



SNOWFLAKE PLATFORM

Technical Deep Dive

May 2024

WHAT DO YOU ALREADY KNOW?



THERE IS NO ENTERPRISE STRATEGY WITHOUT A DATA STRATEGY



Simplify Your Data Foundation

For all data and
workloads



Distribute Analytics

For all users against
enterprise data



Scale with Applications

With streamlined development,
deployment, and productization



Simplify Your Data Foundation

For all data and workloads



Distribute Analytics

For all users against enterprise data



Scale with Applications

With streamlined development, deployment, and productization

REQUIREMENTS FOR YOUR DATA FOUNDATION

Endless silos



Unified data

Hidden costs



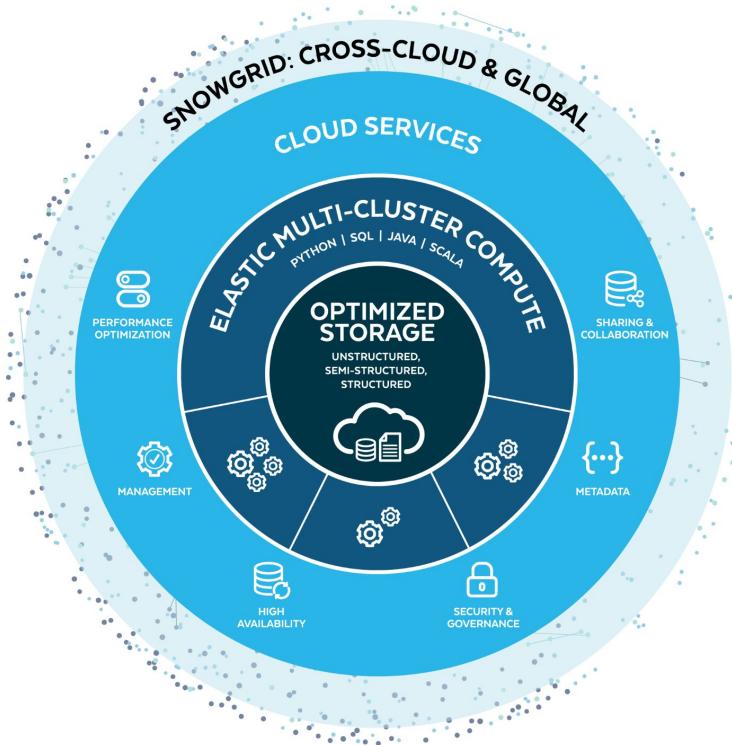
Optimal TCO

Piecemeal policies



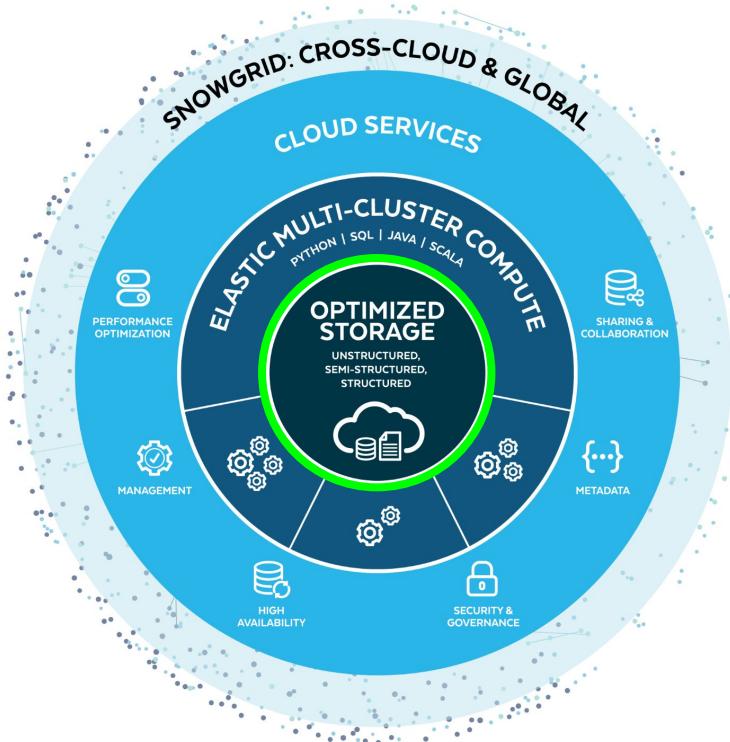
Universal governance

SNOWFLAKE PLATFORM ARCHITECTURE

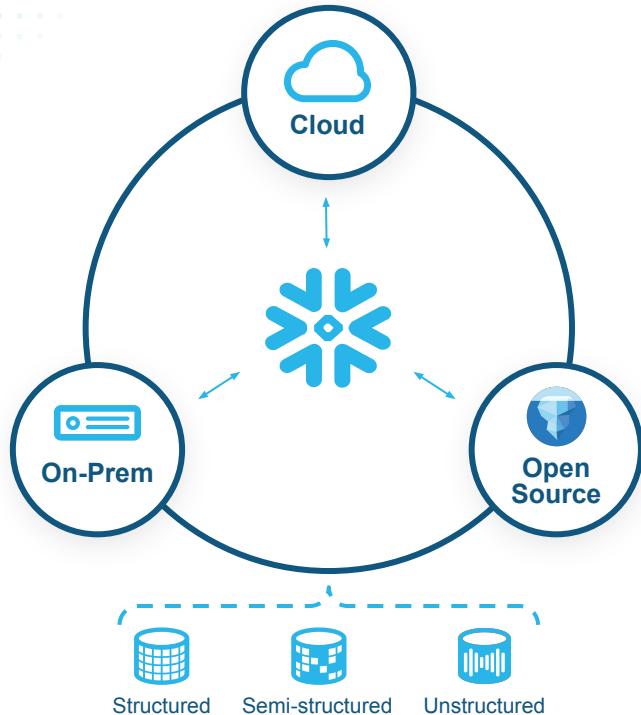


SNOWFLAKE PLATFORM ARCHITECTURE

STORAGE



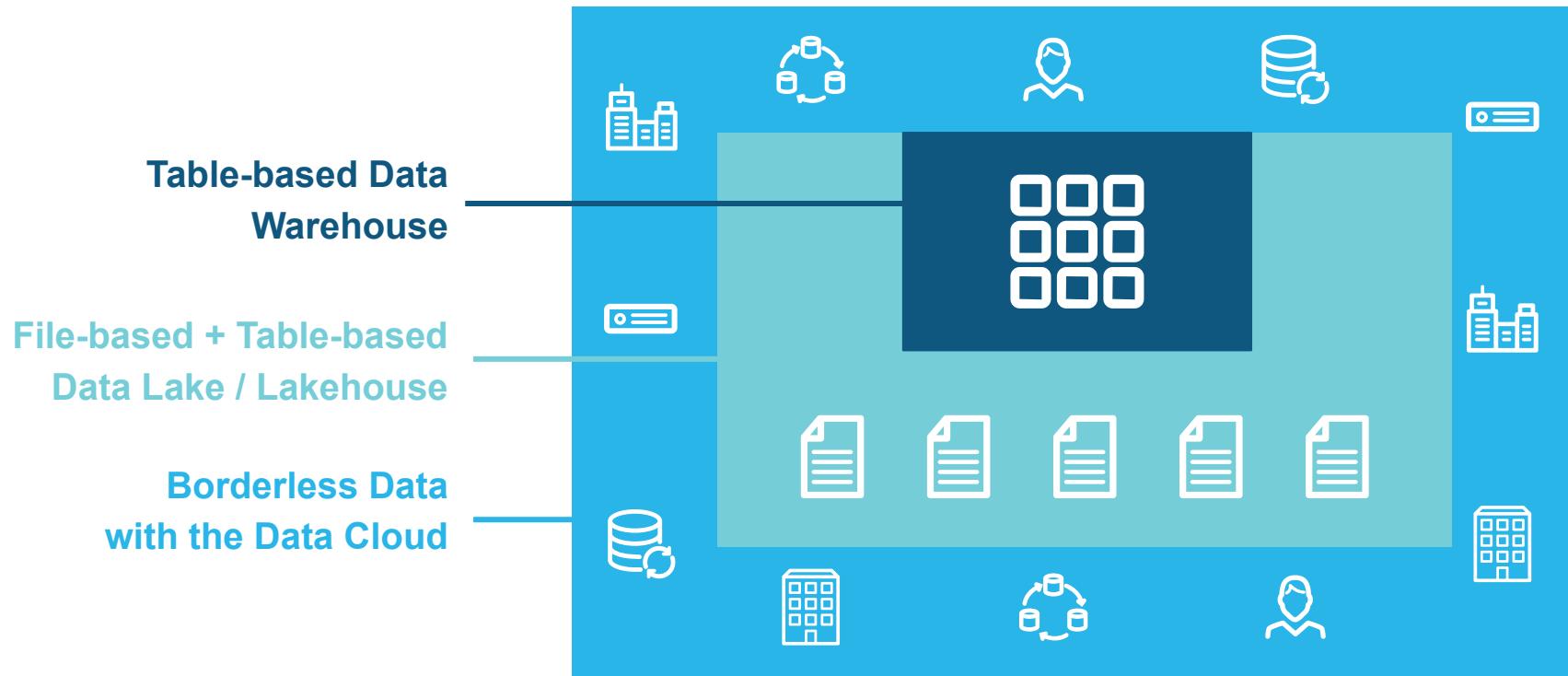
OPTIMIZED STORAGE



- **Un-siloed access to your data**
- **Easily manage data at scale**
- **Flexibility and interoperability**



Evolution of Data Management



Iceberg Tables

Choice to bring Apache Iceberg to the Snowflake platform

ICEBERG TABLES

One object, multiple catalog options



Snowflake-managed

Iceberg Table with
Snowflake-managed catalog.

SDK for Apache Spark™ access



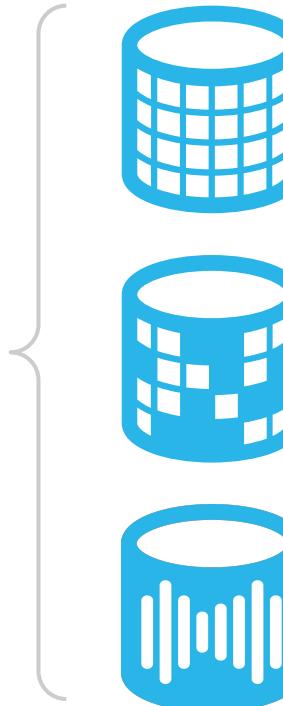
Externally managed

Iceberg Table with **externally
managed** catalog.

