SNOWFLAKE FOR
DATA LAKE

SNOWFLAKE'S PLATFORM

DATA SOURCES

OLTP DATABASES

ENTERPRISE APPLICATIONS

THIRD-PARTY

WEB/LOG DATA

IoT

DATA ENGINEERING

DATA WAREHOUSE

DATA SCIENCE

DATA APPLICATIONS

DATA SHARING

DATA LAKE

Access, performance, and security for your data lake

DATA CONSUMERS

DATA MONETIZATION

OPERATIONAL REPORTING

AD HOC ANALYSIS

REAL-TIME ANALYTICS

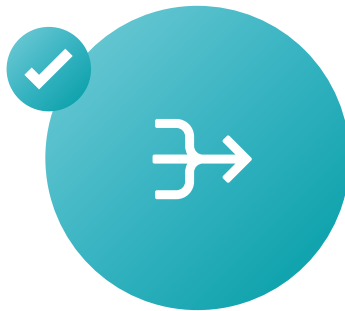Google Cloud   aws   Azure

SNOWFLAKE FOR DATA LAKE

Exceptional Query Performance

Achieve the full potential of your data with accelerated analytics and unlimited scale

Integrated Data Pipelines with Extensibility

Run simple, but reliable and performant data pipelines that meet your unique needs

Governed & Secure Collaboration

Enable collaboration across your data lake

Complement Your Data Lake with Snowflake

Build Your Data Lake with Snowflake

Complement Your Data Lake with Snowflake

Build Your Data Lake with Snowflake

DATA LAKE WITH SNOWFLAKE
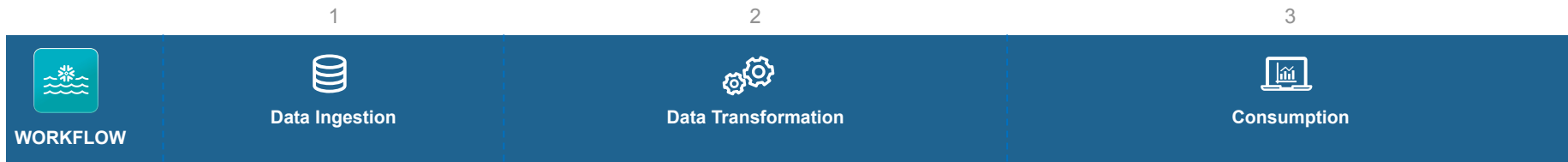
1

2

3

WORKFLOW

Data Ingestion

Data Transformation

Consumption

DATA LAKE WITH SNOWFLAKE

1   2   3

WORKFLOW   **Data Ingestion**   **Data Transformation**   **Consumption**

SNOWFLAKE CLOUD DATA PLATFORM

One platform, one copy of data, many workloads

Secure and governed access to all data

Near-zero maintenance, as a service

Unlimited performance and scale

DATA LAKE WITH SNOWFLAKE

CLICK ON NAV BAR TO GO TO DEEP DIVE

WORKFLOW	1 Data Ingestion	2 Data Transformation	3 Consumption
DATA LAKE FEATURES			Directly query data in your data lake with **External Tables**
			Better query performance using **Materialized Views** over External Tables
			Synchronize External Tables with **Apache Hive** metastore
			Automatically register new files with **auto-refresh of partitions**
			Accelerate data exploration with **Snowsight**
PLATFORM FEATURES			**Elastic Scaling** with multi-clustered warehouse
			Secure **data sharing** & **data exchange**
			Encryption, RBAC, PCI, HIPAA, FedRAMP
			Dynamic **Data Masking** & External **Tokenization**

SNOWFLAKE CLOUD DATA PLATFORM

- One platform, one copy of data, many workloads
- Secure and governed access to all data
- Near-zero maintenance, as a service
- Unlimited performance and scale

DATA LAKE WITH SNOWFLAKE

	1	**2**	**3**
WORKFLOW	**Data Ingestion**	**Data Transformation**	**Consumption**
DATA LAKE FEATURES	Auto load stage data using **Snowpipe** — Highly performant, cloud-based storage ingest — Incremental ingestion from your data lake using **Materialized Views** over External Tables		Directly query data in your data lake with **External Tables** — Better query performance using **Materialized Views** over External Tables — Synchronize External Tables with **Apache Hive** metastore — Automatically register new files with **auto-refresh of partitions** — Accelerate data exploration with **Snowsight**
PLATFORM FEATURES	Unmatched support for **semi-structured data** with **schema-on-read**		**Elastic Scaling** with multi-clustered warehouse — Secure **data sharing** & **data exchange** — Encryption, RBAC, PCI, HIPAA, FedRAMP — Dynamic **Data Masking** & External **Tokenization**

