CLOUD DATA PLATFORM SECURITY

HOW SNOWFLAKE SETS THE STANDARD

snowflake®

WHITE PAPER

The threat of a data security breach, someone gaining unauthorized access to an organization's data, is what keeps CEOs, CISOs, and CIOs awake at night. Such a breach can quickly turn into a public relations nightmare, resulting in lost business and steep fines from regulatory agencies. Snowflake Cloud Data Platform sets the industry standard for data platform security, so you don't have to lose sleep. All aspects of Snowflake's architecture, implementation, and operation are designed to protect customer data in transit and at rest against both current and evolving security threats.

SNOWFLAKE SECURITY FRAMEWORK

Snowflake was built from the ground up to deliver end-to-end data security for all data platform users. It follows best-in-class, standards-based practices for the controls and processes that secure it. As part of its overall security framework, it leverages NIST 800-53 and the CIS Critical Security Controls, a set of controls created by a broad consortium of international security experts to identify the security functions that are effective against real-world threats.

Snowflake comprises a multilayered security architecture to protect customer data and access to that data. This architecture addresses the following:

- External interfaces
- Access control
- Data storage
- Physical infrastructure

This security architecture is complemented by the monitoring, alerts, controls, and processes that are part of Snowflake's comprehensive security framework.

Security for compliance requirements

Snowflake is a multi-tenant service that implements isolation at multiple levels. It runs inside a virtual private cloud (VPC), a logically isolated network section within either Amazon Web Services (AWS), Microsoft Azure (Azure), or Google Cloud Platform (GCP). The dedicated subnet, along with the implementation of security groups, enables Snowflake to isolate and limit access to its internal components. Customers can choose from four Snowflake editions that vary by available features and level of security: Standard edition, Enterprise edition, Business Critical edition, and Virtual Private Snowflake (VPS). The Business Critical edition provides additional security features to support customers who have HIPAA, PCI DSS, or other compliance requirements. In addition, VPS supports customers who have specific regulatory requirements that prevent them from loading their data into a multi-tenant environment. VPS includes the Business Critical edition within a dedicated version of Snowflake. For additional details about the four versions, see the later section, "Four Levels of Snowflake Security."

Snowflake also isolates query processing, which is performed by one or more compute clusters called virtual warehouses. These are multinode compute clusters created by customers using Snowflake-provided interfaces. Snowflake provisions these compute clusters in such a way that the virtual warehouses of each customer are isolated from other customers' virtual warehouses. In addition, virtual warehouses are visible and accessible only to the users within a customer account who have been granted access.

Snowflake also isolates data storage by customer. Each customer's data is always stored in an independent directory and encrypted using customer-specific keys, which are accessible only by that customer.

> **We came to the conclusion that we achieved better security with Snowflake than we could ever do on our own."**
>
> **BOB ASENSIO**
> CIO, CapSpecialty

EXTERNAL INTERFACES

Customers access Snowflake via the internet using only secure protocols. The following drivers and tools may be used to connect to the service:

- Snowflake's command-line interface (CLI) client
- Snowflake's web-based user interface
- Snowflake Connector for Python
- Snowflake Connector for Spark
- Snowflake Connector for Kafka
- The Node.js driver
- The Go Snowflake driver
- The .NET driver
- The JDBC driver
- The ODBC driver

To find more information, see the Connectors & Drivers page in the Snowflake documentation.

All internet communication between users and Snowflake is secured and encrypted using TLS 1.2 or higher. Snowflake also supports IP address whitelisting to enable customers to restrict access to the Snowflake service by only trusted networks.

Customers who prefer to not allow any traffic to traverse the public internet may leverage either AWS PrivateLink (and AWS DirectConnect) or Microsoft Azure Private Link.

ACCESS CONTROL

Authentication

Snowflake employs robust authentication mechanisms, and every request to Snowflake must be authenticated, for example:

- User password hashes are securely stored.
- Strong password policy is enforced.
- Various mechanisms are deployed by Snowflake to thwart brute-force attacks.
- Snowflake also offers built-in multi-factor authentication (MFA), MFA for users with administrative privileges, and key-pair authentication for non-interactive users.
- For customers who want to manage the authentication mechanism for their account, and whose providers support SAML 2.0, Snowflake offers federated authentication.
- System for Cross-domain Identity Management (SCIM) can be leveraged to help facilitate the automated management of user identities and groups (that is, roles) in cloud applications using RESTful APIs.

