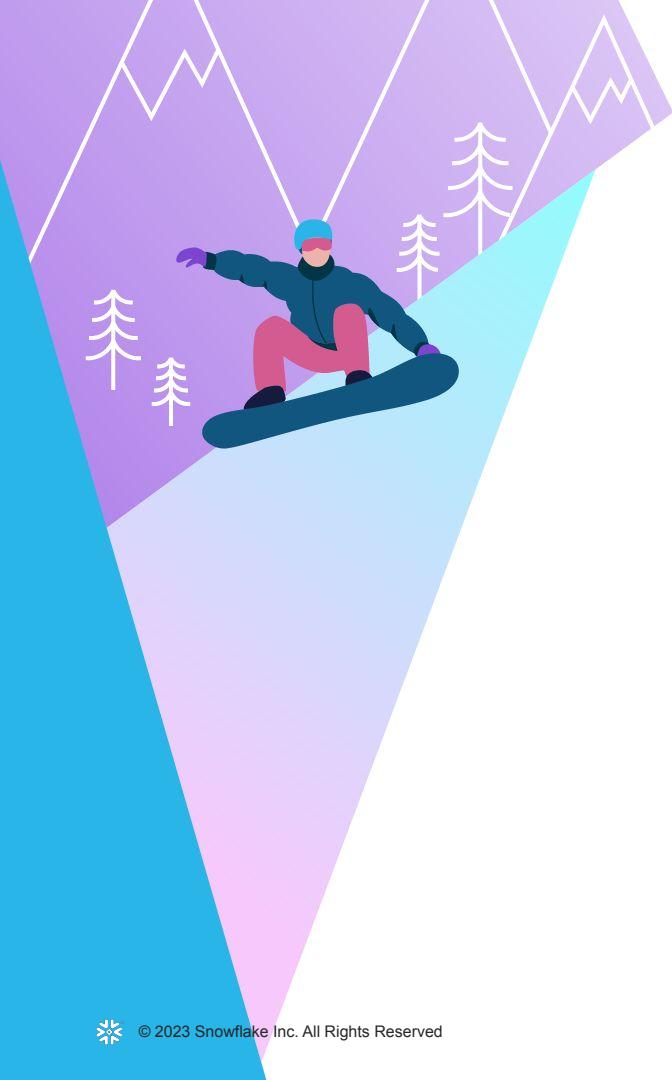




SNOWFLAKE TECH MEETUP

Piotr Pietrzkiewicz
Łukasz Leszewski

September 2023



Piotr Pietrzkiewicz

Piotr.Pietrzkiewicz@snowflake.com
EE Regional Manager
Solutions Engineering

[LinkedIn](#)



Łukasz Leszewski

Lukasz.Leszewski@snowflake.com
Senior Solutions Engineer

[LinkedIn](#)





EMEA
December 5 - 6

[BUILD]

The Dev Conference for AI & Apps

Warsaw
December 7

DIVE DEEP INTO THE WORLD OF APPS
AND GENERATIVE AI

<https://www.snowflake.com/build/>

WHY ATTEND BUILD?

- LEARN
- CODE
- GET INSPIRED

At BUILD 2023, discover and test drive app and genAI innovations from Snowflake through demos, hands-on labs, and product deep dives. Join other developers, engineers, data scientists, and startups for two full days of sessions focused on building and monetizing apps, data pipelines, and ML workflows in this new era.



START YOUR 30-DAY FREE TRIAL

Start your 30-day free Snowflake trial which includes \$400 worth of free usage

<https://signup.snowflake.com/>

The screenshot shows a sign-up form with the following fields:

- First Name*
- Last Name*
- Company Email*
- Company Name*
- Role*: A dropdown menu showing "Poland" as the selected option.
- A checkbox labeled "Yes, I do want Snowflake to send me e-mails about products, services, and events that it thinks may interest me."

Below the form, a note states: "By clicking the button below you understand that Snowflake will process your personal information in accordance with its [Privacy Notice](#)".

At the bottom, there is a large orange "CONTINUE" button and a link "or [sign in to an existing account](#)".

- Gain immediate access to the Data Cloud
- Enable your most critical data workloads
- Scale instantly, elastically, and near-infinitely across public clouds
- Try out Snowflake quickstarts:
<https://quickstarts.snowflake.com/> for example:
 - An Introduction to Tasty Bytes
 - Machine Learning with Snowpark Python: - Credit Card Approval Prediction

