



Session 2- Day2- Snowpro Core certification Snowflake Deep Dive

Santosh Ubale & Robby Thomas GSI Sales Engineer 27 Jul 2021

AGENDA

Day 2

- Data Security & Governance
- Data Masking
- External tokenization
- Secure Data Sharing
- Snowflake Account Options
- Data Movement (Loading & Unloading)
- Snowflake Performance & Tuning
- Built In Monitoring and Management Tools
- Snowflake Editions, pricing and regions

➤ Snowflake Account and Security



SNOWFLAKE SECURITY AT A GLANCE



Access

- All communication secured & encrypted
- TLS 1.2 encryption in both trusted and untrusted networks
- [IP Whitelisting](#)



Authentication

- Password Policy enforcement
- [Multifactor Authentication](#)
- [SAML 2.0 support for Federated Authentication](#)



Application

- Flexible user management
- Role-based access control for granular control
- [RBAC](#) for data and actions



Snowflake Operational Controls

- NIST 800-53
- SOC2 Type 2
- HIPAA
- PCI
- FedRAMP



Data

- [Encrypted at rest](#)
- Hierarchical key model rooted in [Cloud HSM](#)
- Automatic key rotation
- [Time Travel](#) 1-90 days
- Tri-Secret Secure
- [Query statement encryption](#)



Infrastructure

- AWS, Azure Physical Security
- AWS, Azure Redundancy
- Regional Data Centers
 - US
 - EU
 - AP



PROTECTING YOUR DATA IN SNOWFLAKE

End-to-End Encryption

Always-encrypted client communications, plus integration with cloud provider private networking



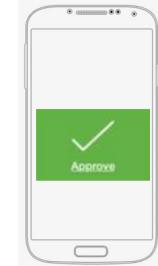
Fully Encrypted Storage

Data at rest is always encrypted while handled by the Snowflake drivers and systems



Strong Authentication

Built in multi-factor, integration with your federated SSO, easy user management



Full Auditing

Track every login, every transaction, every data transfer, and export to your security tools



Role-Based Access Control

All objects, actions, and even compute usage can be controlled with roles



Recovery

We give you options to ensure your data can be recovered in case of an accident or worse



[Snowflake Security Product Documentation](#)



SNOWFLAKE WILL BE 100% TRANSPARENT

Third Party Attestations and Certifications

Snowflake Security and Trust Center: <https://www.snowflake.com/product/security-and-trust-center/>

Snowflake Security Policy: <https://www.snowflake.com/legal/> (first link)



SOC 2 Type II
12 Month Coverage Period
SOC 1 Type II
6 Month Coverage Period



Security Standards Council[®]

PCI-DSS



ISO/IEC 27001



FedRAMP

(Available from OMB MAX)



HIPAA

HITRUST Certified

Self-Assessments
CAIQ, SIG Lite, Pen Test Results

DATA SECURITY & GOVERNANCE

Ensure data privacy and restrict access to source data

SECURE VIEWS



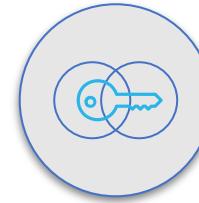
Cell-level security in multi-tenant situations

SECURE UDFs



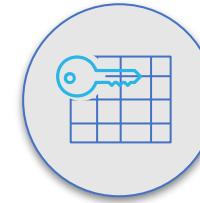
Prevent consumers from seeing underlying data

SECURE JOINS



Securely mask data during join operations

SECURE MATERIALIZED VIEWS

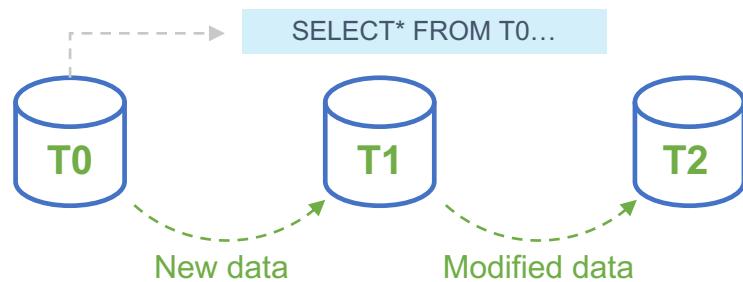


Secure, precomputed views

COMPREHENSIVE DATA PROTECTION



Protection against infrastructure failures
All data transparently & synchronously **replicated**
3+ ways across independent infrastructure



Protection against corruption & user errors
“**Time travel**” feature enables instant roll-back to
any point in time during chosen retention window



Long-term data protection
Zero-copy clones + optional export to cloud
object storage enable user-managed data copies

CONTROL YOUR DATA

Dynamic Data Masking

Dynamically Mask Protected (PII, PHI) Column Data at Query Time

- No change to the stored data
- Mask or partial mask using constant value, hash, and custom functions
- Unmask for authorized users only

Policy Based Control

- Table/View owners and privileged users (such as account admin) unauthorized by default
- Centralized policy mgt

Ease of Management

- Apply single policy to multiple columns
- Prevent secure view explosion



Alex
(Unauthorized)



Jordan
(Authorized)

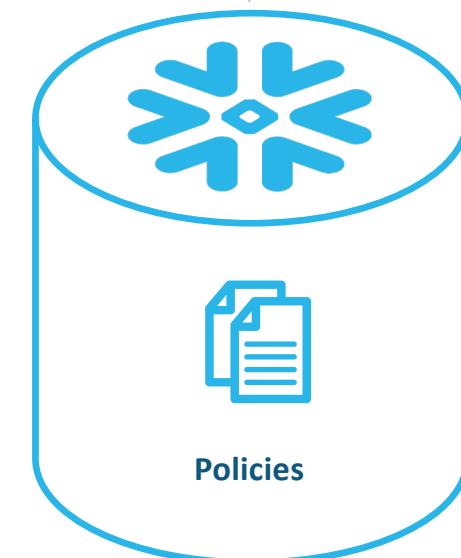
ID	Phone	SSN
----	-------	-----

101	***-***-5534	*****
102	***-***-3564	*****
103	***-***-9787	*****

ID	Phone	SSN
----	-------	-----

101	408-123-5534	*****
102	510-335-3564	*****
103	214-553-9787	*****

Ingest Raw Data



CONTROL YOUR DATA

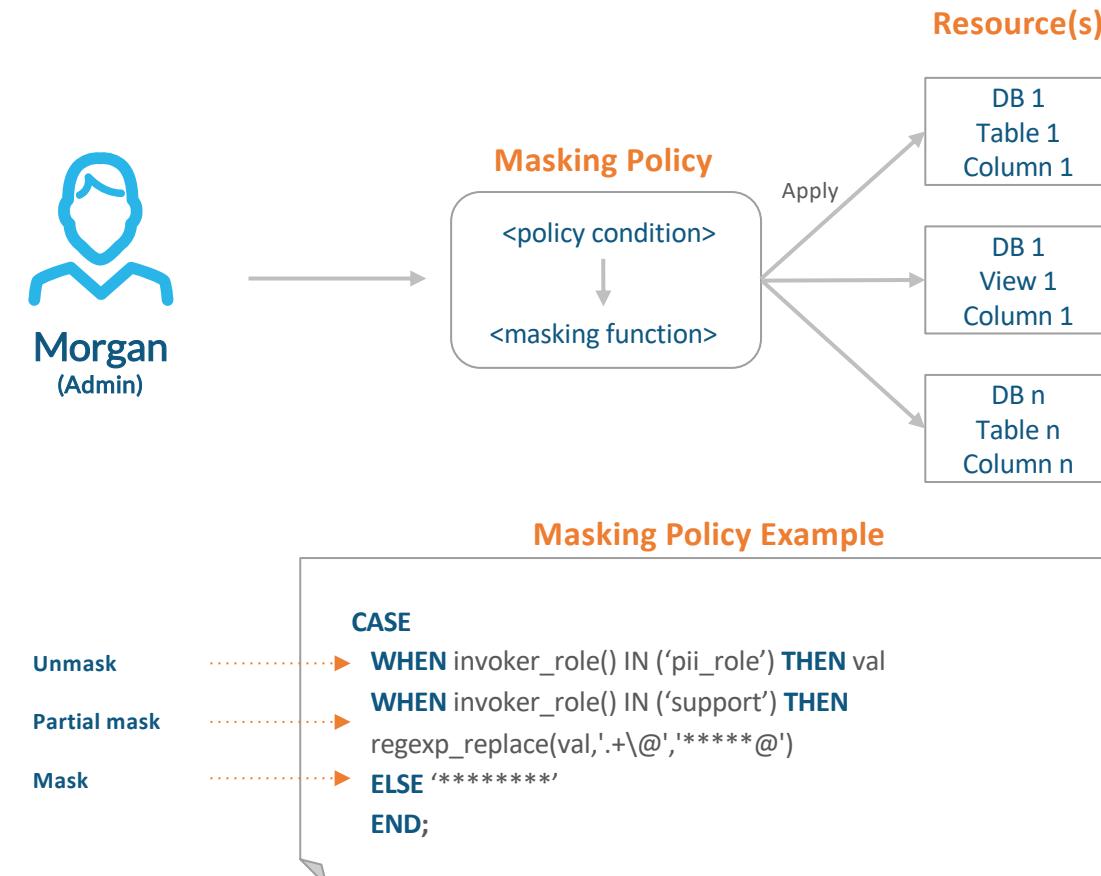
Dynamic Data Masking Policies

Masking Policy

- Policy contains condition(s) and masking function to apply under those conditions
- Policy is applied to one or more table, view, or external table columns in an account
- Nested policy execution for views — policy on table executed before policy on view(s)

Supports:

- All data types
- Data sharing
- Streams
- Clone carries over policy associations