Authorization

Snowflake provides a sophisticated, role-based access control (RBAC) authorization framework to ensure data and information can be accessed or operated on only by authorized users within an organization. Access control is applied to all database objects including tables, schemas, secure views, secure user-defined functions (secure UDFs), and virtual warehouses.

Access control grants determine a user's ability to both view and operate on database objects.

In Snowflake's access control model, users are assigned one or more roles, each of which can be assigned different access privileges. For every access to database objects, Snowflake validates that the necessary privileges have been granted to a role assigned to the user.

Customers can choose from a set of built-in roles or create and define custom roles within the role hierarchy defined by Snowflake.

The OAuth 2.0 authorization framework is also supported.

Encryption everywhere

In Snowflake, all customer data is always encrypted when it is stored on disk, and data is encrypted when it's moved into a Snowflake-provided staging location for loading into Snowflake. Data is also encrypted when it is stored within a database object in Snowflake, when it is cached within a virtual warehouse, and when Snowflake stores a query result.

Data encryption and key management

Snowflake uses strong AES 256-bit encryption with a hierarchical key model rooted in a cluster of hardware security modules. Each customer account has a separate key hierarchy of account-level, table-level, and file-level keys. Snowflake automatically rotates account and table keys on a regular basis. Data encryption and key management are entirely transparent to customers and require no configuration or management.

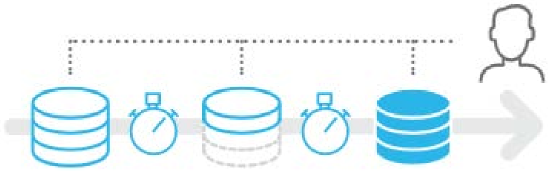
Data protection and recovery through retention and backups

Snowflake was designed from the ground up to be a continuously available cloud service that is resilient to failures to prevent customer disruption and data loss. Its continuous data protection (CDP) capabilities protect against and provide easy self-service recovery from accidental errors, system failures, and malicious acts.

Recovery from accidental errors

The most common cause of data loss or corruption in a database is accidental errors made by a system administrator, a privileged user, or an automated process. Snowflake provides a unique feature called Time Travel that provides easy recovery from such errors.

Time Travel makes it possible to instantly restore or query any previous version of a table or database from an arbitrary past point in time within a retention period.



ACCESS PAST VERSIONS OF DATA

How Time Travel works

Time Travel is made possible by Snowflake's implementation of data manipulation language (DML) operations. Snowflake provides a fully updatable relational database with a complete set of SQL DML operators that support updates to or deletion of rows of data. When any data is modified, Snowflake internally writes those changes to a new storage object and automatically retains the previous storage object for a period of time (the retention period) so that both versions are preserved. When data is deleted or database objects are dropped, Snowflake updates its metadata to reflect that change but keeps the data during the retention period.

During the retention period, all data and data objects are fully recoverable by customers. Using a simple SQL command, users granted administrative privileges can undo a DROP command that removes a database, table, or schema.

Past versions of a data object from any point in time within the retention period can also be accessed via SQL, both for direct access by a SELECT statement as well as for cloning in order to create a copy of a past version of the data object.

After the retention period has passed, Snowflake's Fail-Safe feature provides an additional seven days (the "fail-safe" period) to provide a sufficient length of time during which Snowflake can, at a customer's request, recover any data that was maliciously or inadvertently deleted by human or software error. At the end of that period, an automated process physically deletes the data. Because of this design, it is impossible for the Snowflake service, any Snowflake personnel, or malicious intruders to physically delete data.

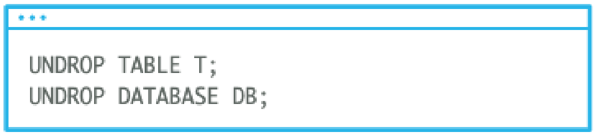
CDP and Time Travel are standard features built into Snowflake. The length of the default retention period is determined by the customer's service agreement. Customers can specify extended retention periods at the time that a new database, table, or schema is created via SQL data definition language (DDL) commands. Extended retention periods incur additional storage costs for the time that Snowflake retains the data during the retention and fail-safe periods.