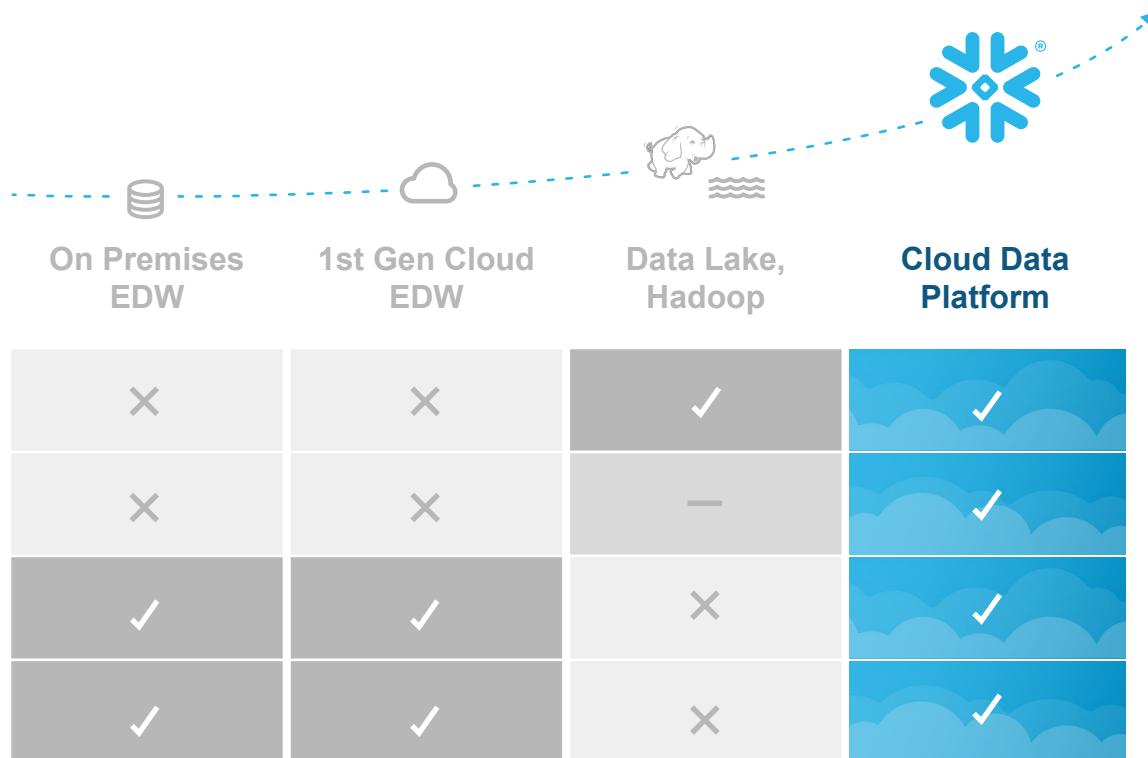


Agenda

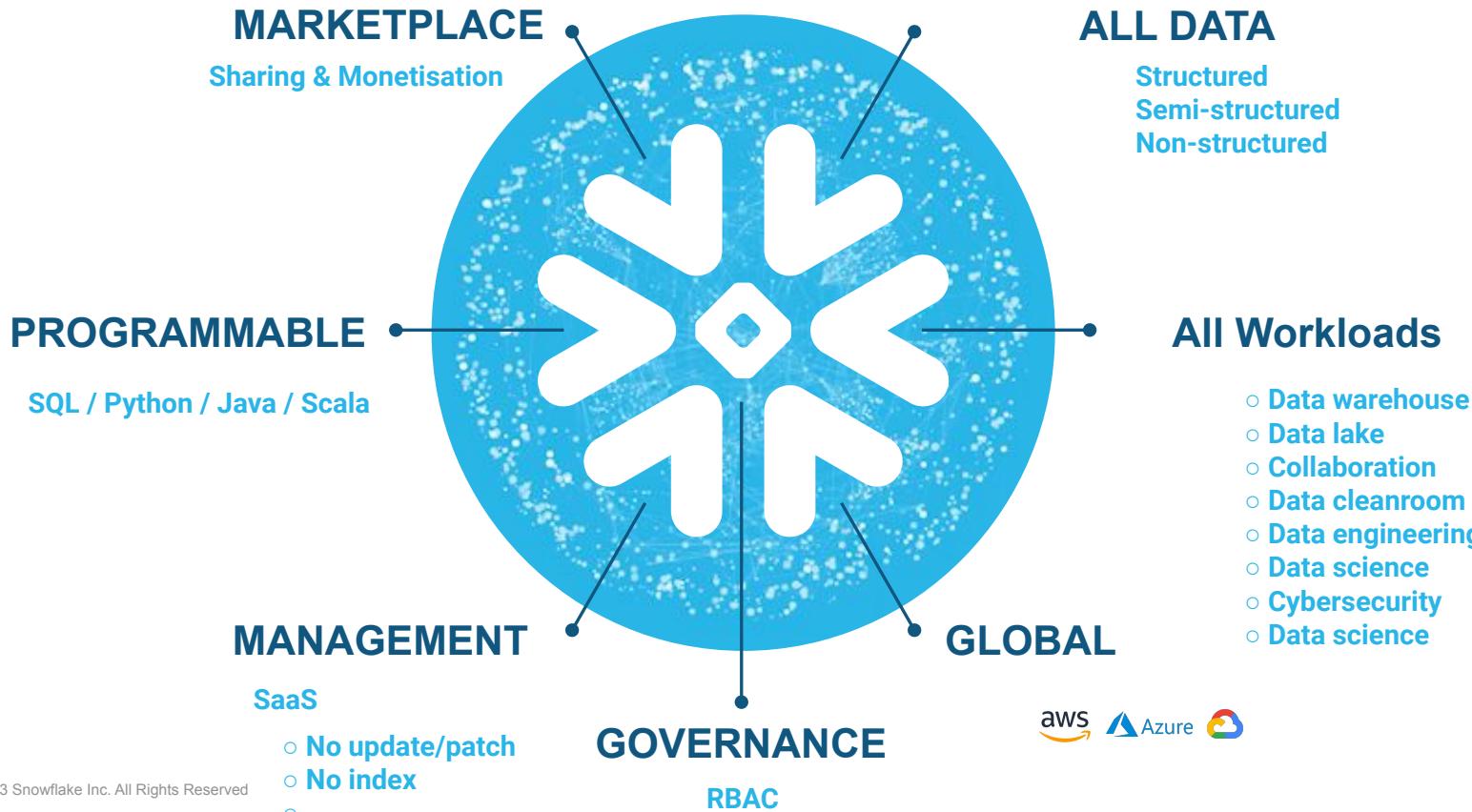
- > What is Snowflake
- > Scalability
- > Data
- > Query & Connectors
- > Snowpark
- > Demos



