



SNOWFLAKE: ONE CLOUD DATA PLATFORM FOR ALL YOUR ANALYTIC NEEDS

INTRODUCTION

By 2023, IDC predicts that 50% of software vendor revenue for data and analytics will come from public clouds, and that cloud implementations will grow eight times faster than on-premises implementations.¹ Gartner goes further, predicting that 75% of all databases will be deployed or migrated to a cloud platform by 2022.²

Forward-looking organizations, however, want a cloud data platform that can enable a long-term strategy for maximizing their data assets, based on an extensible, multi-region and multi-cloud platform that powers all types of data workloads. They do not want to move data and analytic workloads to the cloud haphazardly. In response, Snowflake has created a cloud data platform that enables the best of modern data warehousing, the best of data lakes, the most-advanced data science capabilities, and much more.

Snowflake's cloud data platform allows you to consolidate analytics across data warehouses, data marts, and data lakes into a single source of truth that powers multiple types of business applications. It also streamlines how diverse teams share governed data, internally and externally, in order to collaborate across all data without having to copy data and move it from place to place. This simplifies data sharing, allows you to set up pertinent data access restrictions, and minimizes the data governance and compliance issues of managing multiple copies of the same data sets. Snowflake also allows you to store your data in any region, on any cloud, including popular offerings from Amazon, Microsoft, and Google (see Figure 1). The platform automates everything from data storage and processing to transaction management, security, governance, and metadata management.

