

SNOWFLAKE OVERVIEW



What is Snowflake?

A single, global platform that powers the Data Cloud. Snowflake is uniquely designed to connect businesses globally, across any type or scale of data and many different workloads, and unlock seamless data collaboration.

Performant

Never worry about resource contention again. Snowflake's elastic performance engine delivers instant and near-unlimited scale. Support a virtually unlimited number of concurrent users and workloads ranging from interactive to batch. Bring your code to the data by deploying ML models in Snowflake to access and use all your data in its original format, reducing time to value.

Fully Automated

Trade in manual management and maintenance for automations that enable you to easily operate at scale, optimize costs, and minimize downtime. Snowflake's platform is fully managed to support effortless data management, security, governance, availability, and data resiliency - reducing risk and improving your operational efficiency.

Collaborative

Say goodbye to ETL and data silos. Easily and securely access and share a single copy of your data across your departments, business units, and subsidiaries; with your supply chain and other business ecosystem partners; and with thousands of organizations that make up the Data Cloud. Seamlessly connect cross-cloud and cross-region and enable global governance policies that follow the data.



Solutions Powered by Snowflake

GROWTH



Financial Planning, Analysis &
Budgeting Analytics



Operational & Supply Chain
Innovation with IoT



AI/ML-driven Modeling for
Manufacturing & Production

COST



Reduce IT Overhead &
Increase Time to Market



Inventory, Distribution
& Logistics Optimization



Predictive Maintenance for
Asset Utilization

RISK



Cybersecurity &
Compliance



Increased Visibility in
and Across BUs



Quality Control
Optimization

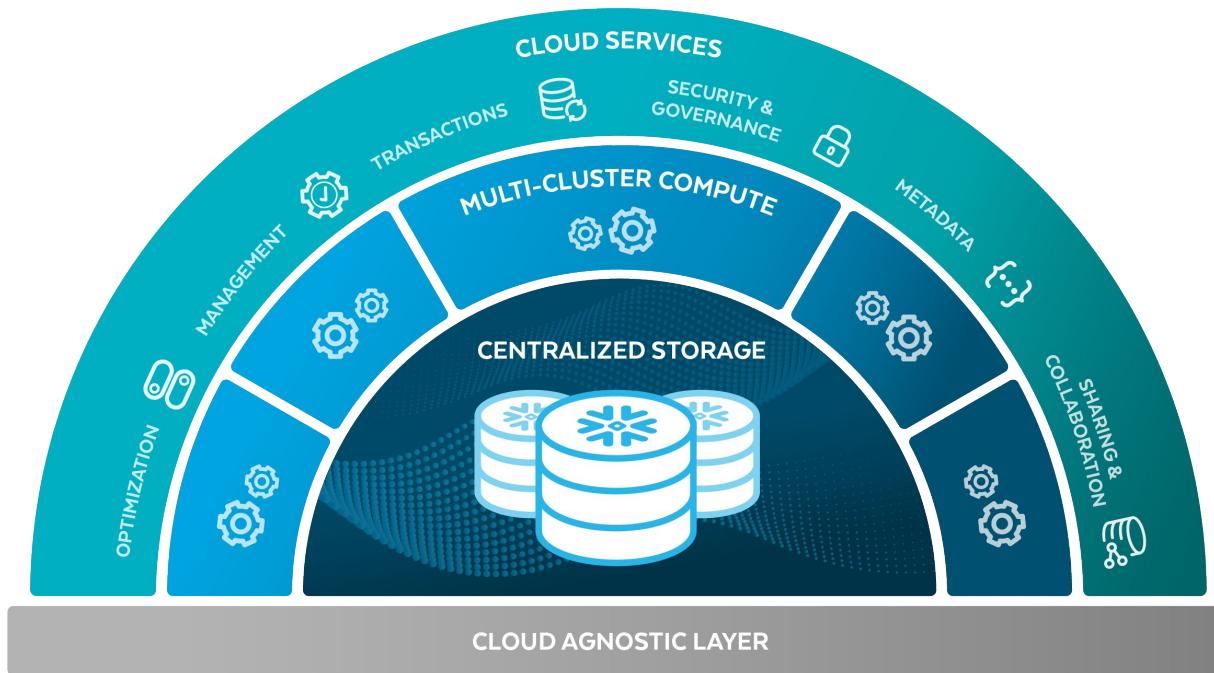


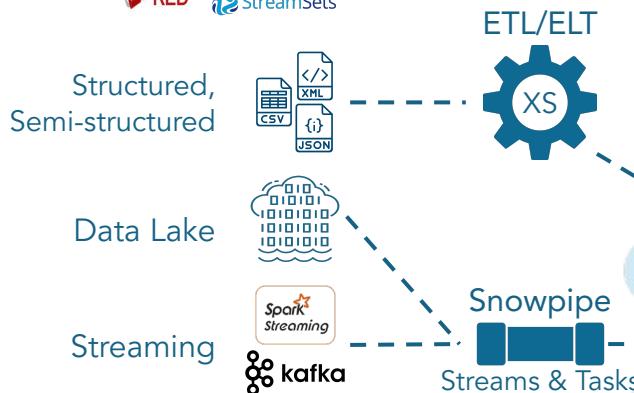
DATA MANAGEMENT WITH SNOWFLAKE

- Scaling, Collaboration, Governance & Security
- Platform Demo



SNOWFLAKE ARCHITECTURE





Global Cloud Services

Metadata Repository
Logical Model
Security
Query Planning & Optimisation
Transactional Control

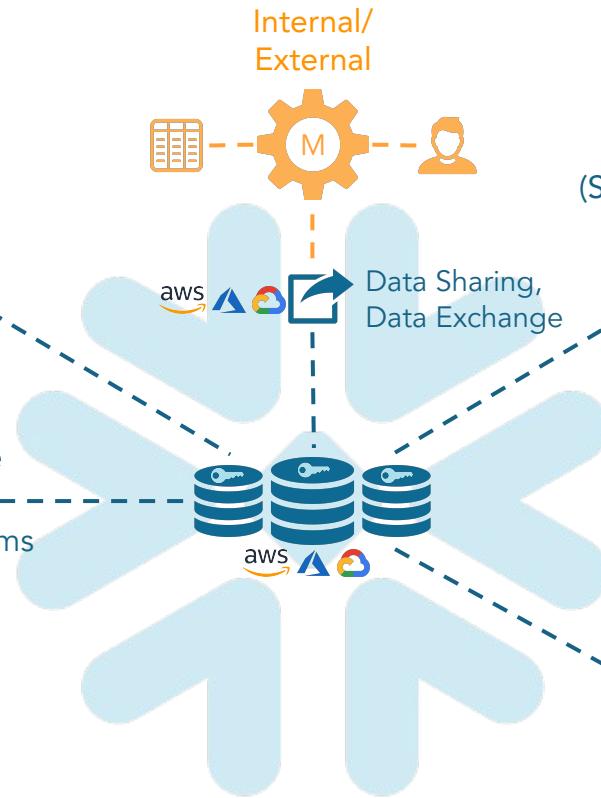
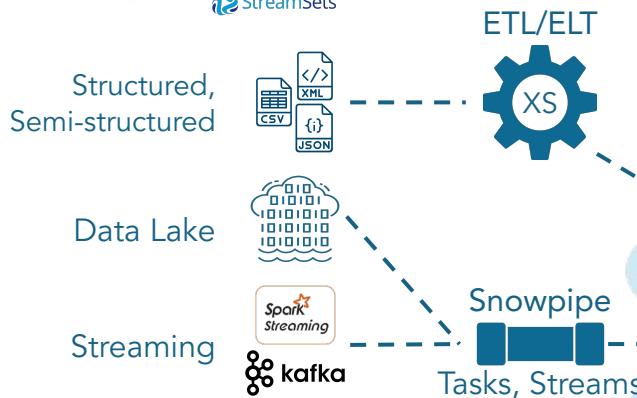