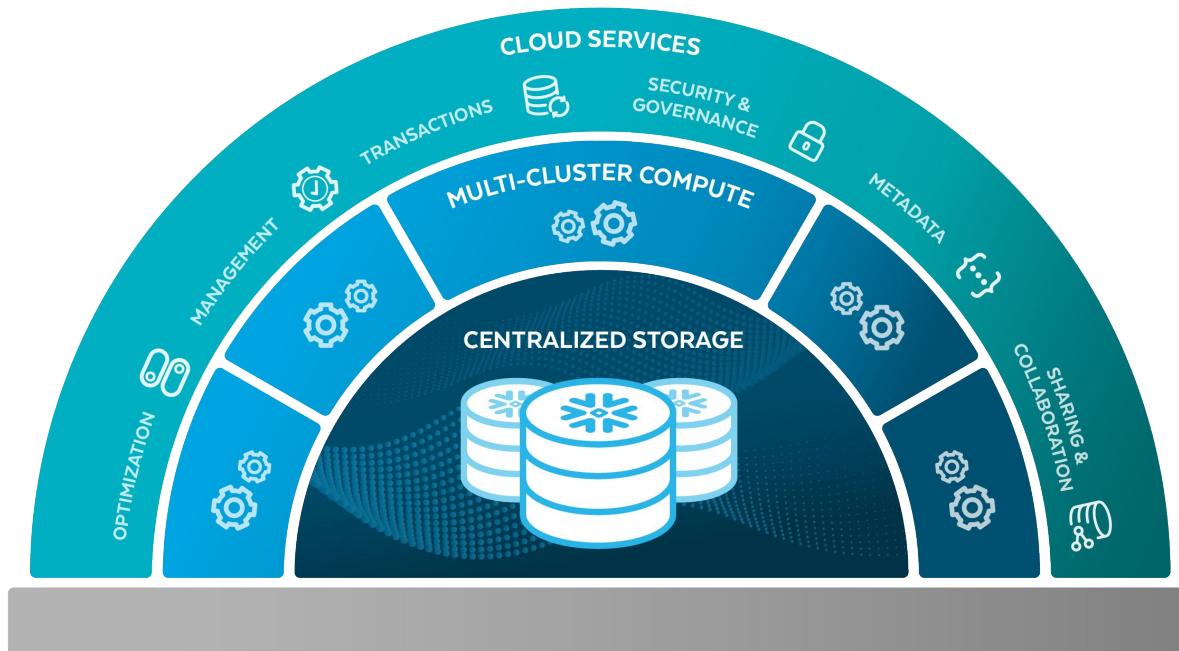
JOURNEY TO A CLOUD DATA PLATFORM



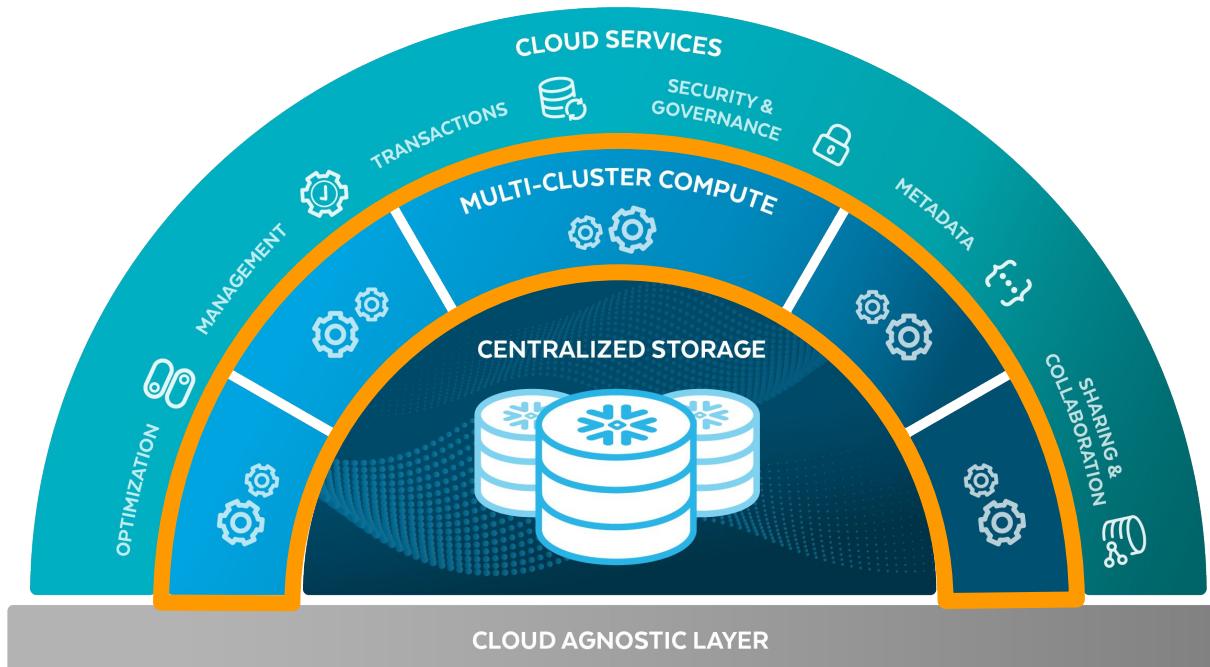
Power your data platform with Snowflake



ENABLED BY SNOWFLAKE'S ARCHITECTURE



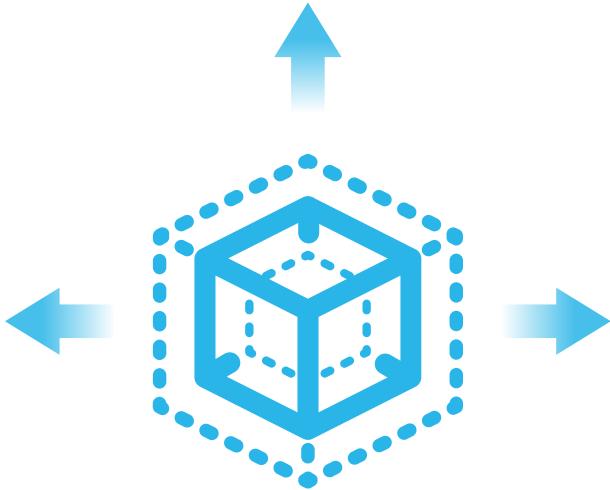
MULTI-CLUSTER COMPUTE



3 Dimensions of Scaling

ACROSS

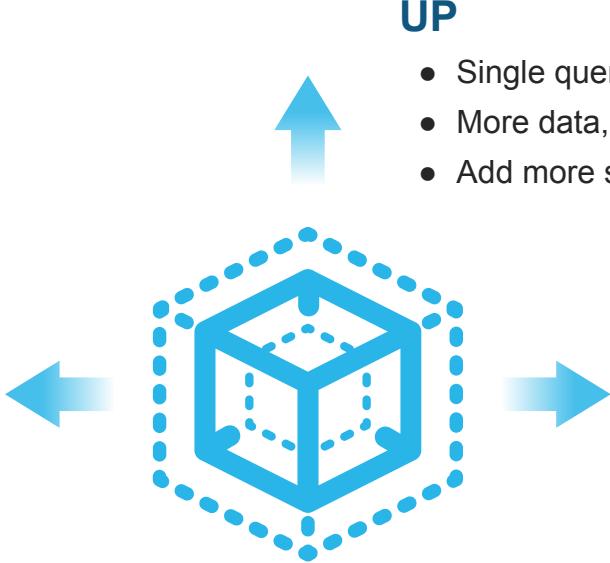
- Many competing workloads
- Resource contention
- Isolate on separate warehouses



3 Dimensions of Scaling

ACROSS

- Many competing workloads
- Resource contention
- Isolate on separate warehouses



UP

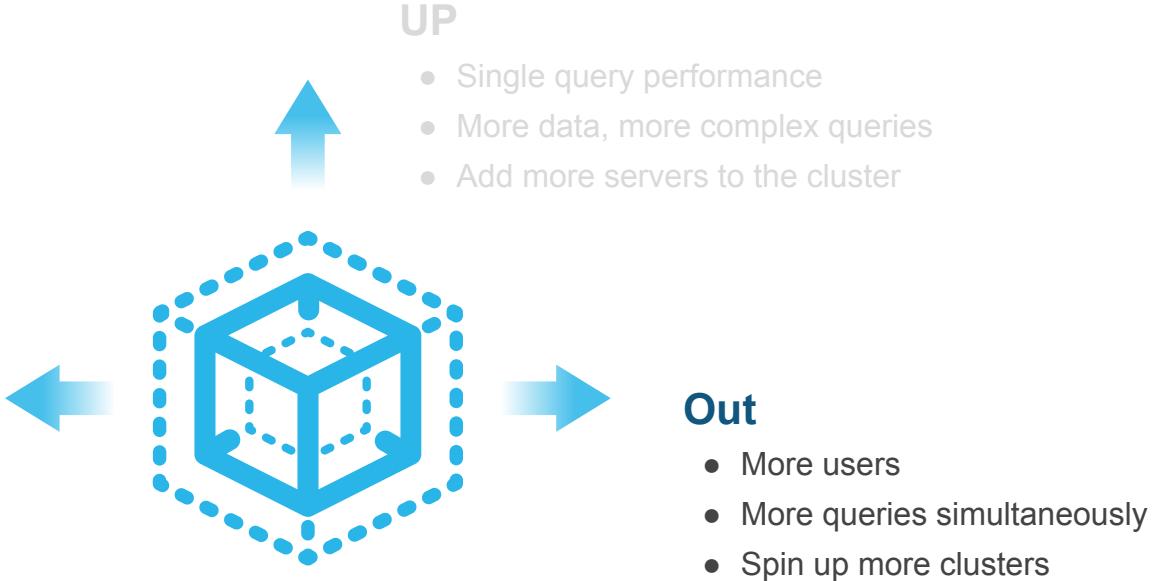
- Single query performance
- More data, more complex queries
- Add more servers to the cluster



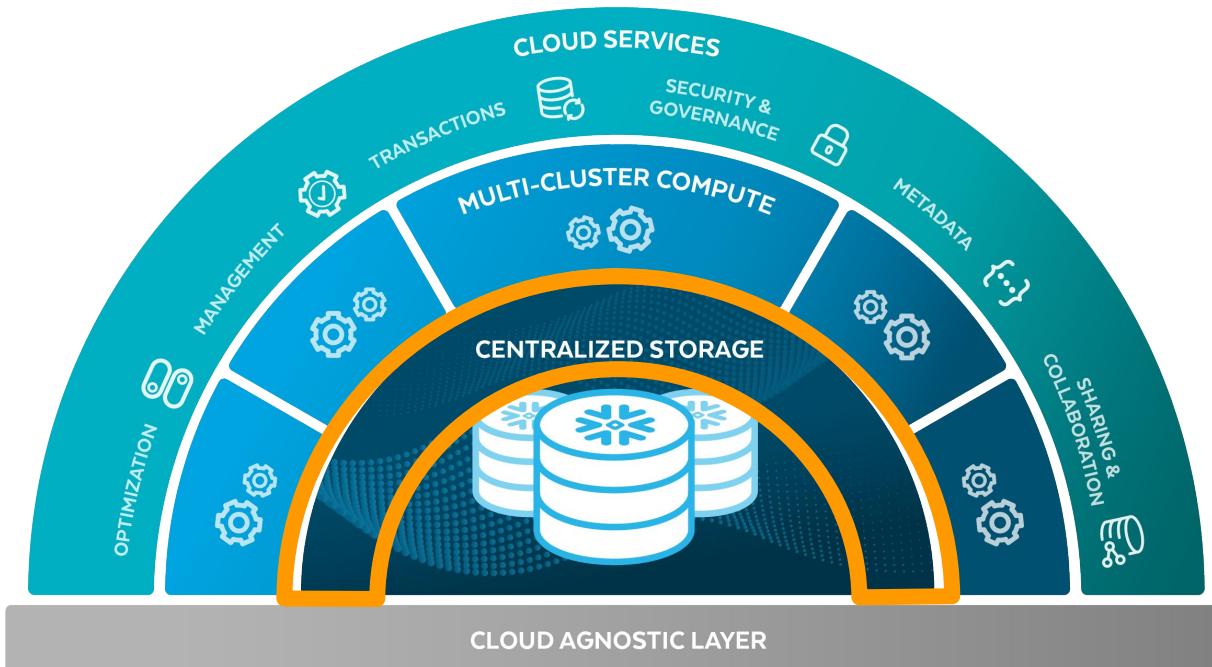
3 Dimensions of Scaling

ACROSS

- Many competing workloads
- Resource contention
- Isolate on separate warehouses



Shared data



WHAT'S A MICRO-PARTITION?