AWS Glue, Object Store

Supporting All Data Types

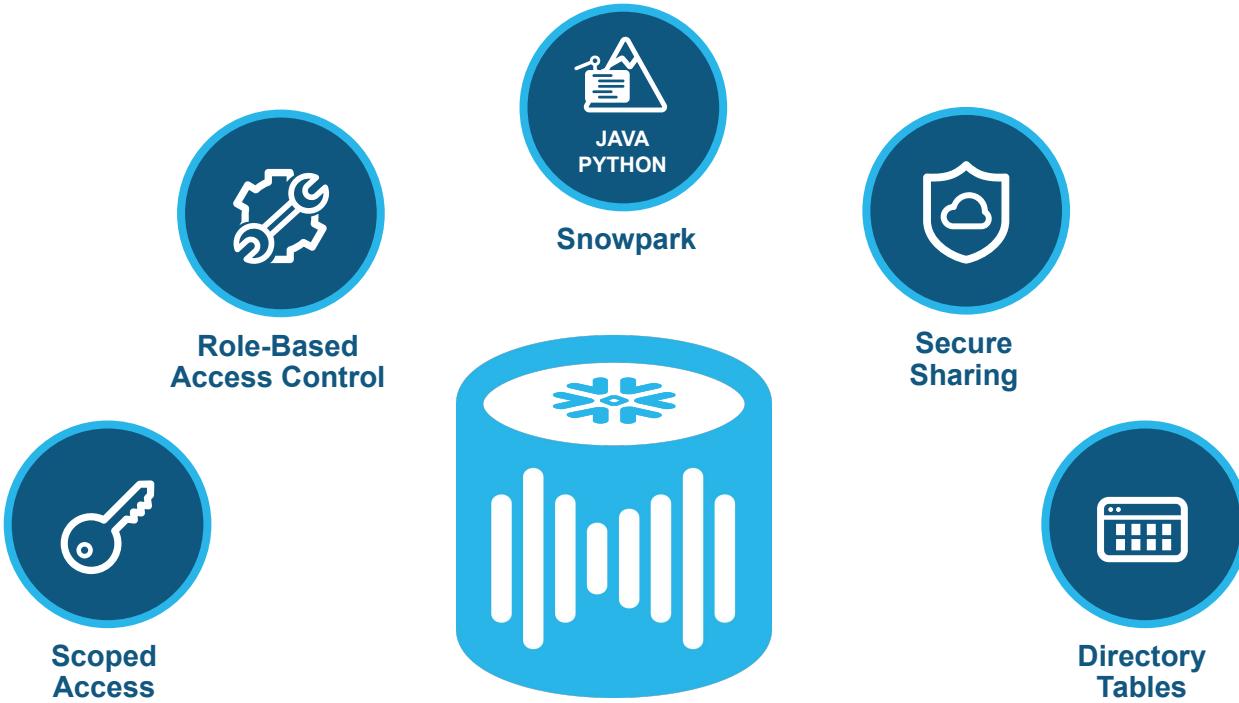
One place to store, govern, process, analyze, and share all data



Structured data

Semi-structured data

Unstructured data



Unstructured data



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

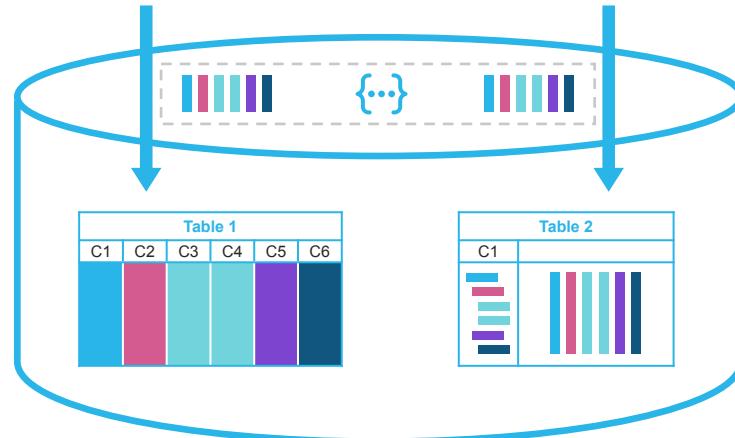
Structured data

(Delimited or fixed width files)



Semi-structured data

(Delimited or fixed width files)

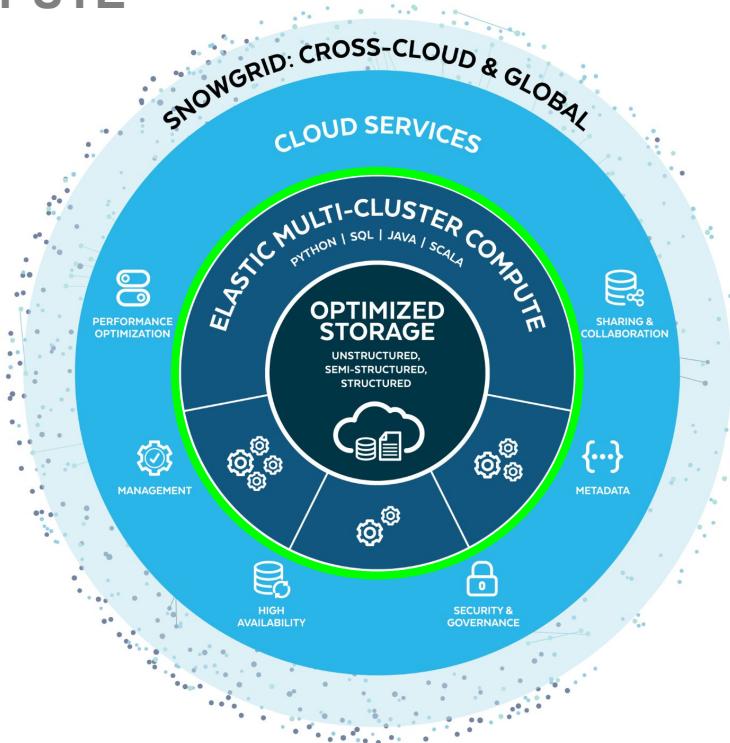


Semi-Structured data

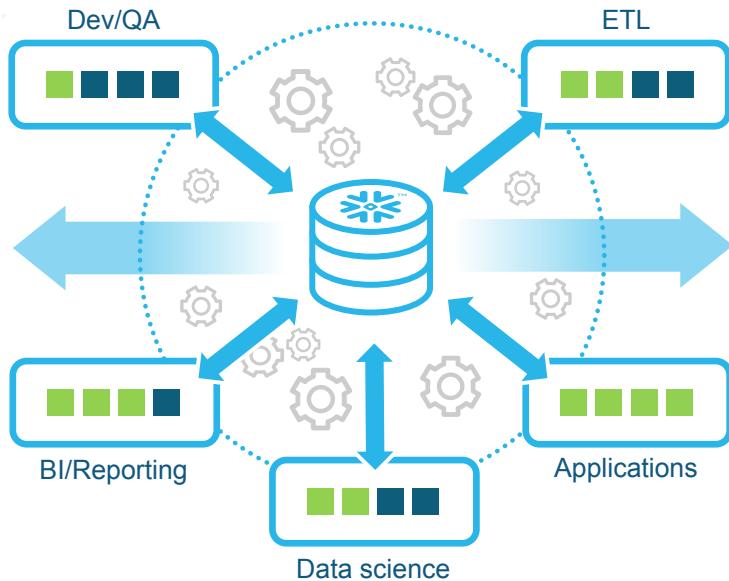


SNOWFLAKE PLATFORM ARCHITECTURE

ELASTIC COMPUTE



ELASTIC MULTI-CLUSTER COMPUTE



- Power AI, apps, and more via a single engine
- Get leading performance and concurrency
- Work in SQL, Python, or Java, and run preferred libraries in Snowpark
- Bring development to the data



THE POWER OF SNOWFLAKE'S ELASTIC SCALABILITY

4.2B

Average daily queries
from January 1, 2024
to
January 31, 2024

82T+*

The number of rows
in the largest
customer table

205K+*

Largest number of
queries being executed
in a one minute interval
by a single customer

186PB*

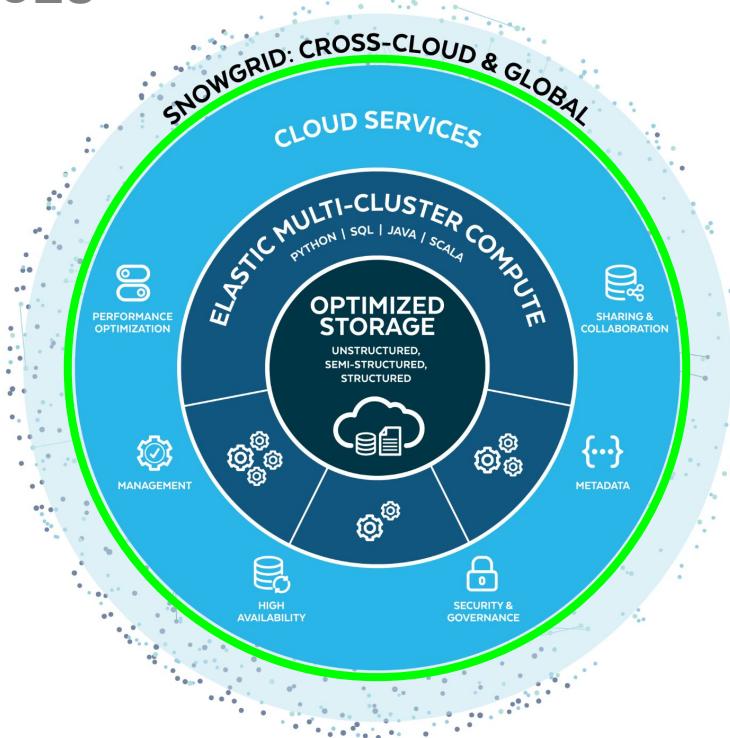
The aggregate amount
of compressed data
stored in Snowflake for
the five largest
customers by data
volume

*Figures as of [January 31 2024](#)