GOVERNANCE AND SECURITY

External tokenization using third party

Ingest Protected (PII/PHI) Data as Externally Tokenized

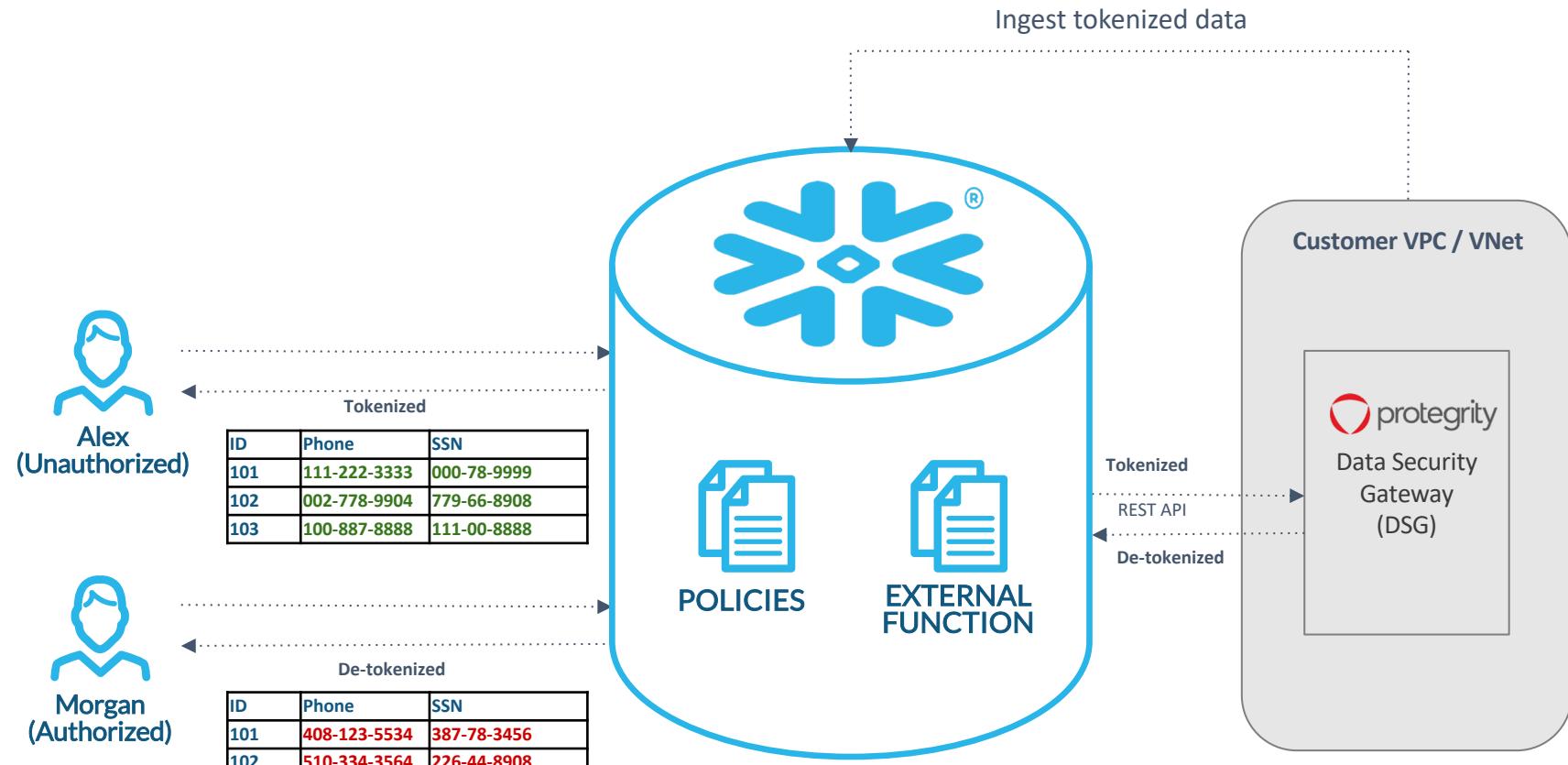
- Using Protegity agents on ETL tools

De-tokenize for Authorized Users at Query Time

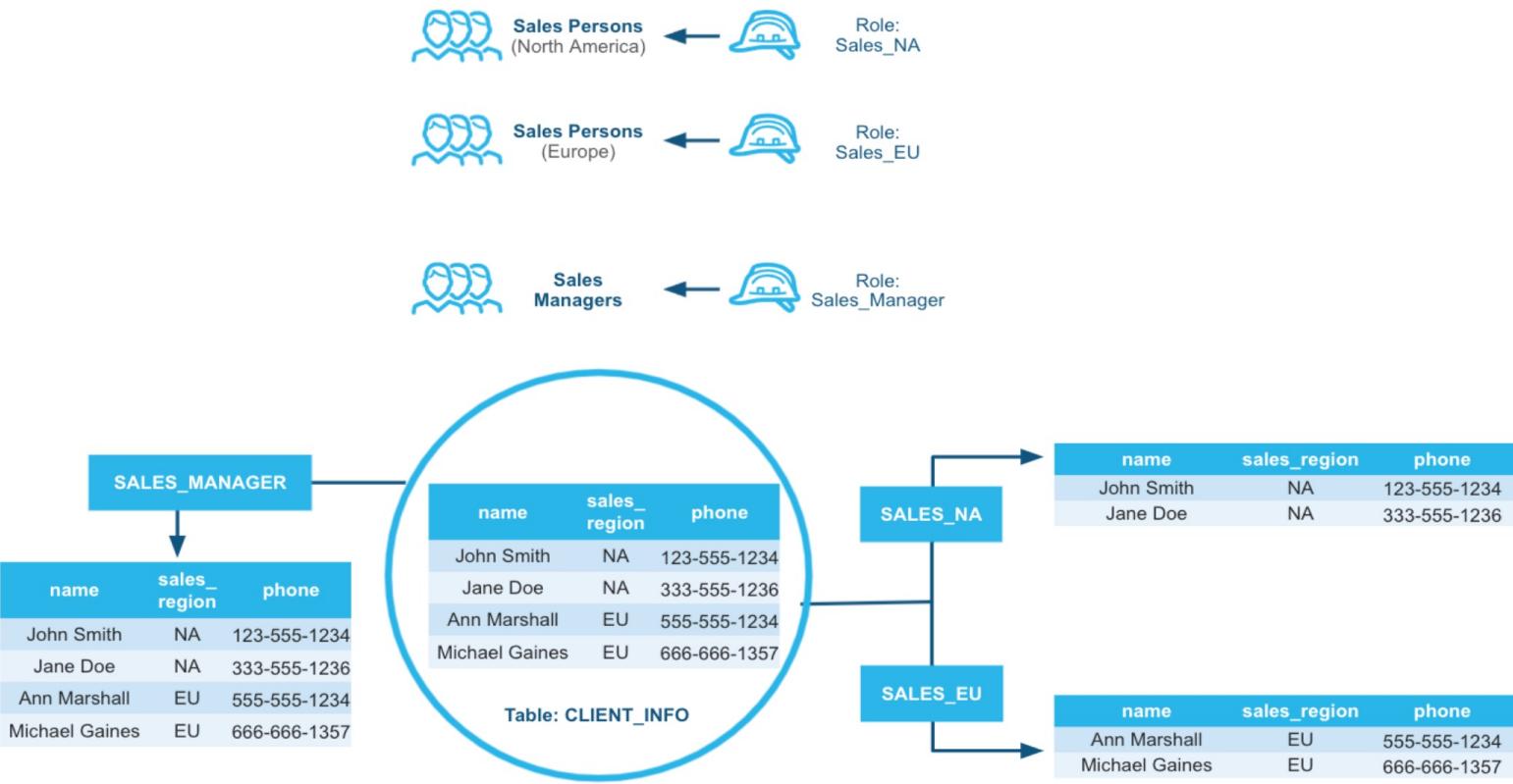
- Protegity DSG called using external functions to de-tokenize data
- For unauthorized users, Protegity DSG is not called

Policy Based Control

- Table/View owners and privileged users (such as accountadmin) unauthorized by default
- Centralized policy mgt



How Row Access Policies Work



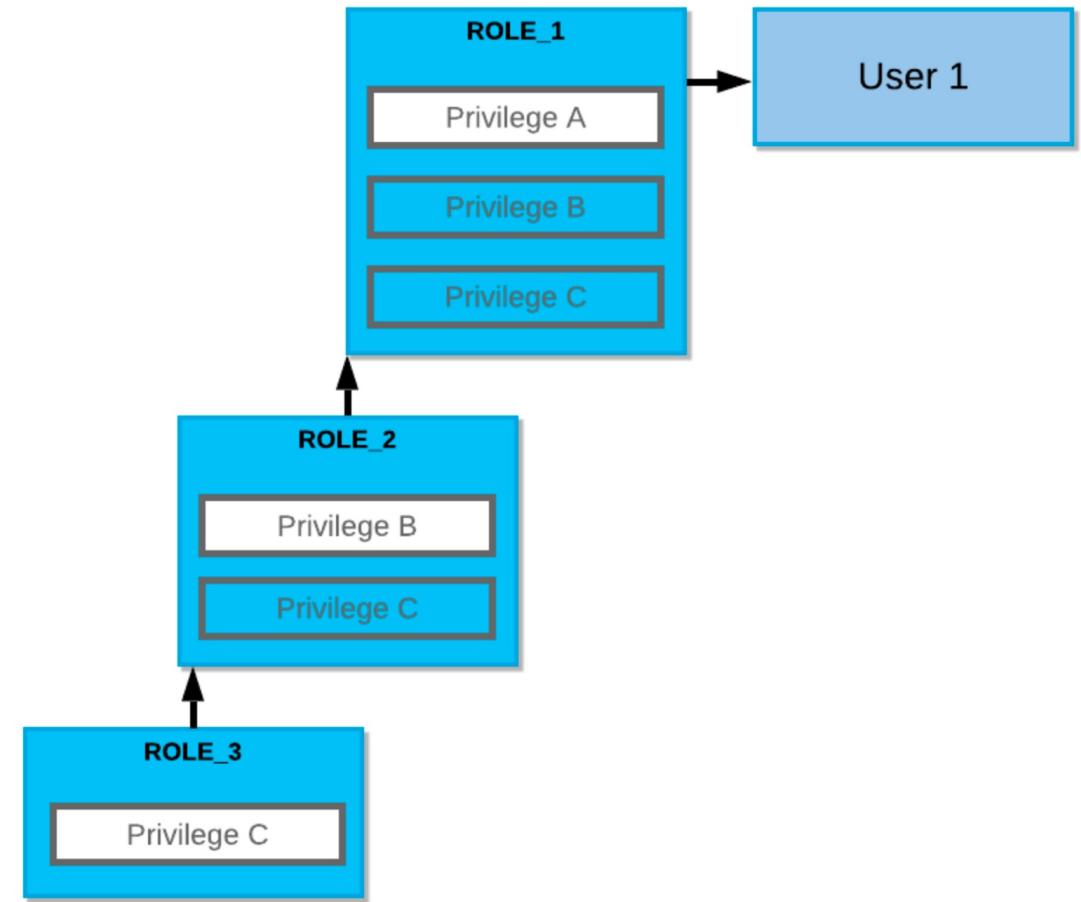
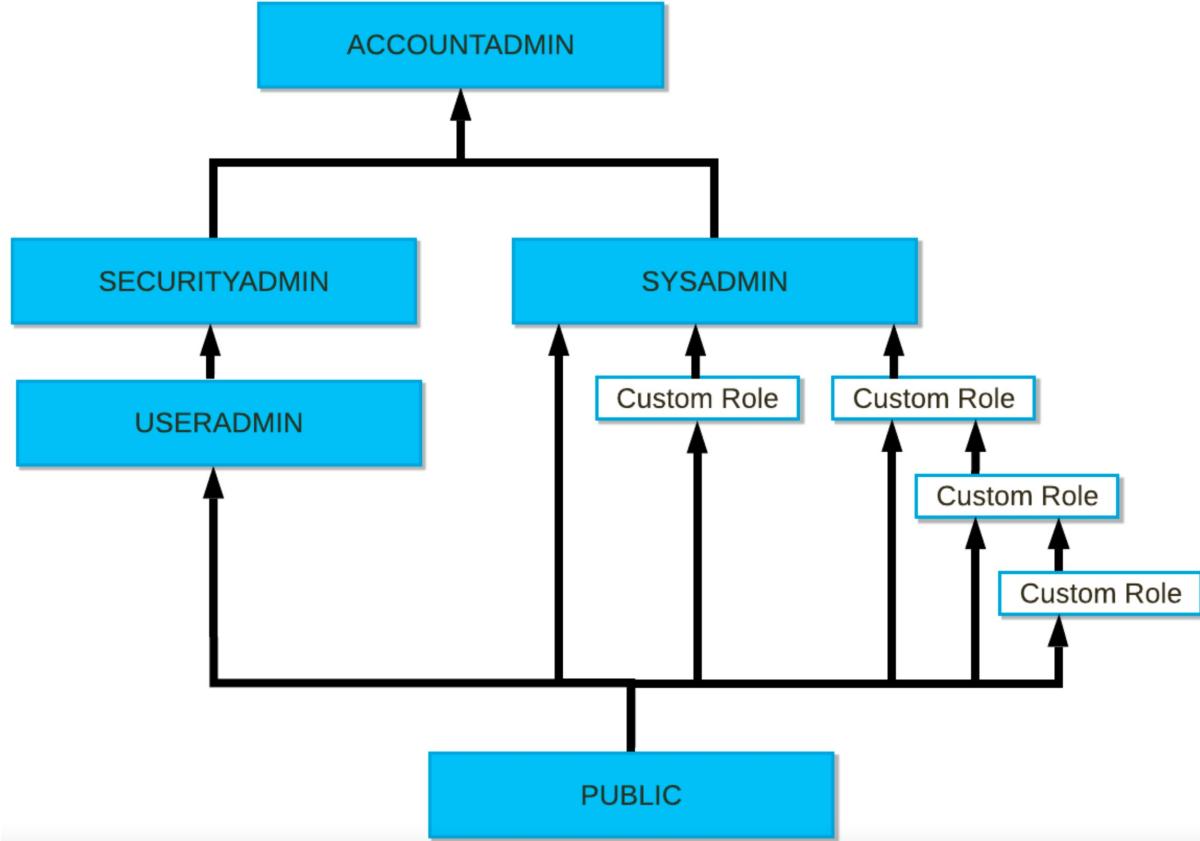
- A single policy can be applied to multiple tables, which greatly simplifies policy administration, change management, and compliance reporting.
- For organizations with many users and complex authorization logic, Snowflake supports a data-driven approach where authorizations are stored in mapping tables.
- Snowflake supports a hybrid policy administration model: Policy administrators can define and manage policies centrally, while data stewards throughout the organization apply these policies to data.
- Row Access Policies can be replicated securely through Snowgrid and applied across multiple accounts and regions. This enables enterprises to easily achieve a consistent security posture worldwide.

You need to perform only three simple steps to use Row Access Policies to accomplish row-level security:

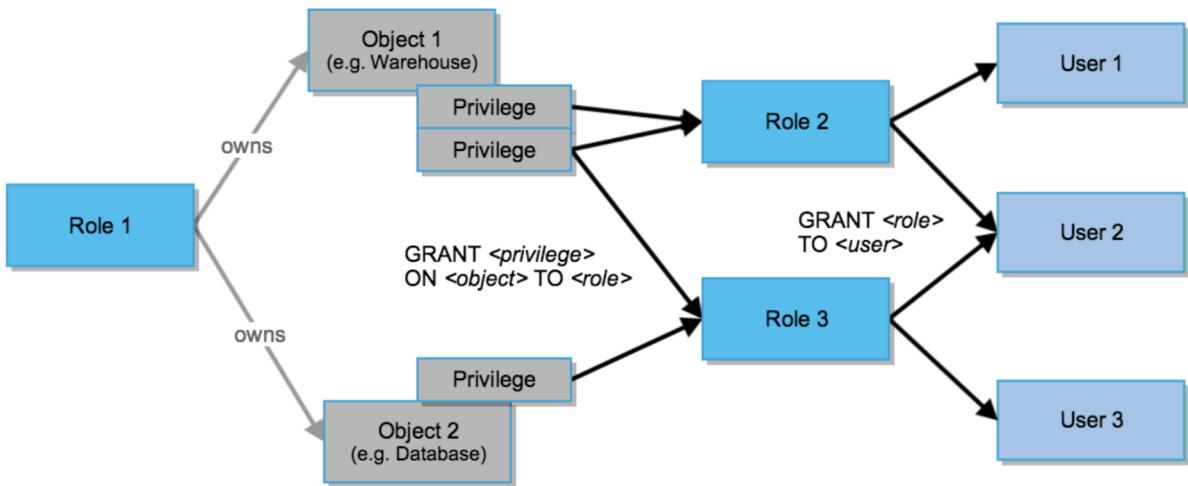
- Define a policy and optionally define a mapping table.
- Apply the policy to one or more tables.
- Query the data.



Role Hierarchy and Privilege Inheritance



ACCESS CONTROL



Granular control over access to objects

Covers who can access what objects, what operations can be performed on those objects, and who can create or alter access control policies

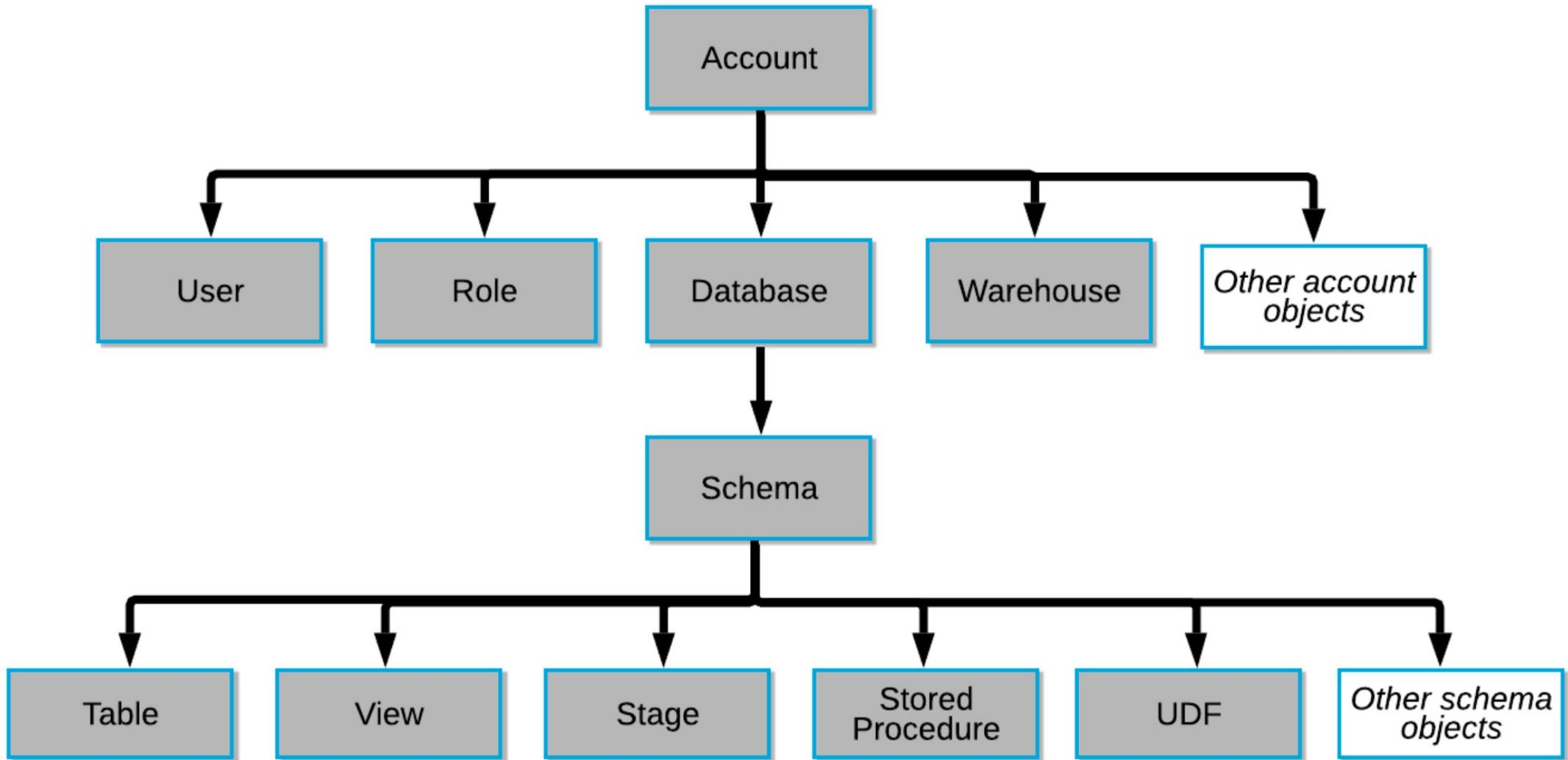
Two modes:

- Role-based Access Control (RBAC)
- Discretionary Access Control (DAC)

All object access authorizations occur through granted role grants

Independent of authentication

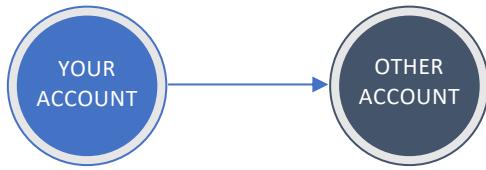
Securable Objects



SECURE DATA SHARING

A single, live copy of the data: no copying, no moving, no delays.