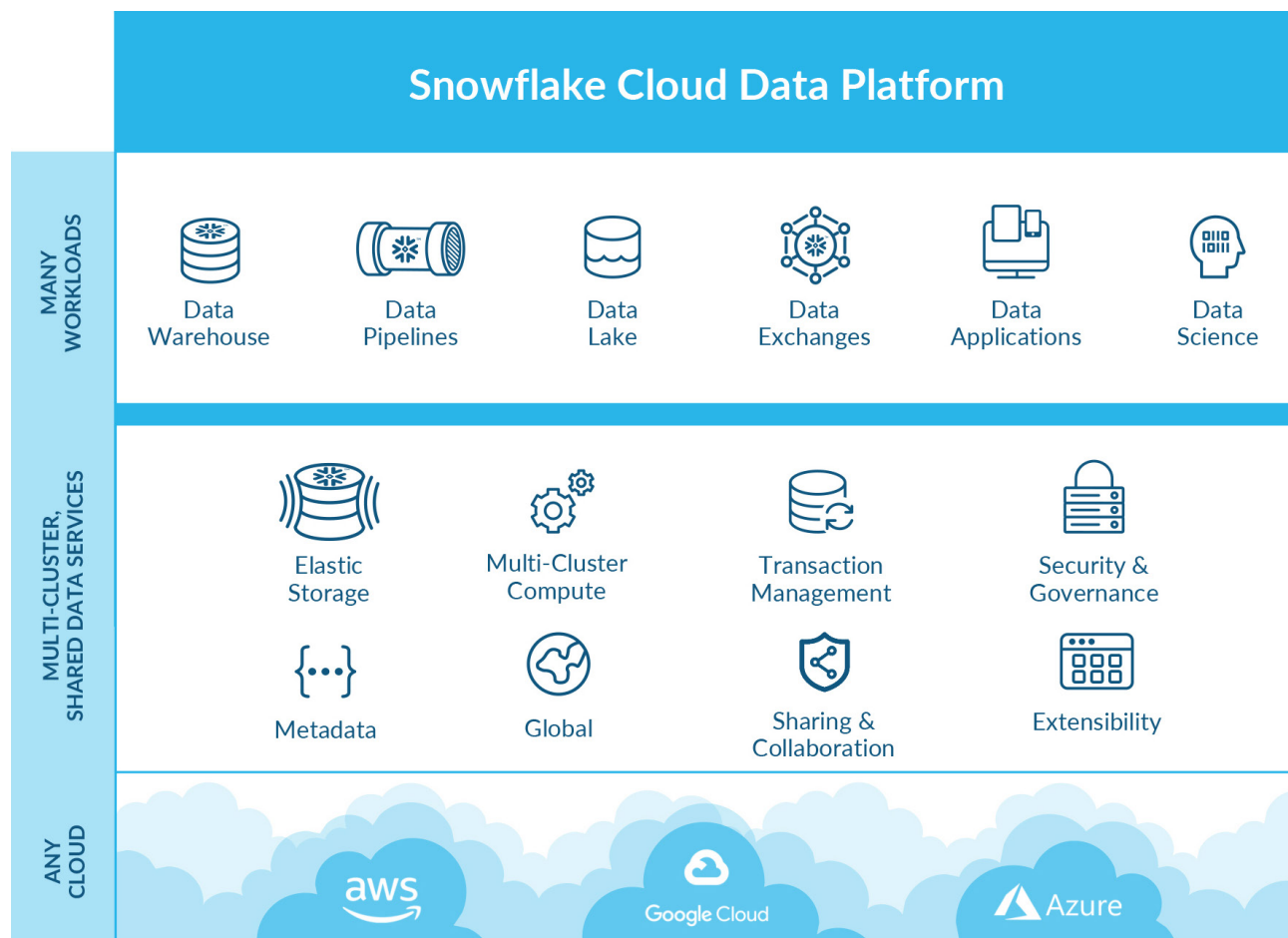


Figure 1: Snowflake's unique multi-cluster architecture works with any cloud, and delivers a host of powerful services to enable a number of modern use cases and workloads.

¹ Worldwide Big Data and Analytics Software Forecast, 2019–2023. IDC. September 2019.

² <https://www.gartner.com/en/newsroom/press-releases/2019-07-01-gartner-says-the-future-of-the-database-market-is-the>

MULTI-CLUSTER, MULTI-CLOUD COMPUTE ARCHITECTURE

While traditional data architectures are difficult to scale, one of Snowflake's core differences allows you to scale the exact amount of compute resources up and down, instantly and independently. All data processing is performed by any amount of compute power you choose for each workload, which you can scale back down just as fast. Snowflake's unique architecture enables you to process enormous quantities of data with maximum speed and efficiency by dedicating independent computing power to each workload, so there's no impact on performance. You can specify the number of compute clusters you want to use or let the service scale automatically.

Designed for efficiency, Snowflake determines precisely what is required from the storage layer, eliminating unnecessary scans to satisfy queries in the most performant manner possible. Data is cached along with the query results to improve the performance of future queries. This architecture allows you to run multiple workloads across multiple teams without resource contention, maximizing performance and efficiency.

BEST PRACTICE

Snowflake's unique architecture separates storage resources from compute resources: all workloads can simultaneously leverage just the compute power they need, when they need it, without contention.

FLEXIBLE, SCALABLE, AND COST-EFFECTIVE

Snowflake's cloud data platform seamlessly scales to support thousands of users, facilitating collaboration across the enterprise. It's an ideal solution for powering self-service analytics, including operational reporting, ad hoc querying, and real-time

decision-making. Resource isolation ensures that all users have a consistent experience, while Snowflake's per-second billing means each user and workgroup pays only for the precise resources used. It's easy to expand and scale each analytic application, yet pay only for the computing power you use and the storage capacity you need.

Delivered as a service, and with consistent functionality across multiple regions and clouds, Snowflake's advanced architecture enables instant elasticity and allows one cohesive platform to serve all types of users and workloads in a consistent way. Behind the scenes, the platform transparently automates server provisioning, data governance, data protection, data security, and performance tuning. This frees your team to focus on gaining the most value from your data.

MULTI-CLUSTER, SHARED DATA PLATFORM SERVICES

Snowflake's multi-cluster shared data services include an integrated set of services to authenticate user sessions, manage resources, enforce security, compile queries, and guarantee the integrity of all transactions.

This services layer is constructed of stateless compute resources running across multiple availability zones. It utilizes a highly available, distributed metadata store for global state management. Sound complicated? It's not. With Snowflake's highly efficient and trouble-free cloud data platform, there is no hardware to purchase, no software to maintain, and no infrastructure to integrate. As your workloads run, the data platform determines the most efficient, cost-effective, and high-performance way to process your data.

Thanks to this modern architecture, the Snowflake cloud data platform can handle many workloads concurrently, even when they reside across multiple clouds and regions, with exceptional performance.

BEST PRACTICE

The Snowflake cloud data platform services layer unifies security, governance, and metadata management, protecting your data, optimizing workload performance, eliminating resource contention, and ensuring transactional consistency.