SNOWFLAKE CLOUD DATA PLATFORM

- One platform, one copy of data, many workloads
- Secure and governed access to all data
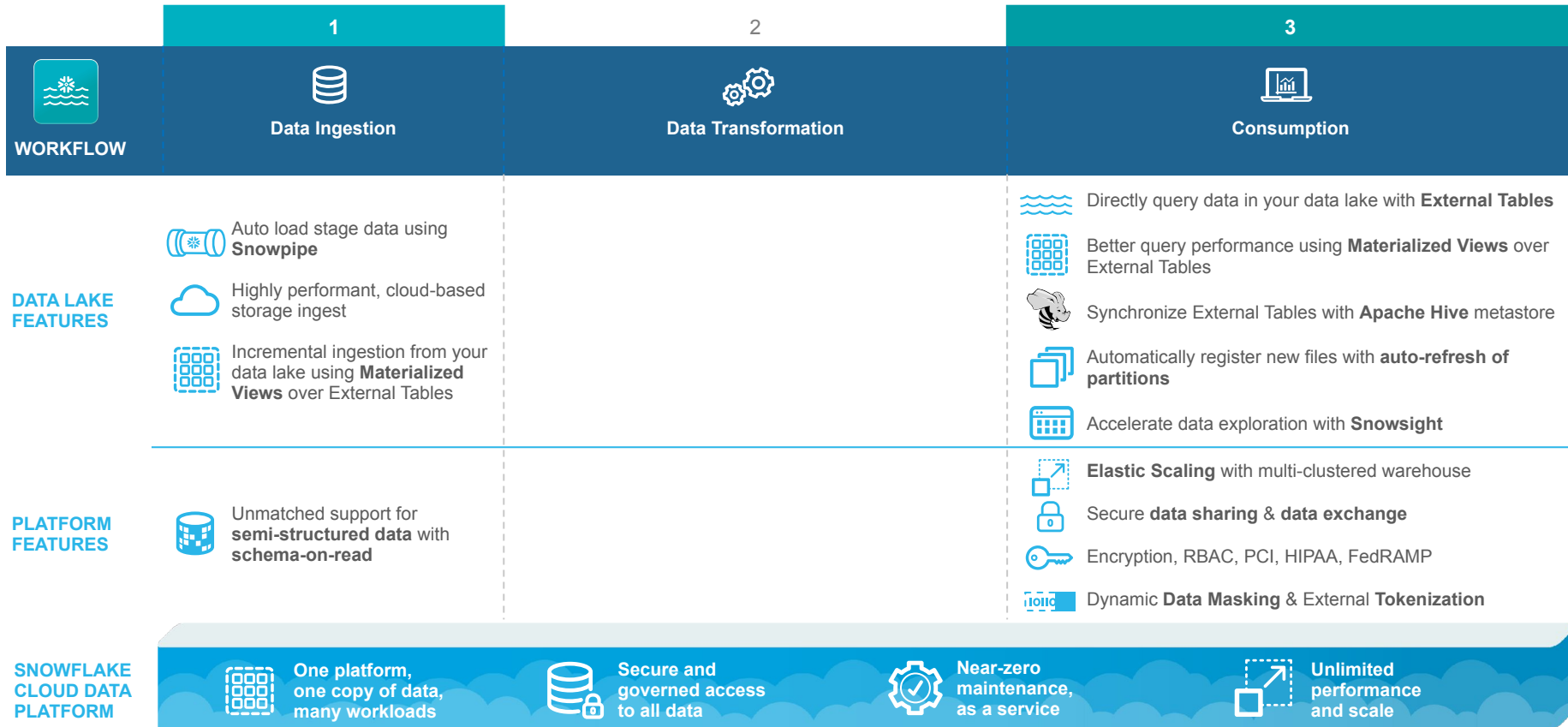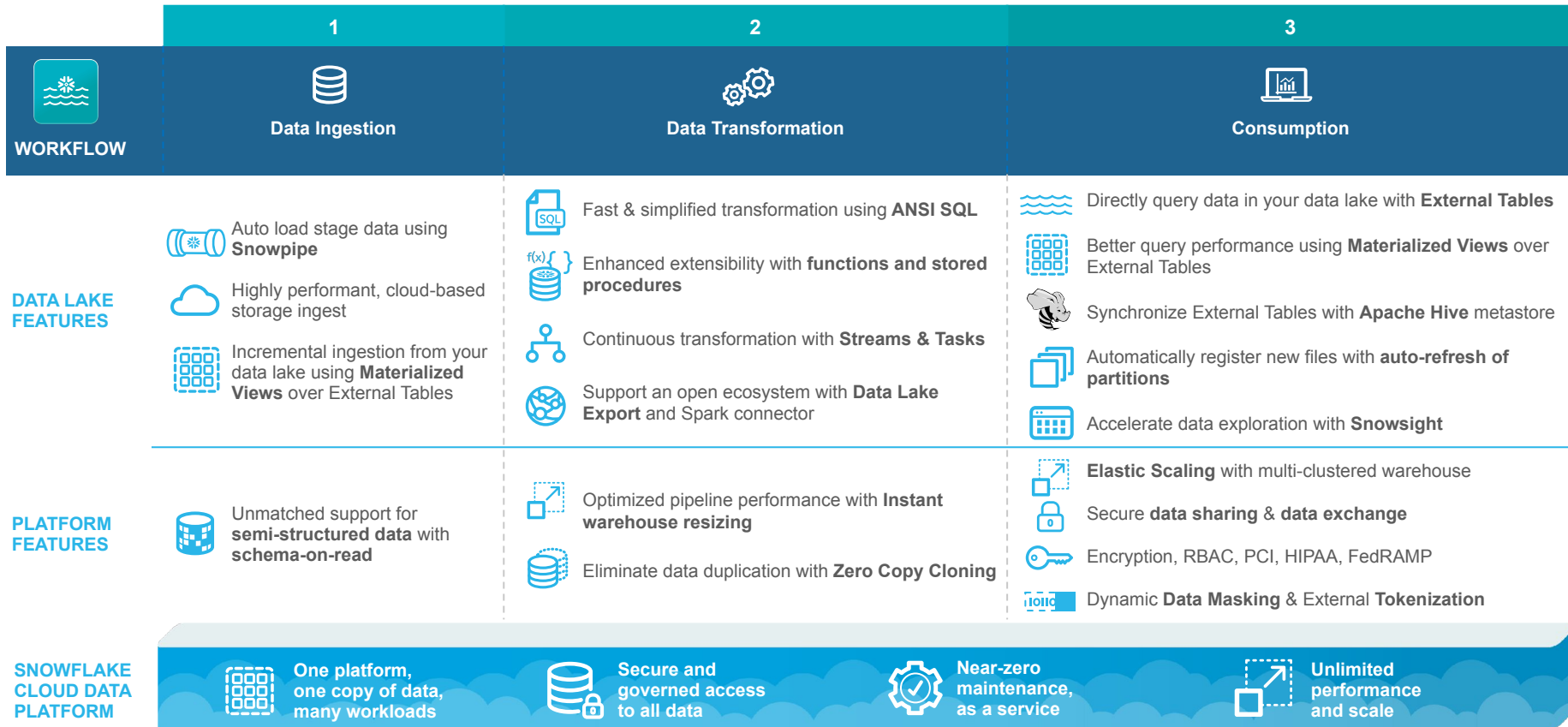- Near-zero maintenance, as a service
- Unlimited performance and scale

