**SNOWFLAKE**

What Snowflake is used for?

🡪Snowflake enables data storage, processing, and analytic solutions that are faster, easier to use, and far more flexible than traditional offerings. The Snowflake data platform is not built on any existing database technology or “big data” software platforms such as Hadoop.

🡪Snowflake supports both ETL and ELT and works with a wide range of data integration tools, including Informatica, Talend, Tableau, Matillion and others.

🡪Is Snowflake SQL or Nosql?

Snowflake is fundamentally built to be a complete SQL database. It is a columnar-stored relational database and works well with Tableau, Excel and many other tools familiar to end users.

🡪Developed in 2012, *Snowflake* is a fully managed SaaS (software as a service) that provides a single platform for data warehousing, data lakes, data engineering, data science, data application development, and secure sharing and consumption of real-time / shared data.

What makes up the Snowflake platform?

How Snowflake is designed is through three main components. These are the foundation for Snowflake’s cloud data platform:

**🡪Cloud services**. Snowflake uses ANSI SQL for cloud services empowering users to optimize their data and manage their infrastructure. Snowflake handles the security and encryption of stored data. They maintain robust data warehousing certifications such as PCI DSS and HIPAA. Services include authentication, infrastructure management, query parsing and optimization, metadata management, and access control.

When data is loaded into Snowflake, Snowflake reorganizes that data into its internal optimized, compressed, columnar format. Snowflake stores this optimized data in cloud storage.

Snowflake manages all aspects of how this data is stored — the organization, file size, structure, compression, metadata, statistics, and other aspects of data storage are handled by Snowflake. The data objects stored by Snowflake are not directly visible nor accessible by customers; they are only accessible through SQL query operations run using Snowflake.

🡪Query processing. The compute layer of Snowflake is made up of virtual cloud data warehouses that let you analyze data through requests. Each Snowflake virtual warehouse is an independent cluster and they do not compete for computing resources nor affect the performance of each other — which means workload concurrency is never a problem.

Query execution is performed in the processing layer. Snowflake processes queries using “virtual warehouses”. Each virtual warehouse is an MPP compute cluster composed of multiple compute nodes allocated by Snowflake from a cloud provider.

Each virtual warehouse is an independent compute cluster that does not share compute resources with other virtual warehouses. As a result, each virtual warehouse has no impact on the performance of other virtual warehouses.

🡪Database storage. A Snowflake database is where an organization’s uploaded structured and semistructured data sets are held for processing and analysis. Snowflake automatically manages all parts of the data storage process, including organization, structure, metadata, file size, compression, and statistics.

The cloud services layer is a collection of services that coordinate activities across Snowflake. These services tie together all of the different components of Snowflake in order to process user requests, from login to query dispatch. The cloud services layer also runs on compute instances provisioned by Snowflake from the cloud provider.

Services managed in this layer include:

* Authentication
* Infrastructure management
* Metadata management
* Query parsing and optimization
* Access control

What is a stage?

A stage specifies where data files are stored, So it can be loaded into the table.

Stage is the location of files, that can be external or internal to snowflake.