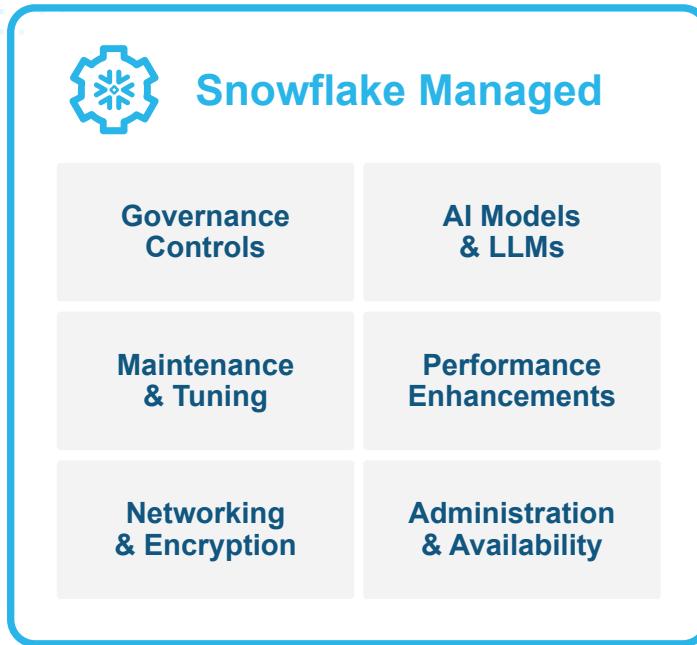


SNOWFLAKE PLATFORM ARCHITECTURE

CLOUD SERVICES



CLOUD SERVICES



- Automate costly and complex operations
- Get unified, built-in governance
- Access built-in AI models, LLM, and vector search functionality



MAXIMIZING COST EFFICIENCY AT SCALE

The screenshot shows the Snowflake Cost Management interface for the account PM_AWS_US_WEST_2. Key metrics displayed include:

- Spending: \$4,793.95 (Spend in USD), 3,053 credits used, \$2.15 price per credit, \$319.59 average daily cost, 203 average daily credits.
- Cost insights: Rarely used auto-clustered tables (85.18 credits), rarely used materialized views (48.03 credits), rarely used search optimization indexes (35.64 credits), rarely used large tables (28.95 TB stored).
- Most expensive queries: A list of SQL queries with their execution times and costs.
- Top warehouses by cost: SERVICENOW_WH (2,456 credits), AX_XS (1,858 credits), APP_WH (1,452 credits).
- Top tags by spend: COST_CENTER (2,456 credits), PRIVACY (1,858 credits), GLOBAL_PII (1,452 credits).

- **Avoid over-provisioning with right-sized resources**
- **Benefit from ongoing, transparent performance enhancements**
- **Control & manage spend via the built-in cost management interface**



SNOWFLAKE HORIZON

Built-in governance solution with a unified set of capabilities

Protect & audit your data

with compliance, business continuity, data quality monitoring, & lineage.



Secure your environment

with continuous risk monitoring & protections, RBAC, & granular authorization policies.

Classify, share, discover & take immediate action

on data, apps and more across regions & clouds.

Unlock the value of your sensitive data

with advanced privacy policies and data clean rooms.

Integrate

with other Apache Iceberg-compatible catalogs & engines, & with data catalog & data governance partners.



SNOWGRID



**Cross-cloud collaboration,
with no ETL**

**Cross-cloud governance to
minimize risk**

**Cross-cloud business
continuity to eliminate
disruptions**

The map is intended for illustrative purposes and does not show all available regions, for a full list of regions see [documentation](#).



© 2024 Snowflake Inc. All Rights Reserved

THE POWER OF SNOWFLAKE'S ELASTIC SCALABILITY

4.2B

Average daily queries
from January 1, 2024
to
January 31, 2024

82T+*

The number of rows
in the largest
customer table

205K+*

Largest number of
queries being executed
in a one minute interval
by a single customer

186PB*

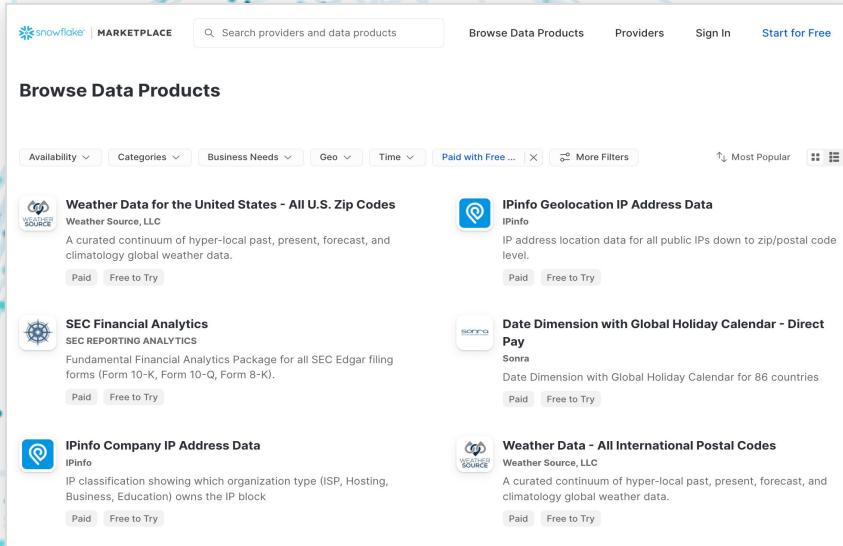
The aggregate amount
of compressed data
stored in Snowflake for
the five largest
customers by data
volume

*Figures as of [January 31 2024](#)



SNOWFLAKE MARKETPLACE

Access data and apps



The screenshot shows the Snowflake Marketplace homepage. At the top, there's a search bar labeled "Search providers and data products". Below it, a navigation bar includes "Browse Data Products", "Providers", "Sign In", and "Start for Free". A large, semi-transparent network graph is overlaid on the page. The main section is titled "Browse Data Products" and features a grid of data product cards. Each card includes a provider logo, the product name, a brief description, and two buttons: "Paid" and "Free to Try".

- Weather Data for the United States - All U.S. Zip Codes**
Weather Source, LLC
A curated continuum of hyper-local past, present, forecast, and climatology global weather data.
[Paid](#) [Free to Try](#)
- IPInfo Geolocation IP Address Data**
IPInfo
IP address location data for all public IPs down to zip/postal code level.
[Paid](#) [Free to Try](#)
- SEC Financial Analytics**
SEC REPORTING ANALYTICS
Fundamental Financial Analytics Package for all SEC Edgar filing forms (Form 10-K, Form 10-Q, Form 8-K).
[Paid](#) [Free to Try](#)
- Date Dimension with Global Holiday Calendar - Direct Pay**
Sonra
Date Dimension with Global Holiday Calendar for 86 countries
[Paid](#) [Free to Try](#)
- IPInfo Company IP Address Data**
IPInfo
IP classification showing which organization type (ISP, Hosting, Business, Education) owns the IP block.
[Paid](#) [Free to Try](#)
- Weather Data - All International Postal Codes**
Weather Source, LLC
A curated continuum of hyper-local past, present, forecast, and climatology global weather data.
[Paid](#) [Free to Try](#)

Curate and discover data and apps

Accelerate procurement and access

Reduce integration costs

>2400

live, ready-to-use data,
services, and Snowflake
Native Apps.*

>540

providers on the
Snowflake Marketplace,
across 18+ categories*

*as of January 31, 2024





Simplify Your Data Foundation

For all data and workloads



Accelerate AI

For all users against enterprise data



Scale with Applications

With streamlined development, deployment, and productization

REQUIREMENTS TO SCALE WITH APPLICATIONS

Architectural complexity



Build many apps with ease

Unpredictable operational burden



Efficient scaling, given apps are the future consumption layer

Limited distribution



Secure, global deployment to run apps where data already is



Simplify Your Data Foundation

For all data and workloads



Accelerate AI

For all users against enterprise data



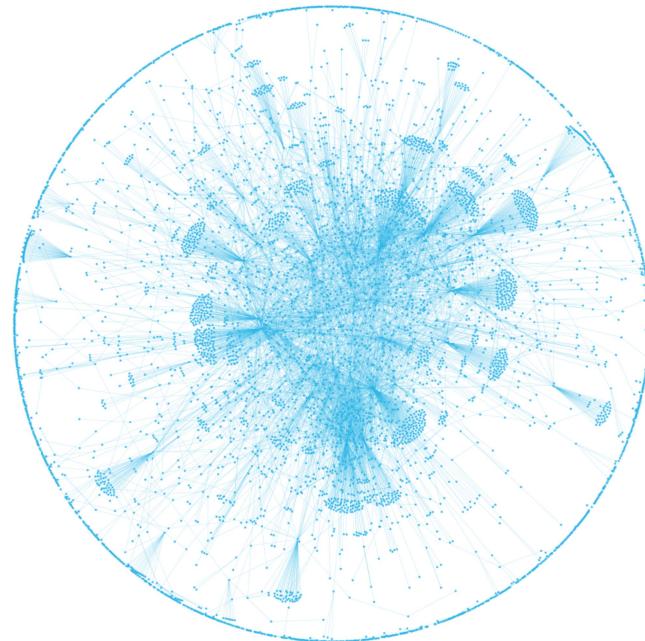
Scale with Applications

With streamlined development, deployment, and productization

DATA CLOUD GROWTH



April 2020



January 2024



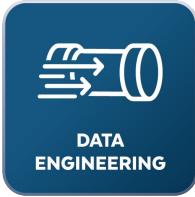
Core Platform Concepts

Powering Many Workloads



COLLABORATION

Discover, access and monetize live data, services and apps in the Data Cloud



DATA ENGINEERING

Build simple, reliable data pipelines at scale in the language of your choice



CYBERSECURITY

Protect your enterprise with near-unlimited visibility, unified data, and powerful analytics



AI/ML

Accelerate your ML workflow with fast access and elastically scalable processing



APPLICATIONS

Build data-intensive applications without operational burden



DATA WAREHOUSE

Accelerate analytics for users and queries with leading price / performance and no complexity



