



# SNOWPARK PRICE PERFORMANCE: CUSTOMER RESULTS



Customers see an average of **4.6x faster performance** and **35% cost savings** with Snowpark over managed Spark\*

## SNOWPARK OVERVIEW

Snowpark is the set of libraries and code execution environments that run Python and other programming languages next to your data in Snowflake. You can use Snowpark to build data pipelines, ML models, apps and other data processing tasks.

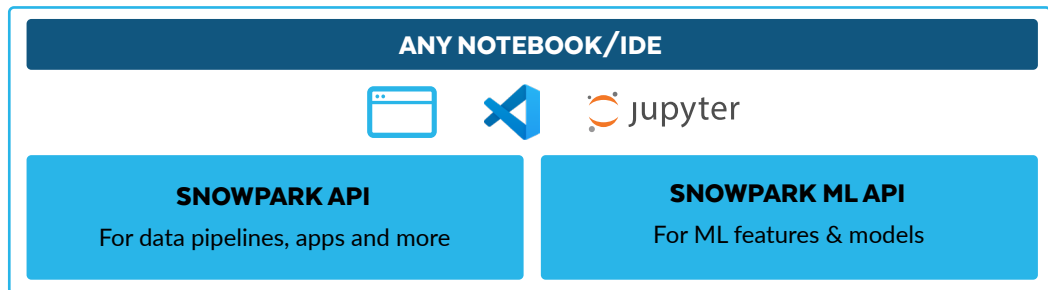
On the client side, Snowpark consists of libraries including the Snowpark DataFrame API and the native Snowpark machine learning (ML) API with components for model development and operations.

On the server side, all Python and other programming code is automatically pushed down for processing in the Snowflake engine's code execution environments. Code execution environments include Python, Java and Scala runtimes in virtual warehouses or any other runtime brought in through Snowpark Container Services (public preview). Using warehouse compute, developers can leverage user-defined functions (UDFs) and stored procedures (sprocs) to bring in and run custom logic. Snowpark Container Services enables customized runtimes and hosting of containers for custom code within Snowflake on a variety of hardware including GPUs.

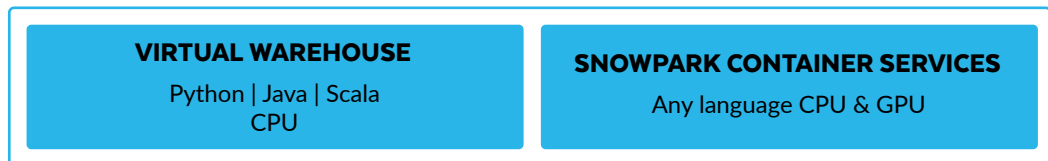
Thousands of organizations are accelerating development and performance of their data engineering and ML workloads with Snowpark for Python. It has continued to see strong growth since its general availability in November 2022, and as of January 2024, over 40% of Snowflake customers use Snowpark on a weekly basis. In January 2024, Snowflake customers ran approximately 38 million queries in Snowpark on average each day (based on January 2024 data).

**40% of Snowflake customers are using Snowpark, running approximately 38 million daily queries**

### CODE DEVELOPMENT & DEPLOYMENT FROM CLIENT-SIDE LIBRARIES



### CODE EXECUTION ENVIRONMENTS IN SNOWFLAKE'S ELASTIC ENGINE



## \*METHODODOLOGY

- SAMPLE SIZE:** 50+ real customer POCs and production results
- INDUSTRIES:** Financial Services, Retail & Consumer Goods, Healthcare & Life Sciences, Manufacturing & Industrial, Media & Entertainment, Technology, Telecom, Consumer Services & Hospitality, Public Sector
- LOCATIONS:** Global
- WORKLOADS USED:** Data engineering, AI/ML

This report summarizes a collection of customer production use cases and proof-of-concept exercises comparing the speed and cost for Snowpark versus managed Spark services between November 2022 and January 2024. **All findings in this report summarize actual customer outcomes with real data and do not represent fabricated datasets used for benchmarks.**

## WORKLOADS EVALUATED:

Based on customer proof-of-concepts and production workload results, we see that Snowpark is primarily used for data engineering and AI/ML workloads through Python. While Java, Scala and Python are all generally available, more than 75% of the sample set in this analysis are consuming Snowpark with Python.

**75% of the sample set in this analysis are consuming Snowpark with Python**

Data engineers are leveraging Snowpark for data transformations, custom business logic and pipelines feeding ML models, reporting dashboards and applications. This includes both ETL, where data is transformed from a data lake and written into Snowflake, and ELT, where Snowflake is the data source and target.

Data scientists and ML engineers have been using Snowpark to perform feature engineering, model training and model inference with large datasets.

## RESULTS:

Snowpark users are seeing significant speed and cost improvements after migrating from Spark.

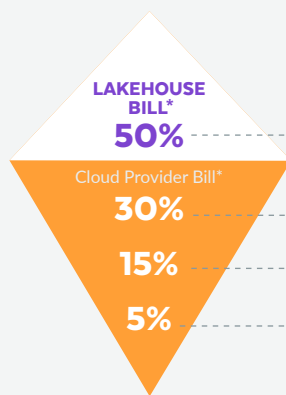
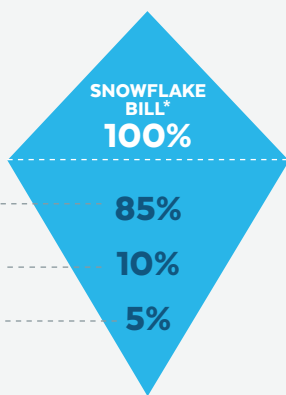
Cost savings results do not reflect the full TCO benefits of Snowpark because they omit the indirect operating costs that are prevalent for Spark solutions. Query optimizations, security and governance often are not pre-configured out of the box and require specialized skill sets to manually tune. This can decrease productivity and lead to long setup and onboarding timelines, queries not performing well and a greater risk of breaches from data not secured properly.