Store data in files called
Micro-partitions

Micro-partitions are small files that
store data in columnar format

Snowflake uses this Micro-partition
metadata for performance

METADATA: MIN / MAX VALUES - PARTITION ELIMINATION

Micro-Partitions

Date	Store	Cust ID	Amount
1-Jan	Ealing	12395	\$40.25
1-Jan	Ealing	93855	\$50.15
1-Jan	Ealing	29625	\$4.99
1-Jan	Ingatestone	87252	\$40.25
1-Jan	Ingatestone	54612	\$2.75

1

2

3

2-Jan	Ealing	19286	\$19.00
2-Jan	Windsor	98262	\$49.99
2-Jan	Windsor	12662	\$17.50

1

2

3

Metadata Entries

	Date	Store	Cust ID	Amount
Min	1-Jan	Ealing	12395	\$2.75
Max	1-Jan	Ingatestone	93855	\$50.15

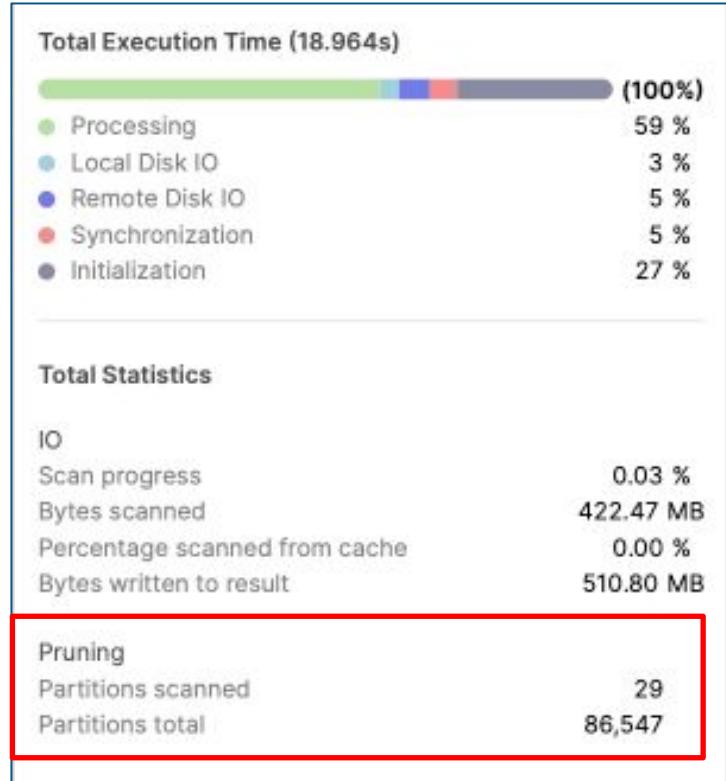
Min	2-Jan	Ealing	12662	\$17.50
Max	2-Jan	Windsor	98262	\$49.99

Min	3-Jan	Ealing	19285	\$7.50
Max	3-Jan	Windsor	86542	\$14.99

Where DATE = 1-JAN or AMOUNT > \$40.00



THE BENEFIT?



```
select ss_date,  
       store,  
       sum(amount)  
  from sales s  
 where ss_date between '01-FEB-19'  
       and '28-FEB-19'  
 group by ss_date,  
         store
```

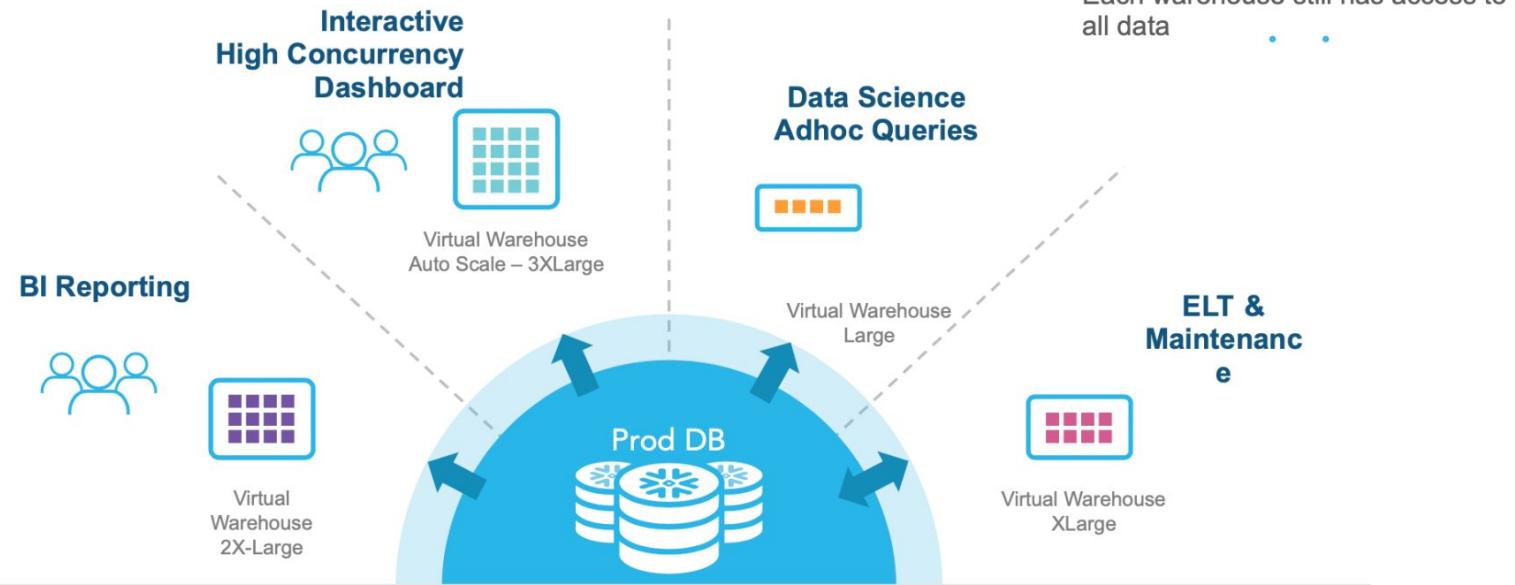
Results in seconds against billions of rows.



THE UNIT OF WORKLOAD MANAGEMENT: VIRTUAL WAREHOUSE

A unit of workload management can be based on type of workload:

- ELT
- Dashboards
- Data Science
- BI Reporting
- and more



Keep relatively homogeneous workload on each warehouse

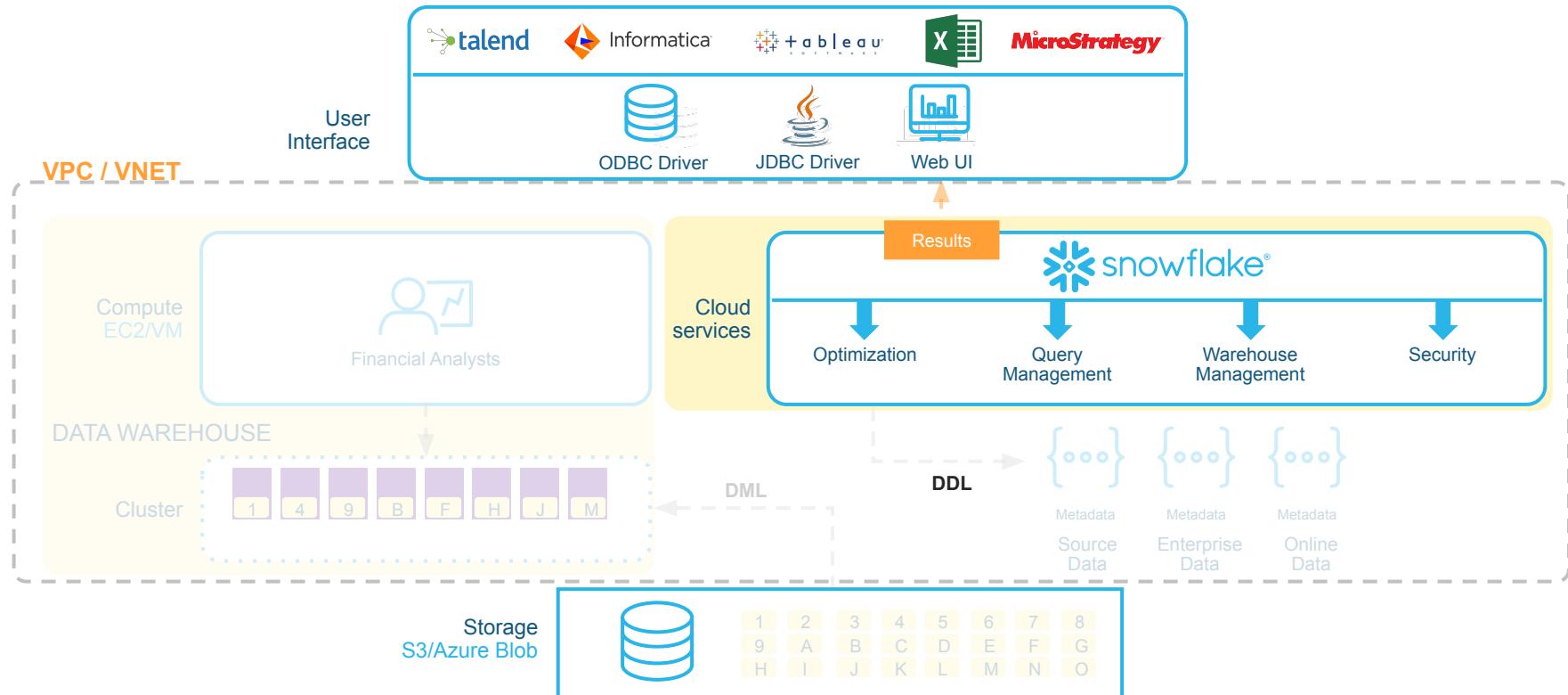
- Amount of data accessed by each query

- Complexity of each query

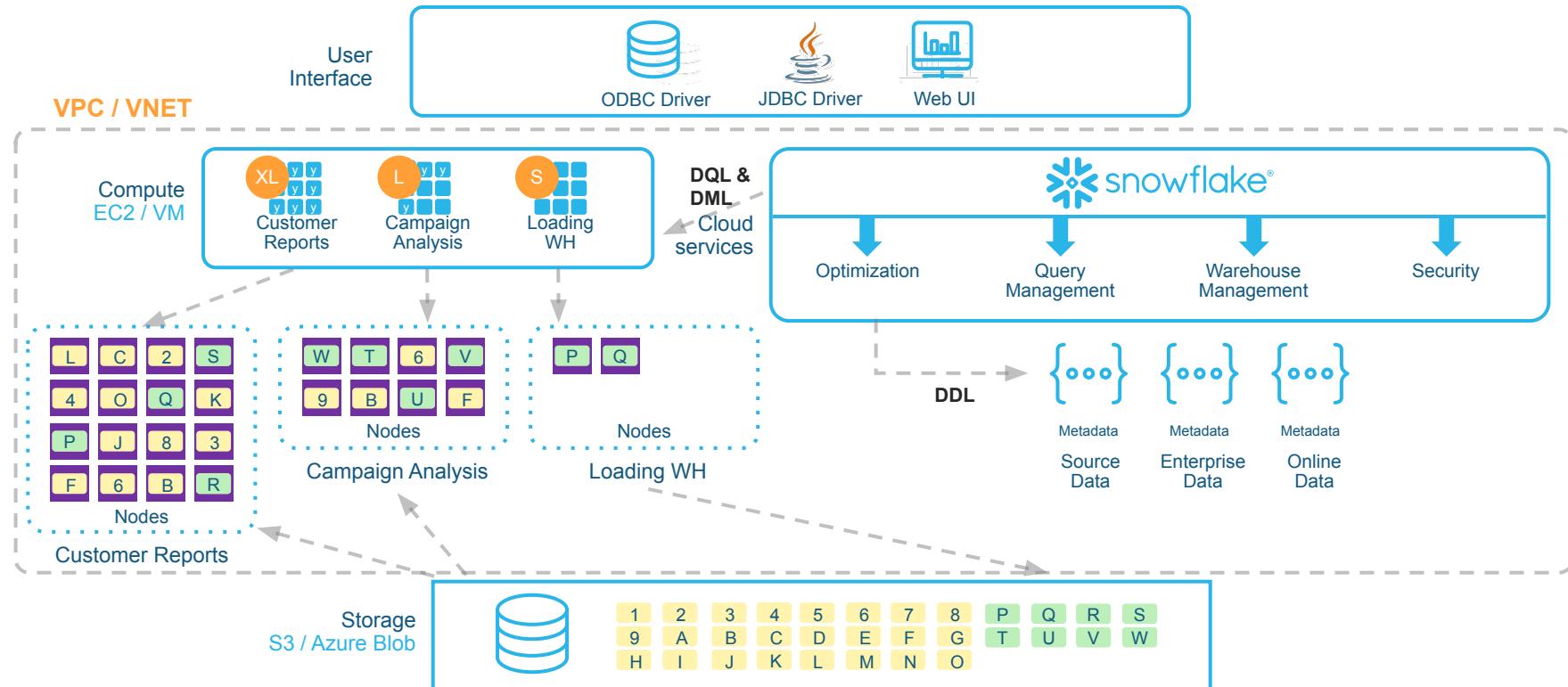
Each warehouse still has access to all data