DATA LAKE

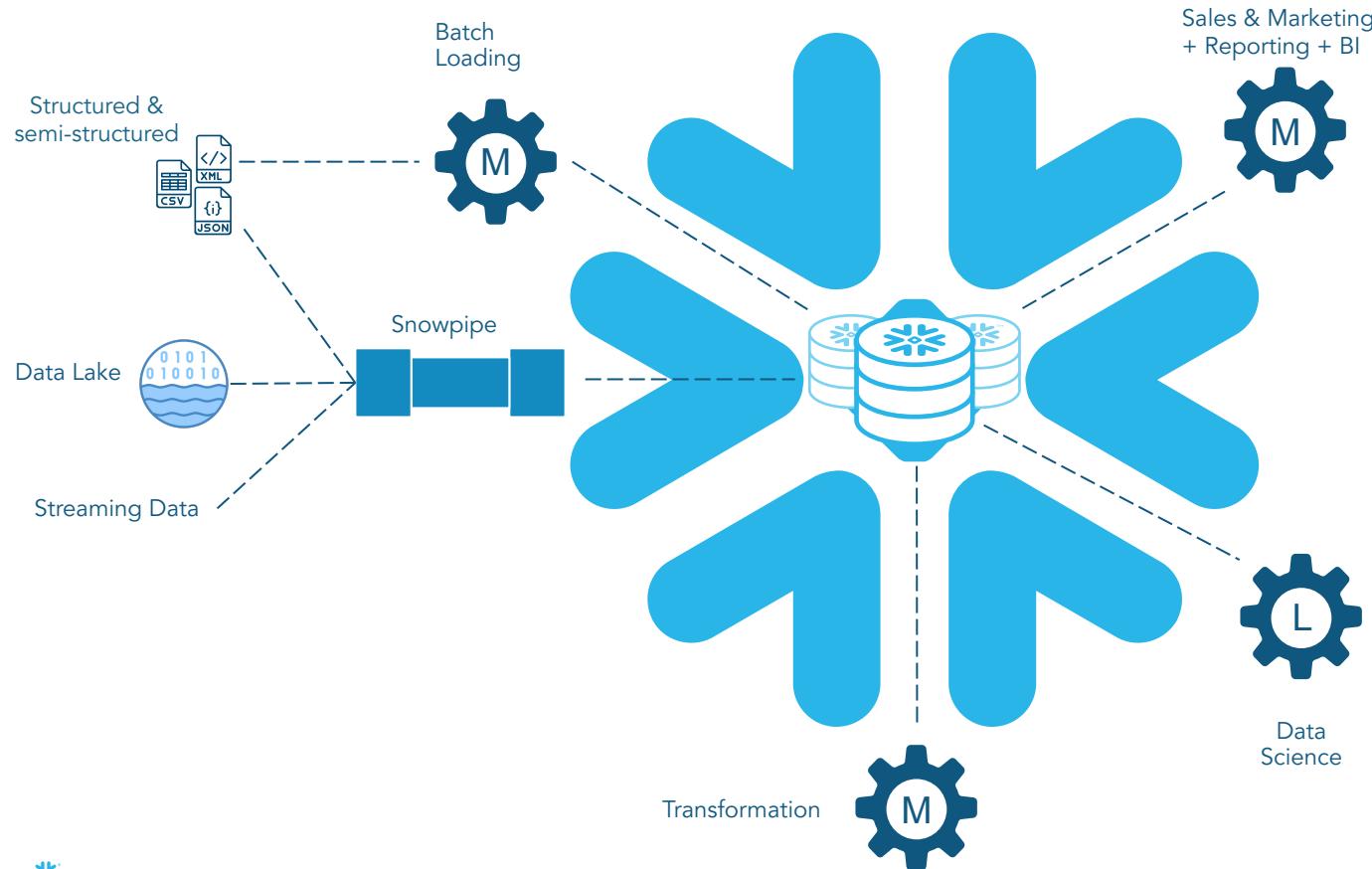
Deploy flexible architectural patterns with governance and optimized storage at scale

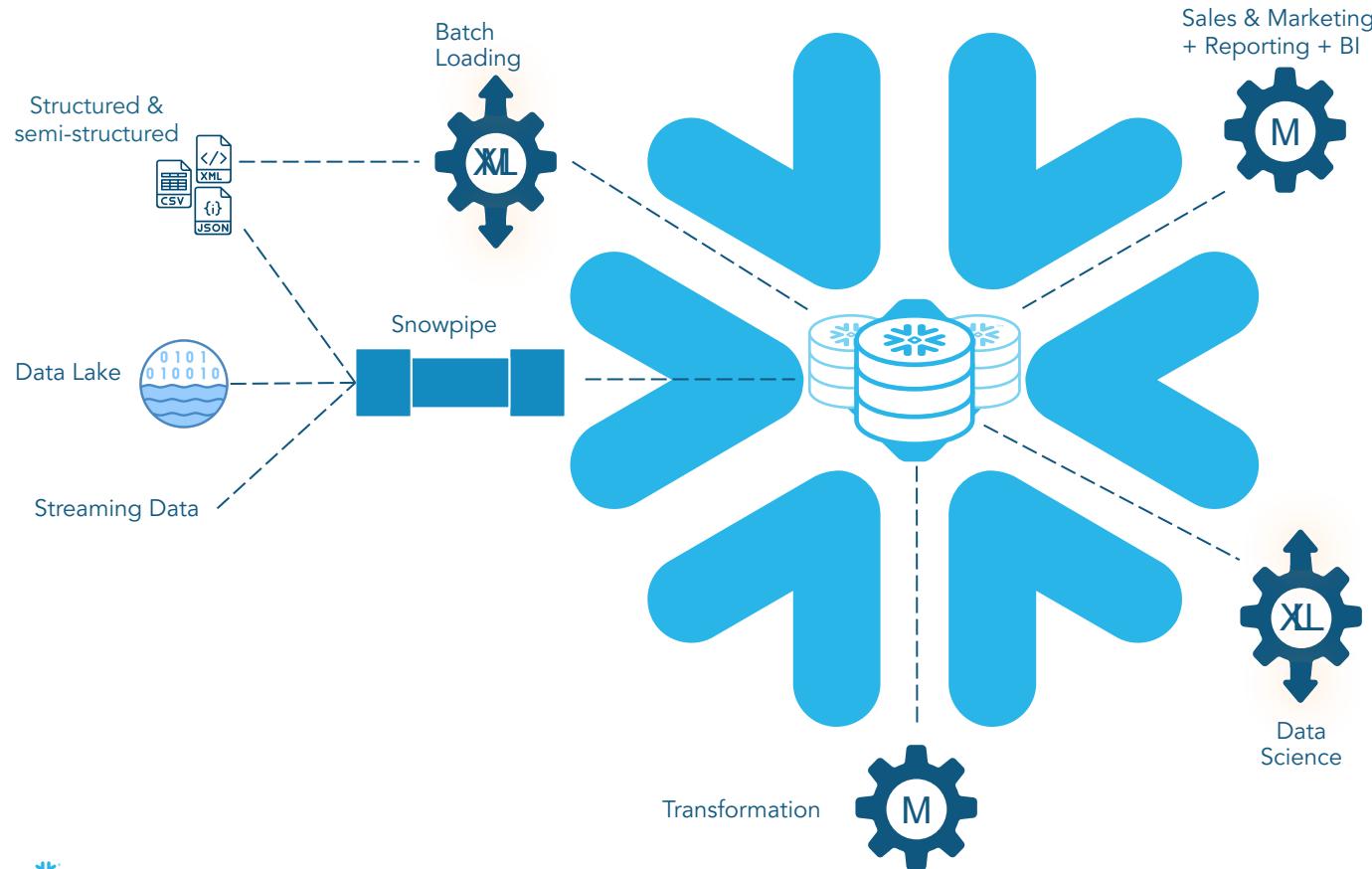


UNISTORE

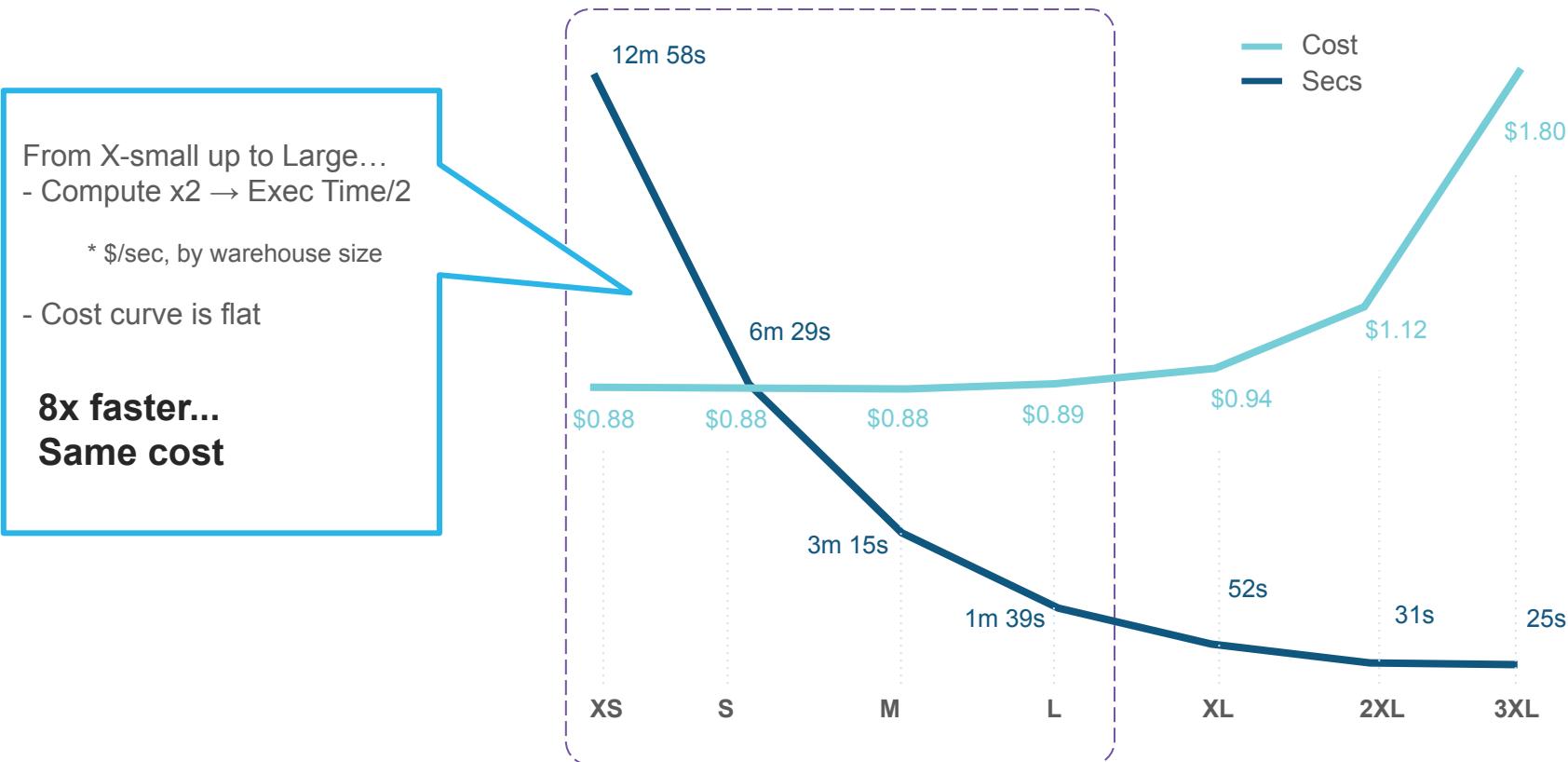
Delivers a modern approach to working with transactional and analytical data together