DIRECT SHARE



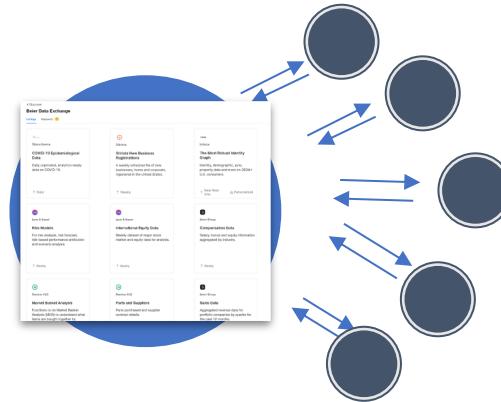
For sharing data 1-to-1



Square sharing marketing data with a partner

DATA EXCHANGE

YOUR DATA
EXCHANGE IN
YOUR
ACCOUNT

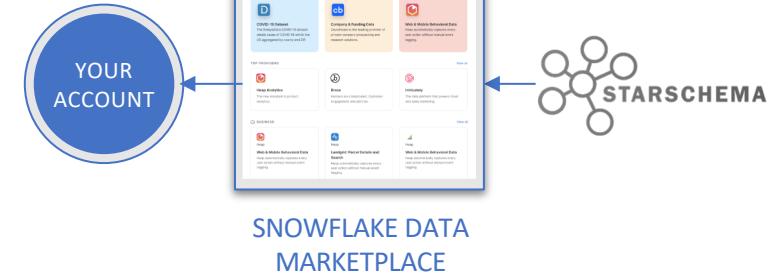


For data sharing AT SCALE:
many parties, many data sets



Cisco's data exchange for its suppliers

SNOWFLAKE DATA MARKETPLACE

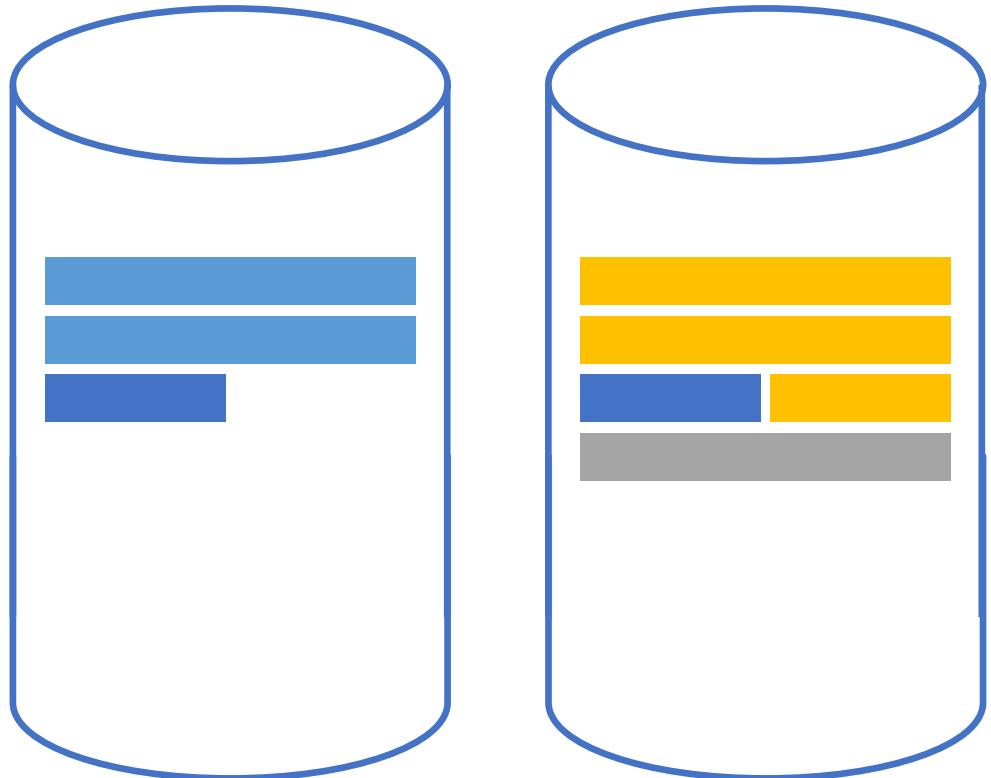


Consuming data from 3rd parties:
Access data from over 150 providers



StarSchema COVID data

DATA SHARING BOTH ON SNOWFLAKE



1. Account A grants share to Account B
2. Account B now has read-only on the share
3. Account B can join with its own data for new business value
4. New business intelligence insights

SNOWFLAKE DATA MARKETPLACE

Snowflake is the only solution for accessing live data and data services globally

The screenshot shows the Snowflake Data Marketplace interface. On the left, a sidebar menu includes options like Worksheets, Dashboards, Data, Databases, Shared Data, Marketplace (which is selected and highlighted in blue), and Admin. Below the sidebar is a "Classic Console" link. The main content area has a header "Discover" and "Snowflake Data Marketplace". It features a search bar and a "Covid Economic Recovery" section with three cards: "COVID-19 Reemergence Tracker - State / Category" (PlaceIQ), "In-Store and Online Purchase Data" (Ibotta), and "AI Powered COVID-19 Insights & Analytics" (Accern). Below this is a "Featured providers" section with cards for S&P Global Market Int. (16 Listings), Experian (24 Listings), and FactSet (33 Listings). The "Most recent" section lists three new listings: "Consumer Data Insights – Platinum – 2020" (Epsilon), "Company Intelligence" (S&P Global Market Intelligence), and "Actionable Weather Forecasts" (AccuWeather).

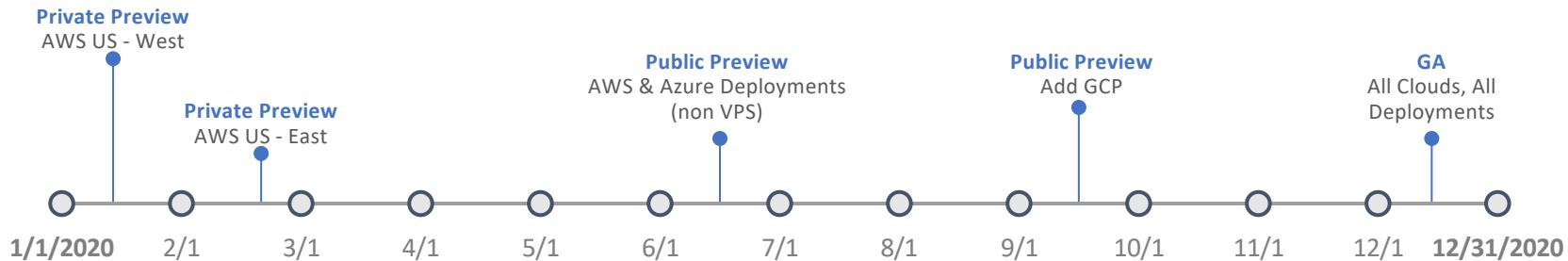
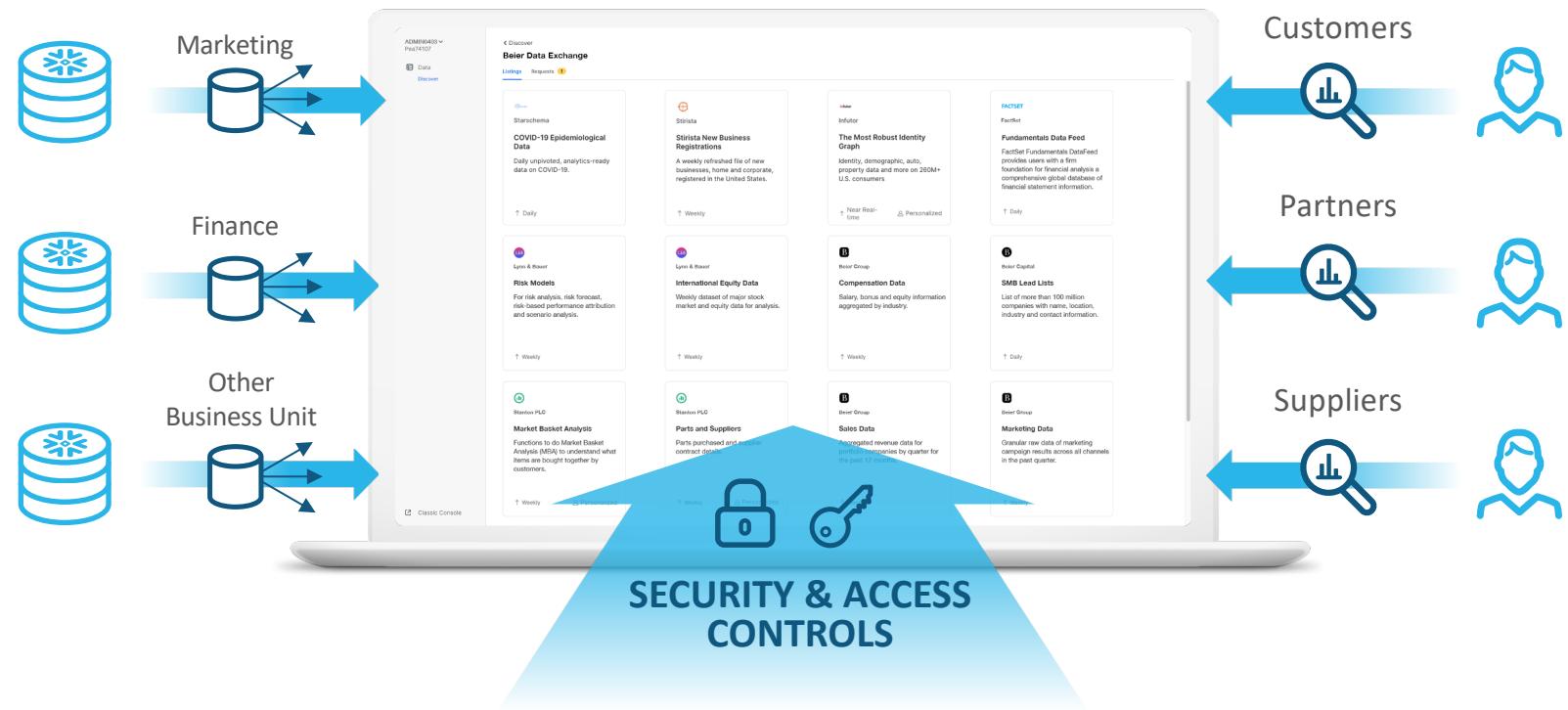
Live, ready-to-query
data; no copying or
moving

Data and data services
from over 150
providers

Globally available,
across clouds

CREATE YOUR OWN DATA EXCHANGE

- Consumers can discover shared datasets within your company or from outside
- Control who can publish/consume on your data exchange; request approval workflow
- Access metrics on who's using the listings
- All works on a single, live copy of data: no ETL, no copying, no moving, no delays.



Snowflake Account Options & Assurances

Account Options

Understanding Snowflake's strict compliance processes and certifications, a potential customer should feel more confident about exploring Snowflake with a trial account.

After entering standard information on the first page of the form, options an enrollee can expect to encounter include:

- Snowflake Edition
- Cloud Platform
- Geographic Deployment Region

The screenshot shows the Snowflake trial sign-up interface. At the top left is the Snowflake logo. To the right, a large blue arrow points right towards a white sign-up box. The box contains the text "Start Your 30-Day Free Trial" and "Receive \$400 of credits to try all Snowflake features". Below this are four input fields: "Florence" and "Neige" (placeholder names), "florence.dneige@gmail.com" (email address), and "Cafe De Mel" (placeholder for a deployment region). A large pink arrow points from the bottom left towards the "CREATE ACCOUNT" button, which is highlighted in blue. To the right of the sign-up box, the text "TEST DRIVE THE DATA WAREHOUSE BUILT FOR THE CLOUD" is displayed with a left-pointing arrow. At the very bottom of the page, it states "Snowflake is HIPAA, PCI DSS, SOC 1 and SOC 2 Type 2 compliant, and FedRAMP Ready" followed by logos for PCI DSS, HIPAA, AICPA SOC 2, and FedRAMP.

Snowflake Account Options & Assurances

Account Options - Snowflake Editions

The second part of the sign-up form asks users to **Choose your Snowflake edition**.

For more information on the various editions:

<https://docs.snowflake.com/en/user-guide/intro-editions.html>

Start Your 30-Day Free Trial

Receive \$400 worth of free usage with sign-up.

Choose your Snowflake edition

Standard

A strong balance between features, level of support, and cost.

