snowflake®

The Little Book of Big Success with Snowflake

DATA LAKES

TABLE OF CONTENTS

USING SNOWFLAKE CLOUD DATA PLATFORM AS A DATA LAKE FOR FASTER, BETTER ANALYTICS

Legacy data lakes typically have significant limitations, including poor analytical capabilities, multiple copies of data, excessive cost and complexity, and lack of built-in security or governance. These limitations vanish with Snowflake Cloud Data Platform because it brings together in a single platform the best storage and analytical attributes of both a data lake and a data warehouse.

With Snowflake, you get a consolidated data store for all your structured and semi-structured data, unlimited scale and speed, simplicity, cost-effectiveness, and built-in security and governance. Snowflake can optionally also complement an existing data lake for better, faster analytics and data transformations.

REIMAGINING DATA LAKES WITH SNOWFLAKE

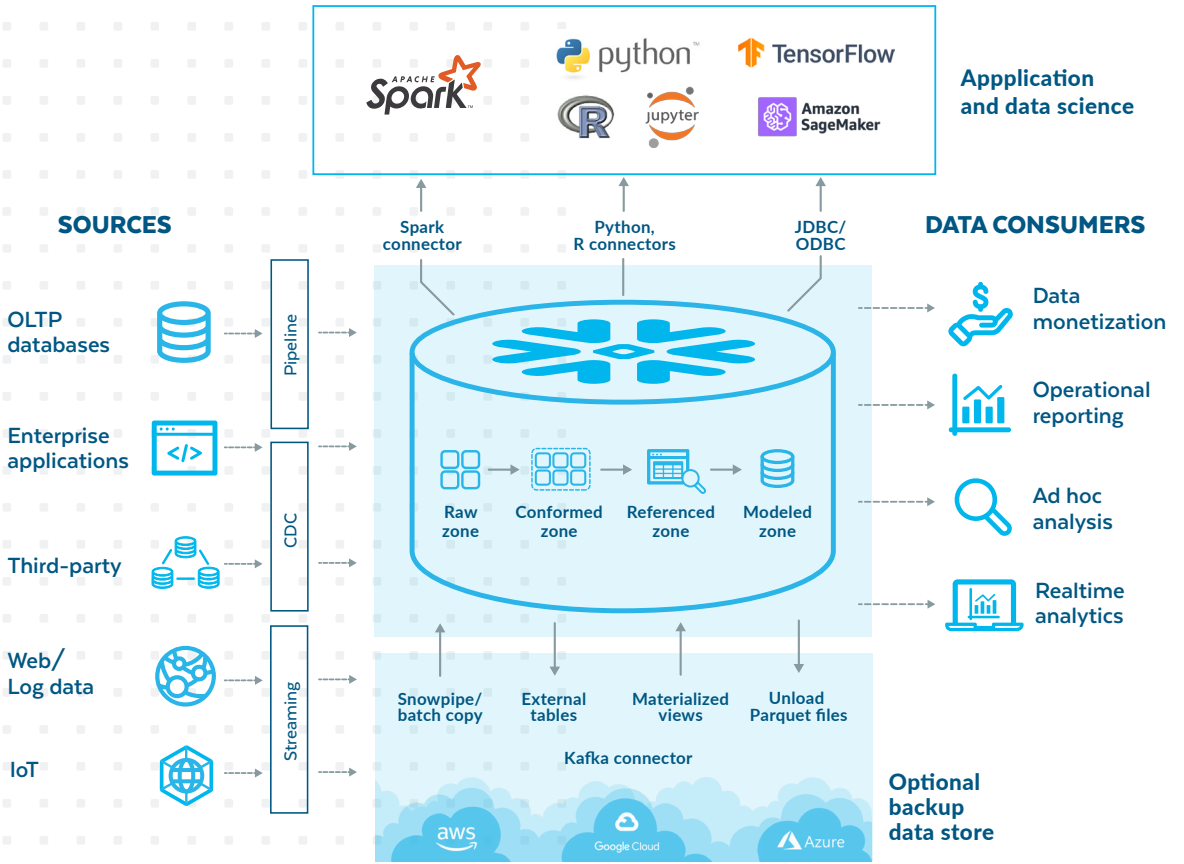
PROVIDE FAST, ACCURATE ANALYTICS FOR ALL USERS
Having the best of a data lake and data warehouse all in one means you'll have a fresh data source, no data silos, and fast analytics, so all users can run any query at any time to make accurate, data-driven decisions.

REDUCE MAINTENANCE WITH A SIMPLE, SECURE SERVICE
As a single platform offered as a service, Snowflake requires near-zero maintenance, uses standard SQL, and provides built-in security, governance, and ACID transaction support.