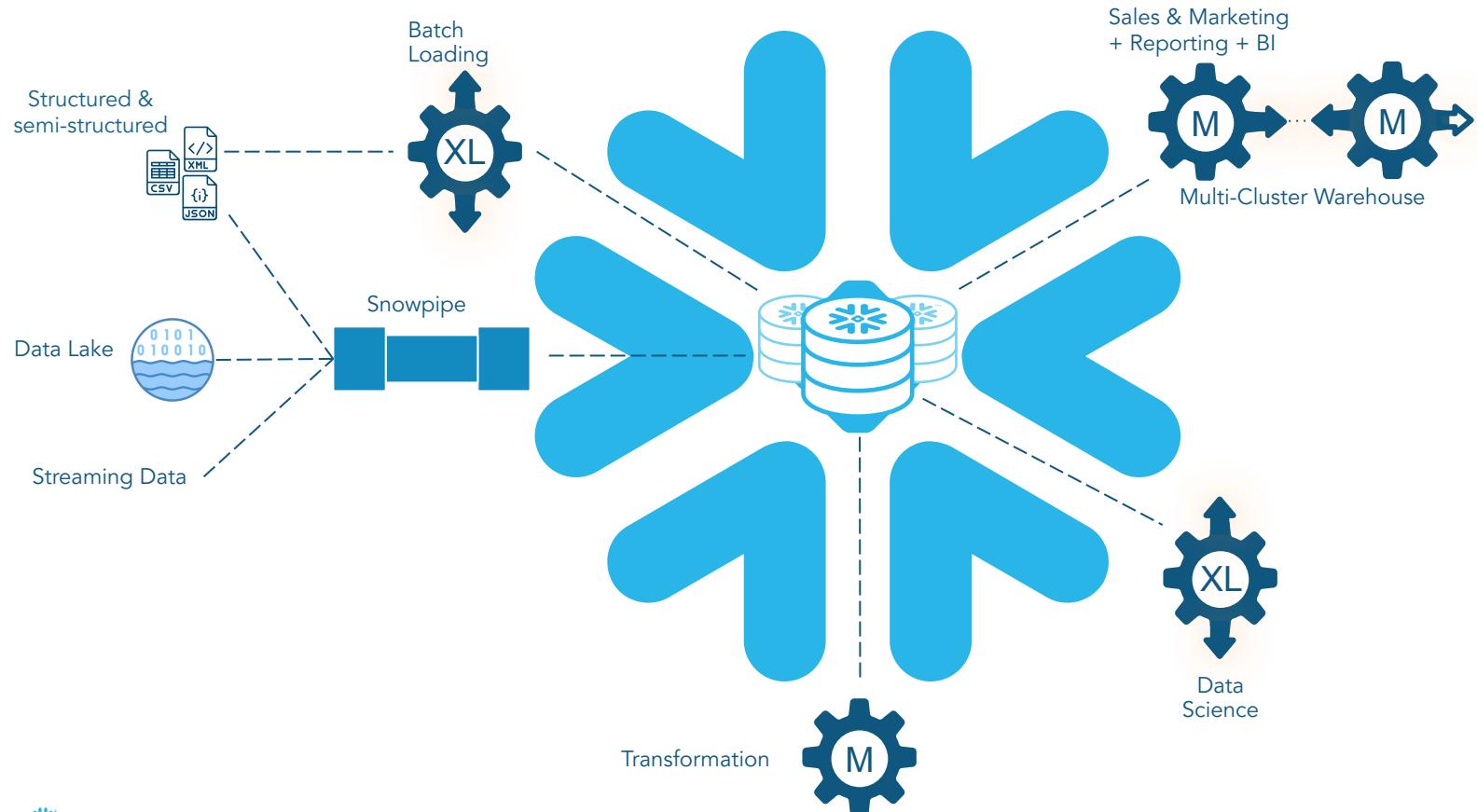




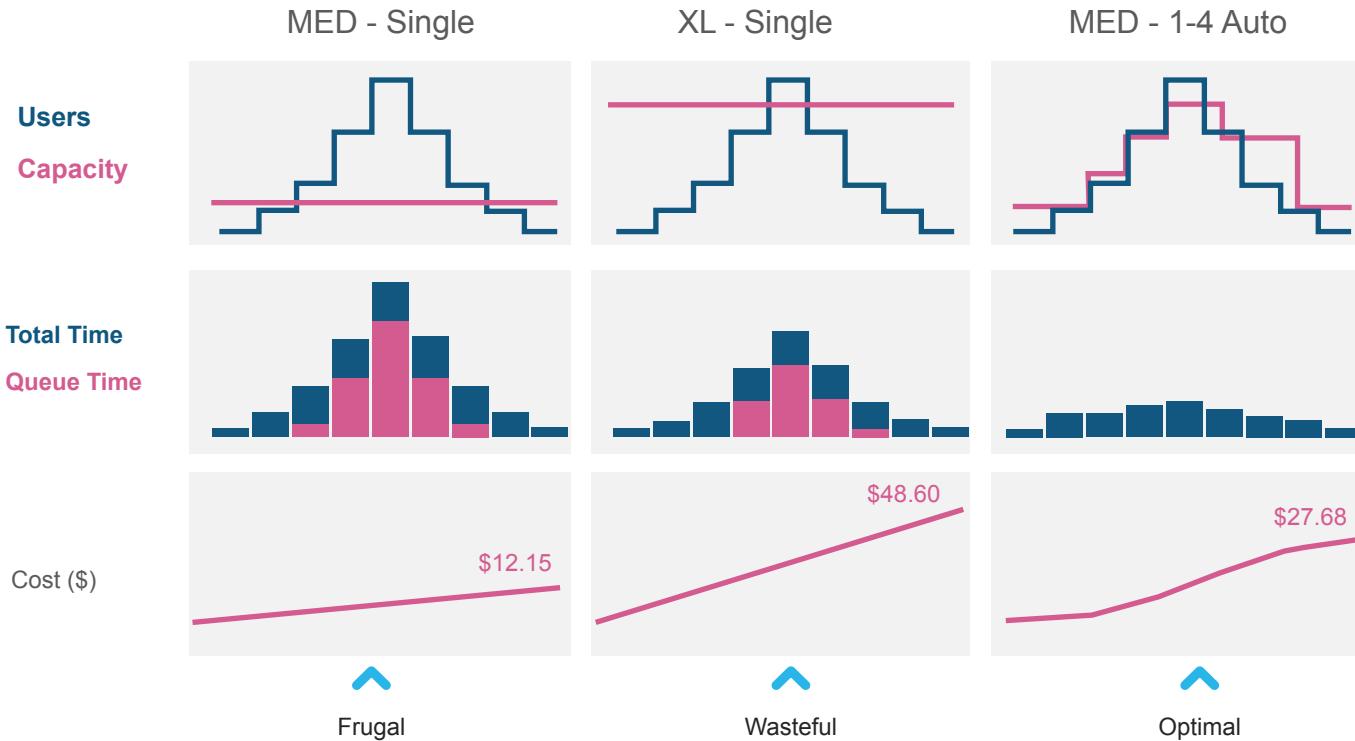


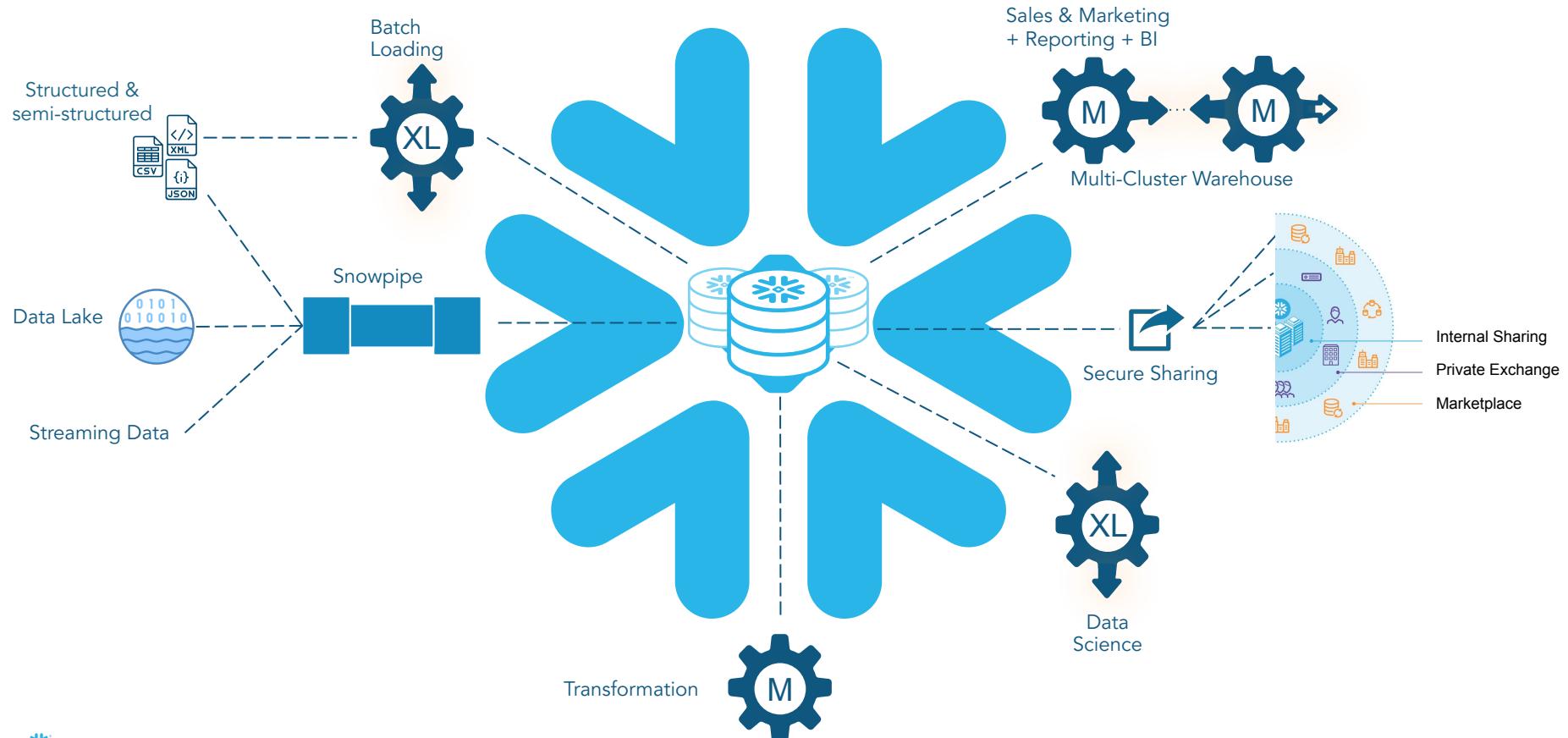
SCALE UP - LOADING 1BN RECORDS



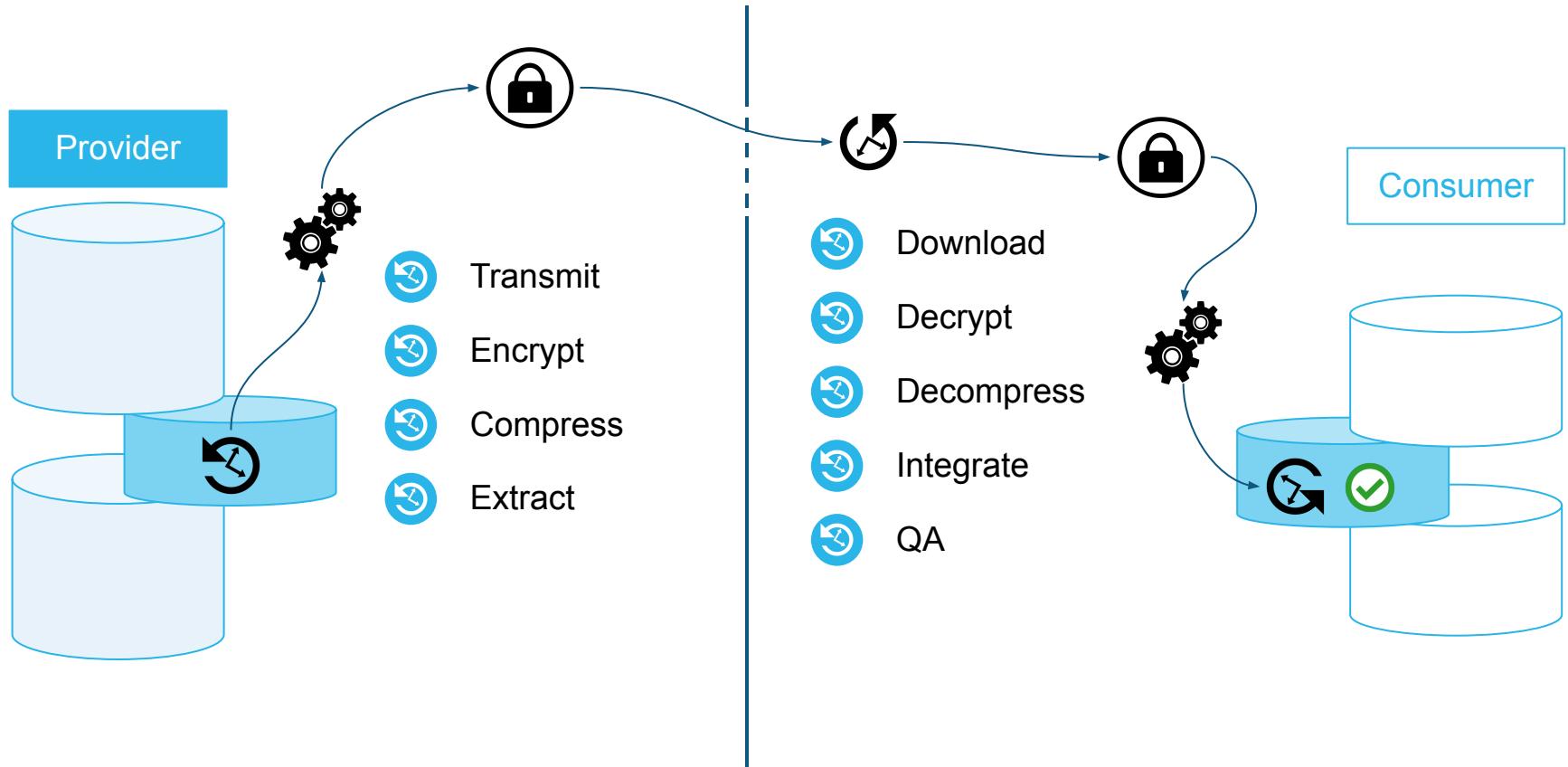


SCALE OUT - MULTI-CLUSTER WAREHOUSES