Enterprise

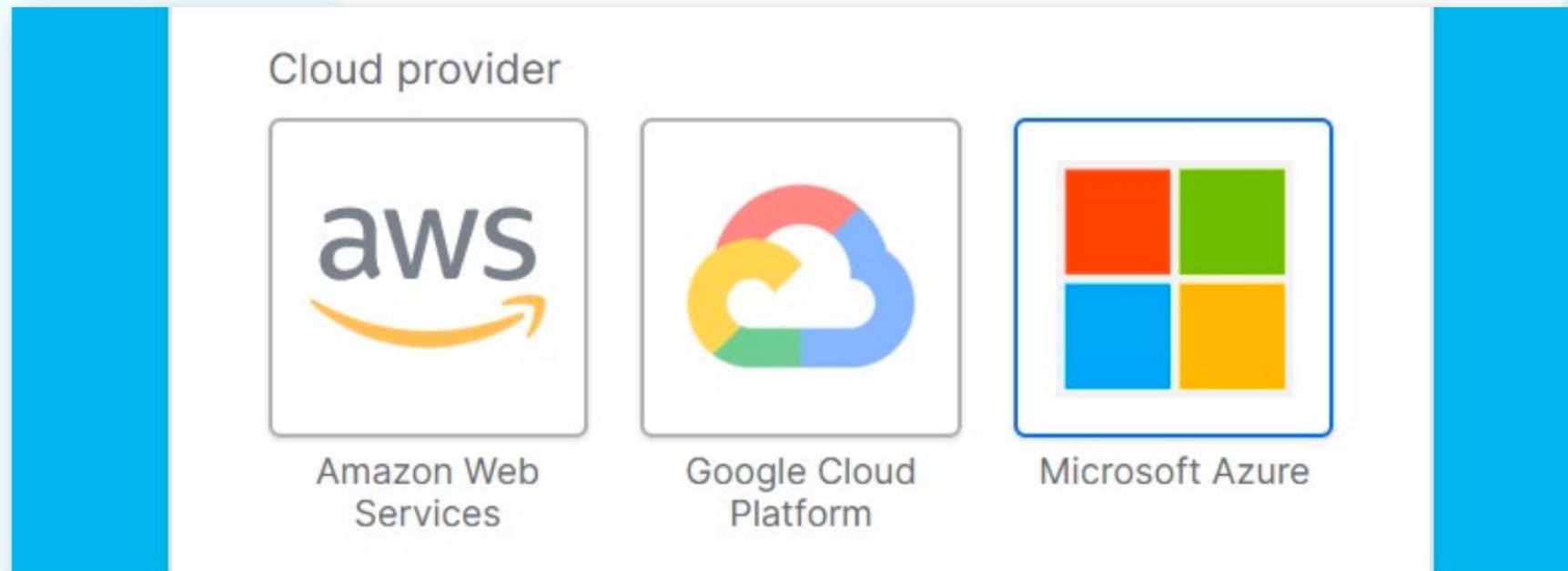
Standard plus 90-day time travel, multi-cluster warehouses, and materialized views.

Business Critical

Enterprise plus enhanced security, data protection, and database failover/fallback.

Snowflake Account Options & Assurances

Cloud Platform Provider



Next, users choose a **Cloud provider**.

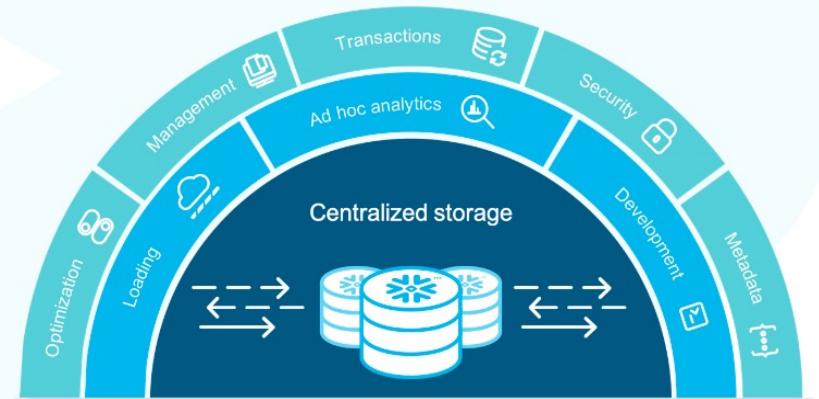
Snowflake Account Options & Assurances

Cloud Platform Provider

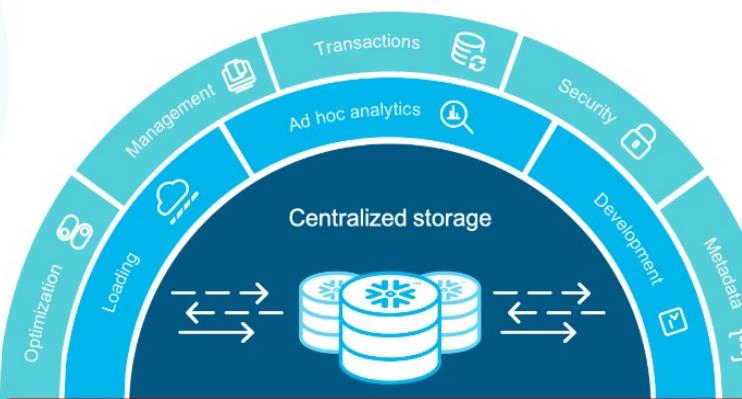
Snowflake could not operate without an underlying cloud provider infrastructure. When you choose a cloud, you choose which provider your Snowflake account sits on top of. However, after choosing your infrastructure provider, you do not interact with the cloud provider and you are not charged by them in the management of your Snowflake account.



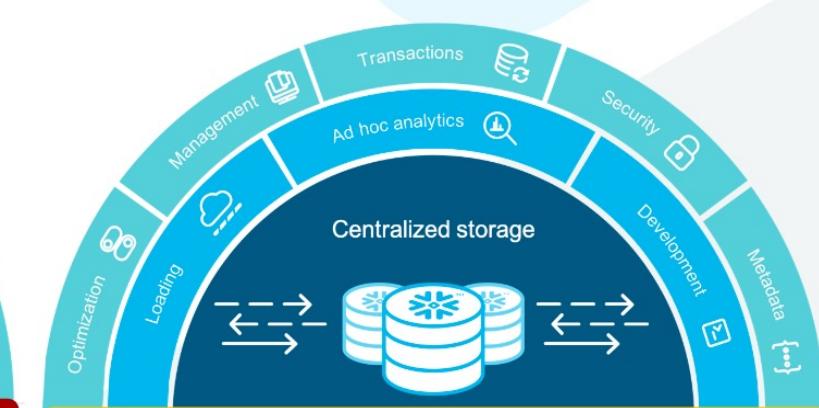
With Snowflake, you interact only with the Service layer. The Service layer interacts with Compute and Storage layers, which in turn interact with the underlying cloud infrastructure provider.



Uses Azure storage and compute resources behind the scenes.



Uses AWS storage and compute resources behind the scenes.



Uses GCP storage and compute resources behind the scenes.



GLOBAL SNOWFLAKE

Available on customer's cloud & region of choice



Generally Available

- US West (Oregon)
- US East (N. Virginia)
- Europe (Frankfurt)
- Europe (Ireland)
- Asia Pacific (Sydney)
- Asia Pacific (Singapore)
- Canada Central (Montreal)
- US East (Ohio)
- Japan (Tokyo)
- India (Mumbai)
- FIPS-ready region in US East (N. Virginia)
- Europe (London)

Coming H1 2021

- GovCloud West

Coming H2 2021

- GovCloud East
- Japan (Osaka)



Generally Available

- East US 2 (Virginia)
- West Europe (Netherlands)
- Australia East (New South Wales)
- US Government (Virginia)
- Canada Central (Toronto)
- Southeast Asia (Singapore)
- West US 2 (Washington)
- Switzerland North (Zurich)

Coming H1 2021

- US Central (Iowa)
- Japan (Tokyo)



Google Cloud Platform

Generally Available

- US Central 1 (Iowa)
- Europe West 4 (Netherlands)
- Europe West 2 (London)



Snowflake Account Options & Assurances

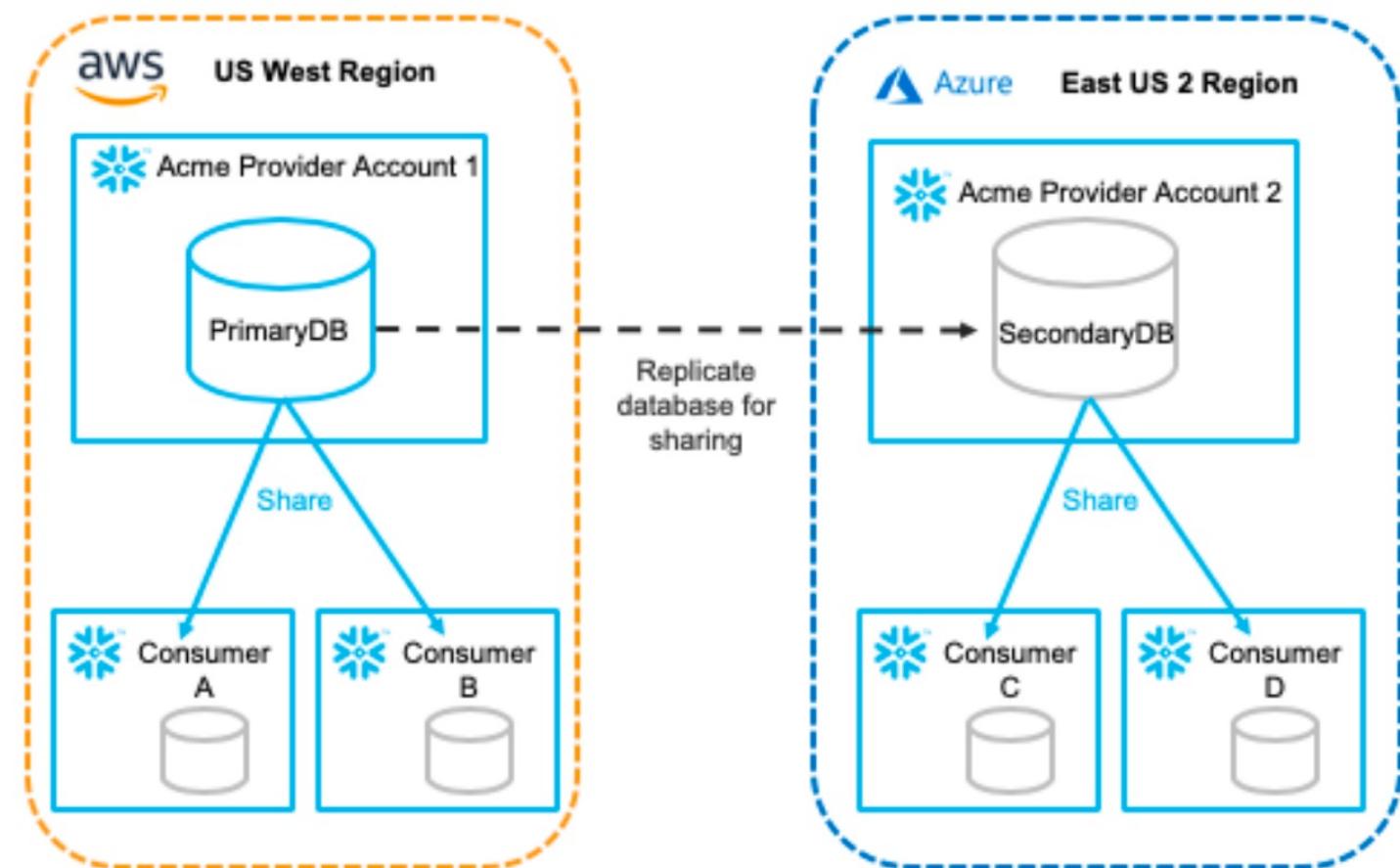
Multi-Region and Multi-Cloud for High Availability

What if a Geographic Region goes down?

Snowflake makes it easy to replicate your data across multiple regions and clouds.

For companies that need to have their data available in more than one region or on more than one cloud platform, Snowflake provides sharing and replication options that make this easy to achieve.

<https://docs.snowflake.net/manuals/user-guide/secure-data-sharing-across-regions-platforms.html>





Data Movement (Loading and Unloading)



HOW TO LOAD DATA INTO SNOWFLAKE

COPY/PUT Commands:

- Designed for batch loads
- Needs to be scheduled
- Needs a warehouse
- Micro-batching up to a point
- Easy to hit concurrency limits

Snowpipe:

- Designed for continuous loads
- No scheduling
- No warehouse needed
- Server-less billing model
- Designed for high concurrency



Batch

Micro-Batch

Continuous



HOW TO LOAD DATA INTO SNOWFLAKE

Bulk Loading (using COPY)

* Virtual Warehouse Required *

Local File System



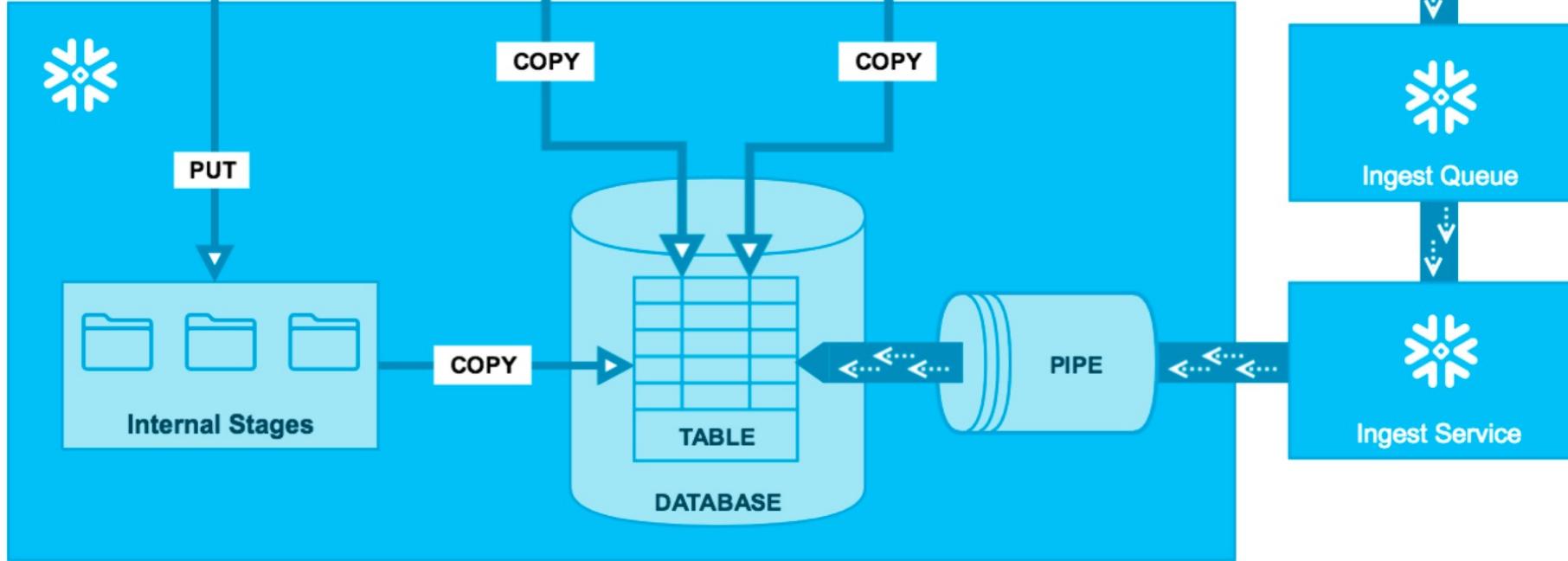
External Stages



Continuous Loading (using Snowpipe)

