing time and it wastes resources

21 with cloud based management, infrastructures is managed by the vender It is accessible globally via internet

Now with on-premise Database, the ~~infrastructur~~ infrastructure is managed by ~~the~~ the organization and it is accessible primarily within the corporate network

22 Data governance is the overall ~~mang~~ management of the availablility, usability, integrity and Security or data used in an enterprise.
It is important because it ensures compliance, data quality improvement and good Security

23 Data integrity refers to the accuracy, completeness and consistency of data throughout its entire lifecycle
It can be maintained by
- implementing constraints
- using data validation checks at the points of data entry
- using foreign keys to ensure relationships between tables are consistent.

24 Data quality is the assessment of data's fitness to serve a specific purpose
It is critical because if data is poor quality we can get inaccurate insights, we get poor decision making and wasted time

25 The role of a data analyst is querying and ~~at~~ data retrieval, Data cleaning and preparation and data analysing, ~~repot~~ reporting

26 A data administrator is ~~it~~ responsible for ~~for~~ performance integrity, and security of database System

27 main Steps involved in designing a data pipeline
are Ingestion / Extraction
transporting
transformation
Loading
monitoring

28 Some of the challenge are:
having to maintain fast query performance despite massive data ~~values~~ volumes and the ability to add resources quickly to handle growing data and user load

29

MySQL	-	Web applications
Snowflake	-	Cloud data warehousing
oracle	-	~~Comp~~ enterprise resource planning
PostgreSQL	-	Complex applications

30 The main ~~st~~ data Storage type use are
CSV — A Simple, text-based, delimited format
JSON - A text-based Semi-structured format that Store data key value pair
Parquet - A columnar Storage format built for efficient data access
~~o~~ Avro - A row-based format that uses a Schema to Store and manage data