Dashboards,
(Self Service) BI



Data
Science

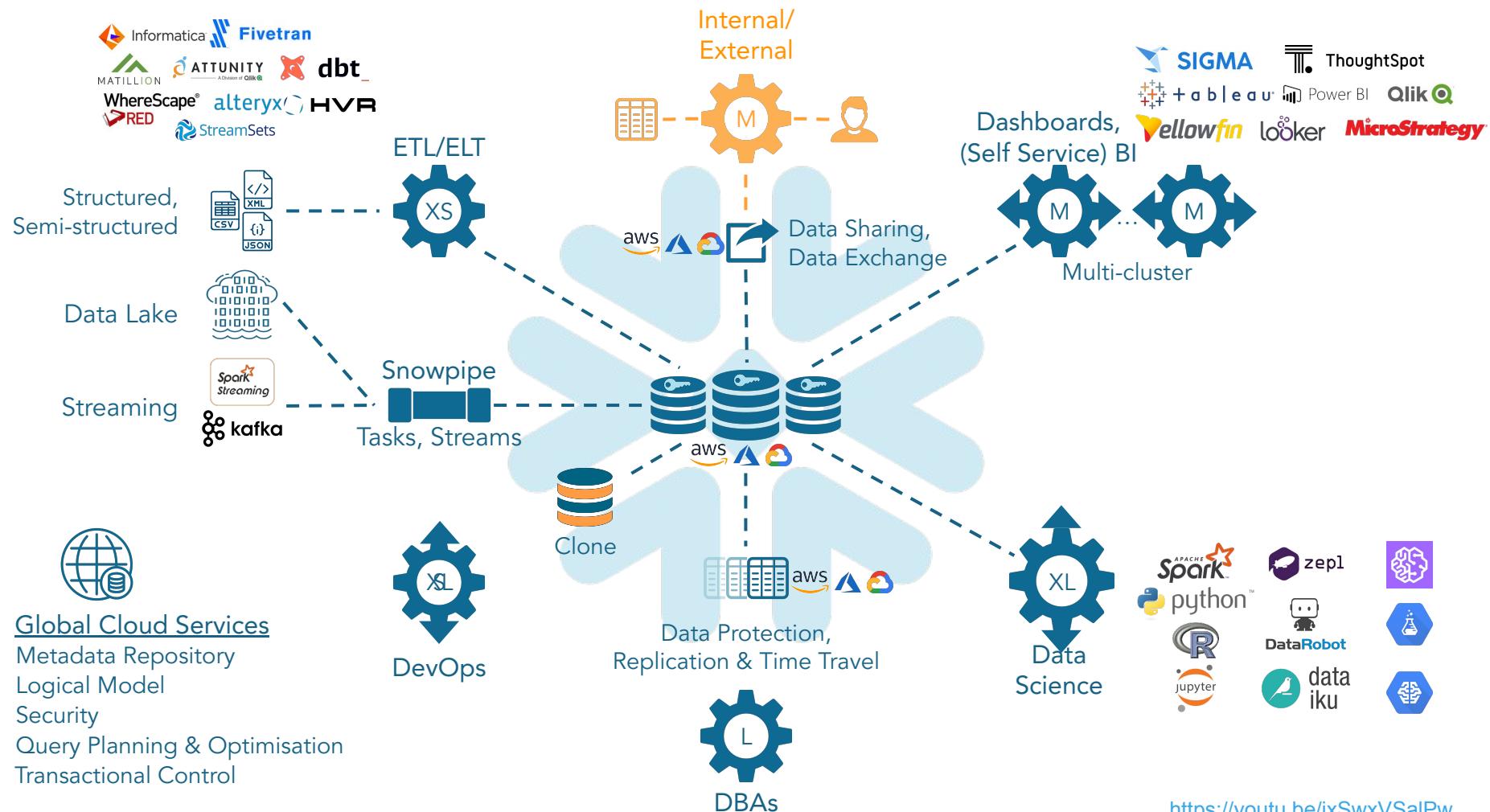




Global Cloud Services

Metadata Repository
Logical Model
Security
Query Planning & Optimisation
Transactional Control





<https://youtu.be/jxSwxVSalPw>

Data Engineering with Snowflake

What's in the toolbox?

1



DATA ENGINEERING FEATURES



- Highly performant, batch & micro-batch data loads using **Virtual Warehouses**
- Load data continuously using **Snowpipe**
- Ingest streaming data using the **Kafka Connector**
- Huge ecosystem

PLATFORM FEATURES



- Native support for structured, semi-structured and unstructured data enabling schema-on-read
- Data Sharing & the Snowflake Marketplace