• •

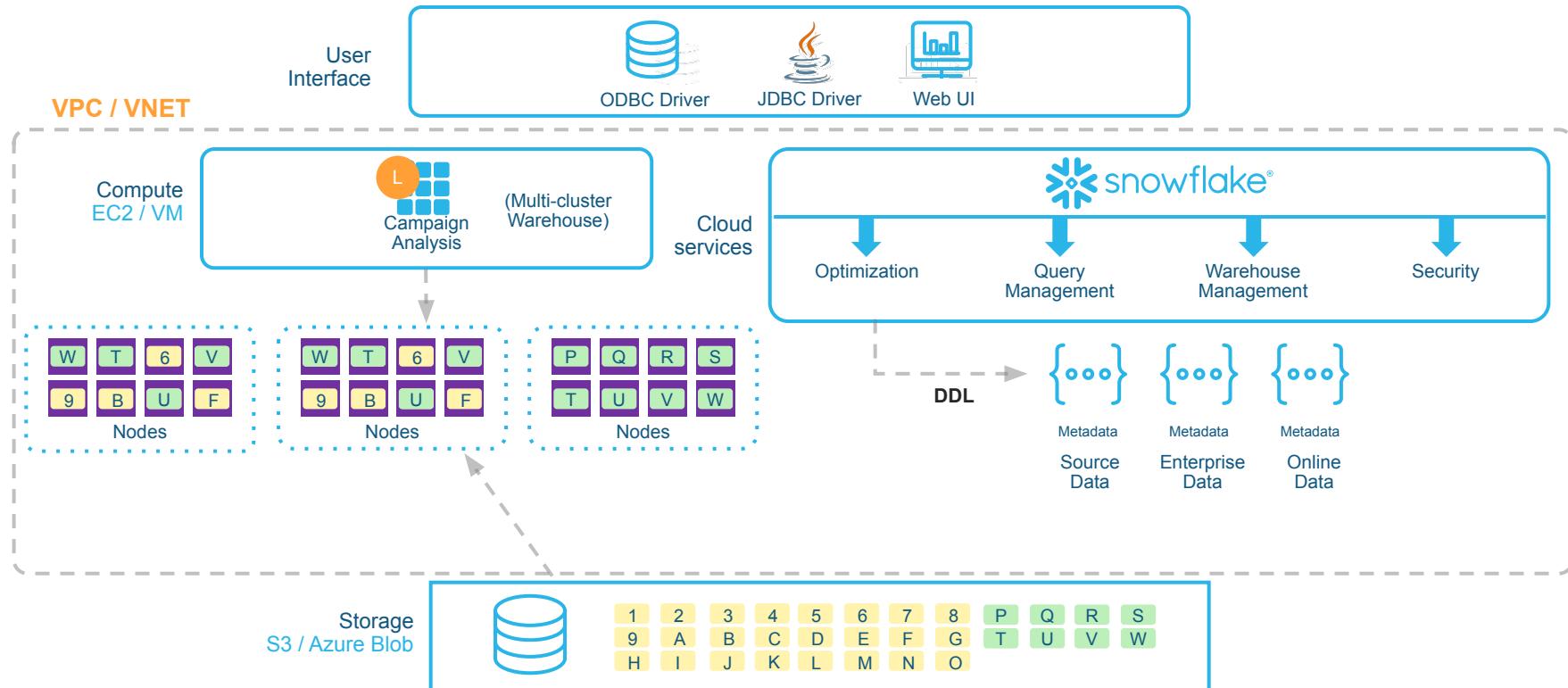
SNOWFLAKE ARCHITECTURE



SNOWFLAKE ARCHITECTURE



SNOWFLAKE ARCHITECTURE



CONNECTIVITY CLIENTS

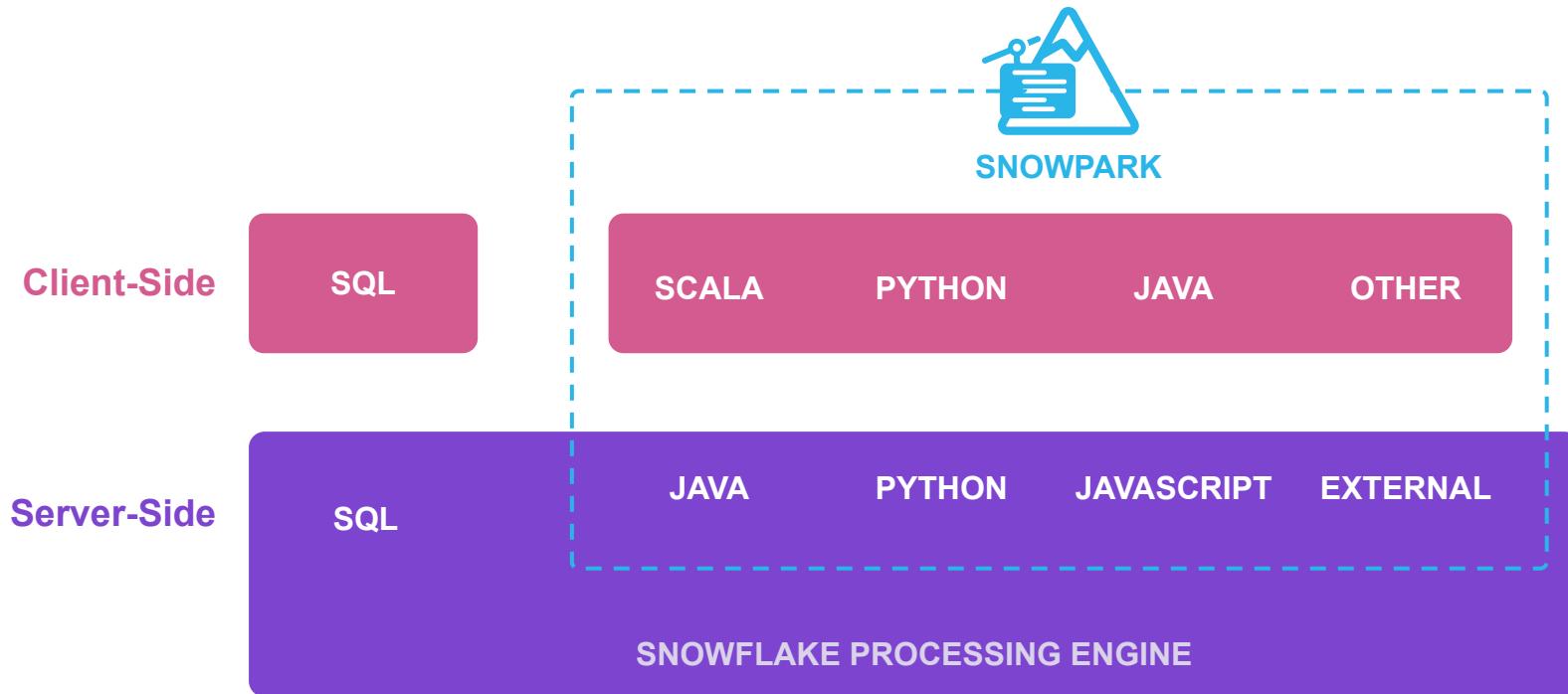
- JDBC driver
- ODBC driver
- Python connector
- Spark connector
- R dplyr
- NodeJS driver
- REST
- Golang driver
- .Net driver
- PHP PDO
- C/C++ API
- Ingest (Snowpipe) SDK
- Django



User guide for all the connectors and drivers:

- <https://docs.snowflake.net/manuals/user-guide-connecting.html>

Code the Same Way, Execute Faster with Snowpark





> Develop in IDE of Choice

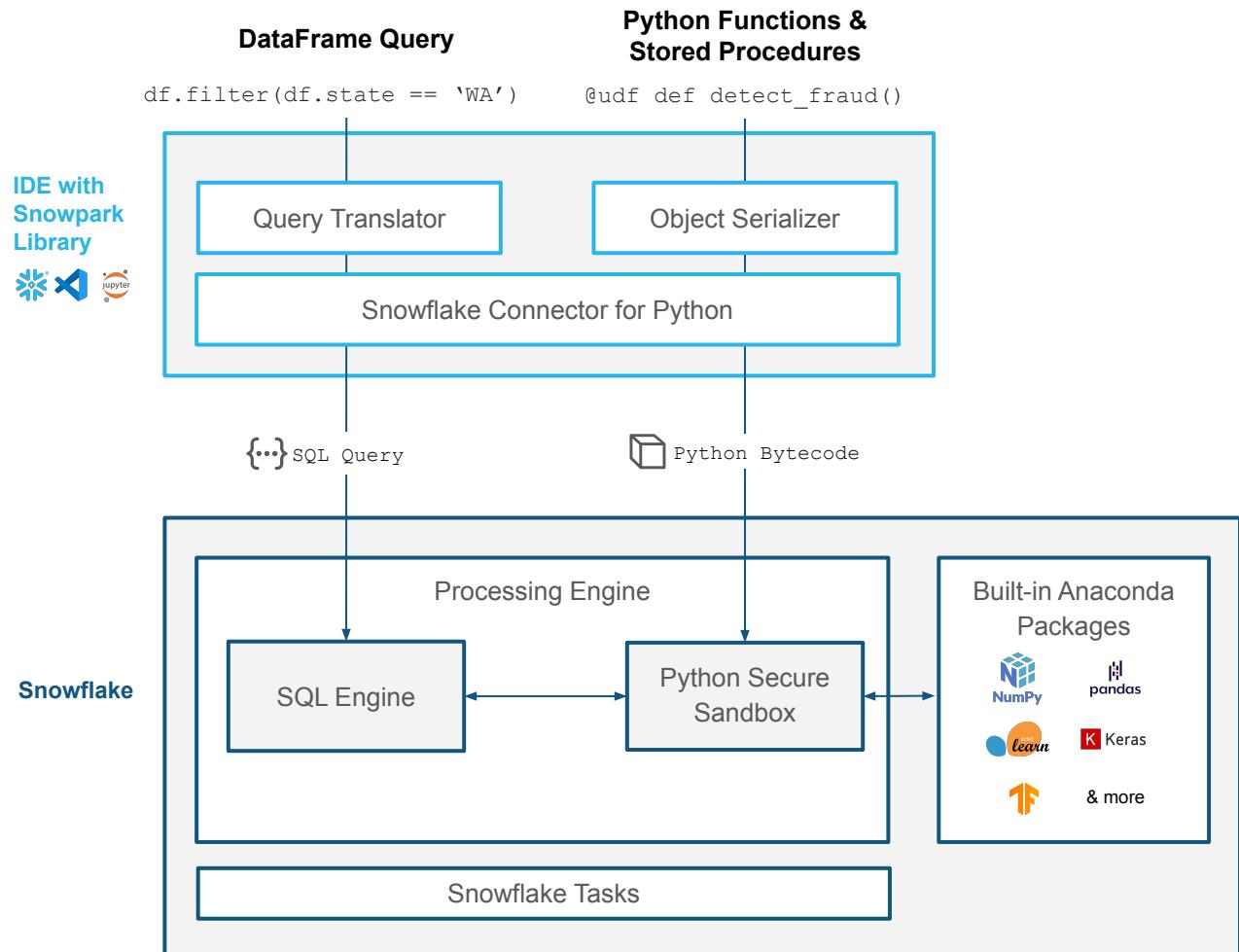
Use Python worksheets, VSCode, Jupyter or any client with a python kernel