DATA LAKE WITH SNOWFLAKE

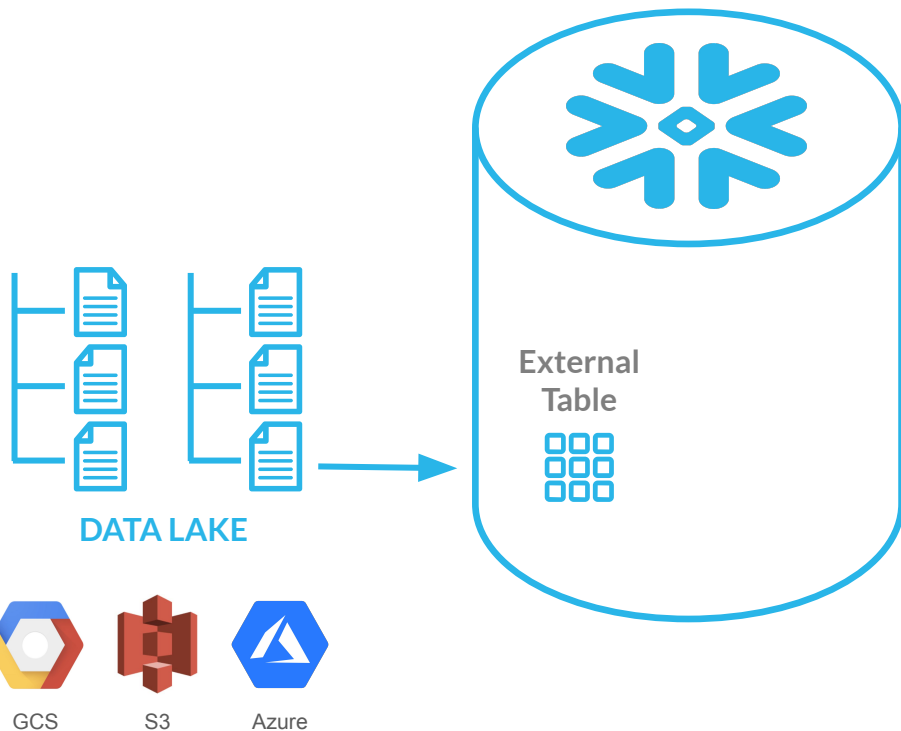|	1	2	3
WORKFLOW	**Data Ingestion**	**Data Transformation**	**Consumption**
DATA LAKE FEATURES	Auto load stage data using **Snowpipe** Highly performant, cloud-based storage ingest Incremental ingestion from your data lake using **Materialized Views** over External Tables	Fast & simplified transformation using **ANSI SQL** Enhanced extensibility with **functions and stored procedures** Continuous transformation with **Streams & Tasks** Support an open ecosystem with **Data Lake Export** and Spark connector	Directly query data in your data lake with **External Tables** Better query performance using **Materialized Views** over External Tables Synchronize External Tables with **Apache Hive** metastore Automatically register new files with **auto-refresh of partitions** Accelerate data exploration with **Snowsight**
PLATFORM FEATURES	Unmatched support for **semi-structured data** with **schema-on-read**	Optimized pipeline performance with **Instant warehouse resizing** Eliminate data duplication with **Zero Copy Cloning**	**Elastic Scaling** with multi-clustered warehouse Secure **data sharing** & **data exchange** Encryption, RBAC, PCI, HIPAA, FedRAMP Dynamic **Data Masking** & External **Tokenization**

SNOWFLAKE CLOUD DATA PLATFORM

One platform, one copy of data, many workloads	**Secure and governed access to all data**	**Near-zero maintenance, as a service**	**Unlimited performance and scale**