CONTINUOUS DATA INGESTION

The Snowflake cloud data platform includes a serverless ingestion service called Snowpipe that asynchronously loads data into your cloud storage environment. Standard connectors and adapters allow you to easily ingest event streams from Kafka and other messaging systems, while Snowflake streams and tasks make it easy to schedule data loads for SQL jobs. You can ingest data into the repository and share it with a global customer base, without having to set up extract, transform, and load (ETL) pipelines or exchange data among regions. The platform automatically transforms data into the type and shape required for each target table. An Apache Kafka connector lets you continuously stream JSON records for storage and analysis.

ROBUST TRANSACTION MANAGEMENT AND DATA PARTITIONING

Guaranteeing the integrity of database transactions is essential in many industries. This is especially important for organizations that handle time-sensitive data, such as financial services companies that conduct monetary transactions and manufacturing firms that run real-time production processes. Snowflake enforces database atomicity, consistency, isolation, and durability (ACID) properties to guarantee transaction consistency even in the event of unforeseen errors, power failures, and other mishaps.

Adherence to ACID properties ensures database accuracy and transactional integrity, while micro-partitioning improves database performance. As data is loaded, Snowflake transparently divides it into micro-partitions—contiguous units of storage organized in a columnar fashion. This structure allows for extremely granular pruning of very large tables, which can be composed of millions of micro-partitions to significantly improve query performance. Snowflake stores partitions in a central storage layer that all compute nodes can access. During a query, it automatically selects the optimal distribution method for the partitions based on the current size of the data set.

BEST PRACTICE

Snowflake's cloud data platform supports data loading and analytics on mixed data formats with complete transactional integrity, guaranteeing accuracy of all database transactions and ensuring optimal performance for all types of queries.

INDUSTRY-LEADING SECURITY AND COMPLIANCE

The Snowflake cloud data platform is built on a multilayered security foundation that includes encryption, access control, network monitoring, and physical security measures, in conjunction with comprehensive monitoring, alerts, and cybersecurity practices. Every aspect of the platform is geared toward protecting your data, both in transit and at rest.

Whether your data resides in one or many locations, you have to protect and control the flow of data at each point. Snowflake takes this seriously: security technologies are woven throughout every facet of the platform (see Figure 2 on the following page). For example, role-based access control ensures users can access only the data they're permitted to see.

How Snowflake Protects Your Cloud Data

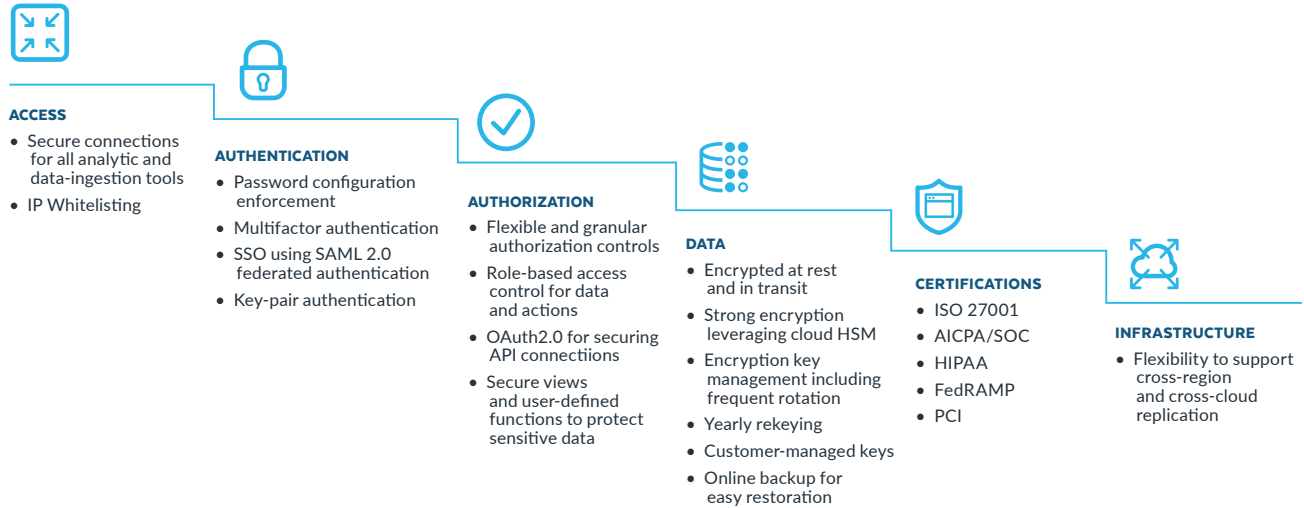


Figure 2: Security technologies are woven throughout every facet of the platform.