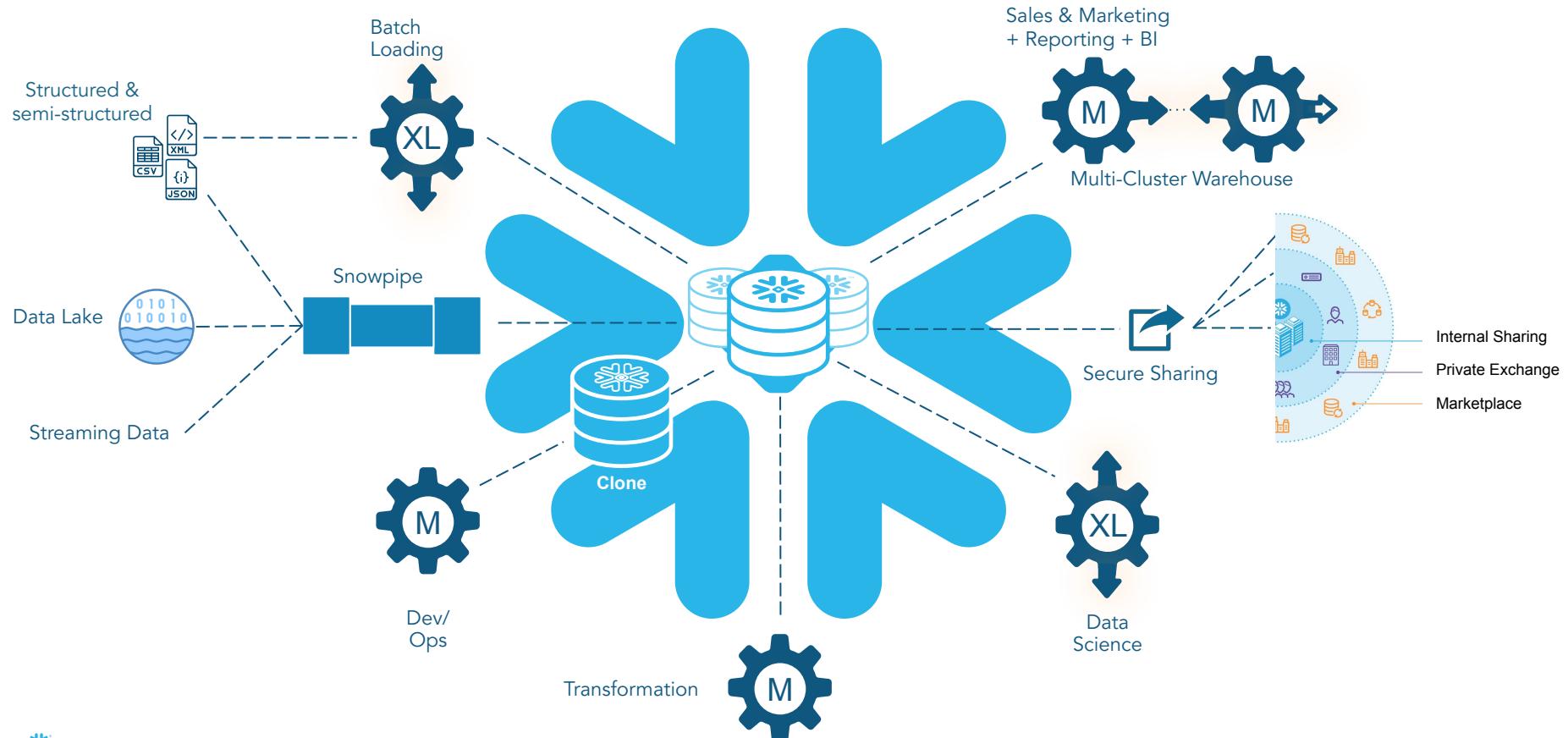
EXCHANGING DATA USED TO BE DIFFICULT



OPERATION

Sharing on Snowflake





Adaptive Caching

Metadata

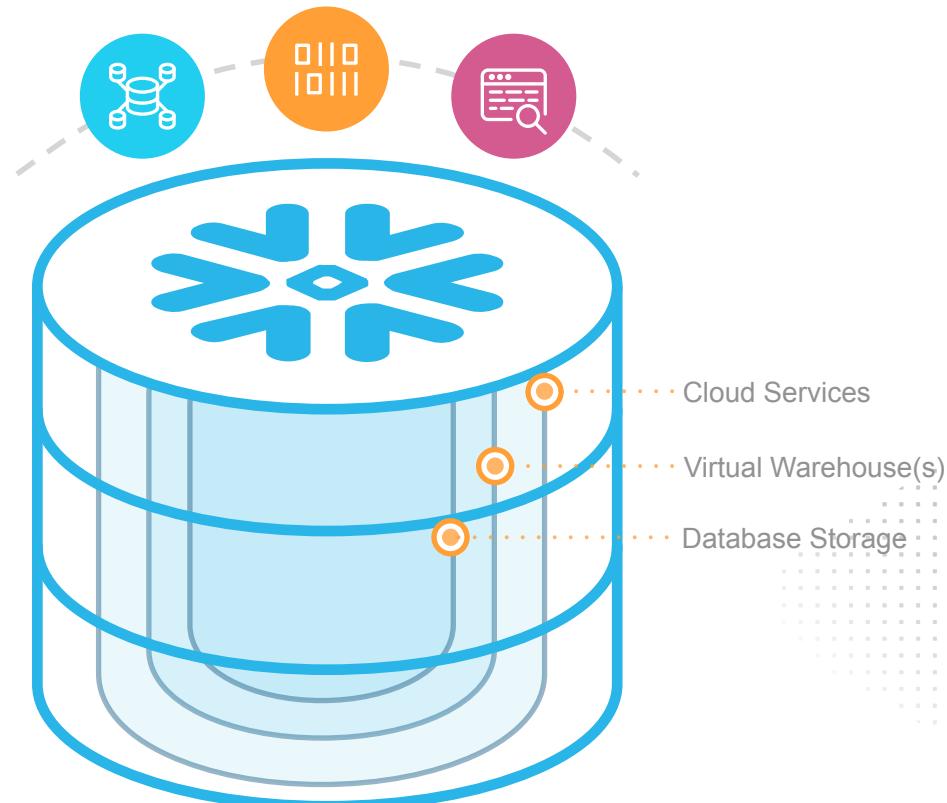
Metadata cached for fast access during query planning