* No Virtual Warehouse Required *

Data Feed



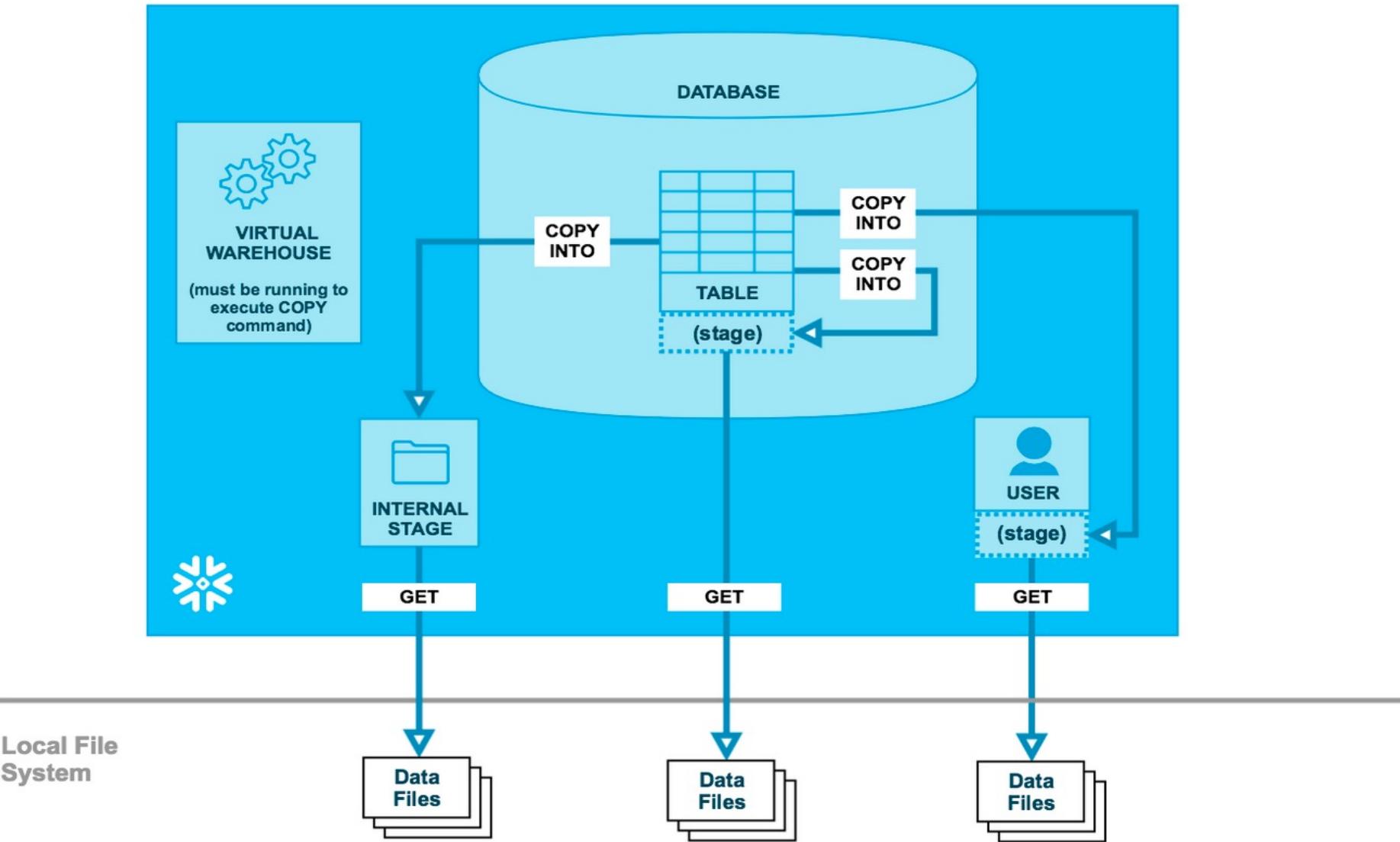
Batch/Bulk Loading:

- From local files
- From existing cloud landing zones

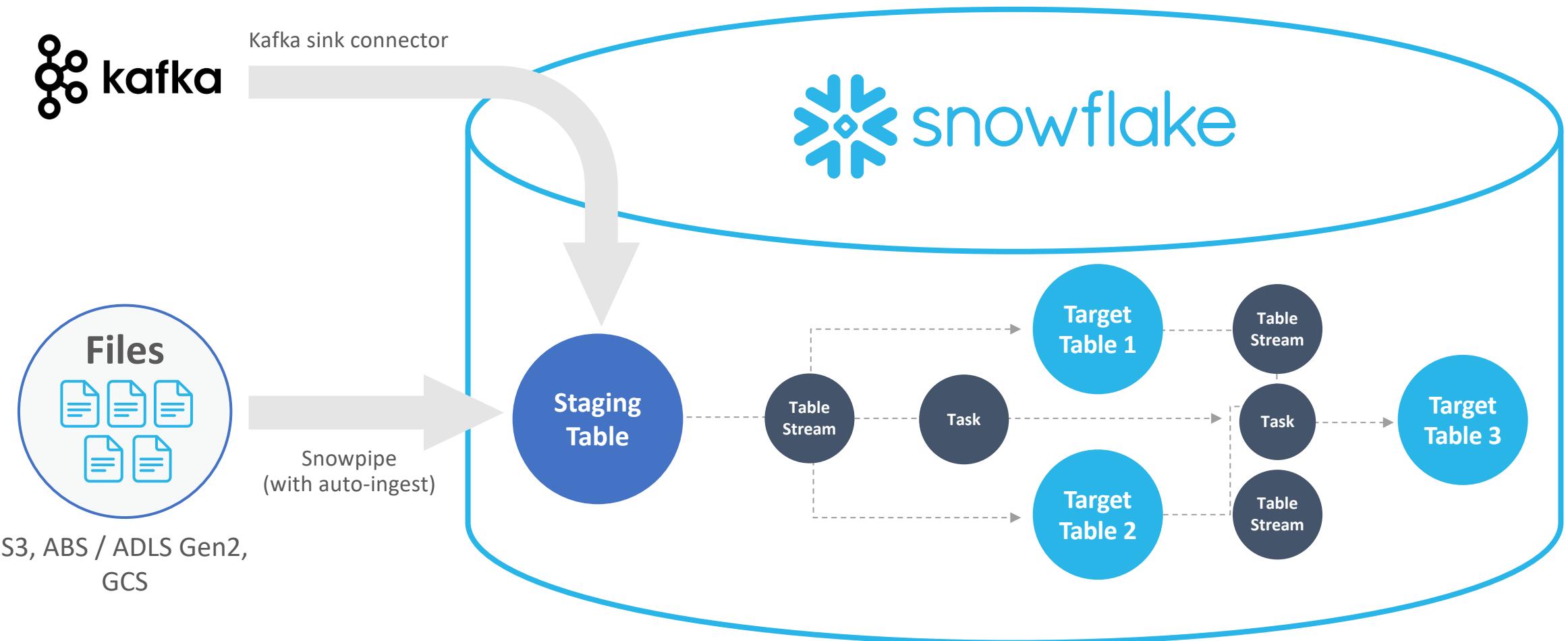
Continuous Loading:

- Snowpipe
- Serverless model (no user-managed warehouse needed)

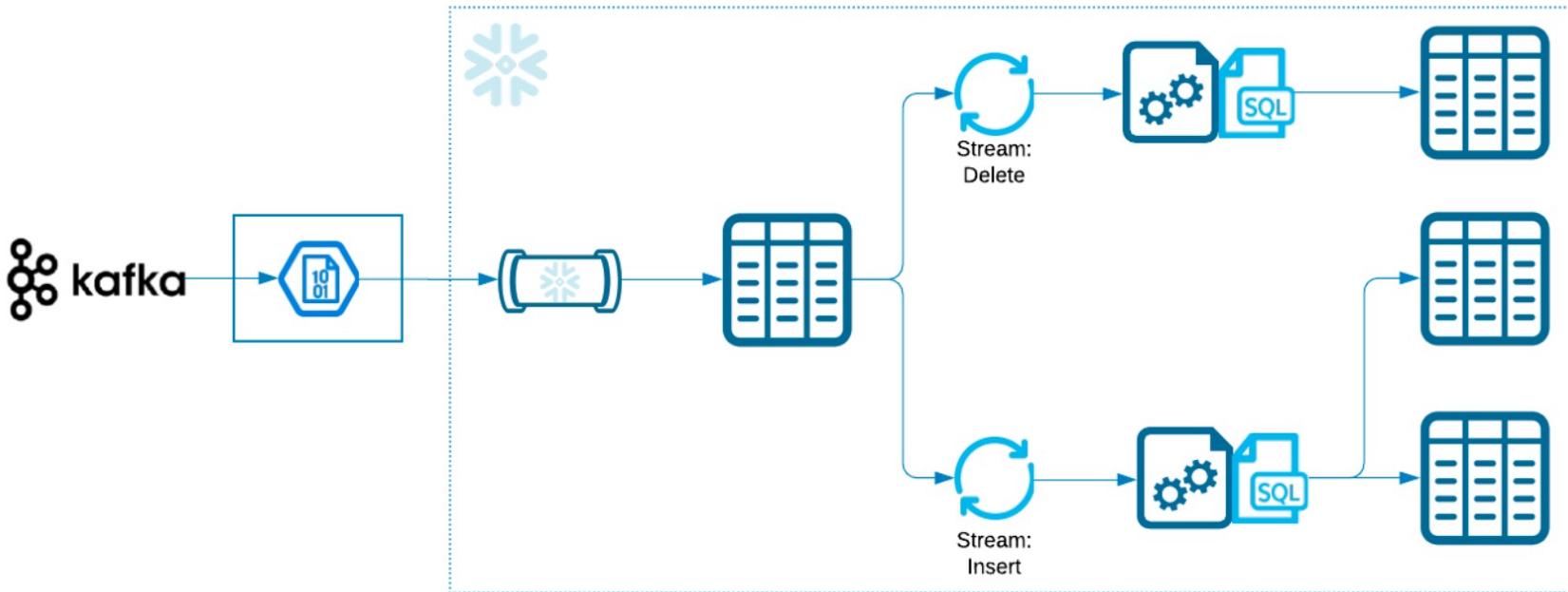
Unloading Data from Snowflake



CONTINUOUS DATA PIPELINES



STREAMS & Task



Multiple streams can be set on a table

CDC can track appends, deletes, or both (updates = inserts + deletes)

Tasks can be scheduled (CRON) to consume changes from a stream

Tasks can execute single SQL statements or call stored procedures

TASKS

Calls a single SQL statement or stored procedure using a stated virtual warehouse in the task configuration

Types:

Interval - repeats every X minutes (special parameter to enable seconds)

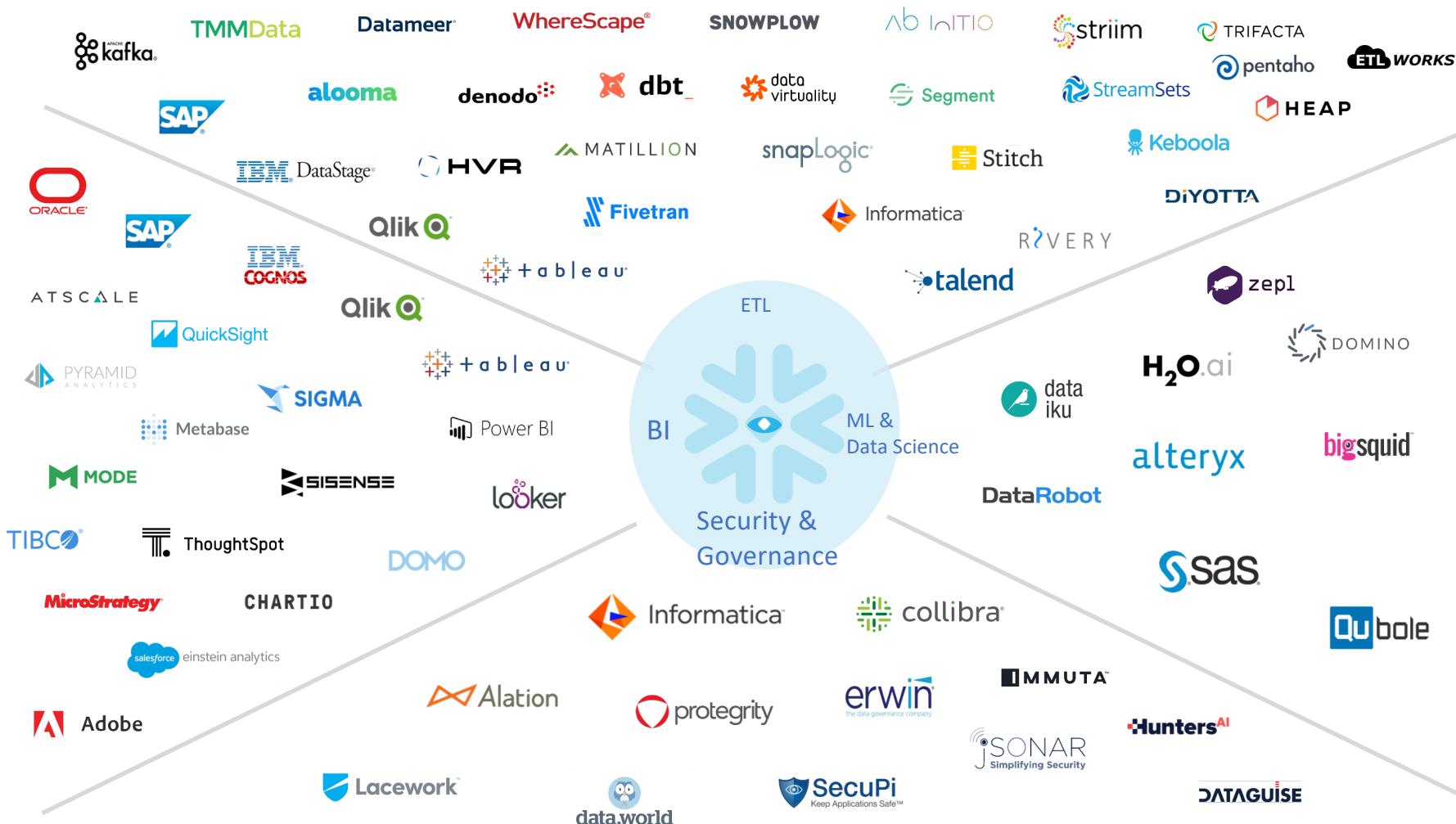
Schedule - repeats on a cron schedule

Predecessor - child task runs after a parent task completes; can daisy chain

All task types can also check for one or more streams to be populated before executing (will add more conditions over time)