These control mechanisms can be applied to all database objects including tables, schemas, and any virtual extensions to a data set. Multi-factor authentication procedures can issue a secondary verification such as a one-time security code sent to a user's mobile phone. However, all this comprehensive security doesn't stand in the way of usability. Single sign-on procedures and federated authentication technologies make it easy for people to log into a data warehouse, data lake, or other analytic service directly from within those applications.

BEST PRACTICE

Snowflake's cloud data platform protects your data both in transit and at rest, with an emphasis on encryption, access control, data storage, and physical infrastructure in conjunction with comprehensive cybersecurity practices.

METADATA MANAGEMENT TO ENFORCE ACCURACY AND CONSISTENCY

Snowflake reduces complexity by consolidating data into a single repository that can accommodate structured or semi-structured data from any source—generated on-premises or in the cloud—including OLTP databases, enterprise applications, third-party applications, web log data, and IoT data. A robust metadata service spans the entire system. Queries are compiled within the services layer, and metadata is used to determine the micro-partition columns that need to be scanned. This makes it possible to track where data is coming from, who touched that data, and how various data sets relate to one another. External tables store file-level metadata such as file paths, version identifiers, and partitioning information. This enables business users to query data via multiple workloads consistently—as if it all emerged from a single database.

ELASTIC, MULTI-FACETED STORAGE

Whether you are importing relational data from enterprise applications, log data from website activities, IoT data from equipment sensors, or social media data from consumer networks, the volume and complexity of today's data sources can quickly overwhelm a conventional data platform. You need an elastic storage layer to easily and efficiently store all your data and make it instantly available, regardless of scale.

Snowflake's cloud data platform handles both structured and semi-structured data and can span multiple regions and clouds, consolidating all your data into a single source of truth. For example, in addition to relational data from enterprise applications, you can use this repository to store semi-structured data such as JSON, Avro, and XML, along with open source data types such as Apache Parquet and ORC, and all at petabyte scale. Built on versatile blob storage, this storage layer holds your data, tables, and query results.

SECURE DATA SHARING AND EXCHANGES

Traditional data sharing methods require any organization that shares data to copy that data and send it to its data consumers. These cumbersome, costly, and risky methods are based on sharing static data, which quickly becomes stale and must be

continually refreshed with current versions. Outdated approaches such as these incur unnecessary costs and require extra effort, yet they lack the ability to secure sensitive information or prevent data breaches.

Snowflake's cloud data platform features Snowflake Data Sharing by enabling authorized members of a cloud ecosystem to tap into live, read-only versions of the data. This allows you to easily and securely share subsets of your data, as well as receive shared data in a secure and governed way. For example, Snowflake Data Sharing enables data providers to instantly share data with vendors, supply chain partners, logistics partners, customers, and many other constituents. You can also set up data exchanges and other data sharing services with Snowflake to monetize your data, turning your data warehouse or data lake into a profit center (see Figure 3).