BENEFITS



One platform, one copy of data



Secure and governed access to all data



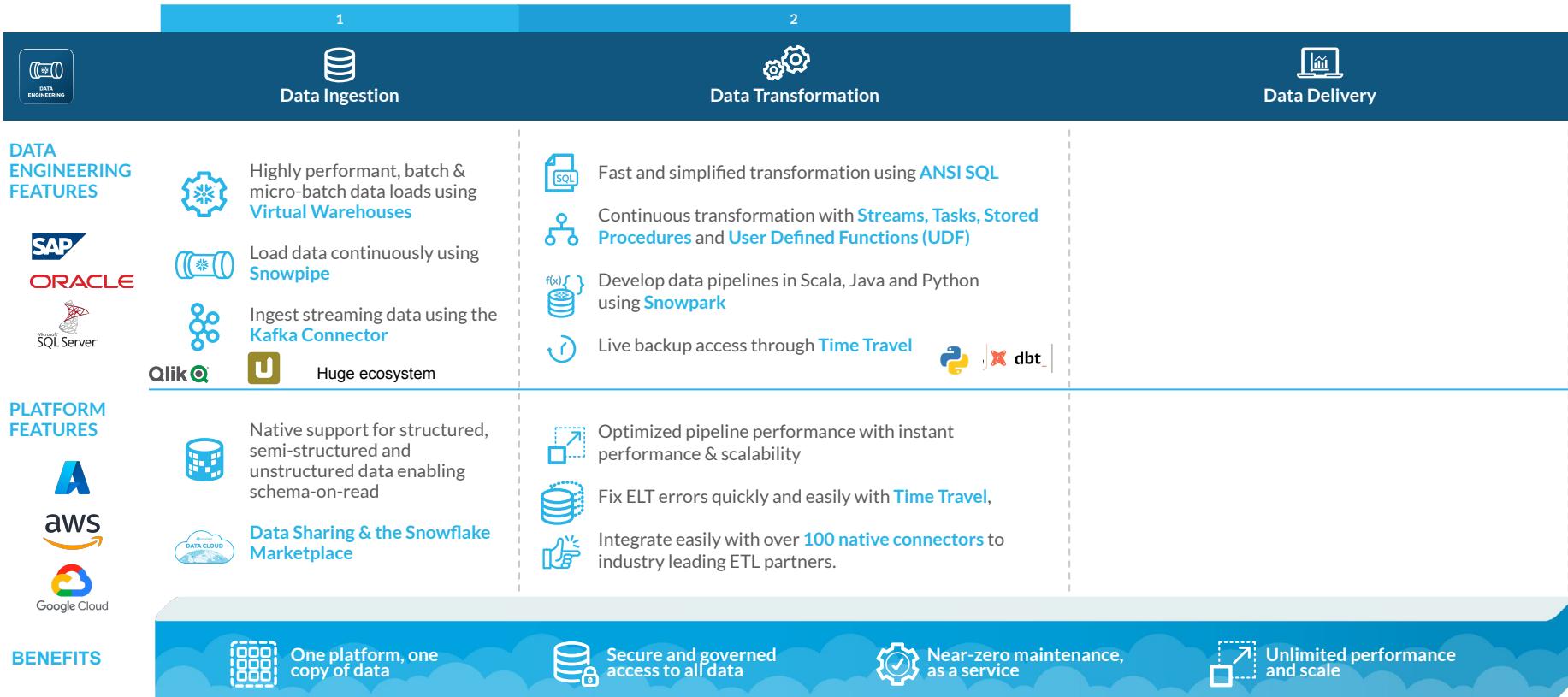
Near-zero maintenance, as a service



Unlimited performance and scale

Data Engineering with Snowflake

What's in the toolbox?

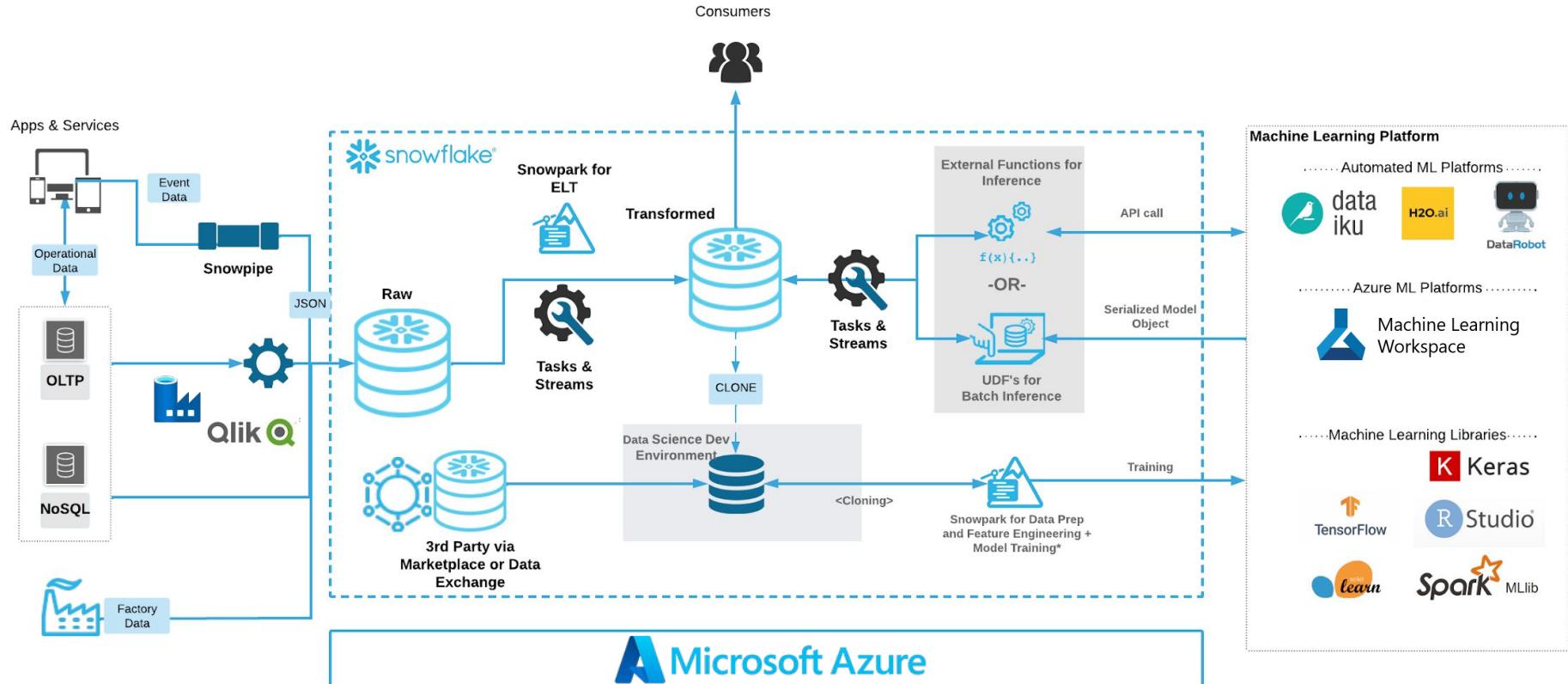


Data Eng with Snowflake for SAP to PowerBI

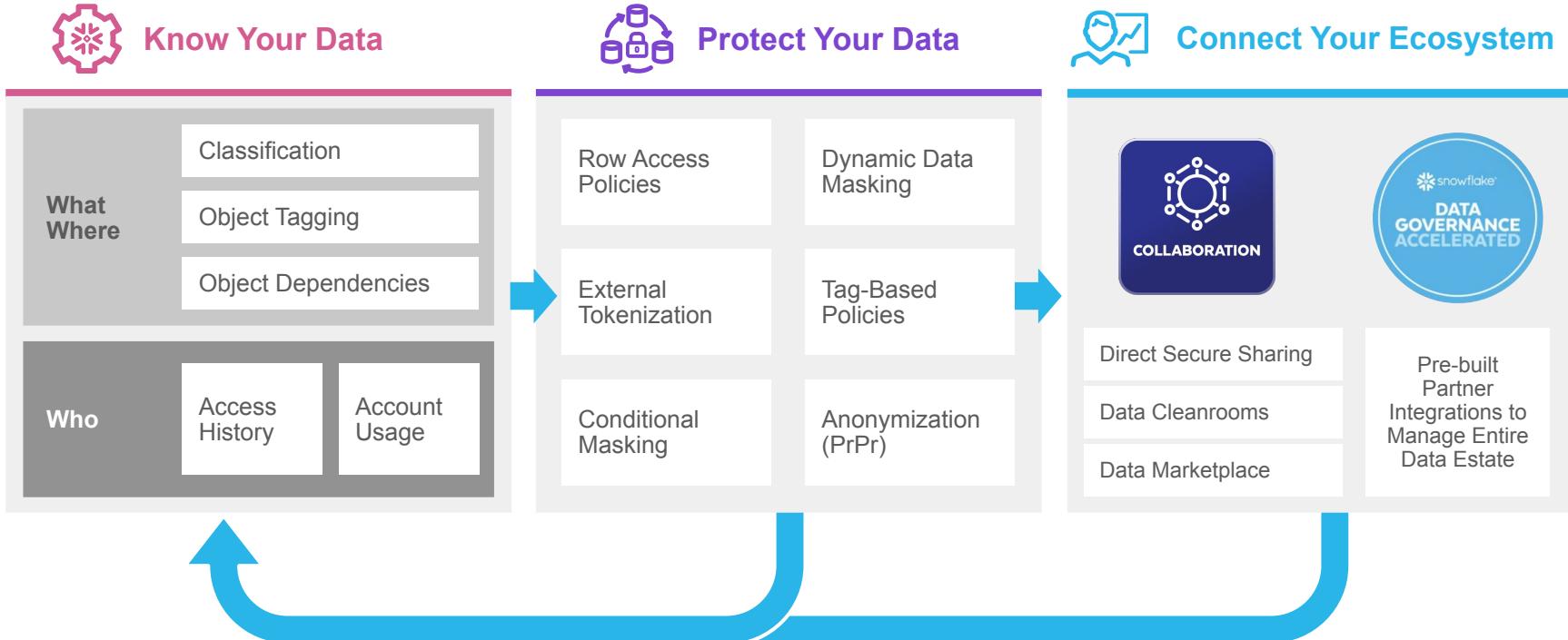
What's in the toolbox?