The Time Travel, Fail-Safe, and CDP features provide customers with an unprecedented ability to recover from accidental errors. For example, if an errant data loading script corrupts a database, it is possible to create a logical duplicate of the database (a clone) from the point in time just prior to the execution of a specific statement.

To illustrate, the next sections provide some examples of statements using the Time Travel feature.

Example of recovering dropped objects

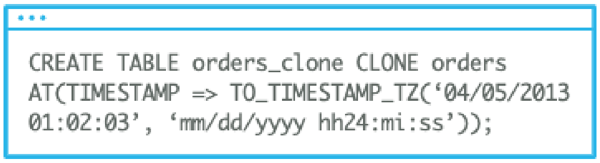
The UNDROP command can be used to recover any dropped object:



```
UNDROP TABLE T;
UNDROP DATABASE DB;
```

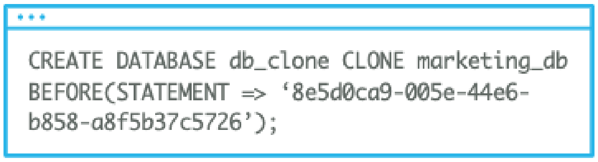
Examples of recovering previous versions

Recovering a previous version of a table can be done by cloning a past version of a table at a specific point in time:



```
CREATE TABLE orders_clone CLONE orders
AT(TIMESTAMP => TO_TIMESTAMP_TZ(‘04/05/2013
01:02:03’, ‘mm/dd/yyyy hh24:mi:ss’));
```

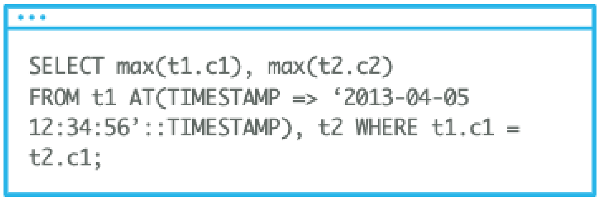
Recovering a previous version of a database can be done by cloning a past version of a database just before a query (identified here by a query ID) was processed:



```
CREATE DATABASE db_clone CLONE marketing_db
BEFORE(STATEMENT => ‘8e5d0ca9-005e-44e6-
b858-a8f5b37c5726’);
```

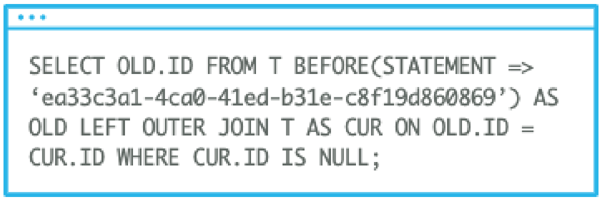
Examples of selecting data from past versions

Selecting data from an arbitrary time during the retention period can be done using the AT or BEFORE command:



```
SELECT max(t1.c1), max(t2.c2)
FROM t1 AT(TIMESTAMP => ‘2013-04-05
12:34:56’::TIMESTAMP), t2 WHERE t1.c1 =
t2.c1;
```

Combining the current state of table “t2” with a historical state of table “t1” as it existed before a previous query (identified by the query ID) can be done like this:



```
SELECT OLD.ID FROM T BEFORE(STATEMENT =>
‘ea33c3a1-4ca0-41ed-b31e-c8f19d860869’) AS
OLD LEFT OUTER JOIN T AS CUR ON OLD.ID =
CUR.ID WHERE CUR.ID IS NULL;
```

Further details about Time Travel and commands are described in the Snowflake documentation.

