**Data Engineering & Analytics | Online Assessment**

Time Limit: 2 Hours

Please answer the following questions and be as detailed as possible in your responses. You can answer the questions based on online research but ensure that you articulate the responses in your own words. Submissions with any answers copied from online/offline resources will be disqualified.

1. What is data warehouse schema? Explain different types of schema.

data warehouse schema describes the logical structure of any data warehouse containing records.Data,Star,Glalaxy and snowflake are the types of schema. Simplest structural description of any data warehouse and is the least complex among all the dimension models i.e. Snowflake and Galaxy schema.

The star schema contains a single fact table that is connected to multiple dimension tables in the form of shape of star. The snowflake schema describes the logical structure in much more detail as compared to star schema. Snowflake schema is more complex than Star schema but less complex than Galaxy Schema. Galaxy schema is also known as fact constellation. Fact constellation refers to combination of fact tables and dimension tables using joins.

1. What is the difference between OLTP and OLAP? Explain their application with the help of one example each.

OLAP vs. OLTP is an online processing system. Opposed to a batch processing system, where transaction details pile up in a stack of documents that are then registered in batches, an online processing system is one that registers ongoing transactions in real-time.OLTP is a transaction processing system i.e.; it manages transaction-based applications over the internet. OLTP systems, for instance, are responsible for providing data to data warehouses.On the other side, OLAP is an analytical processing system. This simply means that it responds to multidimensional analytical queries corresponding to financial reporting, forecasting, etc. The data available to a data warehouse, for example, is analyzed using an OLAP system.One major difference between OLAP vs. OLTP is that while the former is an online database modifying system, the latter is an online database query answering system. However, the difference between the two online processing systems goes way beyond this distinction.

1. Explain Fact Table and Dimension Table.

A Fact Table is a central table in a star schema of a data warehouse.It stores quantitative information for analysis and is often denormalized.It works with dimension tables and it holds the data to be analyzed and a dimension table stores data about the ways in which the data can be analyzed. A Dimension Table is a table in a star schema of a data warehouse;They facilitate the reality table or fact table to gather dimensions on that the measures needs to be taken

1. Compare AWS Redshift and Google BigQuery cloud data warehouses. Which of these is best suited for what purpose?

Amazon Redshift is provisioned on clusters and nodes. Google BigQuery is serverless. Redshift supports 1,600 columns in a single table, BigQuery supports 10,000 columns. Redshift requires periodic management tasks like vacuuming tables, BigQuery has automatic management. RedShift is more economical for everyday data warehouse operations for most businesses. But, BigQuery is better for businesses looking to do data mining or those who deal with extremely variant workloads.

1. What are the benefits of using a data warehouse over a database?

With data warehousing techniques and processes, data can be accessed and analyzed from multiple sources. Thus, the data is not limited to any specific section, which benefits business people to make improved and intelligent business decisions. The data warehouse and related BI processes can also be directly implemented in inventory management, financial management, sales and marketing. It has become need of every individual company to implement data-ware house truly as because of the fact that there are both technical and business reasons are there.

1. Explain the components of a Data Analytics Stack. What is its significance?

There are 4 components : plan, design, build and managing the stack.

Data wrangling: There are a lot of ways that data can be gathered, and even more ways that they can be incomplete, messy or irrelevant. Tools in this space make it easier to manage data sources and clean the recorded data into forms that are more conducive to analysis.

Data storage: The cleaned data needs to be stored somewhere so that it can be referenced later on for analysis. Depending on the size of an organization and how much of the original data is considered important enough to keep, the kind of storage architecture to employ can change greatly.

Data analysis: The data you’ve collected don’t mean anything unless you’re able to draw insights from them. Tools at this point of the pipeline help with further aggregation and summarization of the data along with creation of visualizations that can demonstrate outcomes at a glance.

1. What is the difference between SQL and NoSQL databases? Explain with examples.

SQL databases are relational, NoSQL are non-relational,SQL databases use structured query language and have a predefined schema,NoSQL databases have dynamic schemas for unstructured data,SQL databases are vertically scalable, NoSQL databases are horizontally scalable,SQL databases are table based, while NoSQL databases are document, key-value, graph or wide-column stores, SQL databases are better for multi-row transactions, NoSQL are better for unstructured data like documents or JSON. For high transactional based application: SQL databases are best fit for heavy duty transactional type applications, as it is more stable and promises the atomicity as well as integrity of the data. While you can use NoSQL for transactions purpose, it is still not comparable and sable enough in high load and for complex transactional applications.

For properties: SQL databases emphasizes on ACID properties ( Atomicity, Consistency, Isolation and Durability) whereas the NoSQL database follows the Brewers CAP theorem ( Consistency, Availability and Partition tolerance).