> Process with security & scalability

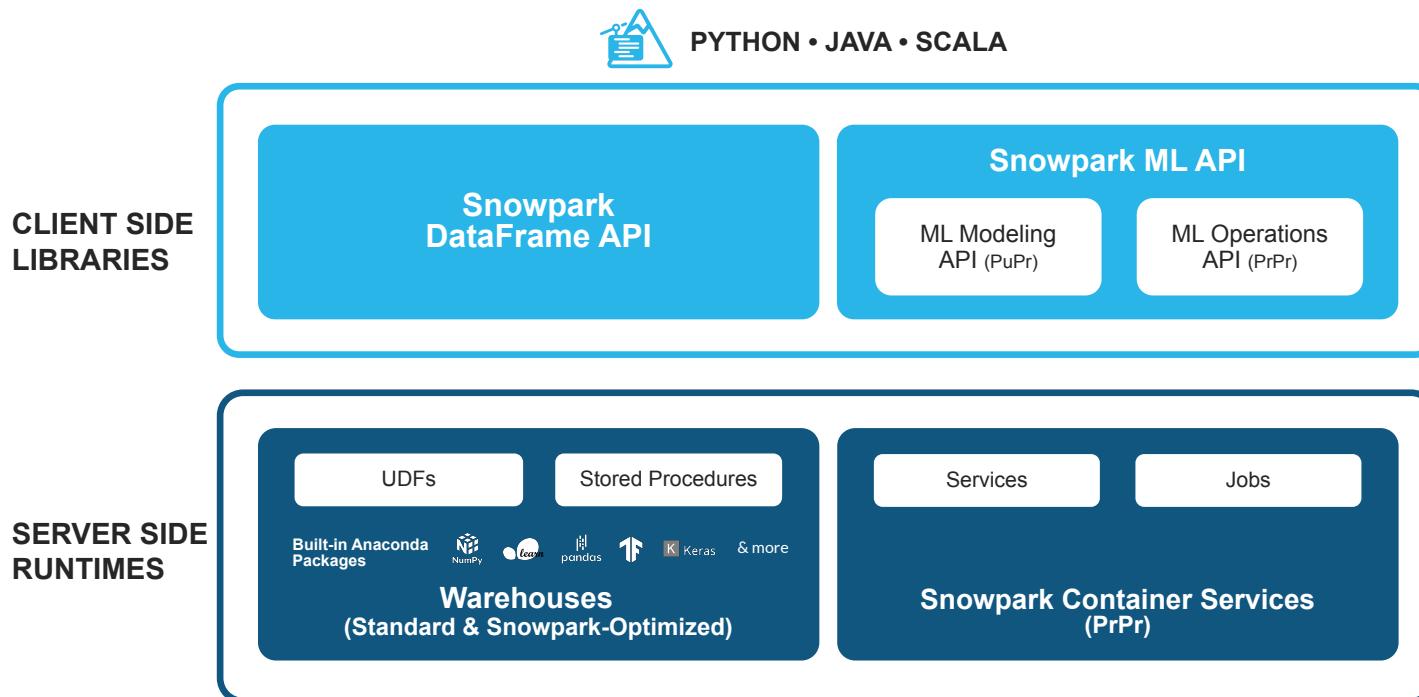
Bring processing to governed data for DataFrame operations and custom Python/Java code

> Automate code execution

Orchestrate your code with DAG representations using SProcs with Snowflake Tasks



Snowpark: Securely Deploy & Process Non-SQL Code



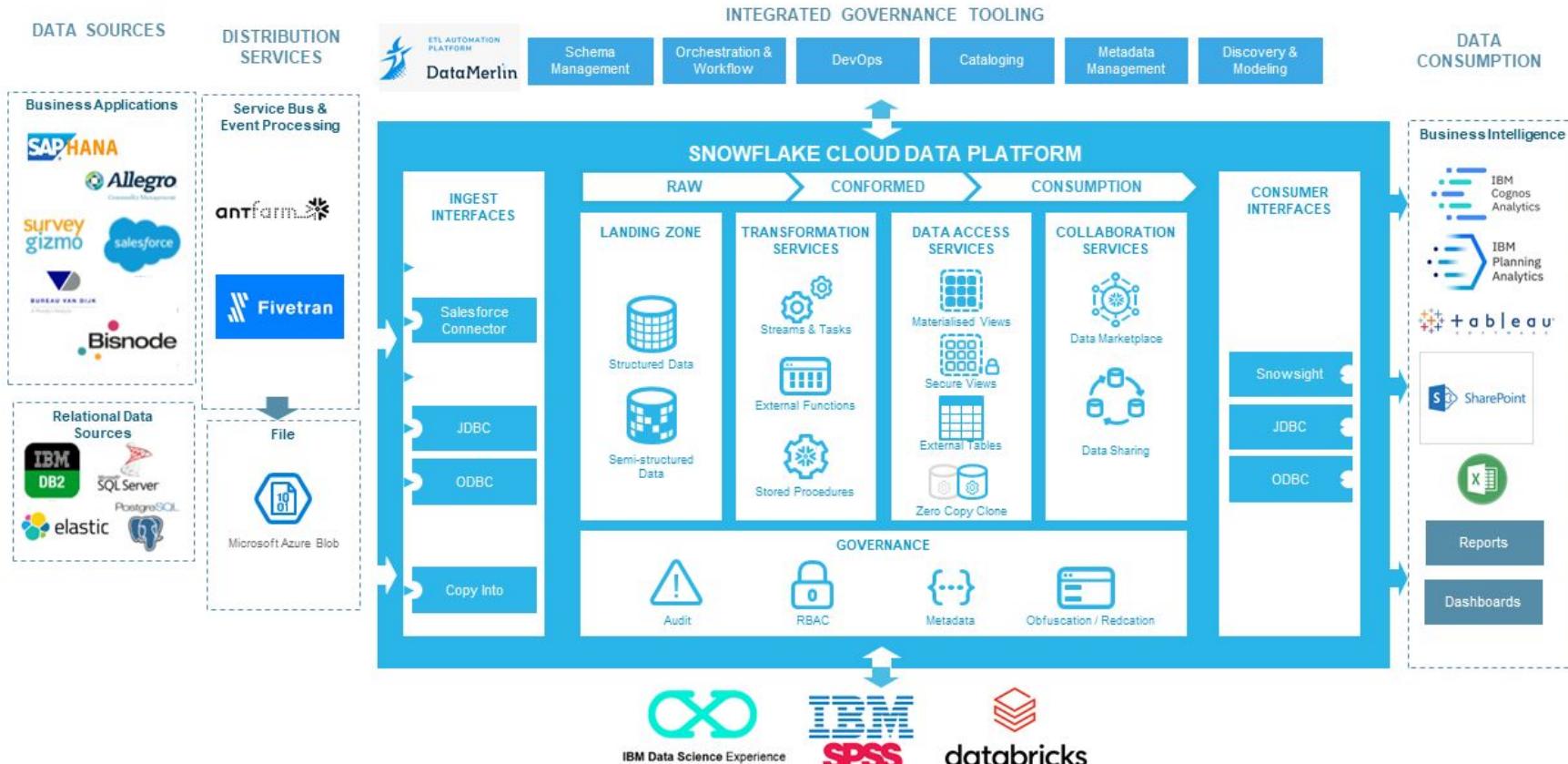
Machine Learning Libraries

Snowflake with the [Anaconda repository](#) includes the most common libraries used for Machine Learning that will be executed in a secure environment:

- Theano
- Keras
- Tensorflow
- NLTK
- Numpy
- Pandas
- PyTorch
- Scikit-learn
- sciPy
- Prophet
- MXNet
- XGBoost
- Statsmodels
- Dask
- Transformers



REFERENCE LOGICAL ARCHITECTURE



DEMO



© 2023 Snowflake Inc. All Rights Reserved

WHO IS TASTY BYTES?

ABOUT US: Global food truck network, localized menu options, 15 countries, 30 major cities, and 15 core brands.

OUR MISSION

We serve to give people unique food options with high quality items in a safe, convenient and cost effective way. We ensure that the ingredients used are of the highest quality from mostly local food vendors to make sure our success has a positive impact on community partners.



OUR VISION

To become the largest food truck network in the world by 2027 that has sustainable profitability with a zero carbon footprint future that our team, customers, and communities are proud of supporting.

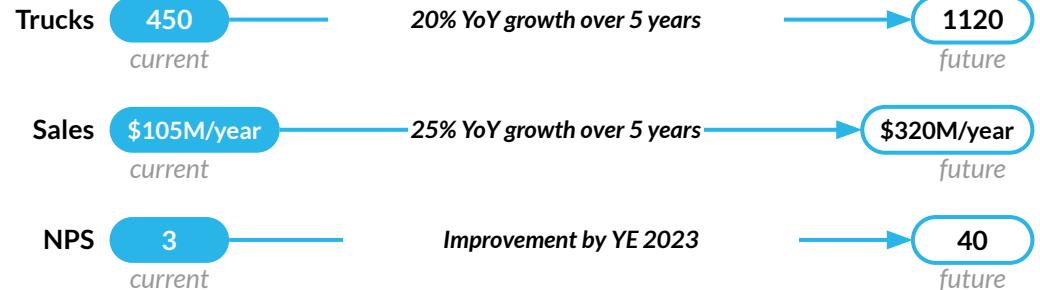


LOCATIONS SERVED

- USA: San Mateo, Denver, Seattle, Boston, New York City
- Canada: Toronto, Vancouver, Montreal
- United Kingdom: London, Manchester
- France: Paris, Nice
- Poland: Warsaw, Krakow
- India: Mumbai, Delhi
- Japan: Tokyo
- South Korea: Seoul
- Australia: Sydney, Melbourne



CURRENT STATE & FUTURE GOALS



Demo overview

Tasty Bytes is one of the largest food truck networks in the world with localized menu options spread across 30 major cities in 15 countries. **Tasty Bytes is aiming to achieve 25% YoY sales growth over 5 years.**

As Tasty Bytes Data Scientists, we have been asked to support this goal by helping our food truck drivers more intelligently pick where to park for shifts.