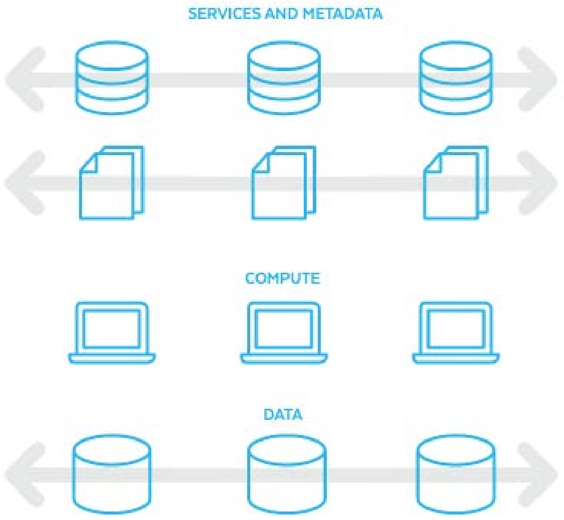
Protection against system failures

The second most common type of data loss is caused by some form of system failure: both software failures and infrastructure failures such as the loss of a disk, a disk array, a server or, most significantly, a data center.

The Snowflake architecture is designed for resilience, without data loss, in the face of such failures. Snowflake, which runs on all the major cloud providers’ platforms (AWS, GCP, and Azure), uses a fully distributed and resilient architecture combined with the resiliency capabilities available in these cloud platforms to protect against a wide array of possible failures.

As illustrated below, the Snowflake architecture consists of three layers, each of which is resilient to failures:

- Data storage layer. Stores all customer data in cloud storage.
- Compute layer. Consists of one or more virtual warehouses, each of which is a multinode compute cluster that processes queries. Virtual warehouses cache data from the data storage layer in encrypted form, but they do not store persistent data.
- Cloud services layer. The brain of the system, this layer manages infrastructure, queries, security, and metadata. The services running in this layer



SNOWFLAKE'S ARCHITECTURE PROTECTS BOTH DATA AND THE ABILITY TO ACCESS THAT DATA.

are implemented as a set of stateless processes.

Each layer in the Snowflake architecture is distributed across availability zones. Because availability zones are geographically separated data centers with independent access to power and networking, operations continue even if one or two availability zones become unavailable. In addition, the database storage layer leverages the cloud provider's resilient storage service to provide highly durable, cost-effective storage. When a transaction is committed in Snowflake, the data is securely stored in the cloud provider's highly durable data storage, which enables data survival in the event of the loss of one or more disks, servers, or even data centers. Amazon S3 synchronously and redundantly stores data across multiple devices in multiple facilities. It is designed for eleven 9s (99.999999999%) of data durability.

INFRASTRUCTURE SECURITY

Threat detection

Snowflake uses advanced threat detection tools to monitor all aspects of its infrastructure. All security logs, including logs and alerts from third-party tools, are centralized in Snowflake's security data lake, where they are aggregated for analysis and alerting. Activities meeting certain criteria generate alerts that are triaged through Snowflake's security incident process. Specific areas of focus include the following:

- File integrity monitoring (FIM) tools are used to ensure that critical system files, such as important system and application executable files, libraries, and configuration files, have not been tampered with. FIM tools use integrity checks to identify any suspicious system alterations, which include owner or permissions changes to files or directories, the use of alternate data streams to hide malicious activities, and the introduction of new files.
- Behavioral monitoring tools monitor network, user, and binary activity against a known baseline to identify anomalous behavior that could be an indication of compromise.
- Snowflake uses threat intelligence feeds to contextualize and correlate security events and harden security controls to counteract malicious tactics, techniques, and procedures (TTPs).

Physical security

Snowflake is hosted in AWS, Azure, or GCP data centers around the world. Snowflake's infrastructure-as-a-service cloud provider partners employ many physical security measures, including biometric access controls and 24-hour armed guards and video surveillance to ensure that no unauthorized access is permitted. Neither Snowflake personnel nor Snowflake customers have access to these data centers. For more specific information on the security controls implemented by Snowflake's cloud provider partners, please refer to the security and compliance documentation provided by your provider.

SECURITY COMPLIANCE

Snowflake's portfolio of security and compliance reports are continuously expanded as customers request reports. The following is the current list of reports available to all customers and prospects who are under a non-disclosure agreement. Please contact Snowflake for copies of the reports applicable to your organization or to find out if a particular certification will soon be available.