	1 Data Ingestion	2 Data Transformation	3 Data Delivery
DATA ENGINEERING FEATURES	 Highly performant, batch & micro-batch data loads using Virtual Warehouses  Load data continuously using Snowpipe  Ingest streaming data using the Kafka Connector   Huge ecosystem	 Fast and simplified transformation using ANSI SQL  Continuous transformation with Streams, Tasks, Stored Procedures and User Defined Functions (UDF)  Develop data pipelines in Scala, Java and Python using Snowpark  Live backup access through Time Travel  	 Easily share or consume data with partners, suppliers or internal departments using Data Sharing  Built-in support for common Connectors & Drivers such as ODBC, JDBC, Python, Spark, etc.  Support an open ecosystem by unloading Data Lake Export
PLATFORM FEATURES	 Native support for structured, semi-structured and unstructured data enabling schema-on-read   Data Sharing & the Snowflake Marketplace	 Optimized pipeline performance with instant performance & scalability  Fix ELT errors quickly and easily with Time Travel ,  Integrate easily with over 100 native connectors to industry leading ETL partners.	 Over 100 native connectors to industry leading tech partners such as PowerBI, Tableau, Qlik, Microstrategy etc   
BENEFITS	 One platform, one copy of data	 Secure and governed access to all data	 Near-zero maintenance, as a service
			 Unlimited performance and scale

Typical Reference Architecture + Data Science



Unified Governance



COLLABORATION IN THE DATA CLOUD

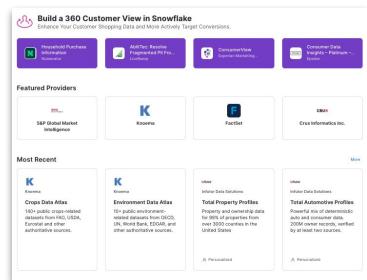
PRIVACY PRESERVING COLLABORATION FOR EVERY SCENARIO

DISCOVER AND MONETIZE IN THE DATA CLOUD

Acquire and directly access data, services and apps from 400+ providers

[SNOWFLAKE MARKETPLACE](#)

Monetize your products with usage based pricing



SHARE WITH COMPANIES NOT YET ON SNOWFLAKE

READER or REFERRAL

Manage Reader accounts on your customer's behalf or easily refer them to Snowflake using a [Listing Referral Link](#).



SHARE ACROSS YOUR BUSINESS ECOSYSTEM

EXTERNAL SHARING

Share with your customers, partners, vendors, suppliers with Snowflake accounts



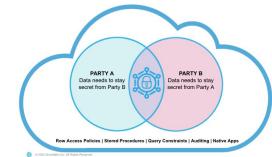
INTERNAL SHARING

Facilitate collaboration and discovery across business units

ANALYZE DATA WITHOUT SHARING IT

DATA CLEAN ROOMS & DATA SERVICES

Collaborate on sensitive data between multiple parties

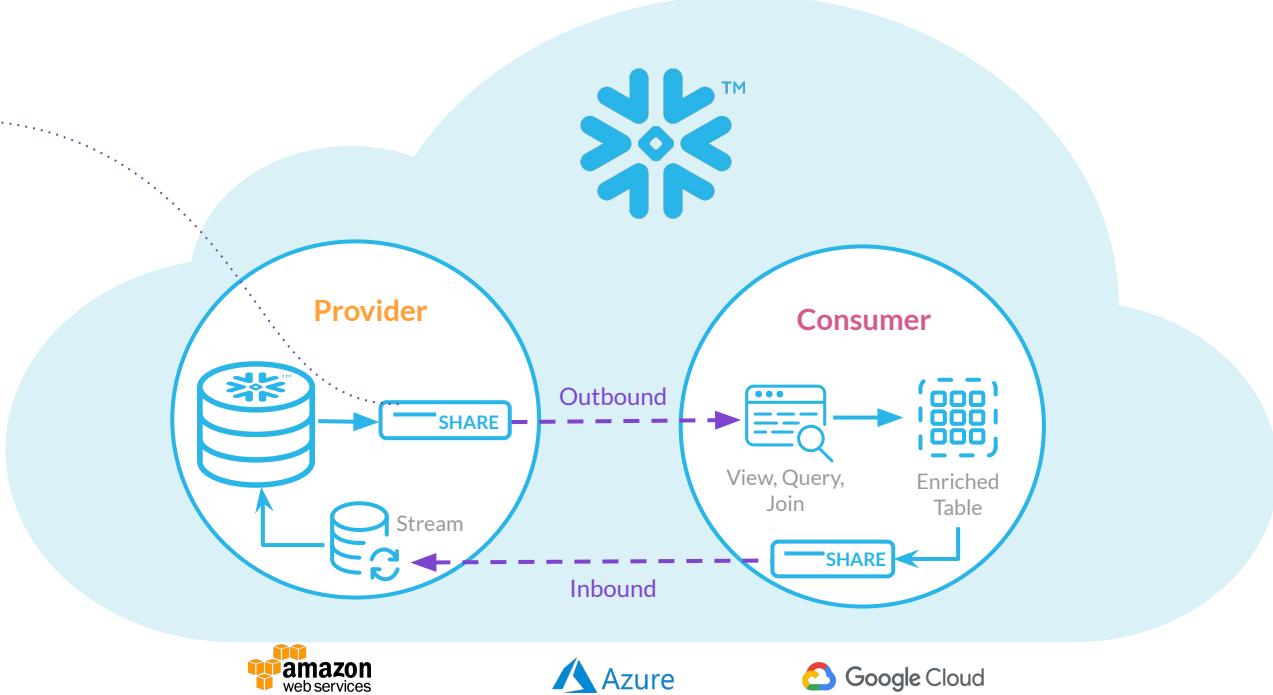


Live data, services and apps cross-cloud/region. **No ETL ever.**

SECURE DATA SHARING

within a Snowflake region on any cloud

Components of a Share	
Tables	External Tables
 Tables	 External Tables
Secure Views	Secure access to data across multiple databases
 Secure Views	 Secure access to data across multiple databases
Secure UDFs	API-style access
 Secure UDFs	 API-style access
Secure Materialized Views	Filters or aggregates pre-computed and updated
 Secure Materialized Views	 Filters or aggregates pre-computed and updated
Data Services	Enrich input data using double-blind secure joins
 Data Services	 Enrich input data using double-blind secure joins

