Snowflake Real Time Data Warehouse Project for Beginners-1

Business Overview

Snowflake's Data Cloud is based on a cutting-edge data platform delivered as a service (SaaS). Snowflake provides data storage, processing, and analytic solutions that are quicker, easier to use, and more versatile than traditional options.

Snowflake isn't based on any current database technology or large data software platforms like Hadoop. Snowflake, on the other hand, combines a brand-new SQL query engine with cutting-edge cloud architecture. Snowflake gives users all the features and capabilities of an enterprise analytic database, plus a lot more.

- There is no hardware to choose, install, configure, or manage (virtual or actual).
- There isn't much to install, set up, or maintain in terms of software.
- Snowflake oversees ongoing maintenance, administration, updates, and tweaking.

Snowflake is entirely based on cloud infrastructure. Except for optional command-line clients, drivers, and connectors, all components of Snowflake's service operate on public cloud infrastructures.

Snowflake's computing demands are met by virtual compute instances, and data is stored permanently via a storage service. Snowflake isn't compatible with private cloud environments (on-premises or hosted).

The architecture of Snowflake is a mix of classic shared-disk and shared-nothing database designs. Snowflake employs a central data repository for persistent data, like shared-disk architectures, available from all compute nodes in the platform. Snowflake executes queries utilizing MPP (massively parallel processing) compute clusters, wherein each node in the cluster stores a part of the whole data set locally, like shared-nothing architectures. This method combines the simplicity of a shared-disk design with the speed and scale-out advantages of a shared-nothing architecture.

Snowflake components include:

- Warehouse/Virtual Warehouse
- Database and Schema
- Table
- View
- Stored procedure
- Snowpipe
- Stream
- Task
- Time Travel

Tech Stack:

➔ Languages: SQL
➔ Services: Amazon S3, Snowflake, SnowSQL, QuickSight

SnowSQL
SnowSQL is a next-generation command-line client for connecting to Snowflake and running SQL queries, and performing all DDL and DML actions, such as loading and unloading data from database tables.

Amazon S3
Amazon S3 is an object storage service that provides manufacturing scalability, data availability, security, and performance. Users may save and retrieve any quantity of data using Amazon S3 at any time and from any location.

QuickSight
Amazon QuickSight is a scalable, serverless, embeddable, machine learning-powered business intelligence (BI) service built for the cloud. It can connect to various sources like Redshift, S3, Dynamo, RDS, files like JSON, text, CSV, TSV, Jira, Salesforce, and on-premises oracle SQL-server. It is the first BI service to offer pay-per-session pricing, where you only pay when your users access their dashboards or reports, making it cost-effective for large-scale deployments.

Key Takeaways

· Introduction to Snowflake
· Understanding Snowflake Architecture
· Understanding Security in Snowflake
· Preparation of files
· Configuration setup for Snowflake
· Loading data through the web interface
· Loading data through SnowSQL
· Loading data using Cloud Provider
· Streaming data using Snowpipe
· Visualization using QuickSight
· Understanding pricing of Snowflake
· Time Travel in Snowflake
· Performance optimization in Snowflake

Architecture