SOC 1 Type 2

The SOC 1 Type 2 report is an independent auditor's attestation of the financial controls that Snowflake had in place during the report's coverage period.

SOC 2 Type 2

The SOC 2 Type 2 report is an independent auditor's attestation of the security controls that Snowflake had in place during the report's coverage period. This report is provided for customers and prospects to review to ensure there are no exceptions to the documented policies and procedures in the policy documentation.

PCI DSS

The Payment Card Industry Data Security Standard is a set of prescriptive requirements to which an organization must adhere in order to be considered compliant. Snowflake's PCI DSS Attestation of Compliance provides an independent auditor's assessment results after testing Snowflake's security controls.

HIPAA

The Health Information Portability and Accountability Act is a law that provides data security and privacy provisions to protect protected health information. Snowflake is able to enter into a business associate agreement (BAA) with any covered entity that requires HIPAA compliance.

ISO/IEC 27001

The International Organization for Standardization provides requirements for establishing, implementing, maintaining, and continually improving an information security management system. Snowflake's ISO certificate is available for download here.

FedRAMP in Process (Moderate)

The Federal Risk and Authorization Management Program, or FedRAMP, is a government-wide program that provides a standardized approach to security. Federal agencies may download Snowflake's FedRAMP package from OMB MAX.

FOUR LEVELS OF SNOWFLAKE SECURITY

Snowflake offers four editions, with varying levels of security. Each subsequent version contains all the capabilities of the preceding versions. For example, the Business Critical edition includes everything the Enterprise edition offers.

	STANDARD EDITION	ENTERPRISE EDITION	BUSINESS CRITICAL EDITION	VPS
All authentication methods (incl. SAML, OAuth)	●	●	●	●
RBAC	●	●	●	●
User and role provisioning using SCIM	●	●	●	●
Network policies	●	●	●	●
AWS PrivateLink and Azure Private Link			●	●
Annual rekeying of data			●	●
Tri-Secret Secure (customer-managed key)			●	●
Tri-Secret Secure (customer-managed key)			●	●
HIPAA compliance			●	●
PCI DSS compliance			●	●
Operational visibility				●

Enterprise edition

All data is re-encrypted annually. Federated authentication is also available so users can access Snowflake with secure single sign-on capability. Snowflake's unique data protection feature, Time Travel, enables deleted or modified data to be restored to its original state for up to 90 days. Cross-region replication is also available in the Enterprise edition, making it possible to add additional redundancy to Snowflake's standard in-region replication.

Business critical edition

The Business Critical edition is Snowflake's solution for customers who have specific compliance requirements. It includes HIPAA support, is PCI DSS compliant, and features an enhanced security policy. This edition enables customers to use Tri-Secret Secure, which provides split encryption keys for multiple layers of data security.

When a customer uses Tri-Secret Secure, access to the customer's data requires the combination of the Snowflake encryption key, the customer encryption key (which is wholly owned by the customer), and valid customer credentials with role-based access to the data.

Because the data is encrypted with split keys, it is impossible for anyone other than the customer, including Amazon, to gain access to the underlying data. Snowflake can gain access to the data only if the customer key and access credentials are provided to Snowflake. This ensures that only the customer can respond to demands for data access, regardless of where they come from.

Virtual Private Snowflake (VPS)

VPS represents the most sophisticated solution for customers with sensitive data. It differs from other Snowflake editions in a number of important ways.