Snowflake Technology Partner Ecosystem



➤ Snowflake Performance , Tuning & Semi- Structured



SANOFLAKE PERFORMANCE TUNNING

Cloud/Snowflake Architecture

Cloud provider and Snowflake-specific concepts

Snowflake Table Fundamentals

Logical and physical structures; Statistics

Query Processing

Caches; Cost-based optimizer

SQL Optimization

Standards; Filtering; Joins; Optimization techniques

Query Profiler

Interpreting graphical execution plan

Data Clustering

Concepts; Design considerations; Live demo

Internal Tooling (Profiler,Snovi)

Interpreting graphical execution plan (internal tool)



SNOWFLAKE QUERY PROCESSING - CACHING

Result Set Cache

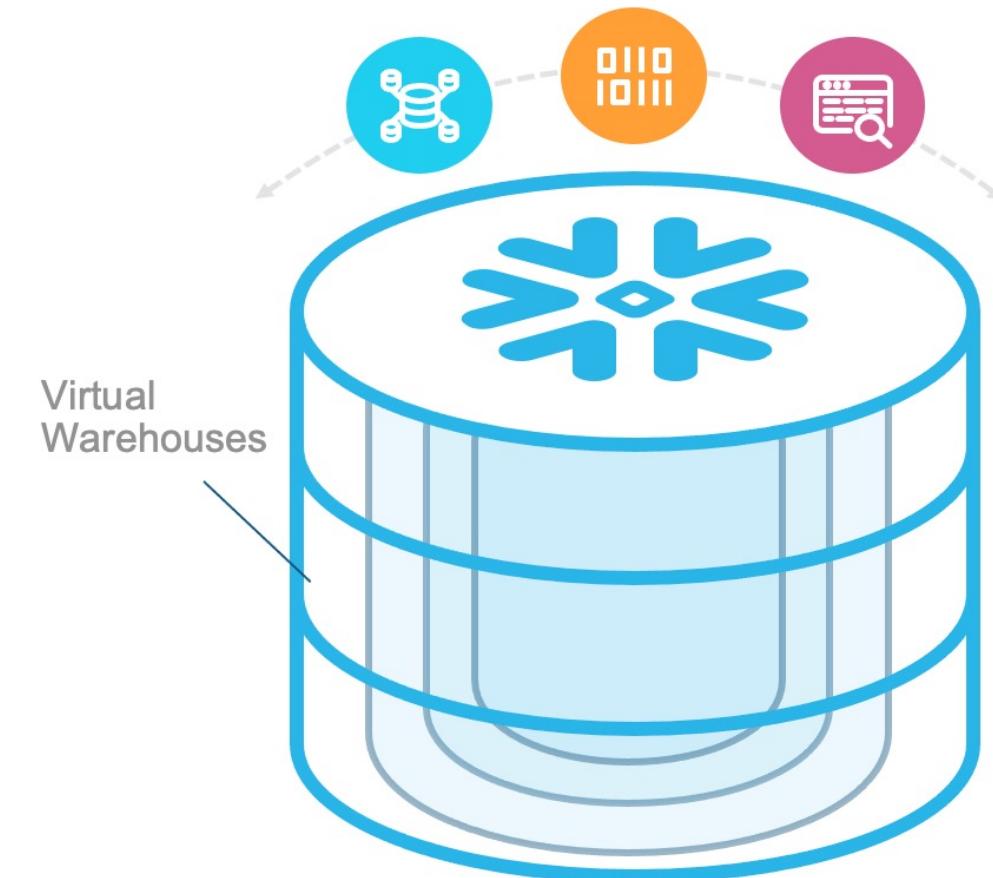
- Fastest way to retrieve data from Snowflake service
- Managed entirely within Global Services layer
- Result sets are cached in Cloud Storage layer
- No resources from Virtual Warehouse layer are utilized
 - **NOTE: compilation is STILL REQUIRED**
- **Eligibility requirements** for query to use result set cache...
 - Exact same SQL query (* except maybe whitespace)
 - Result must be deterministic (eg. no random function)
 - Recently extended to allow CURRENT_DATE()
 - Changes CAN be made to source table(s), but only if NONE affect any micro-partitions relevant to query
- Result sets are stored for 24 hours; counter resets each time matching query is re-used



SNOWFLAKE QUERY PROCESSING - CACHING

Data Cache

- Implemented within Virtual Warehouse layer
- Nodes in XP cluster have attached SSD storage, used for caching micro-partitions that have been pulled over the network from cloud storage service
- Reading from local SSDs is MUCH faster than reading from the cloud storage service; XP cluster will first read any micro-partitions available ‘locally’, then will read others from ‘remote’ cloud storage
- Cache stores FILE HEADERS and specific COLUMN data
- Cache has a finite size, and uses LRU (Least Recently Used) approach to flush out data that has not been accessed recently



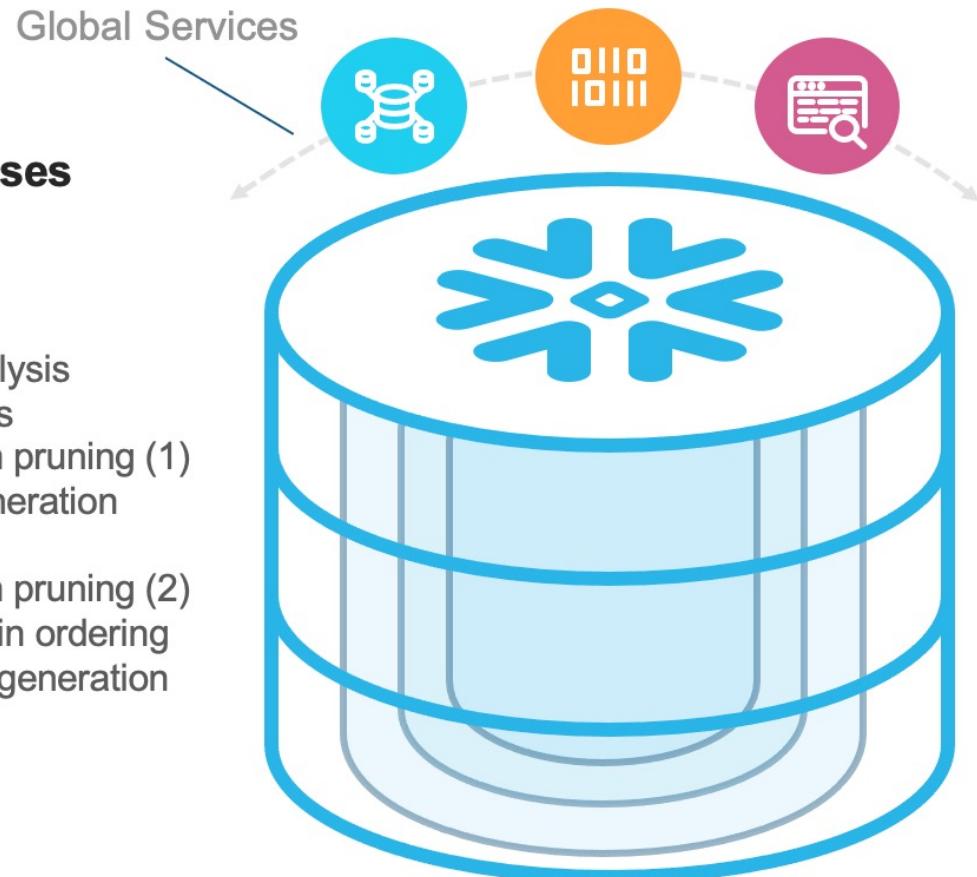
SNOWFLAKE QUERY PROCESSING - COST-BASED OPTIMIZER (CBO)

Cost-Based Optimizer

- Snowflake employs a powerful CBO in the Global Services layer that determines optimal plan for processing SQL statement
- CBO utilizes metadata statistics to prune unnecessary micro-partitions, minimizing I/O
- Execution of CBO happens EVERY TIME a query is submitted to service
- Sometimes compilation time > execution time
- Compilation times are affected by:
 - Query complexity
 - Massive tables, since applying filtering (pruning) against an extremely high number of micro-partitions can be expensive
- When troubleshooting performance issues, it is important to monitor query compilation time

Optimization Phases (in sequence):

- Parser
- Semantic Analysis
- Logical Writers
- Micro-partition pruning (1)
- Initial plan generation
- Plan rewrites
- Micro-partition pruning (2)
- Cost-based join ordering
- Physical plan generation



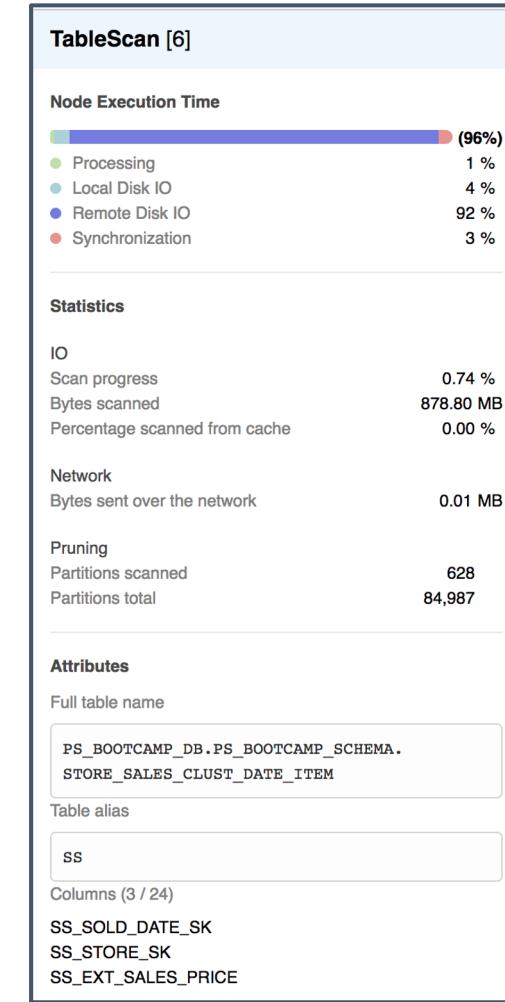
SNOWFLAKE OPTIMIZATION - QUERY PROFILER

Breakdown (%) of Node Execution Time...

- **Initialization** = setup activities prior to processing
- **Processing** = CPU data processing
- **Local Disk I/O** = blocked on (read from/write to) local SSD on node
- **Remote Disk I/O** = blocked on (read from/write to) remote cloud storage
- **Network Communication** = blocked on network data transfer (read from/write to)
- **Synchronization** = various sync activities between processes

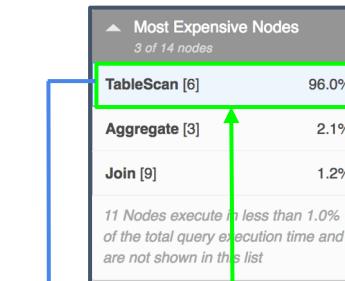
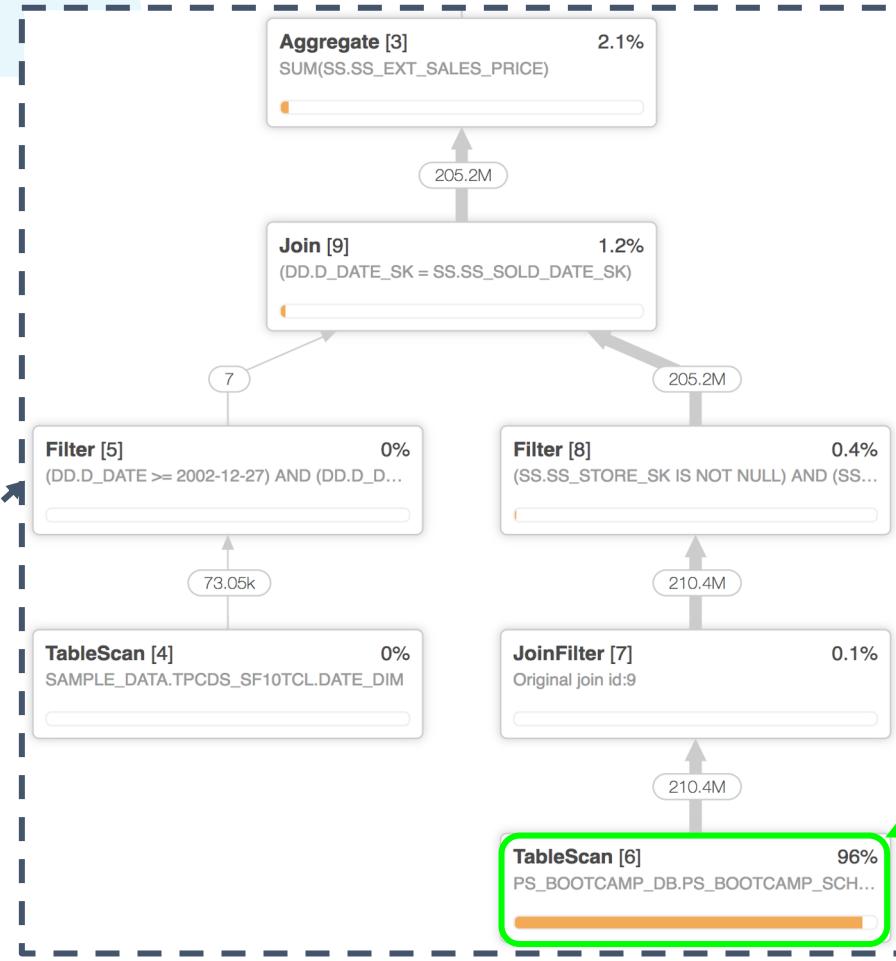
Statistics...