Data

Active working set transparently cached on virtual warehouse SSD

Query results

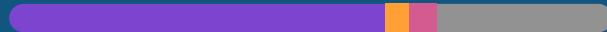
Results sets cached for reuse without requiring compute (e.g., static dashboard queries)



Micro-partitions

Date	Product	Customer	Amount
Feb 14	Boots	Frank	\$150
Feb 14	Boots	Benoit	\$150
Feb 14	Skis	Thierry	\$300
Feb 14	Snowboard	Mike	\$250
Feb 15	Boots	Chris D	\$150
Feb 15	Skis	Denise	\$600
Feb 15	Snowboard	Shelly	\$250
Feb 16	Boots	Rob	\$150
Feb 16	Skis	Sunny	\$600
Feb 16	Snowboard	Chris K	\$250
Feb 16	Snowboard	Greg	\$750
Feb 16	Snowboard	Matt	\$750

Profile Overview (Finished)



- Processing (100%)
- Local Disk IO (63%)
- Synchronization (3%)
- Initialization (31%)

100%

63%

3%

3%

31%

100.00%

179.58 MB

100.00%

25.33 MB

Scan progress

Bytes scanned

Percentage scanned from cache

Bytes written to result

Partitions scanned

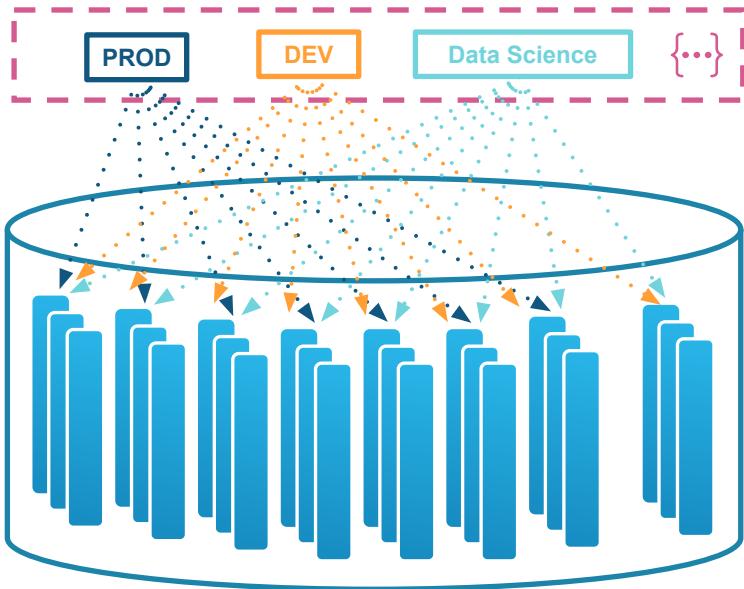
13

Partitions total

73



Zero-Copy Cloning



The Metadata layer keeps track of every micro-partition file in every customer database.

Creating a DEV environment usually means copying the PROD database

Limited to subset of full Prod

Up to 2x storage requirement

Periodic refreshes

Snowflake Zero-Copy Clones

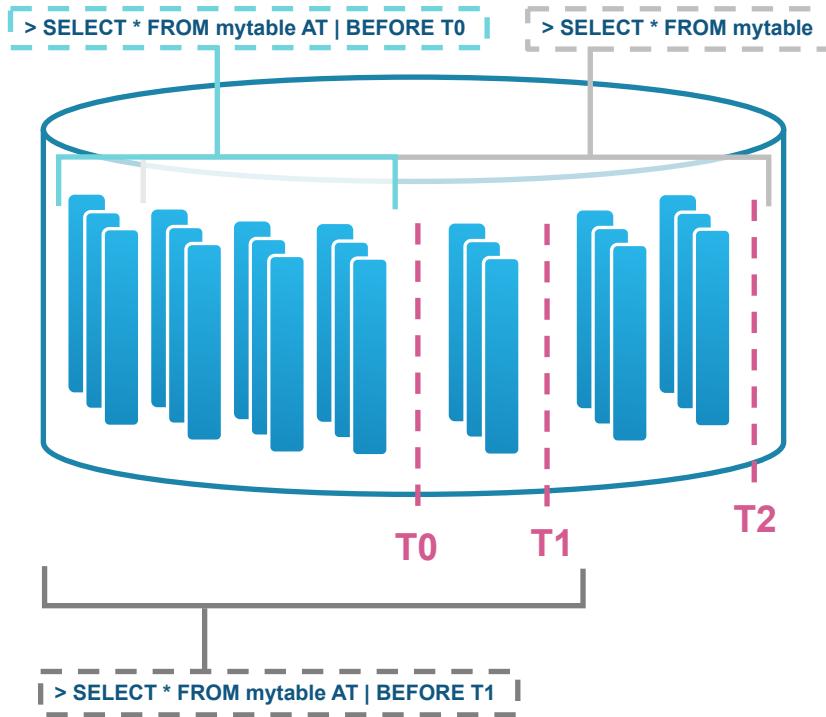
Simply “point” to the same files

Consumes zero additional storage

Changes to either DB are isolated



Time Travel



T0 – Initial state of database

T1 – update myTable set
colX = Y where...

T2 – ELT job loads new data

**Previous versions of data
automatically retained**

AT | BEFORE [timestamp | statement | offset]

CLONE AT | BEFORE to recreate a prior version

UNDROP recovers from accidental deletion

Accessed via SQL extensions

AT | BEFORE [timestamp | statement | offset]

CLONE AT | BEFORE to recreate a prior version

UNDROP recovers from accidental deletion



Snowflake Performance Index

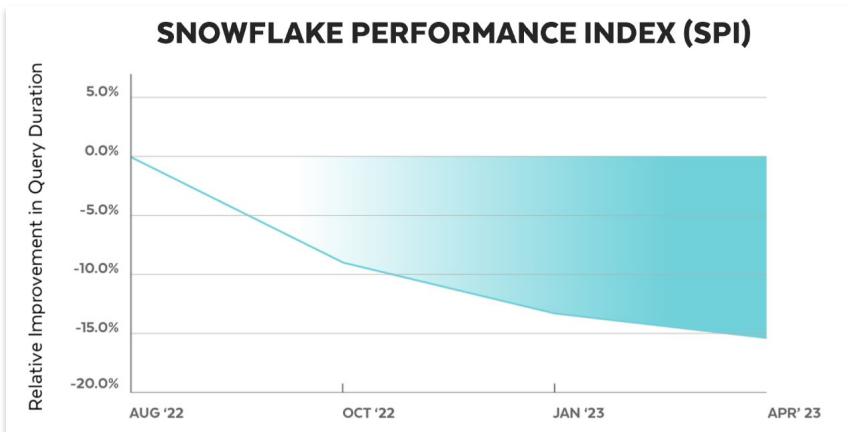
Aggregate index for measuring improvements to performance experienced by customers over time

Results

Query duration for customers' recurring workloads improves by **15%** since we started tracking metric*

How it's calculated

The SPI is calculated based on a comparable cohort of stable customer workloads consisting of millions of jobs each month



* Based on internal Snowflake data from August 25, 2022 to April 30, 2023. To calculate SPI, we identify a group of customer workloads that are stable and comparable in both amount of queries and data processed over the period presented. Reduction in query duration resulted from a combination of factors, including hardware and software improvements and customer optimizations.



SQL Delight Features

Description

Series of SQL improvements and functions, enabling customers to write more concise, efficient queries

- **Select*** now includes **Exclude** and **Rename** keywords
- **MIN_BY** and **MAX_BY** aggregate functions
- Banker's rounding
- **Group By All**

Value

Reduce lengthy queries and unnecessary joins, so you can concentrate on core business logic instead

```
SELECT * EXCLUDE first_name RENAME (department_id AS department,
employee_id AS id)
FROM employee_table;

SELECT MIN_BY (employee_id, salary, 3), MAX_BY (employee_id, salary, 3)
FROM employees;

SELECT State, City. SUM(retail_price * quantity) AS gross_revenue
FROM sales
GROUP BY ALL;
```



Speed Up a Variety of Workloads

Employ key serverless services



Auto Clustering

Reorganize table data to enable efficient pruning of micro partitions



Search Optimization

Speed up point lookup searches that locate limited subset of rows from large table



Materialized Views

Pre-compute expensive calculations for fast access



Serverless Tasks

Schedule SQL statement to be executed as a recurring event



Query Acceleration

Add an extra burst to large scale, exploratory analysis



Search Optimization Service Updates

Description

An extensive set of predicates for filtering data to significantly improve performance of certain types of lookup and analytical queries that use

Value

Run queries in structured, semi-structured, and geography data faster and more cost effectively

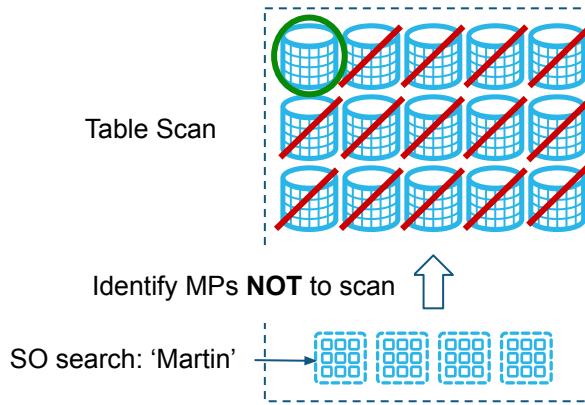
Functionality

Search Optimization now supports the following features:

- Ability to select specific columns for search optimization
- Faster searches in semi-structured data and geography data types
- Faster substring searches in string columns
- Improved performance with Search Optimization and Query Acceleration Service working together

```
SELECT * FROM TRANSACTIONS WHERE Name='Martin'
```

Table Scan



Query Acceleration Service

Elastic scale out of compute resources without changing size of virtual warehouse

What It Does

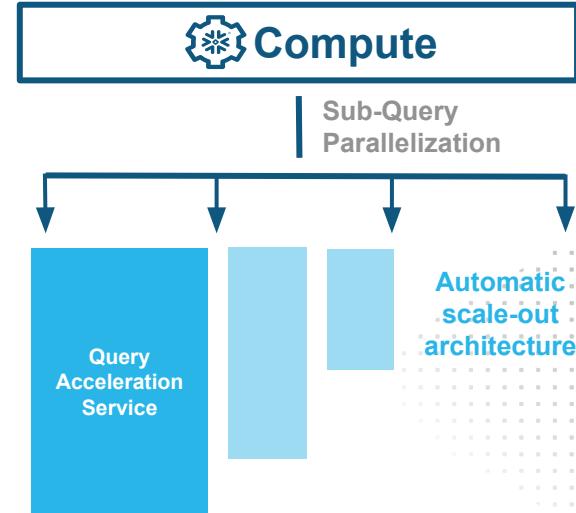
Improved performance with existing compute resources for spikes in workload compute needs.