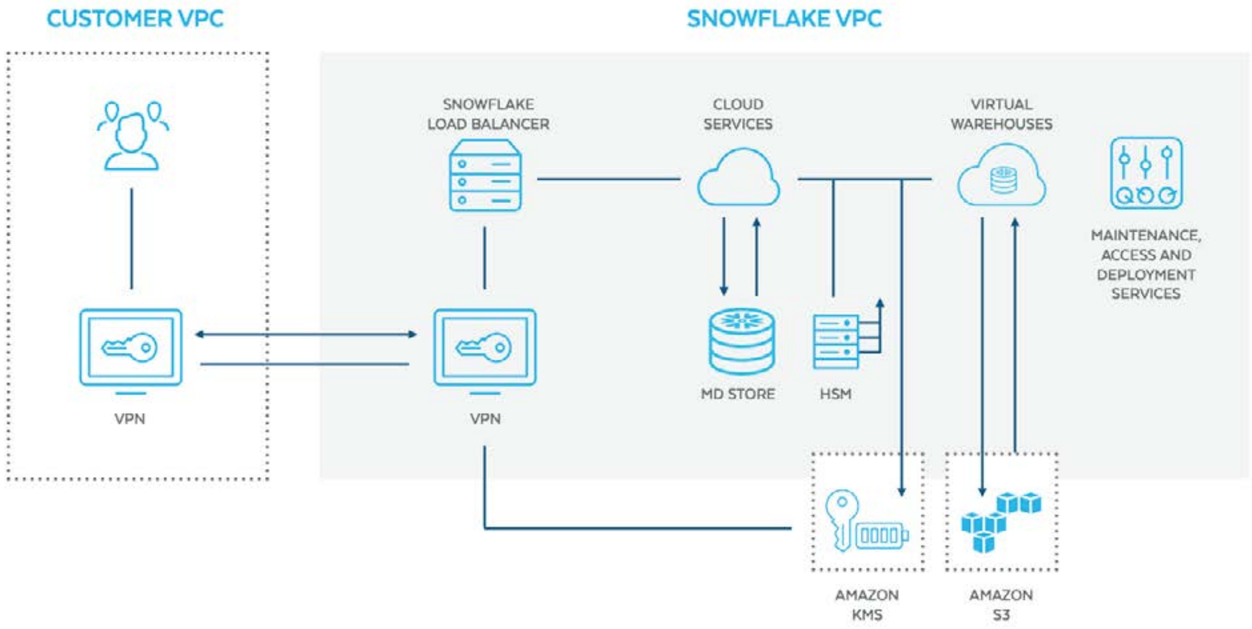
With VPS, all of the servers that contain in-memory encryption keys are unique to each customer. Each VPS customer has their own dedicated virtual servers, load balancer, and metadata store.

STANDARD EDITION	ENTERPRISE EDITION	BUSINESS CRITICAL EDITION	VPS
• Complete SQL data warehouse • Secure data sharing across regions/clouds • Premier support 24x365 • 1 day of time travel • Always-on enterprise-grade encryption of data in transit and at rest • Customer-dedicated virtual warehouses • Federated authentication • Database replication	Standard edition + • Multi-cluster warehouse • Up to 90 days of time travel • Annual rekeying of all encrypted data • Materialized views	Enterprise edition + • HIPAA support • PCI DSS compliance • Data encryption everywhere • Enhanced security policy • Customer-managed encryption keys	Business Critical edition + • Customer-dedicated virtual servers wherever the encryption key is in-memory • Customer-dedicated metadata store

There are also dedicated virtual private networks (VPNs) or virtual private cloud (VPC) bridges from a customer's own VPC to the Snowflake VPC. These dedicated services ensure that the most sensitive components of the customer's data warehouse are completely separate from those of other customers. In addition, VPS is designed to preserve Snowflake's unique ease of use and low burden of management.

Even with VPS, Snowflake's hardware security module and its maintenance, access, and deployment services are still shared services. These components are secure by design, even in a multi-tenant model. For instance, the hierarchical security module (HSM) is configured with a completely separate partition dedicated to the customer. All data is stored in Amazon S3 within a separately provisioned AWS account.

As shown is the following diagram, this design makes it possible for even the most security conscious customers to trust VPS as a comprehensively secure solution for their data.



CONCLUSION

All Snowflake Cloud Data Platform editions provide a secure and protected environment for customer data, protecting data in transit and at rest from current and evolving threats. The features built into Snowflake deliver enterprise-class security by default, without the additional burdens of complexity and management that traditional solutions force customers to endure.

Snowflake is ANSI SQL compliant and designed from the ground up for the cloud and for modern data analytics. Built with a unique new architecture, and provided as an enterprise-class software-as-a-service (SaaS) offering, Snowflake delivers instant elasticity, native support for diverse data, and per-second pricing. Security is fundamental to Snowflake's architecture, implementation, and operation. Every aspect of Snowflake is designed and operated to protect customer data.

ABOUT SNOWFLAKE

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**