- **Scan progress** = % of data scanned for table thus far
- **Bytes scanned** = local + remote I/O
- **Percentage scanned from cache** = local / (local + remote)
- **Bytes written** = bytes into Snowflake (table)
- **Bytes written to result** = bytes to result object
- **Bytes read from result** = bytes from result object
- **External bytes scanned** = bytes from external object (stage)
- **Bytes sent over the network** = peer-peer data exchange
- **Partitions scanned** = number of micro-partitions read (local + remote)
- **Partitions total** = number of micro-partitions in table
- **Bytes spilled to local storage** = bytes written to local SSD on node (*insufficient memory*)
- **Bytes spilled to remote storage** = bytes written to remote cloud storage (*insufficient local SSD*)



SNOWFLAKE OPTIMIZATION - QUERY PROFILER

Details Profile



Click
to
filter
statistics
on right

TableScan [6]	
Node Execution Time	
Processing	(96%)
Local Disk IO	1 %
Remote Disk IO	4 %
Synchronization	92 %
	3 %
Statistics	
IO	0.74 %
Scan progress	878.80 MB
Bytes scanned	0.00 %
Percentage scanned from cache	
Network	0.01 MB
Bytes sent over the network	
Pruning	628
Partitions scanned	84,987
Partitions total	
Attributes	
Full table name	PS_BOOTCAMP_DB.PS_BOOTCAMP_SCHEMA.
	STORE_SALES_CLUST_DATE_ITEM
Table alias	SS
Columns (3 / 24)	
SS_SOLD_DATE_SK	
SS_STORE_SK	
SS_EXT_SALES_PRICE	



SNOWFLAKE OPTIMIZATION - QUERY PROFILER

Some basics before we dive in... SQL History tab

- Provides a tabular, high-level view of each query
- Includes basic performance metrics (duration, bytes scanned, rows)
- Color-coded to provide quick insights

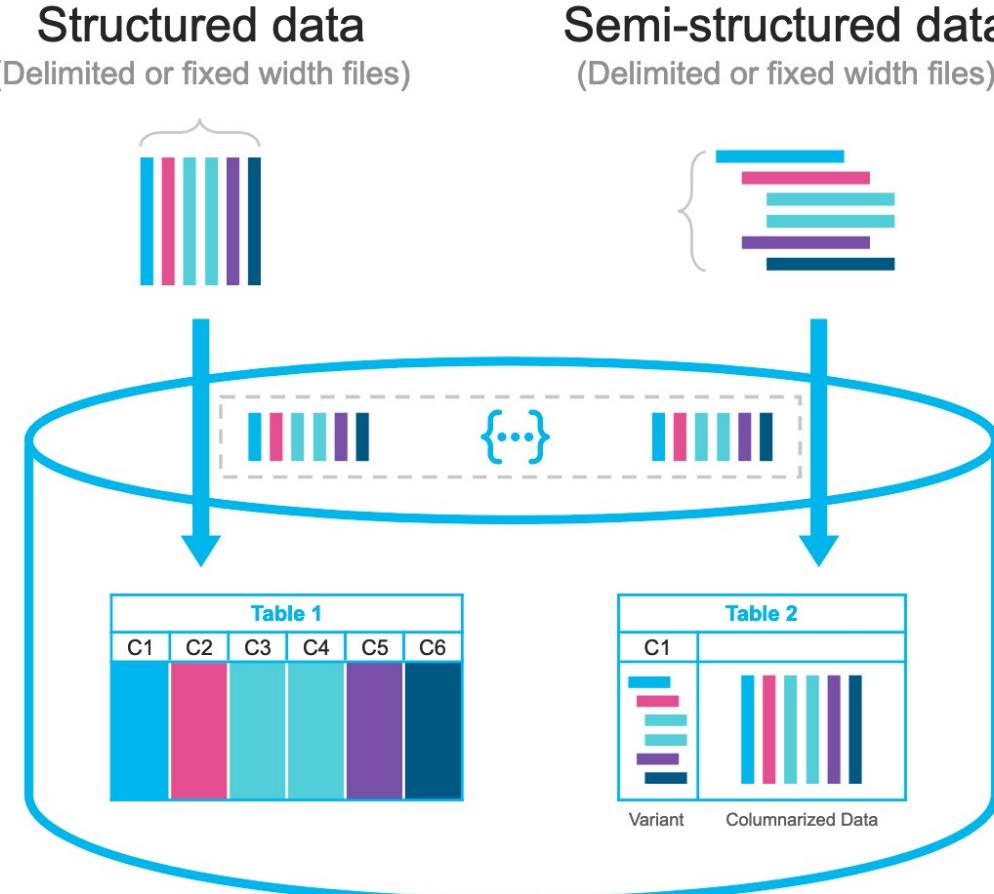
The screenshot shows the Snowflake web interface with the 'History' tab selected. The top navigation bar includes links for Databases, Shares, Warehouses, Worksheets, History (which is highlighted with a green box), and Account. On the right, there are icons for Partner Connect, Help, Notifications, and a user account labeled 'DGARDNER ACCOUNTADMIN'. Below the navigation is a search bar with filters for 'Hide Filters', 'View SQL', and 'Abort...'. The main area displays a table of query history. The table has columns for Status, Query ID, SQL Text, User, Warehouse, Clust..., Size, Session ID, Start Time, End Time, Total Duration, Bytes Scanned, Rows, and Query Tag. The 'Bytes Scanned' and 'Rows' columns use color-coded bars to represent their values. A blue box highlights the last two columns of the table.

Status	Query ID	SQL Text	User	Warehouse	Clust...	Size	Session ID	Start Time	End Time	Total Duration	Bytes Scanned	Rows	Query Tag
✓	8b92ffbf-73...	ALTER WAREHOUSE PS_BOOTCAMP_WH SET WAR...	DGARDNER	PS_BOOTC...			7384337285...	5:40:18 PM	5:40:18 PM	119ms			
✓	1d912b31-...	ALTER WAREHOUSE PS_BOOTCAMP_WH SUSPEND ;	DGARDNER	PS_BOOTC...			7384337285...	5:40:18 PM	5:40:18 PM	108ms			
✓	6c8b59d5-...	ALTER SESSION SET USE_CACHED_RESULT = FALSE;	DGARDNER	PS_BOOTC...			7384337285...	5:40:18 PM	5:40:18 PM	36ms			
✓	0c3ada06-...	WITH X AS (SELECT SS.SS_STORE_SK ,SUM(SS.SS...	DGARDNER	PS_BOOTC...	1	X-Large	7384337285...	5:40:10 PM	5:40:18 PM	7.9s	37.0GB	10	
✓	9259f513-a...	WITH X AS (SELECT SS.SS_STORE_SK ,SUM(SS.SS...	DGARDNER	PS_BOOTC...	1	X-Large	7384337285...	5:39:55 PM	5:40:10 PM	14.5s	37.1GB	10	
✓	8fe8a02e-6...	ALTER SESSION SET USE_CACHED_RESULT = TRUE ;	DGARDNER	PS_BOOTC...			7384337285...	5:39:55 PM	5:39:55 PM	46ms			
✓	ae9eb44b-...	WITH X AS (SELECT SS.SS_STORE_SK ,SUM(SS.SS...	DGARDNER	PS_BOOTC...	1	X-Large	7384337285...	5:39:47 PM	5:39:55 PM	7.9s	37.1GB	10	
✓	4866fb0-8...	WITH X AS (SELECT SS.SS_STORE_SK ,SUM(SS.SS...	DGARDNER	PS_BOOTC...	1	X-Large	7384337285...	5:38:28 PM	5:39:47 PM	1m 19s	46.0GB	10	
✓	dd1182cd-...	ALTER WAREHOUSE PS_BOOTCAMP_WH RESUME ;	DGARDNER	PS_BOOTC...			7384337285...	5:38:28 PM	5:38:28 PM	174ms			
✓	67769372-...	ALTER WAREHOUSE PS_BOOTCAMP_WH SET WAR...	DGARDNER	PS_BOOTC...			7384337285...	5:38:28 PM	5:38:28 PM	121ms			



SEMI-STRUCTURED DATA

Native support for JSON, XML, Avro, Parquet, OR



Structured formats (CSV, TSV, ...)

- Strongly-typed “columns”
- Typically map 1:1 to table columns
- Ingestion process generates important metadata

Semi-structured formats (JSON, XML, ...)

- Traditional DBs require complex transforms to “flatten” data into structures.
- Snowflake has a VARIANT datatype
- Stores original document as-is
- During ingestion, data is columnized and metadata collected
- SQL syntax is a simple dot notation



➤ BUILT IN MONITORING AND MANAGEMENT TOOLS



RESOURCE MONITORING & MANAGEMENT

5 Fundamentals

Management

Monitoring

Auto Suspend
Policies

Admin & Usage
Dashboards

Account_Usage &
Information_Schem
a

Resource Monitors

Warehouse Load
Charts

BI Partner
Dashboards



MANAGING RESOURCES

Virtual Warehouses

AUTO SUSPEND POLICIES

Turning virtual warehouses off quickly after they go idle saves customers money.

- Use cases that take advantage of query caching should be left on longer
- ELT and data loading that don't use query caching should be turned off immediately

Configure Warehouse

Name ADMIN_WH

Size X-Small (1 credit / hour)

Learn more about virtual warehouse sizes [here](#)

Maximum Clusters 1

Multi-cluster warehouses improve the query throughput for high concurrency workloads.

Scaling Policy Standard

The policy used to automatically start up and shut down clusters.

Auto Suspend 3 minutes - custom

The maximum idle time before the warehouse will be automatically suspended.

Auto Resume [?](#)

Comment

Show SQL

Cancel

Finish



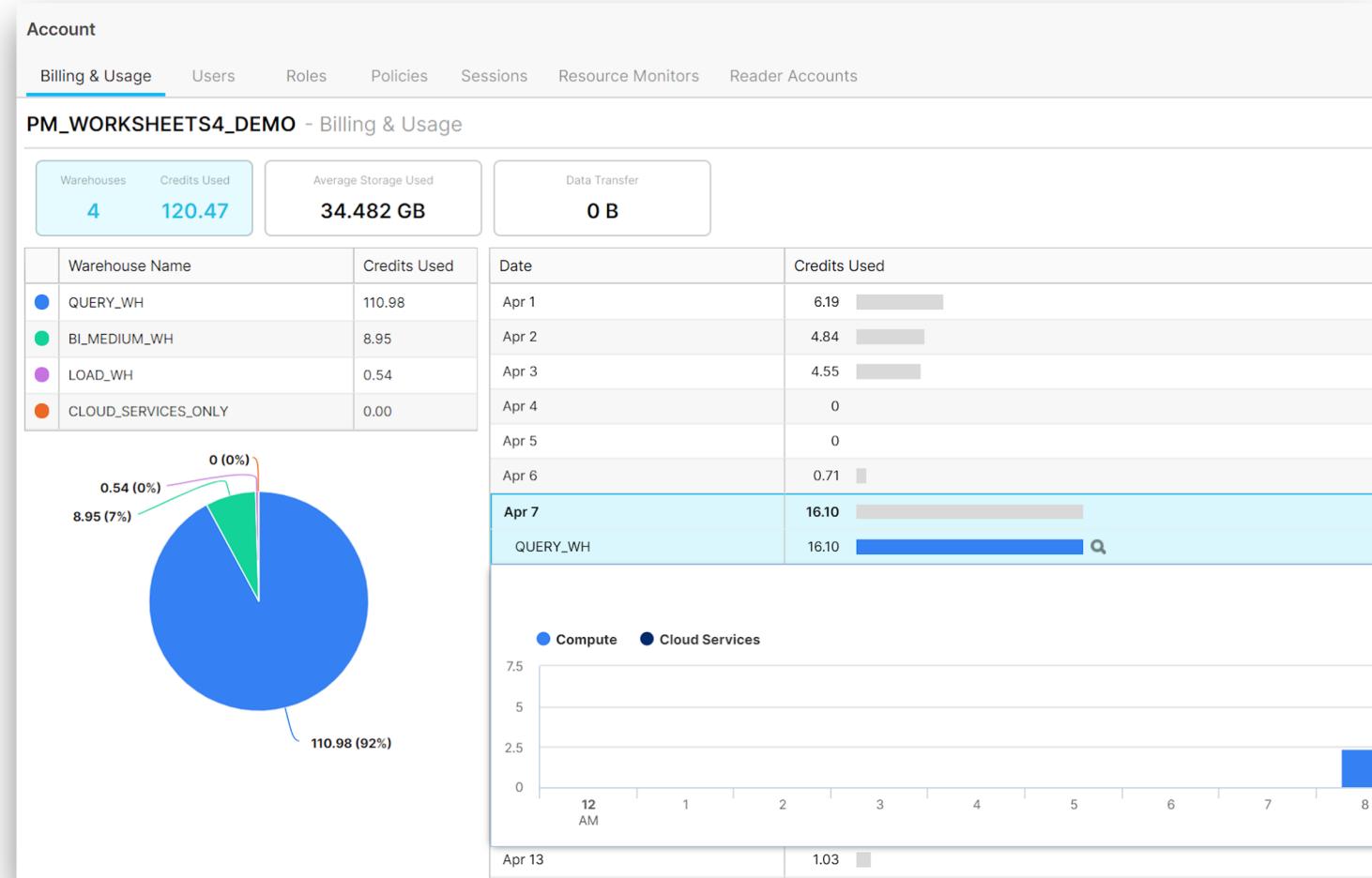
MONITORING RESOURCES

Easy Usage Metrics

ADMIN BILLING AND USAGE

Monitor daily warehouse and cloud services usage and overall storage utilization.