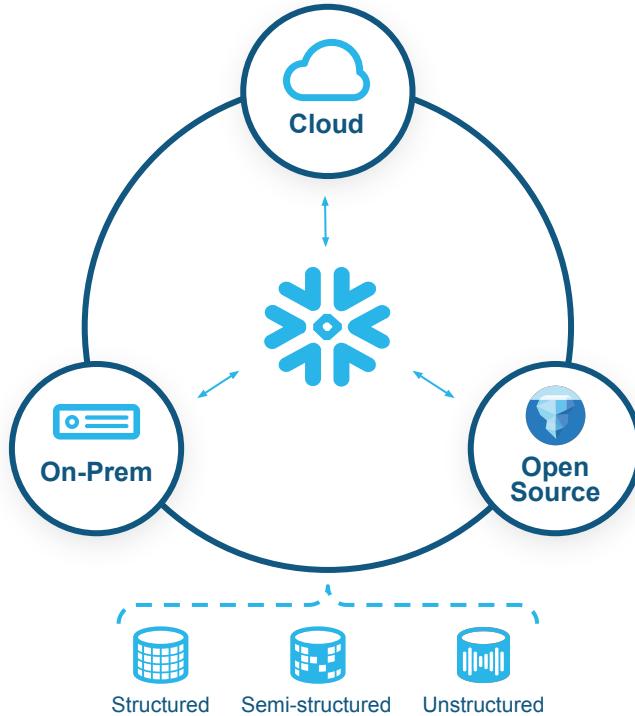


- NO data movement, NO data latency within same cloud region
- Provider can instantly revoke object access or drop share
- Shared Table/View Stream for change capture, historical snapshots

TECHNICAL DEEP DIVE



Optimized Centralized Storage



Unsiloed access to your data

Unstructured, semi-structured, and structured data together with near-infinite scale.

Easily manage data at scale

Fast and efficient access, optimized compression, and secure data - all automated.

Flexibility & interoperability

Work with data on-premises or in open table formats to remove lock-in and adapt to new data patterns.

Support to All Data Types

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

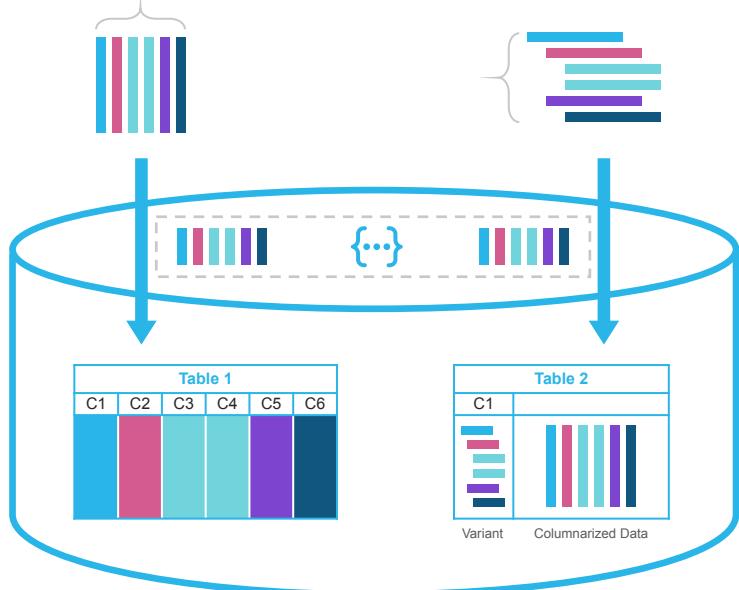


Semi-Structured Data

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

Structured data

(Delimited or fixed width files)



Structured formats (CSV, TSV, ...)

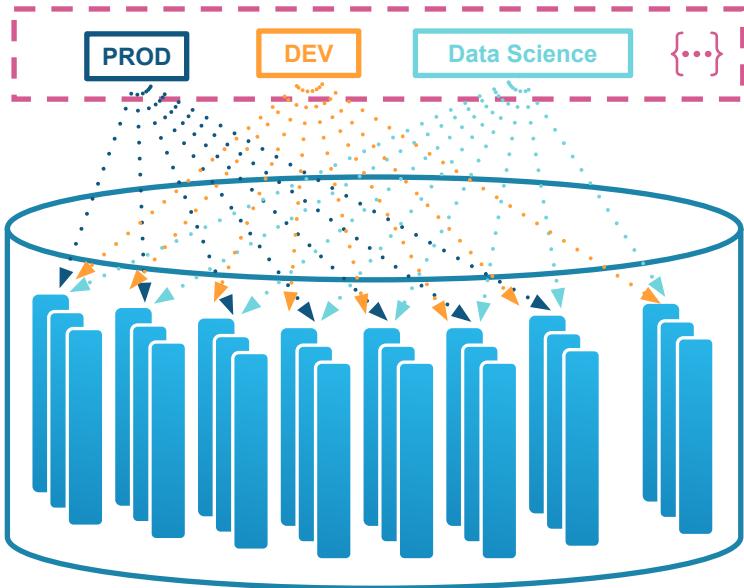
- Strongly-typed “columns”
- Typically map 1:1 to table columns
- Ingestion process generates important metadata