EXTERNAL TABLES

DATA LAKE

GCS   S3   Azure

External Table

Use Case
Read-only tables over files on external cloud storage. Queries over archived or historical cloud data, ad-hoc queries over files.
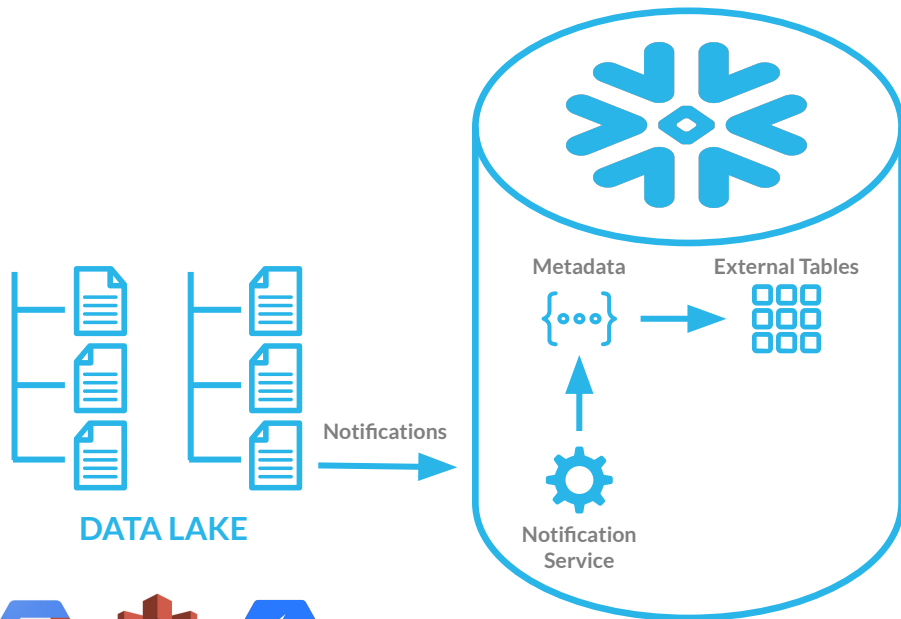
Cost
Small maintenance overhead charged for manually refreshing the external table metadata (using ALTER EXTERNAL TABLE … REFRESH).

Offset by Benefits
Query and access all data from a single governed data platform.

AUTO-REFRESH EXTERNAL TABLE PARTITIONS

Metadata

External Tables

Notifications

DATA LAKE

Notification Service

GCS   S3   Azure

Use Case
Refresh external table metadata automatically using your clouds notification service

- New files in the path are added to the table metadata
- Changes to files in the path are updated in the table metadata.
- Files no longer in the path are removed from the table metadata.

Cost
Small maintenance overhead charged for notification event management and refreshing table metadata.

Offset by Benefits
External data is always up to date and available to users and one location to access all data.

MATERIALIZED VIEWS

DATA LAKE

GCS   S3   Azure

Metadata   External Tables

Notifications

Notification Service

Materialized View
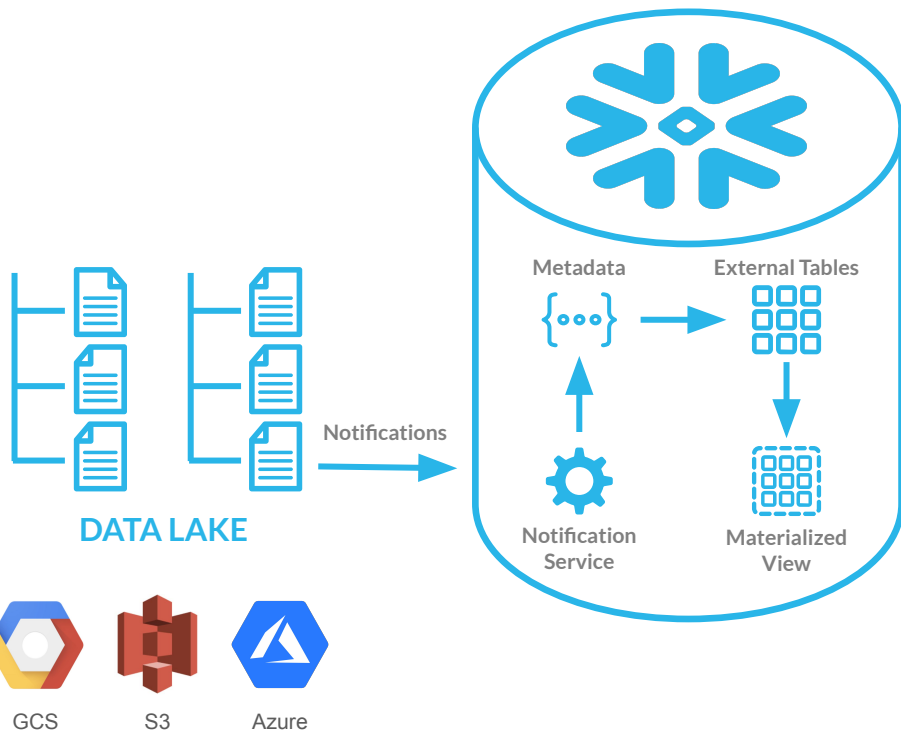
Use Case
Physically persist commonly aggregated data