- Analyze each day's hourly spend by warehouse
- View cloud services spend
- Admins can download data to view historical trends and forecast usage



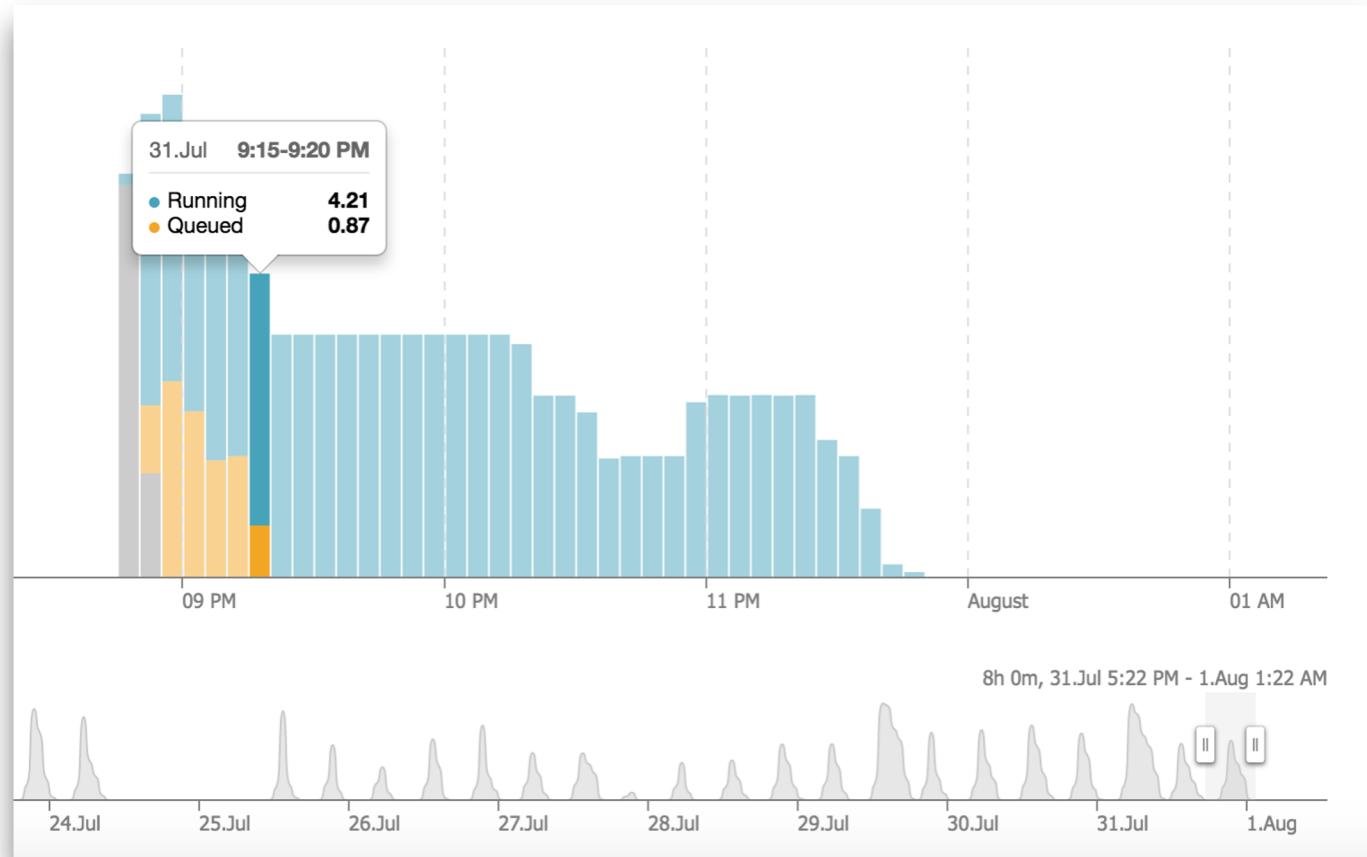
MONITORING RESOURCES

View Load on Your Warehouses

WAREHOUSE LOAD CHARTS

The load monitoring charts show current and historic usage patterns for warehouses, helping determine proper sizing and performance .

- View the performance status of queries on warehouses
- Identify trends in usage to improve query performance
- Adopt multi-cluster warehouse scaling strategies



docs.snowflake.com/en/user-guide/warehouses-load-monitoring.html

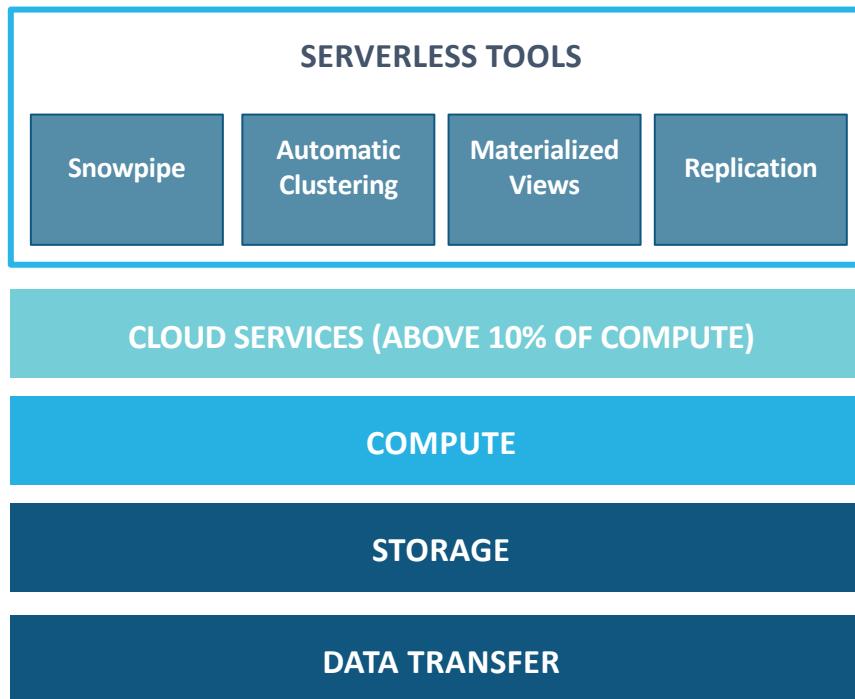


© 2020 Snowflake Inc. All Rights Reserved

CLOUD SERVICES

You are charged for Cloud Services (CS) only if...

CS compute is above 10%+ of your total Snowflake usage, calculated on the daily basis



What are some things that drive high cloud services usage?

Class	Description
COPY	Copy commands with poorly formed buckets (poor selectivity) spend a lot of time listing files from S3. Customers should use appropriate prefixes within the bucket to minimize the number of files listed.
DDL Operations	DDL operations, particularly cloning, are taxing on Cloud Services. When large schemas are created and dropped, this uses a lot of Cloud Services resources. Cloning a database for backup or as an alternative to time travel are expensive actions.
SHOW Commands	SHOW commands use only Cloud Services resources. Applications built on top of Snowflake that use a lot of SHOW commands are going to have heavy Services usage.
INFORMATION_SCHEMA queries	Like SHOW commands. These are resource-intensive for Cloud Services and FDB, but they also have a compute component.
High Frequency Queries	Single queries that aren't bad themselves, but are problematic at high frequencies such as single row insert statements.



SNOWPIPE

File size and number of files are crucial to data load performance and **cost optimization**

- Split large files before loading into Snowflake
- Recommended: **10MB to 100MB (Compressed)**

Semi-structured data being loaded into VARIANT has a 16MB size limit on individual rows

- We recommend enabling the STRIP_OUTER_ARRAY file format option for the COPY INTO <table> command to remove the outer array structure and load the records into separate table rows

Isolate ingestion workloads into own compute cluster

Organize data in logical - granular paths (eg. subject area and create date)

/system/market/daily/2018/09/05/

Name file prefix and then wildcard to reduce scanning cloud storage:

```
copy into t1  
from /system/market/daily/2018/*
```

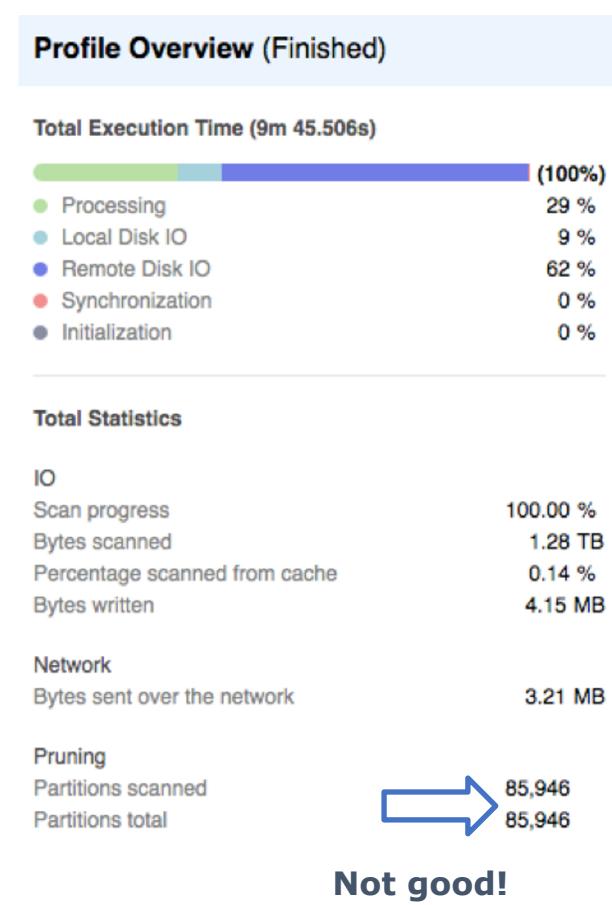
Type	Cost
Compute	0.125 credits per core per hour
Notifications	0.06 credits per 1,000



AUTOMATIC CLUSTERING

When to leverage auto-clustering?

- Table is large (*1TB+*)
 - As the number of partitions are added to a table, clustering becomes more important!
- Natural ingestion order (e.g. *time*) is not sufficient for varying important query workloads
 - Columns used in FILTER, JOIN, or GROUP BY clauses show poor partition pruning (*as shown in query profile*)
 - Clustering quality is poor for query access columns used in FILTER, or JOIN, or GROUP BY
- For best cost performance, **choose the highest cardinality keys and tables that are not very wide (the less columns, the better!)**



MATERIALIZED VIEWS

- Cost for materialized views depends on how many views are created and the view's definition (*number of rows and columns*)
- You can *control* costs by suspending or resuming the materialized view; however, suspending maintenance typically only defers costs, rather than reducing them. The longer that maintenance has been deferred, the more maintenance there is to do.
- Avoid clustering the base table if the data accessed is primarily through the materialized view

Viewing costs

```
select *
  from table(information_schema.materialized_view_refresh_history(
    date_range_start=>'2019-07-19 19:00:00.000',
    date_range_end=>'2019-07-20 20:00:00.000'));
```

Here is sample output:

START_TIME	END_TIME	CREDITS_USED	MATERIALIZED_VIEW_NAME
2019-07-19 19:00:00.000 -0700	2019-07-20 20:00:00.000 -0700	0.223276651	TEST_DB.TEST_SCHEMA.MATERIALIZED_VIEW_1



DATA REPLICATION

How to Estimate Data Replication Cost

- 1) **Data transfer** (egress) is charged based on the rate of the source region. (*e.g. \$20/TB between AWS US-East and AWS US-West*)
- 1) **Replication compute** used to copy data across regions. **The estimated compute cost is ~10-13 credits per TB.**
- 1) **Storage** for the secondary databases shows up as the storage line item (*\$23/compressed TB*)





➤ **Snowflake Editions,
pricing and regions**

What's Sizing and Pricing?

Sizing and pricing is the act of accurately gauging the required Snowflake environment, based on the prospect or customer's needs, and calculating the associated costs which will be delivered in the form of a quote.



Understand Snowflake Pricing



Storage

All customers are charged a monthly fee for the data they store in Snowflake. **Storage cost is measured using the average amount of storage used per month**, after compression, for all customer data stored in Snowflake.



Virtual Warehouses (Compute)

A virtual warehouse is one or more compute clusters that enable customers to load data and perform queries. Customers pay for virtual warehouses using Snowflake credits.



Cloud Services

The cloud services layer provides all permanent state management and overall coordination of Snowflake. Customers pay for cloud services using Snowflake credits.

- **Snowpipe**
- **Database Replication**
- **Materialized Views Maintenance**
- **Automatic Clustering**
- **Search Optimization Service**

There are two ways to buy the Snowflake Service: On Demand or pre-paid Capacity.

On Demand : The easiest and most flexible way to purchase the Snowflake Service is On Demand. Similar to Amazon AWS, customers are charged a fixed rate for the services that are consumed and are billed in arrears every month.

Pre-Purchased Capacity: Snowflake provides customers the option to pre-purchase Capacity. A capacity purchase is a specific dollar commitment to Snowflake. The Capacity purchased is then consumed on a monthly basis. Capacity purchases provide more service options, lower prices, and a long-term price guarantee.



SNOWFLAKE EDITIONS

Standard

- Complete SQL Data Warehouse
- Secure Data Sharing across regions / clouds
- Premier Support 24 x 365
- 1 day of time travel
- Always-on enterprise grade encryption in transit and at rest
- Customer dedicated virtual warehouses
- Federated authentication
- Database Replication
- External Functions
- Snowsight
- Create your own Data Exchange
- Data Marketplace Access

Enterprise

Standard +

- Multi-Cluster warehouse
- Up to 90 days of time travel
- Annual rekey of all encrypted data
- Materialized Views
- Search Optimization Service
- Dynamic Data Masking
- External Data Tokenization

Business Critical

Enterprise +

- HIPAA support
- PCI compliance
- Data encryption everywhere
- Tri-Secret Secure using customer-managed keys (AWS)
- AWS PrivateLink support
- Azure PrivateLink support
- Database Failover and Failback for business continuity
- External Functions - AWS API Gateway Private Endpoints support

Virtual Private Snowflake (VPS)

Business Critical +

- Customer dedicated virtual servers wherever the encryption key is in memory
- Customer dedicated metadata store



Pricing Example

Customer has purchased the Snowflake Capacity Standard Service with Premier Support in the United States. They load data every night and have 8 users using the database during business hours. This customer:

- Stores an average of 4 TB's of data.
- Runs a batch loading task every night for 2.5 hours, using a Small Standard Virtual Warehouse.
- Enables eight users to work 5 days a week from 8am until 6pm using a Medium Standard Virtual Warehouse.

Data Loading Requirements

Parameter	Customer Requirements	Proposed Setup
Data Load Frequency	Daily	Small Standard Virtual Warehouse (2 Credits per hour)
Loading Window	2.5 hours	(2 Credits x 2.5 hours per day x 31 days per month)
Credits Required per month	155	

Storage Requirements

Data Set size (per month)	4TB (After Compression)
---------------------------	-------------------------

User Requirements

Parameter	Customer Requirements	Proposed Setup
No. of Users	8 Users	Medium Virtual Warehouse (4 Credits per hour)
Time Slot	8AM-6PM (10 hours)	
Credits Required per month	800	(4 Credits x 10 hours per day x 20 days per month)
Credits Required per month	955	(800 Credits for Users + 155 for Data Loading)

Total Pricing

Data storage price (per year)	\$1,104	4TB x \$23 (per TB per month, based on Storage table) x 12 months
Virtual Warehouse (per year)	\$21,774	955 (Credits per month) x \$2 (per credit) x 0.95 (5% sample discount) x 12 months
Total Price (per year)	\$22,878	Data Set (per year) + Virtual Warehouse (per year)



