and its APIs

Asma Ahmed
Mhea Bustria
Justin McCarthy
Mane Telpian

What is Snowflake?

Key Concepts

- Analytic data warehouse

Data Warehouse as a Cloud Service

- Easy and faster to access

Data Lifecycle

- Databases, Schema, Table

Continuous Data Protection

- Features for data protection

Key Concepts

- Analytic data warehouse architecture
  - Uses Software-as-a-Service (SaaS) distribution model
- Faster and more flexible than its competitors
- Uses SQL database engine, with its own architecture

Data Warehouse as a Cloud Service

- No hardware to select, install, configure, or manage
- No software to install, configure, or manage
- Maintenance and management are handled by Snowflake
- Run on public cloud infrastructure
- No need to install or manage anything!

Data Lifecycle

- Data shown as tables that are quired and accessed using standard SQL interfaces
- Each table belongs to a schema which belongs to a database
- No limit to how many databases, schemas, or tables can be created
- Data is inserted directly and can be accessed using SQL commands
    - i.e SELECT, DELETE etc

Continuous Data Protection (CDP)

- Feature provided by Snowflake that protects data from:
  - Human Error
  - Malicious acts
  - Software and hardware failures
- Allows for data to be recovered
- Features of CDP:
  - Policies that restrict or grant users access
  - Verification/authentication for account access
  - Security roles that control user access
  - Encryption of data and files
  - Maintenance of historical data and failsafe

What does Snowflake have to offer?

- **Architecture: All three layers run on cloud infrastructure**

  → No additional hardware or software required.

- **Automation**

  → Save time

- **Pay as you need**

  → Save money

Architecture

- All **three layers** can be deployed and managed on a cloud platform.
  - **Centralized storage layer**
    - → All workloads can access the same data
  - **Multi-cluster computing layer**
    - → Workloads are isolated from each other
  - **Cloud services layer**
    - → Manage aspects such as security, client sessions, and query optimization.

Automation

- Multi-cluster virtual warehouses

  → Automatic scaling, and workloads can run without interfering with each other.

- Self-organizing cloud storage

  → Automatic clustering

- Materialized views

  → Automatically maintain accuracy as the data is modified

Pay as you need

- Most features can be modified, suspended, and resumed as needed.
  - **Example:** Having virtual warehouses means that a data warehouse does not need to be running at full capacity at all times.
  - **Example:** While automatic clustering is being done in the background, cluster keys can be modified at any time. Automatic clustering can also be suspended and resumed.
  - **Example:** Materialized views can be suspended at any time so that resources are not being used unnecessarily.

How to use Snowflake?