BENEFIT FROM UNLIMITED SCALE AND SPEED
Separation of storage and compute and elastic, automatic scale means all ETL and query operations get the isolated resources they need to be performant even during spikes in concurrency or demand.

Snowflake has become a key technology for organizations intent on getting the most value out of their data. By ingesting structured and semi-structured data from any data source, Snowflake solves the issue of scattered, siloed data and serves as a centralized data repository for the entire organization. Snowflake's elastic architecture automatically scales up and down, so all workloads have dedicated, appropriate compute resources, ensuring better performance and short query times. Snowflake is ANSI SQL compliant and works with essentially any data loading or analytical product, making it easier for departments and users to run any analytical workload while using fresh, complete data for more insightful analytics and decision-making. Snowflake can also complement an existing data lake.

ALL YOUR DATA QUICKLY ACCESSIBLE TO ALL YOUR USERS



HOW ORGANIZATIONS ARE USING SNOWFLAKE TO COMPLEMENT DATA LAKES

Fast performance and analytics

- Quickly load structured or semi-structured (JSON, Avro, XML Parquet, and so on) data from your data lake into Snowflake for faster SQL queries and analytics.
- Optionally use Snowflake's External Tables feature to directly query data in place in your data lake.

Simpler, faster transformations

- Use Snowflake's Streams and Tasks capabilities to perform automated data transformations directly in Snowflake instead of a data lake.
- Use Snowflake's elastic compute resources for faster and more cost-effective transformations.

Output transformed data back to the data lake

- Use Snowflake to transform data, including possibly transforming data to file types such as Parquet that are better for analytics.
- Export the resulting data back to the data lake for analysis by other tools.

CUSTOMER STORIES

- Devon Energy
- KIXEYE
- Blackboard
- Zeeto
- Siemens

CONSOLIDATING MANY SYSTEMS TO ONE MODERN DATA LAKE

Devon Energy is a leading independent oil and natural gas exploration and production company working to meet the world's growing energy demands. Using Snowflake, it consolidated multiple inefficient, low-performing systems into a modern data lake in the cloud.

GOAL

Consolidate an underused data lake, a data warehouse, and unstable enterprise data sets into a single cloud data platform.

PAIN POINT BEFORE SNOWFLAKE

Devon tried to build an enterprise data warehouse three times, and each time the company encountered challenges.

SCENARIO BEFORE SNOWFLAKE

- The first system was not scalable at high data and query volumes, and its capacity limit was reached in the first six months.
- The second system had too many points of failure and required full-time support, and the third system provided no visibility into what users were doing.

RESULTS WITH SNOWFLAKE

- Devon has democratized data access so all employees can query 95% of the company's data.
- A regulatory report that took 48 hours in SQL Server now runs in just minutes.
- A process that prepares data for loading into an application took 15 hours before but now runs in 30 minutes.
- Snowflake returns results in under 10 seconds for 2,000 simultaneous, random queries against a 40-billion-record table.

> **"We opened our data to everyone in the enterprise. If you have an employee badge and can get in the door, you can see all of the data, unless there's a specific reason to restrict access, such as personally identifiable information."**
>
> **LARRY QUERBACH**
> Enterprise Data Architect, Devon

BUILDING A CLOUD DATA LAKE

KIXEYE is a mobile and online gaming company that is redefining the intersection of fidelity and accessibility to create innovative experiences for competitive gamers. Using Snowflake, its teams work with all their data in one system, without worrying that one group or job will impact another.

GOAL

Simplify a complex, expensive, and slow data processing pipeline.

PAIN POINT BEFORE SNOWFLAKE

Making data accessible with good performance to SQL analysts and tools such as Tableau required transforming the data into aggregated, relationally structured data using multiple steps.

SCENARIO BEFORE SNOWFLAKE

- The previous architecture, which used a Kafka pipeline, a Hadoop cluster, and Hive tables, was complex, introduced a lot of latency, and wasn't scalable.
- It was expensive to maintain all the infrastructure needed to summarize, aggregate, and load data into a SQL database to support querying, reporting, and visualization by SQL analysts and BI tools such as Tableau.

RESULTS WITH SNOWFLAKE

- KIXEYE outputs JSON event data directly from Kafka to Amazon S3, from which it is loaded into Snowflake for processing and analysis.
- KIXEYE now makes data available for analysis an order of magnitude faster than before.
- Snowflake's elasticity gives each group the resources it needs without breaking the bank.
- The company loads JSON data "as is" into event tables and then parses it into views that Tableau understands.

Snowflake has made it easy to get data to analysts faster. Its unique architecture makes it possible to have all of our teams working with data in one system, without performance impact and without needing to be database administrators."