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

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



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

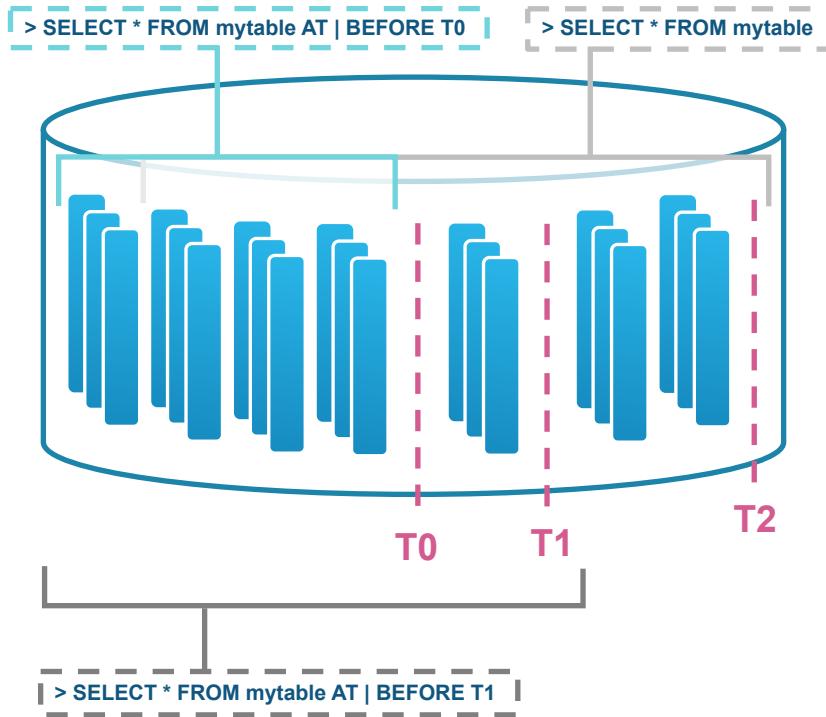
Simply “point” to the same files

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

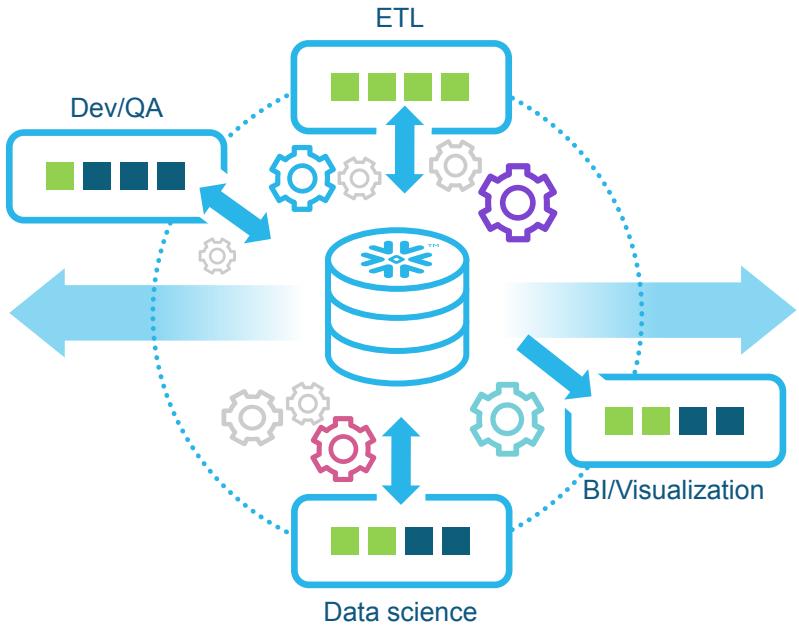
AT | BEFORE [timestamp | statement | offset]

CLONE AT | BEFORE to recreate a prior version

UNDROP recovers from accidental deletion



Elastic Performance Engine



One engine for every workload

Simplify your architecture. Power complex pipelines, analytics, data science, interactive applications, and more.

Leading performance and concurrency

Fast, reliable performance for virtually all users and jobs with no tuning or contention.

Accessible & programmable

Work in SQL, Python, Java, and more, and run your preferred tools and libraries - without moving data.



Intelligent Infrastructure



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.



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

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



Materialized Views

Source Table

col ₁	col ₂	...	col _n
	M		
	F		
	F		
	F		

Partition 7

col ₁	col ₂	...	col _n
	M		
	F		
	F		
	F		

Partition 12

col ₁	col ₂	...	col _n
	M		
	F		
	F		
	F		

Partition 35

Materialized View

Partition	Table	col ₂	SUM(col ₁)
7	243	M	50017
7	243	F	37565
12	243	M	43090
12	243	F	27001
35	243	M	34234
35	243	F	35743

Partition 1

What is it?

Store frequently used projections and aggregations to avoid wasting time and money re-computing

Results from MV queries guaranteed to be up-to-date

Value

Run faster queries, spend fewer compute credits

Asynchronous, incremental maintenance

No impact on DML performance

MV on External Tables

- Fast, governed access to external storage
- Benefit from Snowflake performance while maintaining source of truth outside

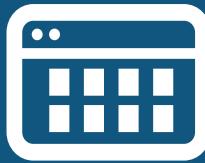
Cost

Serverless refresh, additional storage

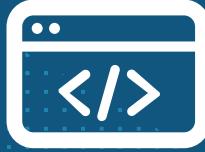


Interactive Use Cases

- Deliver consistent low latency (less than 1 sec)
- Provide cost effective way to deliver high throughput (100+ QPS)



**Interactive
Dashboards**



**Embedded
Analytics & Data
Applications**

Relevant Key features



Match Recognize

Easily find patterns using SQL

1	7	4	2	3	4	5
6	5	9	7	8	9	1
6	8	4	6	8	5	2
4	5	3	2	6	1	2
1	5	4	2	4	2	3
5	8	8	1	5	9	3
7	1	5	9	8	7	6
5	7	2	6	4	2	9
5	2	4	2	3	5	1
2	5	2	1	4	5	9
1	4	2	3	3	4	5
9	4	9	1	3	4	8
2	8	5	4	3	7	6
6	1	3	2	6	4	3
8	3	8	2	7	1	5
7	1	4	4	5	6	5
2	6	2	9	5	8	2
7	3	2	2	9	4	5
1	5	4	2	3	4	5
9	1	2	6	2	4	1
2	9	3	8	5	2	3
7	7	5	5	6	1	9
3	4	4	2	7	8	2
2	3	3	4	6	3	4
3	8	9	5	3	9	7
1	1	4	2	3	4	5
2	7	2	2	5	7	6
9	1	5	7	4	2	3
5	7	2	2	6	2	8

Define the pattern
you're looking for

Identify where that
patterns happen
within your data

Construct your
questions

Analyze your results

Find and analyze patterns that help you
make better business decisions.

Develop new types of analysis that are
otherwise difficult to express in SQL across:

- Financial markets
- IoT sensor data
- Clickstream data
- Customer support
- Fleet telemetry
- User behavior
- Time series



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

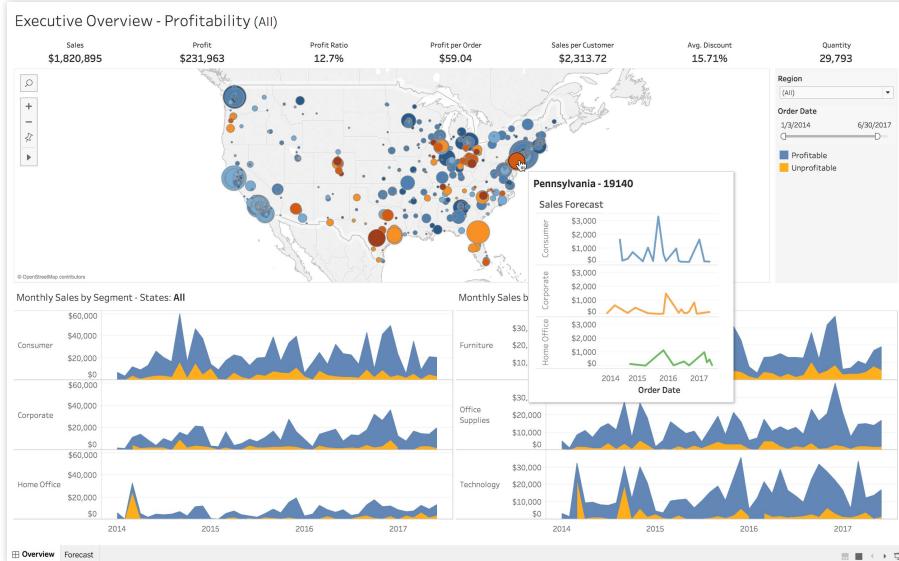
- CONTAINS, INTERSECTS, DISTANCE, DWITHIN, and more

Performance

- Pruning and joins on geospatial predicates (Q2)