DEMO

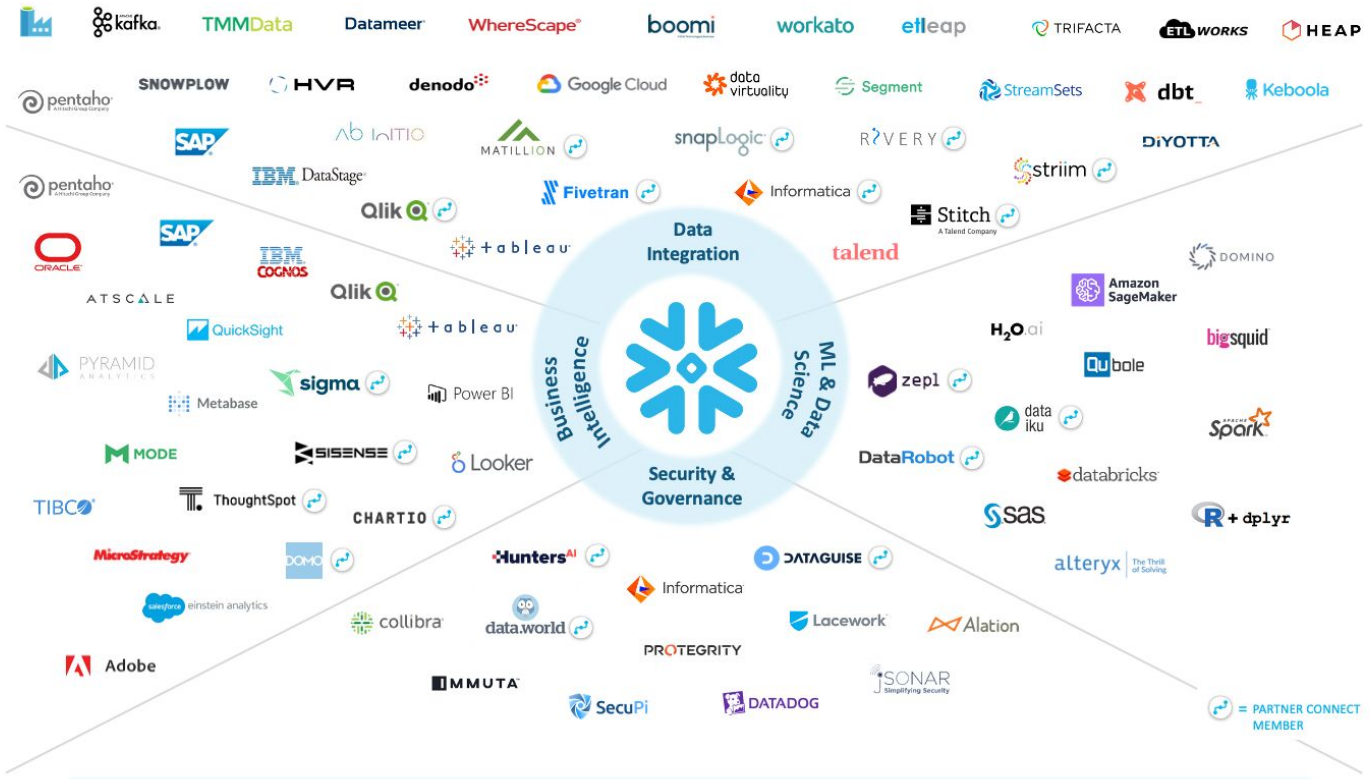
APIs

⁎ Connectors/Drivers-

* SnowSQL (CLI Client)-

```
D:\Snowflake\Test_scripts>snowsql -f test.sql -D id=1;
* SnowSQL * v1.2.1
Type SQL statements or !help
+----+------+--------+
| ID | NAME | CITY   |
|----+------+--------|
|  1 | AAA  | London |
+----+------+--------+
1 Row(s) produced. Time Elapsed: 0.406s
```

Snowflake Ecosystem:



SQL Development & Management

erwin · SqlDBM · DataOps · agnirty · Agile Data ENGINE

Worksheets (UI) · SnowSQL (CLI) · SEEKWELL · DBeaver · SQL Workbench/J

Native Programmatic Interfaces

python + SQLAlchemy · node · GO · .NET

JDBC · ODBC

Competitors

Main competitors and alternatives chosen by developers

→ Amazon Redshift

→ Google BigQuery

→ Microsoft Azure Synapse

→ IBM Db2 Warehouse

IBM **Db2**

Competitors

Amazon Redshift

→ Speed: Columnar storage and Massively Parallel Processing

Google BigQuery

→ Cost: allows users to run analytics environment with three-year TCO, around 34% cheaper than most cloud data warehouses

Microsoft Azure Synapse

→ Offers: workload optimization, business intelligence and machine learning tool integration, Cloud-native HTAP integration, and security features

IBM Db2 Warehouse

→ Offers: in-memory technology, advanced management and development tools, storage optimization, workload management, actional compression, and endless data availability

Advantages of Snowflake

Allows per-second pricing

Performance and Speed

Deals with concurrency issues

Seamless data sharing

Tolerates component and network failures

Thank you!

A PDF of these slides and the URL to our video can be found at
https://github.com/bustr003/cs446_snowflake