**We want to direct our trucks to locations that are expected to have the highest sales on a given shift.
This will maximize our daily revenue across our fleet of trucks.**

To provide this insight, we will use historical shift sales at each location to build a model. This data has been made available to us in Snowflake.

Our model will provide the predicted sales at each location for the upcoming shift.

This is an introduction to Snowpark for Snowflake. I will use Snowpark to:

- Explore the data
- Perform feature engineering
- Train a model
- Deploy the model in Snowflake

Why Snowpark?

- No copies or movement of data
- Maintain governance
- Leverage Snowflake scalable compute
- ...and more!

Snowpark 101 for Data Science

The screenshot shows a video player interface with the following details:

- Video title: "Tasty Bytes - Snowpark 101 for Data Science"
- Duration: "41 mins remaining"
- Step 1: "Overview" (highlighted)
- Step 2: "Setting up the Data in Snowflake"
- Step 3: "Data Science with Snowpark"
- Step 4: "Application Prototype with Streamlit"
- Step 5: "Clean Up"
- Step 6: "Conclusion and Next Steps"

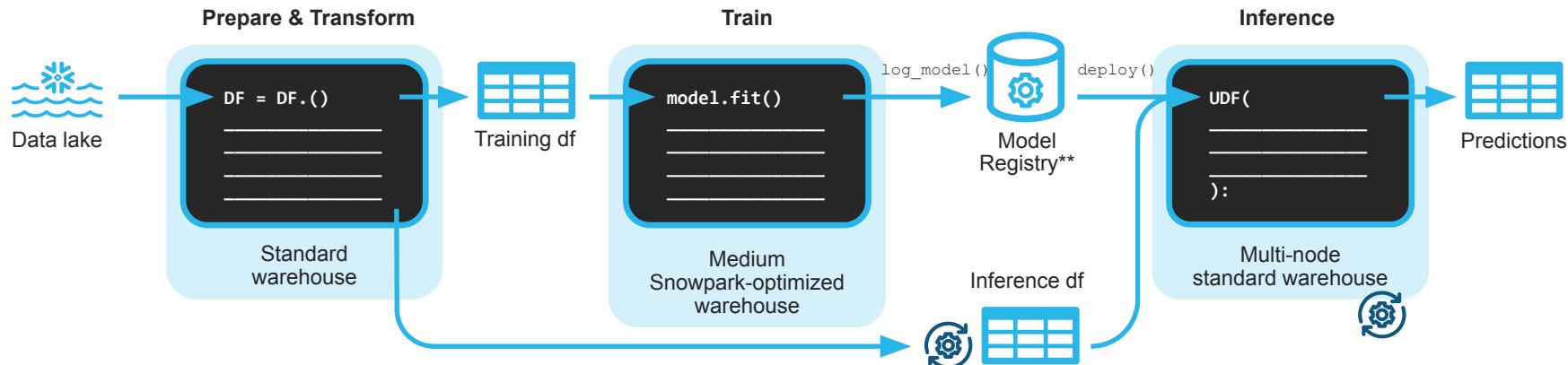
The video player also displays a thumbnail for the first step, which features a blue-toned image of two people smiling and the text "frostbyte." and "TASTY BYTES". A descriptive text below the thumbnail reads: "In this Snowpark 101 for Data Science Quickstart guide, you will be help the fictitious food truck company, Tasty Bytes, to increase sales by training a model to provide location recommendations to truck drivers. You will use Snowpark for Python to prepare data, train a model, and deploy the model. Once deployed, you will create an application prototype using Streamlit to demonstrate how truck drivers could use the model to find the best location to park for an upcoming shift."

[Link to the quickstart](#)

frostbyte.

Example End-to-End Machine Learning

Effortless, scalable and secure processing without data movement across compute environments



Development

`snowflake.ml.modeling.preprocessing*`

`snowflake.ml.modeling*`

Deployment

`snowflake.ml.registry**`

Orchestration

`snowflake.core.task**`





> Develop in IDE of Choice

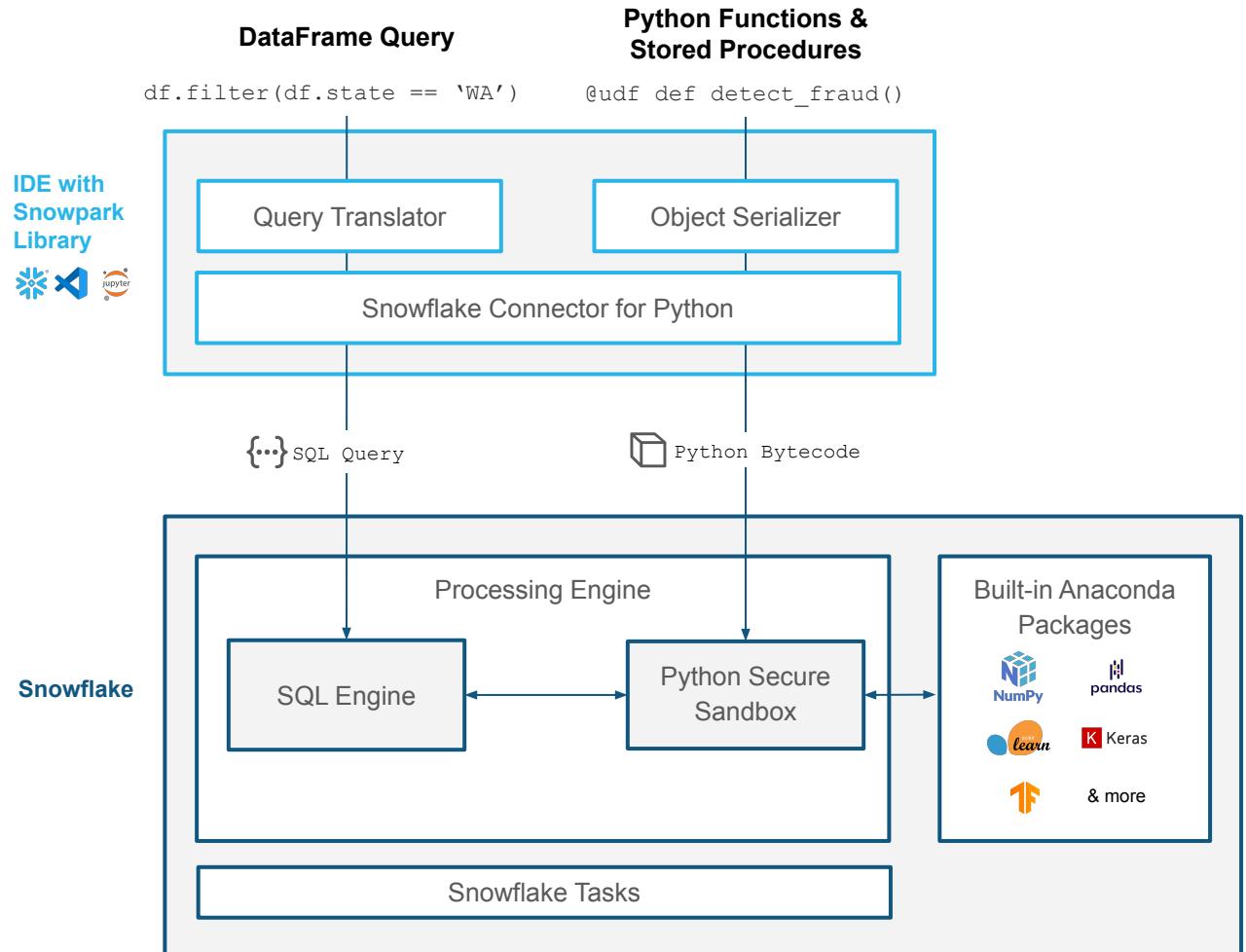
Use Python worksheets, VSCode, Jupyter or any client with a python kernel

> Process with security & scalability

Bring processing to governed data for DataFrame operations and custom Python/Java code

> Automate code execution

Orchestrate your code with DAG representations using SProcs with Snowflake Tasks



QUESTIONS



THANK YOU



© 2023 Snowflake Inc. All Rights Reserved