MV on a Snowflake table guaranteed correct

Cost
Serverless refresh, additional storage

Offset by Benefits
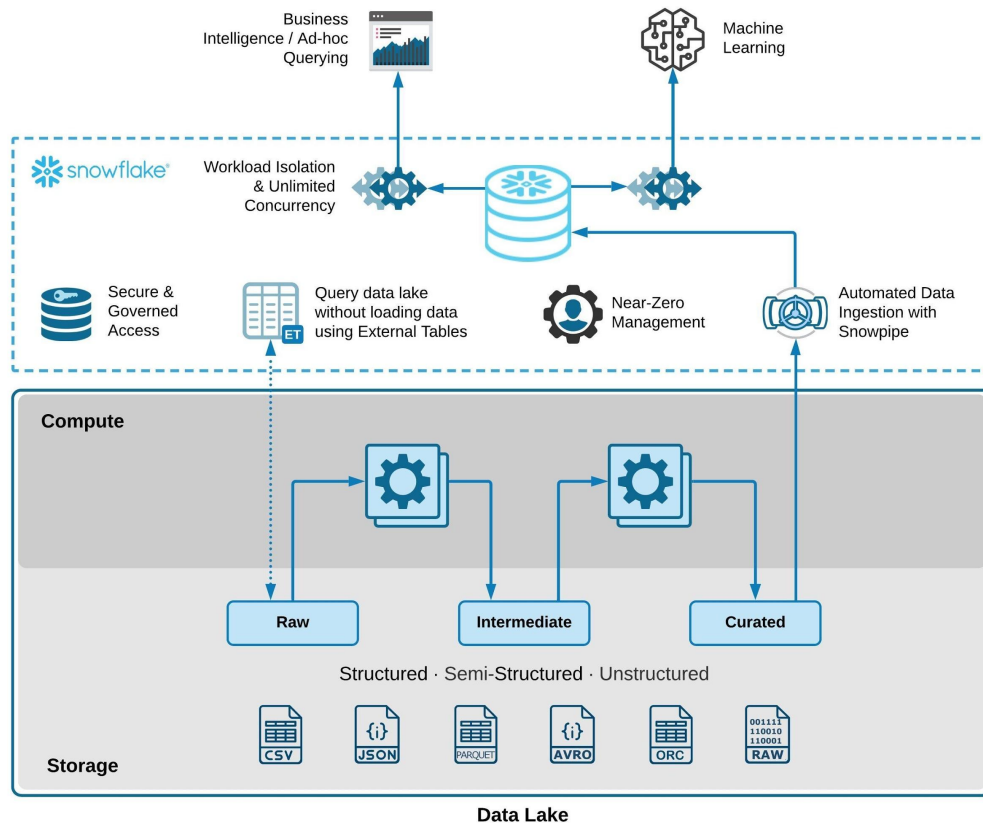Faster queries, fewer compute credits

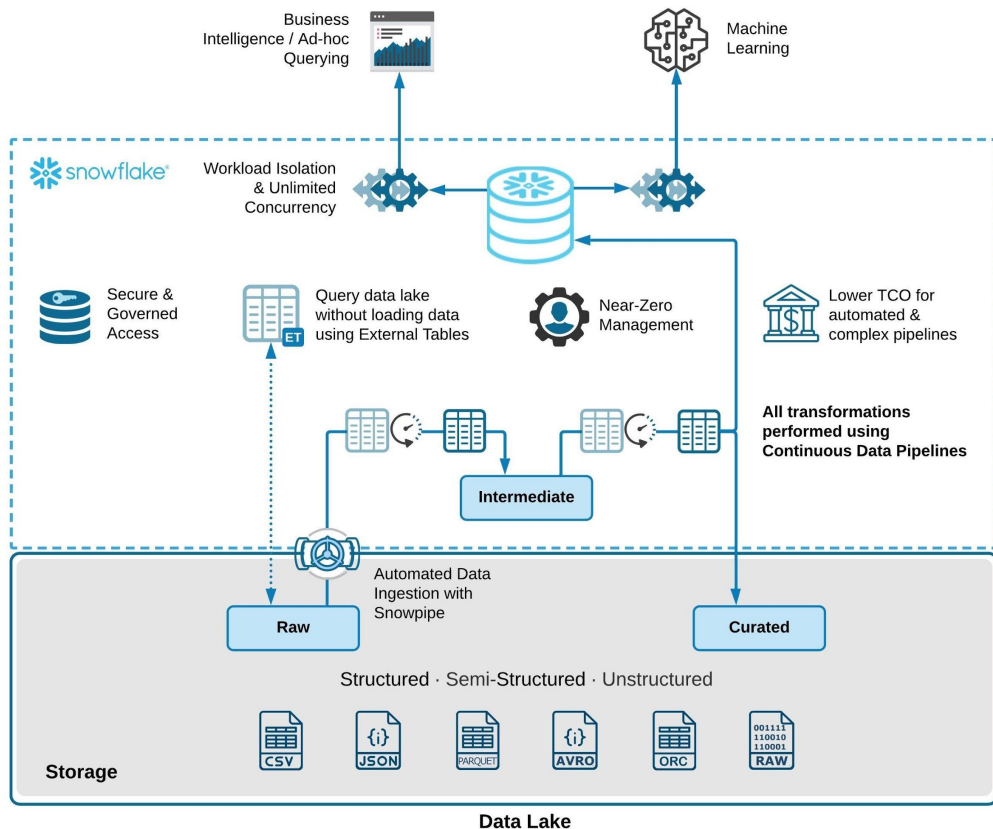**MV on External Tables**
Fast, governed access to Data Lake *

*Must automate / schedule a refresh

REFERENCE ARCHITECTURE #1:
Snowflake as Fast Query Layer

REFERENCE ARCHITECTURE #2:
Snowflake as Fast Query Layer + Transformation

Complement Your Data Lake with Snowflake

Build Your Data Lake with Snowflake

DATA LAKE WITH SNOWFLAKE

1	2	3

WORKFLOW

Data Ingestion

Data Transformation

Consumption

DATA LAKE WITH SNOWFLAKE

1   2   3

WORKFLOW

Data Ingestion   **Data Transformation**   **Consumption**
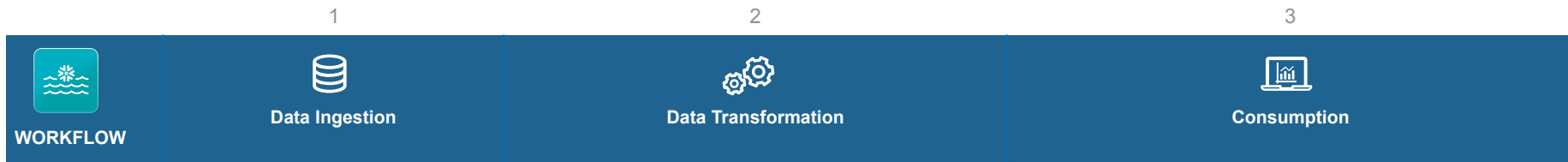
SNOWFLAKE CLOUD DATA PLATFORM