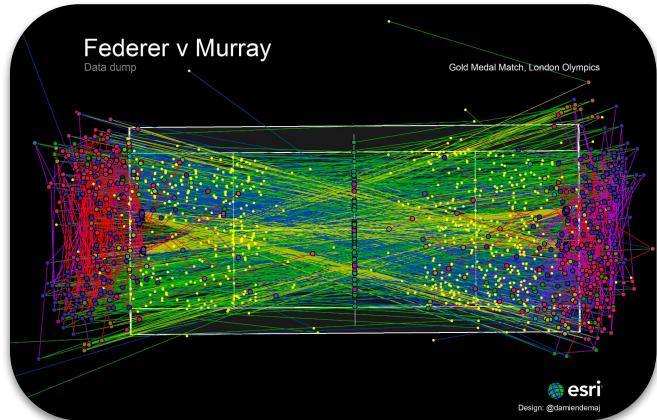
Ecosystem

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



GEOMETRY Data Type Support

- Describes objects on a Cartesian plane
- Coordinates are unit agnostic
- The shortest path between two points is a straight line
- All measurements are calculated in units
- Allows for analyzing data in very specific spatial contexts



Source: esri.com



Source: carto.com

Snowsight



DISCOVER

Find data, results, and queries from across the Data Cloud



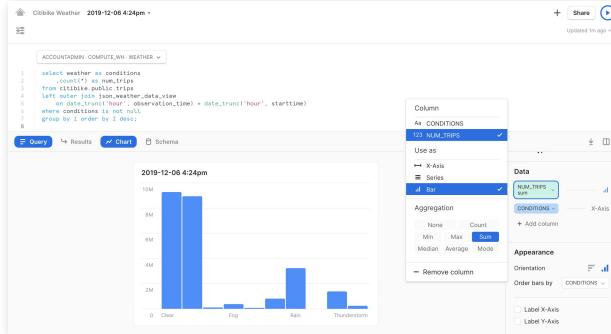
ANALYZE

Analyze with fast and dynamic querying

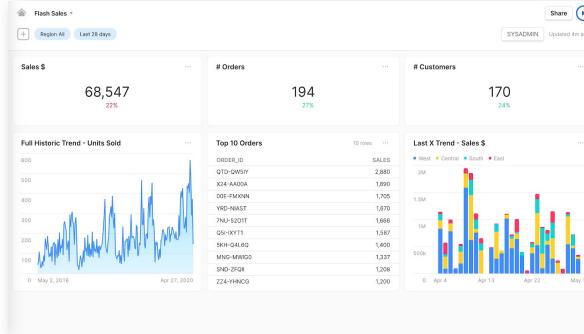


MANAGE

Enable data exploration and insight discovery



Interactive SQL Editor With Charts



Dashboards



Snowsight

Understand and analyze your data directly from the Snowflake UI

Fast and Responsive Querying

A fast, desktop-quality editor in the web

Smart Autocomplete

Contextual suggestions based on your query and SQL dialect, such as aliases and functions

Write Less SQL

Use a date picker to simplify date selection

Interactive Results

Preview data fast no matter how many rows a query returns

Automatic Stats

Interactive stats for all columns help you catch errors and spot trends without follow-up queries

Beautiful Charts

Create charts that look great on any device at any size

Modern Dashboards

Simple drag and drop interface for creating dashboards

Dynamic Data Filters

Use parameters to set dates, customers and more

Share Your Work

Private Links

Send colleagues queries or dashboards with a link to view, run, or edit the contents



Why Snowpark



Language of Choice

Enable all data users to bring their work to a single platform with **native Python, Java, Scala and SQL support**



No Governance Trade-offs

Apply consistent controls **trusted by over 500 of the Forbes Global 2000** across all workloads



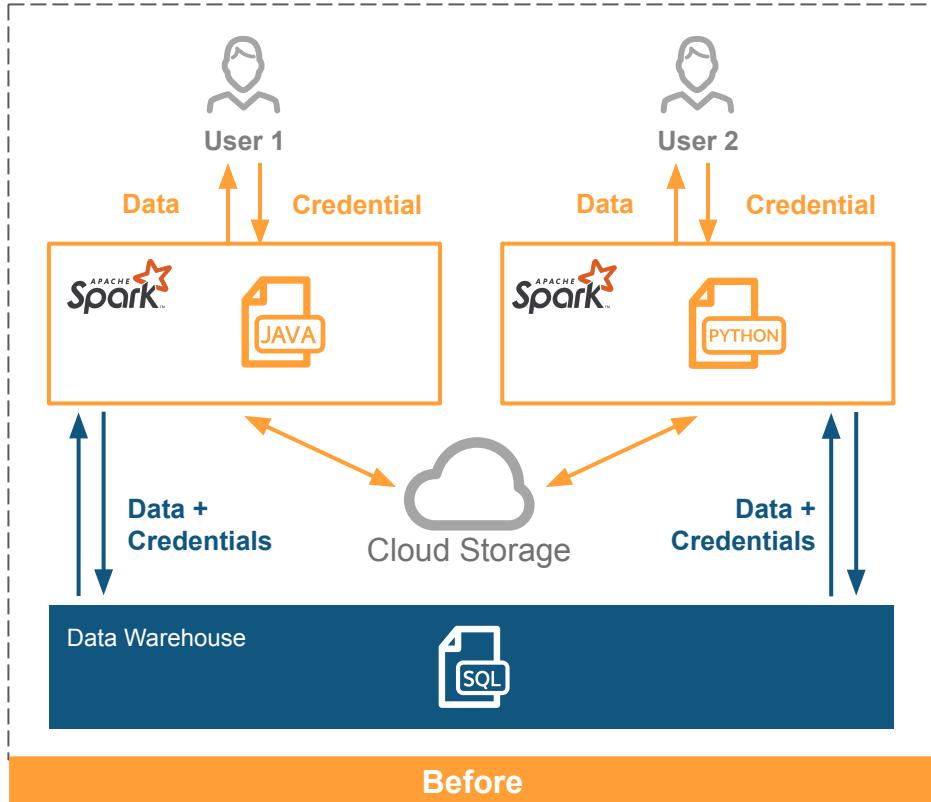
Faster & Cheaper Pipelines

Migrate existing Spark pipelines with minimal code change, better price-performance, transparent cost and less ops overhead

Customers transitioning from Spark-based pipelines are reporting that Snowpark is up to 2-3x faster at 30-50% of the cost.