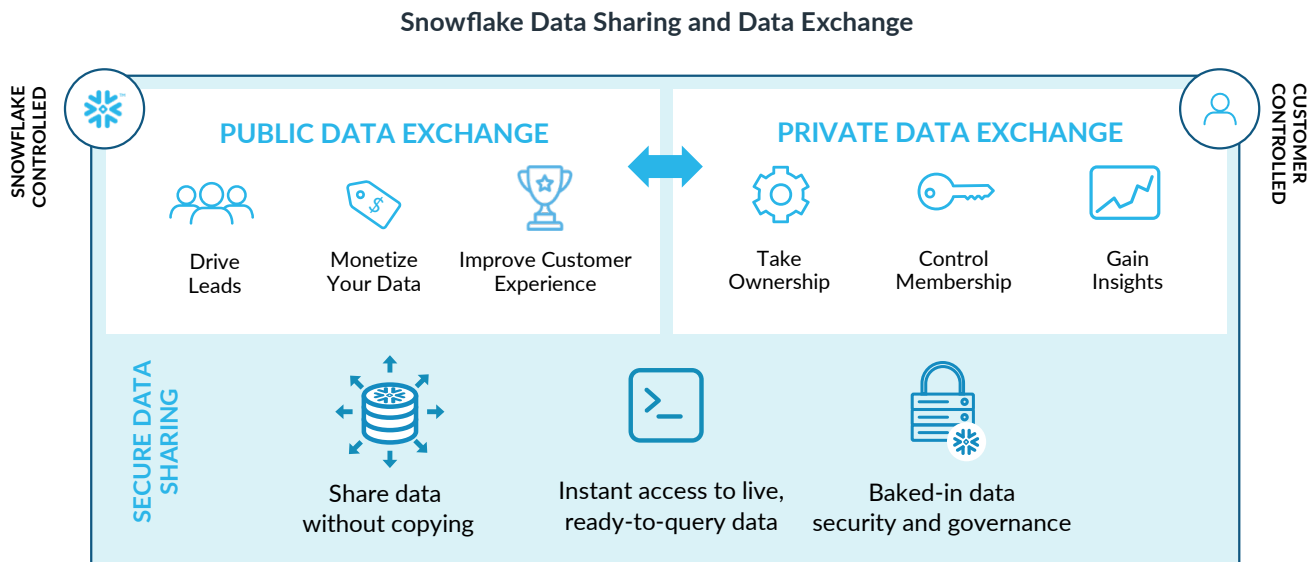


Figure 3: Snowflake Data Sharing enables you to share data externally via the Public Data Exchange, and create your own Private Data Exchange with customers, suppliers, and other business partners.

HIGH PERFORMANCE AND MULTI-WORKLOAD

Snowflake's external tables work with data directly stored in your Amazon S3, Azure Blob Storage, or Google Cloud Platform data lake. Materialized views on these external tables let you materialize all or just the portion of the data set that you use most frequently, eliminating the need to build an ETL layer or orchestration pipeline.

Rather than physically transferring data to internal or external consumers, Snowflake's data sharing technology enables read-only access to a governed portion of the live data set via SQL. The shared data is immediately available to be queried without requiring transformation. The data never leaves the data provider's account. Data consumers don't incur storage costs, and they don't need to worry about ETL processes to make use of the data.

Snowflake streams and tasks make it easy to schedule data loads for SQL jobs. The platform automatically transforms data into the type and shape required for each target table. An Apache Kafka connector lets you continuously stream JSON records for storage and analysis.

BEST PRACTICE

With Snowflake's secure data sharing capabilities, you don't have to copy or move your data. Data is live, ready to use, and always fresh.

GLOBAL DATA REPLICATION AMONG MULTIPLE CLOUDS

Having a global footprint helps ensure disaster recovery and improve business continuity practices. It also expands your options for sharing data with a global customer base. Snowflake's data platform makes it easy to share data among geographic regions and across multiple clouds and regions, including those offered by Amazon Web Services, Microsoft Azure, and Google Cloud Platform. These global replication capabilities make it painless for

your business units, and your partners, to collaborate with each other via a global ecosystem for analytics and data sharing. Snowflake's advanced replication technology allows you to easily share data across regions and clouds without setting up data pipelines, copying data, or resolving differences in security protocols between different clouds.

Replicating data across multiple regions and clouds not only simplifies data sharing but facilitates global expansion and bolsters business continuity. The platform automatically replicates databases and keeps them synchronized across regions and clouds, boosting availability, automating failover, and guaranteeing instant access and recovery for databases of any size.

BEST PRACTICE

Snowflake's cross-cloud data platform enables the free and secure movement of data anywhere in the world while also allowing you to select cloud storage vendors that meet the needs of each application or business unit.

EXTENSIBILITY FOR THE FUTURE

The Snowflake cloud data platform is extensible, which means it is accessible to other sanctioned solutions. The platform provides a solid foundation for many types of data and applications. As an extensible repository and unified data service, the platform can work with data where it resides, eliminating the need to move data from a data warehouse or a data lake. This allows you to manage data in many disparate places and from many diverse workloads but from one single service, regardless of the source and type of data each solution uses or how it is loaded, integrated, and transformed. This versatile cloud service maximizes the value of your data and all the services that connect to it, now and in the future (see Figure 4 on the following page).