One platform, one copy of data, many workloads

Secure and governed access to all data

Near-zero maintenance, as a service

Unlimited performance and scale

DATA LAKE WITH SNOWFLAKE

CLICK ON NAV BAR TO GO TO DEEP DIVE

	1	**2**	**3**
WORKFLOW	**Data Ingestion**	**Data Transformation**	**Consumption**

DATA LAKE FEATURES

- Auto load stage data using **Snowpipe**
- Highly performant, cloud-based storage ingest
- Ingest streaming data using **Kafka connector**
- **Storage compression**

PLATFORM FEATURES

- Unmatched support for **semi-structured data** with **schema-on-read**

SNOWFLAKE CLOUD DATA PLATFORM

- **One platform, one copy of data, many workloads**
- **Secure and governed access to all data**
- **Near-zero maintenance, as a service**
- **Unlimited performance and scale**

DATA LAKE WITH SNOWFLAKE

WORKFLOW	1 Data Ingestion	2 Data Transformation	3 Consumption
DATA LAKE FEATURES	Auto load stage data using **Snowpipe** Highly performant, cloud-based storage ingest Ingest streaming data using **Kafka connector** **Storage compression**	Fast & simplified transformation using **ANSI SQL** Continuous transformation with **Streams & Tasks** Enhanced extensibility with **functions and stored procedures**	
PLATFORM FEATURES	Unmatched support for **semi-structured data** with **schema-on-read**	Optimized pipeline performance with **Instant warehouse resizing** Eliminate data duplication with **Zero Copy Cloning**	