Performance Improvements

How It Works

Enables vertical scaling on a query by creating sub-query parallelization.

Reduce execution times



Geospatial Support

Geography Type

- Store and analyze POINTs, LINESTRINGS, POLYGONs
- Input/Output as GeoJSON, Well-Known Text (WKT), and Well-Known Binary (WKB)
- Spherical model of earth

OGC-Compliant Functions

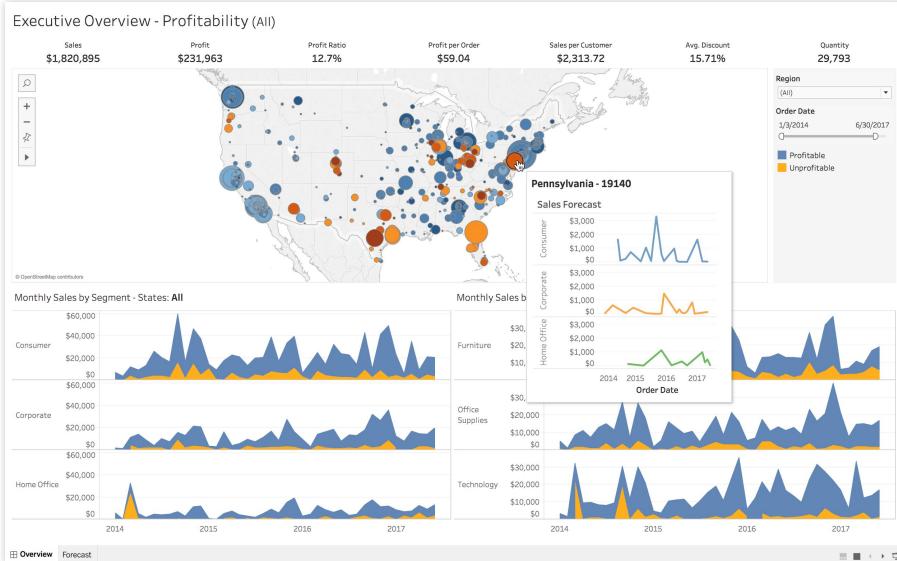
- CONTAINS, INTERSECTS, DISTANCE, DWITHIN, and more

Performance

- Pruning and joins on geospatial predicates (Q2)

Ecosystem

- Integrates with BI tools for visualization
- Works with spatial ETL tools for data integration



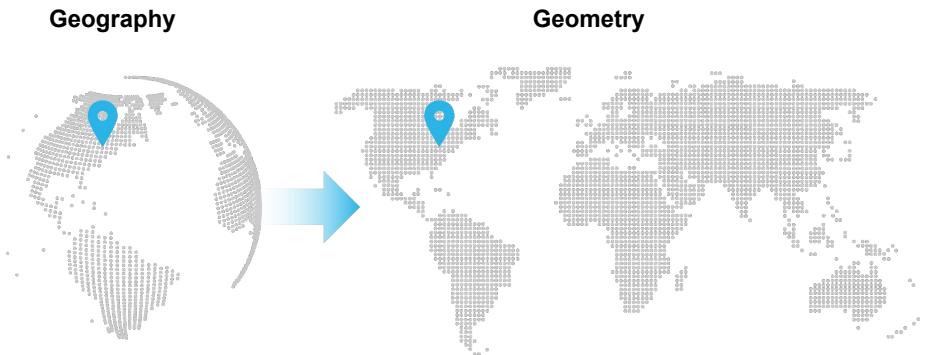
Geospatial Innovations

What is new?

- The new **Geometry** Data Type supports multiple spatial reference systems
- Easily switch between spatial systems with **ST_Transformation** (in PuPr)
- We are also announcing support for invalid shapes for both **Geometry** and **Geography** and new functions for the **Geography** data type

Value

Enable geospatial use cases in manufacturing, government, and infrastructure. Snowflake can now ingest any type of spatial vector object, allowing customers to process all types of vector geospatial data



INTELLIGENT INFRASTRUCTURE



Cloud Services



Snowflake Managed

Maintenance & Tuning

Multi-Cluster Compute Resources

Administration

Availability

Networking & Encryption

Query Design & Tuning

Self-managed

Automate encryption, access controls, availability, tuning, maintenance, and more to keep operations simple and smooth.

Transparent improvements

Continually benefit from the latest performance enhancements and economics - no action required.

Optimized resources & costs

Only pay for what you used and get full visibility and cost governance controls to right-size costs.



Snowflake Horizon

Built-in governance solution with a unified set of capabilities



Compliance

Protect & audit your data with compliance, business continuity, data quality monitoring, & lineage.

Security

Secure your environment with continuous risk monitoring & protections, RBAC, & granular authorization policies.

Privacy

Unlock the value of your sensitive data with advanced privacy policies & data clean rooms.

Interoperability

Integrate with other Apache Iceberg- compatible catalogs & engines, & with data catalog & data governance partners.

Access

Classify, share, discover & take immediate action on data, apps and more across regions & clouds.



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](#)



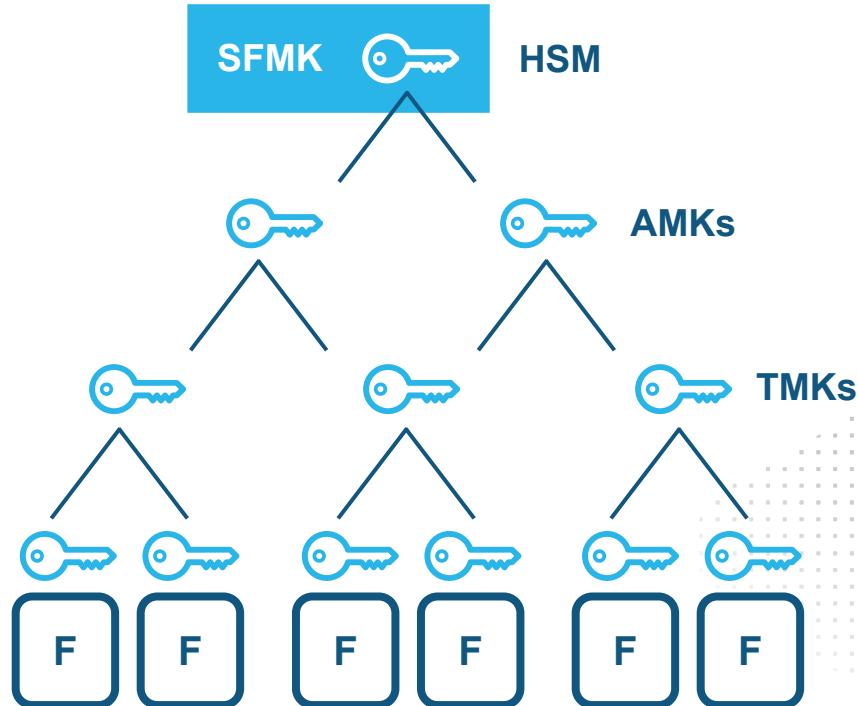
© 2023 Snowflake Inc. All Rights Reserved

Tri-Secret Secure

Encrypt using customer managed key (CMK)

Customer Key

- Bring your own key (BYOK)
- Revoke at any time → no access to data
- Keys are rotated every 30 days
- Customers setup re-keying

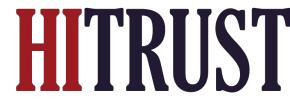


Snowflake Will Be 100% Transparent

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)



Self-Assessments
CAIQ, SIG Lite, Pen Test Results



Snowflake Supported Regions

Available on customer's cloud & region of choice



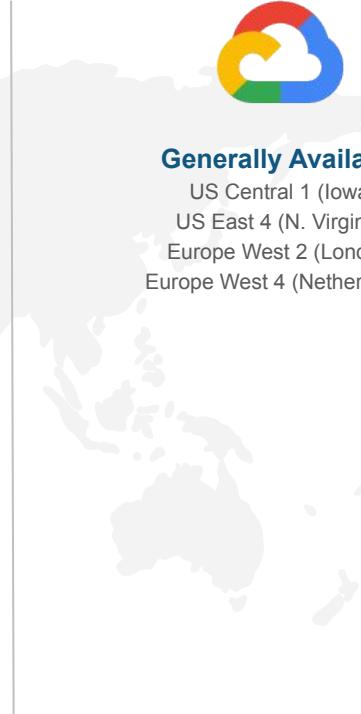
Generally Available

- US West (Oregon)
- US East (Ohio)
- US East (N. Virginia)
- Canada (Central)
- US Gov West 1*
- US Gov East 1*
- US East (Commercial Gov – N Virginia)**
- EU (Ireland)
- Europe (London)
- EU (Paris)
- EU (Frankfurt)
- EU (Stockholm)
- Asia Pacific (Tokyo)
- Asia Pacific (Osaka)
- Asia Pacific (Seoul)
- AWS Pacific (Mumbai)
- Asia Pacific (Singapore)
- Asia Pacific (Sydney)
- Asia Pacific (Jakarta)
- South American (Sao Paulo)



Generally Available

- West US 2 (Washington)
- Central US (Iowa)
- South Central US (Texas)
- East US 2 (Virginia)
- Canada Central (Toronto)
- US Gov (Virginia)**
- UK South (London)
- North Europe (Ireland)
- West Europe (Netherlands)
- Switzerland North (Zurich)
- UAE North (Dubai)
- Central India (Pune)
- Japan East (Tokyo)
- Southeast Asia (Singapore)
- Australia East (New South Wales)



Generally Available

- US Central 1 (Iowa)
- US East 4 (N. Virginia)
- Europe West 2 (London)
- Europe West 4 (Netherlands)



THANK YOU



© 2023 Snowflake Inc. All Rights Reserved