JOSH MCDONALD
Director of Analytics Engineering, KIXEYE

ACCELERATING PERFORMANCE AND PREDICTIVE DATA MODELING

Blackboard is a leading educational technology company that provides software tools and services to educational institutions. Using Snowflake, it uses data to do predictive modeling efficiently to project student outcomes.

GOAL

Use all its data without being limited by the volume of data or the data throughput.

PAIN POINT BEFORE SNOWFLAKE

Researchers spent 80% of their time accessing and cleaning data.

SCENARIO BEFORE SNOWFLAKE

- Blackboard needed a team of engineers to maintain its previous system.
- The system worked reasonably well, but the company didn't realize there was a faster and cheaper way to get results.

RESULTS WITH SNOWFLAKE

- Blackboard accelerated performance 16x by scaling storage, compute, and users independently and rapidly.
- The company now can seamlessly scale up and down just in time, as needed.
- Blackboard ingests JSON data directly and runs queries against huge data sets, all in one place without needing a team of engineers to maintain the system.

One of the things about Snowflake that really differentiated it for us was the ability to scale storage, compute, and users all independently and rapidly."

JAY WHITE
Director of Software Engineering, Blackboard

OPTIMIZING DIGITAL ADVERTISING VIA ANALYTICS

Zeeto's real-time bidding platform surveys website visitors and uses the insights gained to help customers bid on highly targeted advertising. Using Snowflake, its account managers use data to coach advertisers on their bidding strategies and click-through rates.

GOAL

Expand data monetization channels while using less infrastructure.

PAIN POINT BEFORE SNOWFLAKE

As Zeeto added customers, it had to purge critical business performance data because customer demand was outpacing the capacity of its rigid infrastructure.

SCENARIO BEFORE SNOWFLAKE

- Zeeto used a microservices architecture to stream and query all platform events, which amounted to billions of rows of data each year, but the system could not model data quickly and iterate.
- Zeeto's engineering team took up to five days to design and deliver reports to the management team.

RESULTS WITH SNOWFLAKE

- Business users now produce reports in a matter of hours from idea to launch, which streamlines data-driven decisions.
- The CEO can access updated system reports every 15 minutes across all publishers and advertisers.
- Business users receive data and insights they were unable to acquire with Zeeto's legacy system.
- Zeeto saves thousands of dollars per month with Snowflake compared to the cost of its legacy system.

> **"Being able to iterate and produce results in hours instead of days, and days instead of weeks, has been transformational."**
>
> **MATT FERGUSON**
> Chief Technology Officer, Zeeto

SIEMENS

CONNECTING GAS AND POWER DATA TO A DATA LAKE

Siemens, Europe's largest industrial manufacturing company, uses vast amounts of data from the gas and power industry for various groups within the company. With Snowflake, it can store as much data as it would like at a lower cost, while having extreme flexibility with compute clusters.

GOAL

Remove limits on storage and compute power by moving to a cloud-built system.

PAIN POINT BEFORE SNOWFLAKE

Loading large amounts of data in Siemens' previous database management system was too slow.

SCENARIO BEFORE SNOWFLAKE

- Siemens' engineering, sales, and app development groups had different requirements for compute power, which led to times when the cluster was overloaded as well as times when the cluster was idle.
- The company's previous system was expensive.

RESULTS WITH SNOWFLAKE

- The company can store as much data as it wants and storage costs are low.
- Snowflake provides extreme flexibility for compute clusters, so Siemens' various groups have different compute clusters that don't conflict with each other.
- Managers appreciate that the costs associated with each Siemens group are transparent.

Administration has become a lot easier, so we have more time for cleaning data and connecting new data sources to our data lake."

MARCEL GNOTH
Data Lead, Siemens

THE DATA-DRIVEN FUTURE OF DATA LAKES

Great hype erupted when the data lake first emerged. But that hype soon subsided because it was nearly impossible to gain insights from all that data. Snowflake delivers on the promise of the data lake. Using Snowflake with a data lake allows you to work with more-diverse data sets without requiring you to copy or move data or manipulate it for specific use cases, all while you pay for only what you need, when you need it.

ABOUT SNOWFLAKE

The Snowflake Cloud Data Platform shatters the barriers that prevent organizations from unleashing the true value from their data. Thousands of customers deploy Snowflake to advance their businesses beyond what was once possible by deriving all the insights from all their data by all their business users. Snowflake equips organizations with a single, integrated platform that offers the only data warehouse built for any cloud; instant, secure, and governed access to their entire network of data; and a core architecture to enable many other types of data workloads, including a single platform for developing modern data applications. Snowflake: Data without limits. Find out more at **snowflake.com**.