SNOWFLAKE CLOUD DATA PLATFORM

- **One platform, one copy of data, many workloads**
- **Secure and governed access to all data**
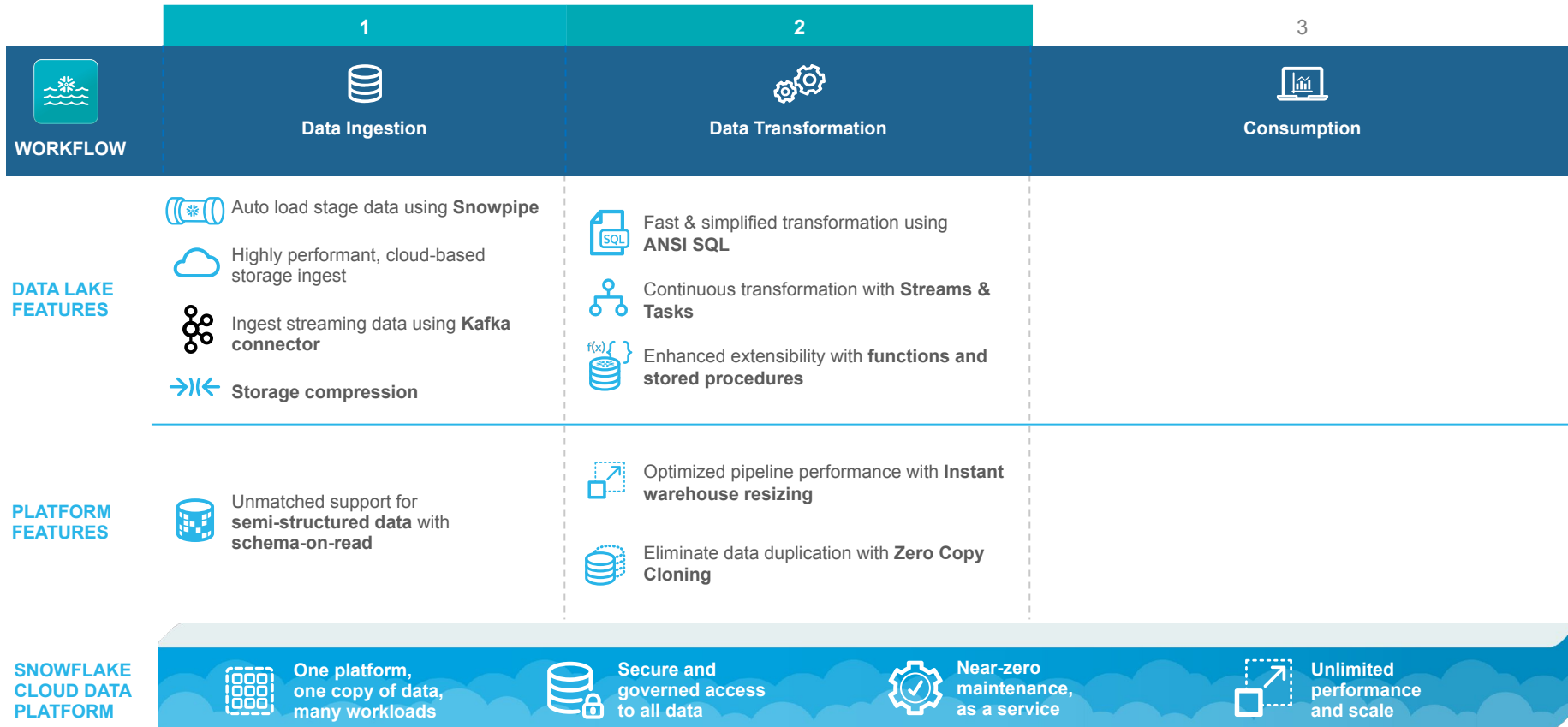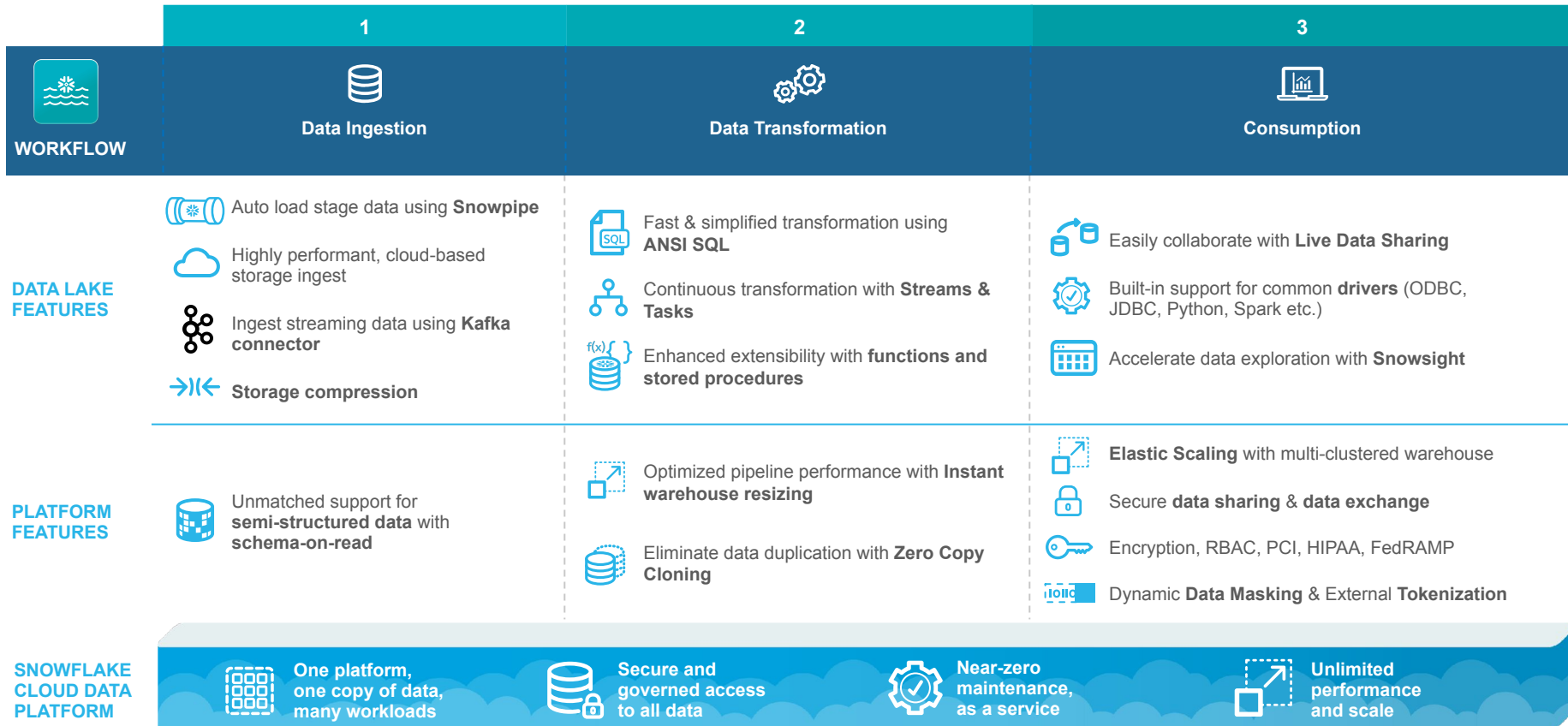- **Near-zero maintenance, as a service**
- **Unlimited performance and scale**

DATA LAKE WITH SNOWFLAKE

	1	2	3
WORKFLOW	**Data Ingestion**	**Data Transformation**	**Consumption**
DATA LAKE FEATURES	Auto load stage data using **Snowpipe** Highly performant, cloud-based storage ingest Ingest streaming data using **Kafka connector** **Storage compression**	Fast & simplified transformation using **ANSI SQL** Continuous transformation with **Streams & Tasks** Enhanced extensibility with **functions and stored procedures**	Easily collaborate with **Live Data Sharing** Built-in support for common **drivers** (ODBC, JDBC, Python, Spark etc.) Accelerate data exploration with **Snowsight**
PLATFORM FEATURES	Unmatched support for **semi-structured data** with **schema-on-read**	Optimized pipeline performance with **Instant warehouse resizing** Eliminate data duplication with **Zero Copy Cloning**	**Elastic Scaling** with multi-clustered warehouse Secure **data sharing** & **data exchange** Encryption, RBAC, PCI, HIPAA, FedRAMP Dynamic **Data Masking** & External **Tokenization**

SNOWFLAKE CLOUD DATA PLATFORM

- **One platform, one copy of data, many workloads**
- **Secure and governed access to all data**
- **Near-zero maintenance, as a service**
- **Unlimited performance and scale**

SHARE YOUR LAKE!