Modern Data Architecture with Snowflake

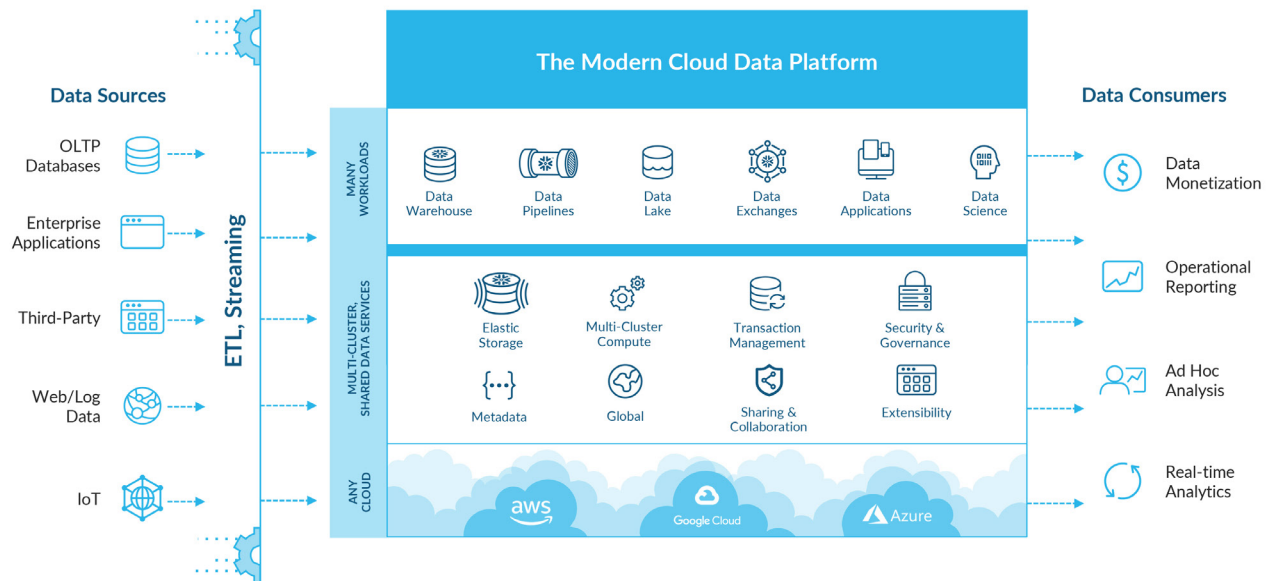


Figure 4: Snowflake powers and extends your data architecture, so you can easily get all your data into a single location and get all the insights from that data.

A SOLID FOUNDATION FOR ANY WORKLOAD

Most companies wrestle with tremendous data diversity. IT professionals must rationalize data coming in from a variety of sources, regions, and clouds. Some of the data is structured, some is semi-structured, and there is no single source of truth from which they can reliably correlate analytics.

Snowflake's cloud data platform brings order to this chaotic landscape. It accommodates all of these data types, at unlimited scale, and it allows different data types to be stored in their native formats, without complex transformations. Snowflake supports real-time streaming as well as batch uploads. This

architecture offers a solid foundation for running any workload, including data warehouses, data lakes, and many types of data engineering and data science applications.

Having a unified repository powered by a comprehensive layer of services brings consistency to all types of analytic projects. Snowflake consolidates all the necessary resources into one cohesive platform to enable many data workloads that organizations need in order to streamline their business, best serve their customers, and reveal new market opportunities. (see Figure 5 below).

What Snowflake's Cloud Data Platform Enables



Figure 5: Snowflake enables many use cases and workloads so your organization can leverage the power of a single solution for data analytics.

BEST PRACTICE

Snowflake lets you store diverse data types in their native forms, without creating new data silos.

Automatic and near-infinite elasticity releases the resources you need, when you need them.



CONCLUSION

With unified security and governance, the ability to store data types in their native format, and built-in transactional security, standardizing on Snowflake's cloud data platform allows you to shift your focus from managing infrastructure to managing data, and to obtaining every possible insight from that data. Delivered as a service, and with consistent functionality across multiple regions and clouds, Snowflake's advanced architecture enables instant and near-infinite elasticity and provides one cohesive platform to serve all types of users and workloads in a consistent way. With a single platform for deriving all the insights from all your data, your organization can make data-driven decisions that best serve your customers and your business—now and in the future.

To learn more, visit snowflake.com/data-platform.

ABOUT SNOWFLAKE

Snowflake's cloud data platform shatters the barriers that have prevented organizations of all sizes from unleashing the true value from their data. More than 2,000 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 the cloud; instant, secure, and governed access to their entire network of data; and a core architecture to enable many types of data workloads, including a single platform for developing modern data applications. Snowflake: Data without limits. Find out more at snowflake.com.