Additionally, TCO can be more difficult to track overall with managed Spark solutions because there are separate bills for cloud infrastructure and Spark-based compute.

**With Snowpark, customers saw an average of 4.6x faster performance and 35% lower cost compared to Spark solutions.**

## UNDERSTANDING THE FULL COST STORY: SEARCH FOR HIDDEN COSTS

Snowflake Compute ..... 85%  
 Optional Features: Snowpipe, Materialized Views, Auto Cluster, Query Acceleration, Serverless Tasks, etc. .... 10%  
 Storage ..... 5%



Spark Compute Costs  
 Cloud Compute (EC2, Azure Virtual Machines, Google Compute)  
 Cloud Storage & Cost of Retrieval (S3, Blob, API Requests)  
 Cloud Data Transfer (network, spark connector compute, etc.)

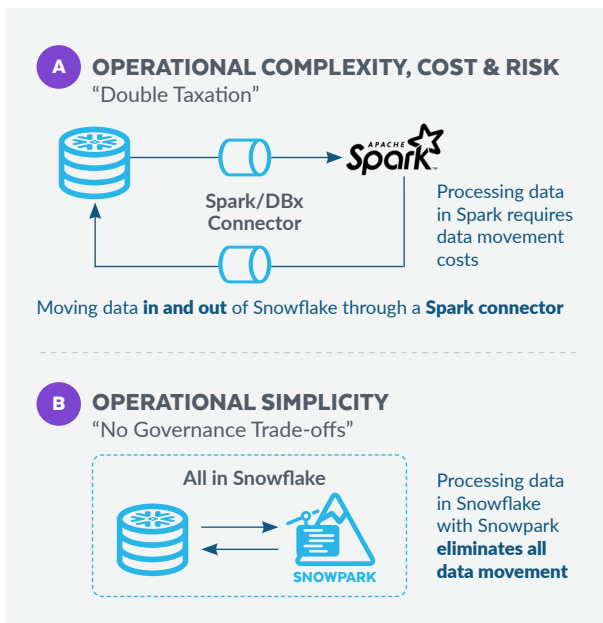
\*%s of actual split may vary depending on workloads and customer environment.

## SNOWPARK IS FASTER THAN SPARK – HERE’S WHY

Spark and Snowpark are both used for data processing at scale – and Snowpark customers are seeing significant price and performance benefits over managed and open source Spark solutions. Why? It all boils down to Snowflake’s engine and the elimination of data movement.

Snowflake’s distributed engine features logically integrated but physically separated storage and compute. It was built using a multi-clustered, shared data architecture that plans and optimizes the execution of concurrent workloads. SQL developers were the first to benefit from this engine, which comes with many built-in optimizations such as auto-clustering and micro-partitioning. Snowpark extends Snowflake’s engine beyond SQL to include Python, Java and Scala developers.

Another significant source of performance and cost gains comes from elimination of external data processing engines for pipelines built in multiple languages, which often requires converting, egressing, scanning and loading data from its source. With Snowpark, data does not need to be moved across clusters for different languages and petabytes of data can be processed using the developer’s preferred language on one platform.



## CUSTOMER STORIES

Snowpark helps customers across industries improve performance while optimizing costs.



### DATA TRANSFORMATIONS: OPENSTORE

Migrated from EMR to Snowpark for data pipelines feeding internal functions.

RESULTS	<b>100%</b> Achieved 100% parity of the prior PySpark workloads.	<b>80%</b> 80% reduction in engineering maintenance hours.	<b>87%</b> Openstore decreased E2E pipeline runtime by 87%.	<b>20X</b> improved ease of scaling and ongoing reliability after switching from PySpark to Snowpark while operating at more than 20x the data volume.
---------	---	---	--	---

### ML MODELING: EDF ENERGY

Migrated from Spark to Snowpark for ML development and deployment.

RESULTS	 Deployed models faster. Reduced time-to-deploy from months to days.	 Released customer-facing products in weeks instead of months.	<b>3-4X</b> Product output has increased by 3-4x.
---------	---	---	--

**“The speed our team can work at has increased significantly. Before, we could support maybe six new data products over the course of a year—I think we could triple or quadruple that now.”**

— DATA SCIENCE LEAD, EDF

### CUSTOM BUSINESS LOGIC: INTERCONTINENTAL EXCHANGE

Migrated from EMR to Snowpark for data pipelines for regulatory reporting

RESULTS	<b>80%</b> Up to 80% time reduction for ad hoc queries.	 Faster data security and compliance reporting.	<b>360°</b> 360-degree view of all business and operational metrics.
---------	--	--	---

**“Snowpark can really help optimize costs while improving performance”**

— DATA, ANALYTICS & GOVERNANCE HEAD, ICE

Snowflake libraries and runtimes makes it easy for all Python data scientists and data engineers to build in Snowflake.

Ready to learn more? Check out [www.snowflake.com/snowpark](http://www.snowflake.com/snowpark).

#### ABOUT SNOWFLAKE

Snowflake enables every organization to mobilize their data with Snowflake's Data Cloud. Customers use the Data Cloud to unite siloed data, discover and securely share data, and execute diverse artificial intelligence (AI) / machine learning (ML) and analytic workloads. Wherever data or users live, Snowflake delivers a single data experience that spans multiple clouds and geographies. Thousands of customers across many industries, including 691 of the 2023 Forbes Global 2000 (G2K) as of January 31, 2024, use the Snowflake Data Cloud to power their businesses.

Learn more at [snowflake.com](http://snowflake.com)