DATA MARKETPLACE

List

Request

① Share with Snowflake Customers via **Data Exchange**

Data Consumer

BI, Query Tools

External Tables

SHARE

View, Query, Join

DATA LAKE

Materialized View

SHARE

② Share with Non-Snowflake Customers

Reader Account

BI, Query Tools

REFERENCE ARCHITECTURE #1:
Snowflake Platform

REFERENCE ARCHITECTURE

Apps & Services

Analytical Queries

Operational Data

Event Data

Database

OLTP

NoSQL

MySQL
PostgreSQL
mongoDB
cassandra

Streaming Service

kafka
Amazon Kinesis
Azure Event Hub
Cloud Pub/Sub

Cloud Object Storage

Google Cloud Storage
Amazon S3
Azure Blob Storage

ETL

snowflake

Transformation Using Streams & Tasks

Query object storage using External Tables

Materialize the external table to improve query performance

Unload to object storage with Partition By

Business Intelligence / Ad-hoc Querying

Machine Learning

MV

ET

| Unlimited performance and scale | Near-zero maintenance, as a service | Workload Isolation ensures unlimitied concurrency |

Azure   aws   [Google Cloud]

SNOWFLAKE: SECURITY AT MULTIPLE LEVELS

You have security features built in to protect the data you load in and use in Snowflake

Snowflake uses the industry standard "shared responsibility" model

Snowflake uses many sophisticated mechanisms to keep the platform safe and stable

Orchestrated security built into the fabric of the platform, automated controls in place for all functions, constant monitoring and analysis to detect and mitigate threats quickly

PROTECTING YOUR DATA IN SNOWFLAKE

End-to-End Encryption

Always-encrypted client communications, plus integration with cloud provider private networking

Fully Encrypted Storage

Data at rest is always encrypted while handled by the Snowflake drivers and systems

Strong Authentication

Built in multi-factor, integration with your federated SSO, easy user management

Full Auditing

Track every login, every transaction, every data transfer, and export to your security tools

Role-Based Access Control

All objects, actions, and even compute usage can be controlled with roles

Recovery

We give you options to ensure your data can be recovered in case of an accident or worse

Snowflake Security Product Documentation

SNOWFLAKE WILL BE 100% TRANSPARENT

Third Party Attestations and Certifications

Snowflake Security and Trust Center: https://www.snowflake.com/product/security-and-trust-center/
Snowflake Security Policy: https://www.snowflake.com/legal/ (first link)

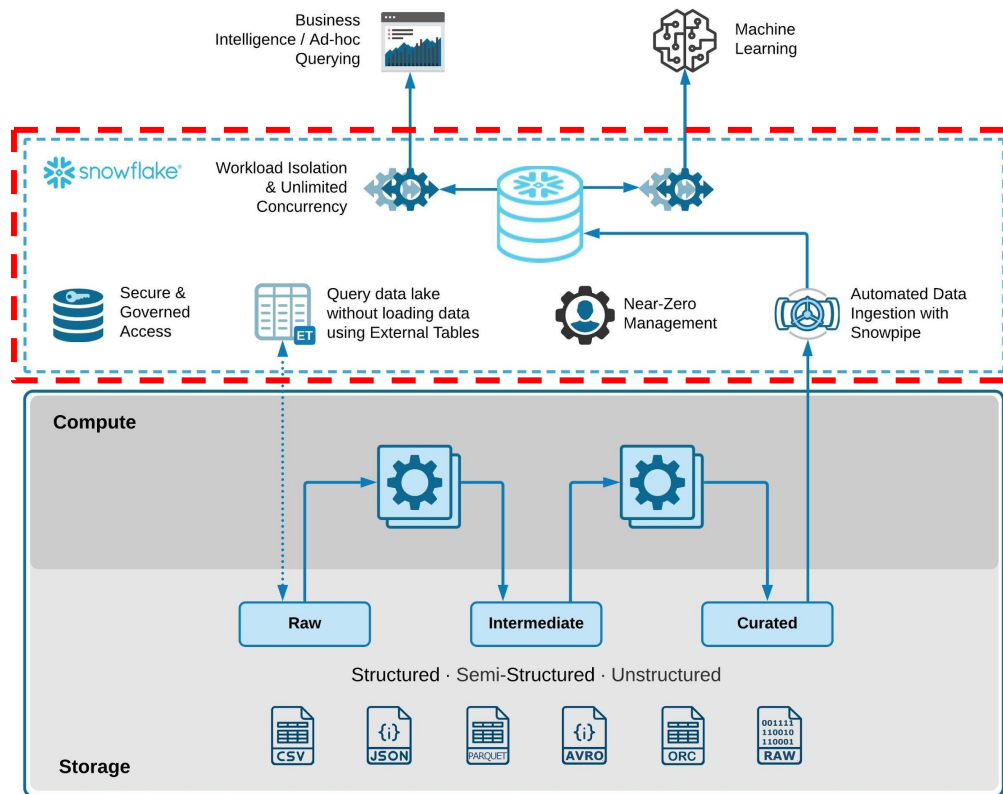| SOC 2 Type II	PCI-DSS	ISO/IEC 27001	FedRAMP	HIPAA
12 Month Coverage Period			(Available from OMB MAX)	**HI**TRUST Certified
SOC 1 Type II				
6 Month Coverage Period				

Self-Assessments
CAIQ, SIG Lite, Pen Test Results

PCI, HIPAA, ISO 27001 etc.

PCI, HIPAA, ISO 27001 etc.

Business Intelligence / Ad-hoc Querying

Machine Learning

snowflake®

Workload Isolation & Unlimited Concurrency

Secure & Governed Access

Cost effective Storage

Near-Zero Management

Lower TCO for automated & complex pipelines

All transformations performed using Continuous Data Pipelines

Raw

Intermediate

Curated

Structured · Semi-Structured · Unstructured

CSV

{i} JSON

PARQUET

{i} AVRO

ORC

001111 110010 110001 RAW

Storage

Batch

Streaming