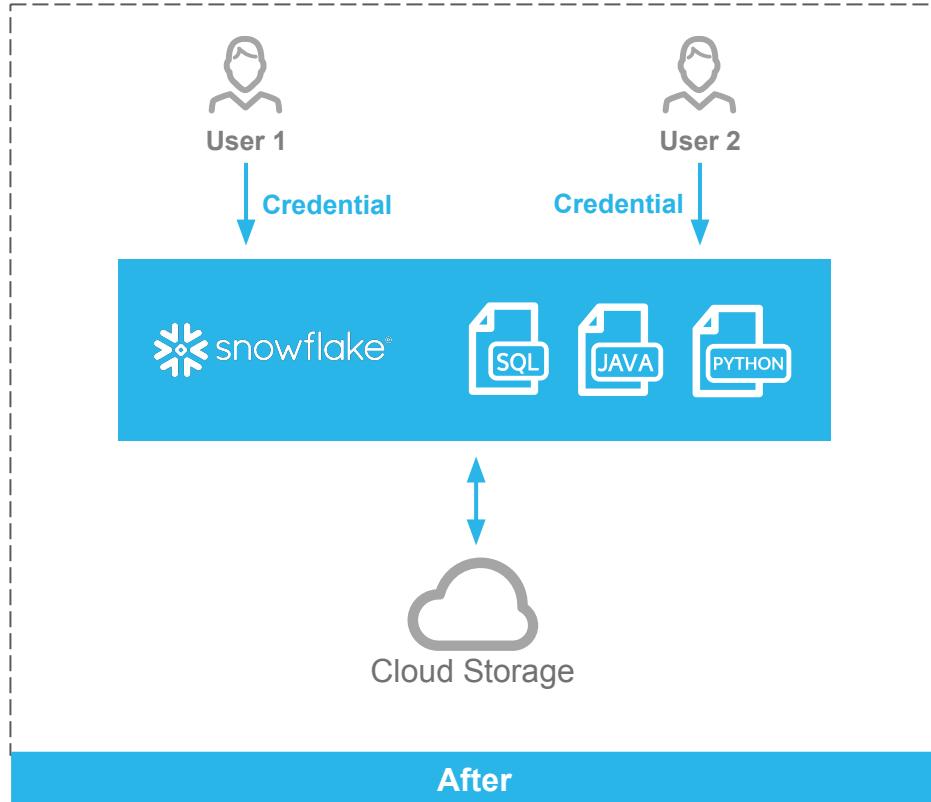
Complexity With Traditional Approach



- Customers often run separate processing clusters for different languages
- Complex capacity management & resource sizing
- Lots of data movement and data silos
- Loose governance control and security loopholes



Streamlined Architecture With Snowflake



- One single platform with native support for different languages
- Simpler capacity management & resource sizing
- Streamline architecture and collaborate on the same data
- Consistent governance and security policies



Snowpark for Python



Familiar Programming Constructs

Use familiar syntax with DataFrame abstraction



Rich Ecosystem

Easy access to hundreds of packages with automated dependency management



Secure Processing

Build with confidence in a highly secure, sandboxed environment



Streamlit Open Source Python Library

```
import streamlit as st
import pandas as pd
import altair as alt

#st.cache
def get_IMI_data():
    AWS_BUCKET_URL = "https://kreamlil-dump-data.s3.us-west-2.amazonaws.com"
    df = pd.read_csv(AWS_BUCKET_URL + "/agri.csv.gz")
    return df.set_index("Region")

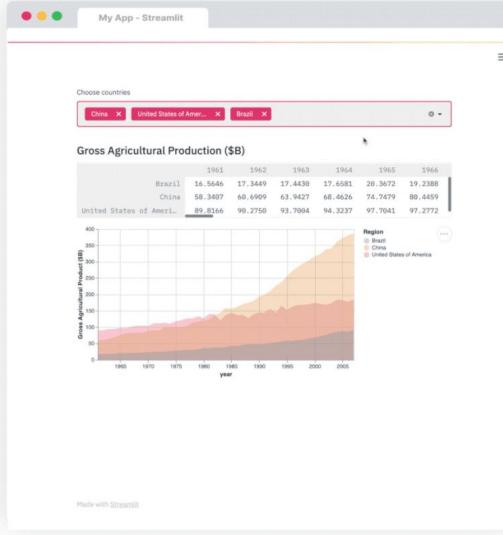
df = get_IMI_data()

countries = st.multiselect(
    "Choose countries", list(df.index), ["China", "United States of America"])

data = df.loc[countries]
data /= 1000000.0
st.write("## Gross Agricultural Production ($B)", data.sort_index())

data = data.T.reset_index()
data = pd.melt(data, id_vars=["index"]).rename(
    columns={"index": "year", "value": "Gross Agricultural Product ($B)"})
data["year"] = data["year"].str[-4:]

chart = alt.Chart(data)
.chart(data)
.mark_area(opacity=0.3)
.encode(
    x="year:T",
    y=alt.Y("Gross Agricultural Product ($B):Q", stack=None),
    color="Region:N",
)
chart
st.altair_chart(chart, use_container_width=True
```



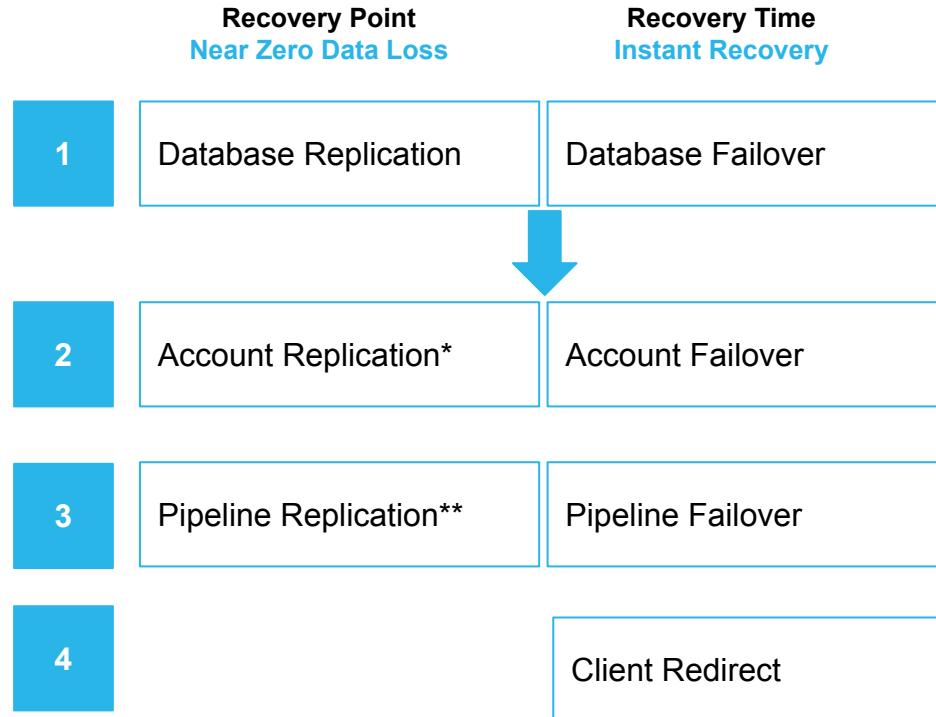
Bring your data to life in minutes by building an interactive app – all in Python

Get started today

- Install Streamlit
 - Connect to Snowflake with
Snowflake's Python connector
 - Create!

Cross-Cloud Business Continuity

Replicate more than just data



Client Redirect

Redirect client connections to the region and cloud of your choice with a single server-side command



BENEFITS

- **New Connection URL**
Can be failed over across regions and clouds
- **Supports all clients**
SnowSQL, Python, JDBC, ODBC, Go, Node.js, .NET, Snowflake UI
- **Redirects instantly**
Connections redirect within 30-45 seconds
- **Supports PrivateLink**
 - Private connections routed via customer's DNS
 - Customer creates & updates CNAME



THANK YOU



© 2023 Snowflake Inc. All Rights Reserved