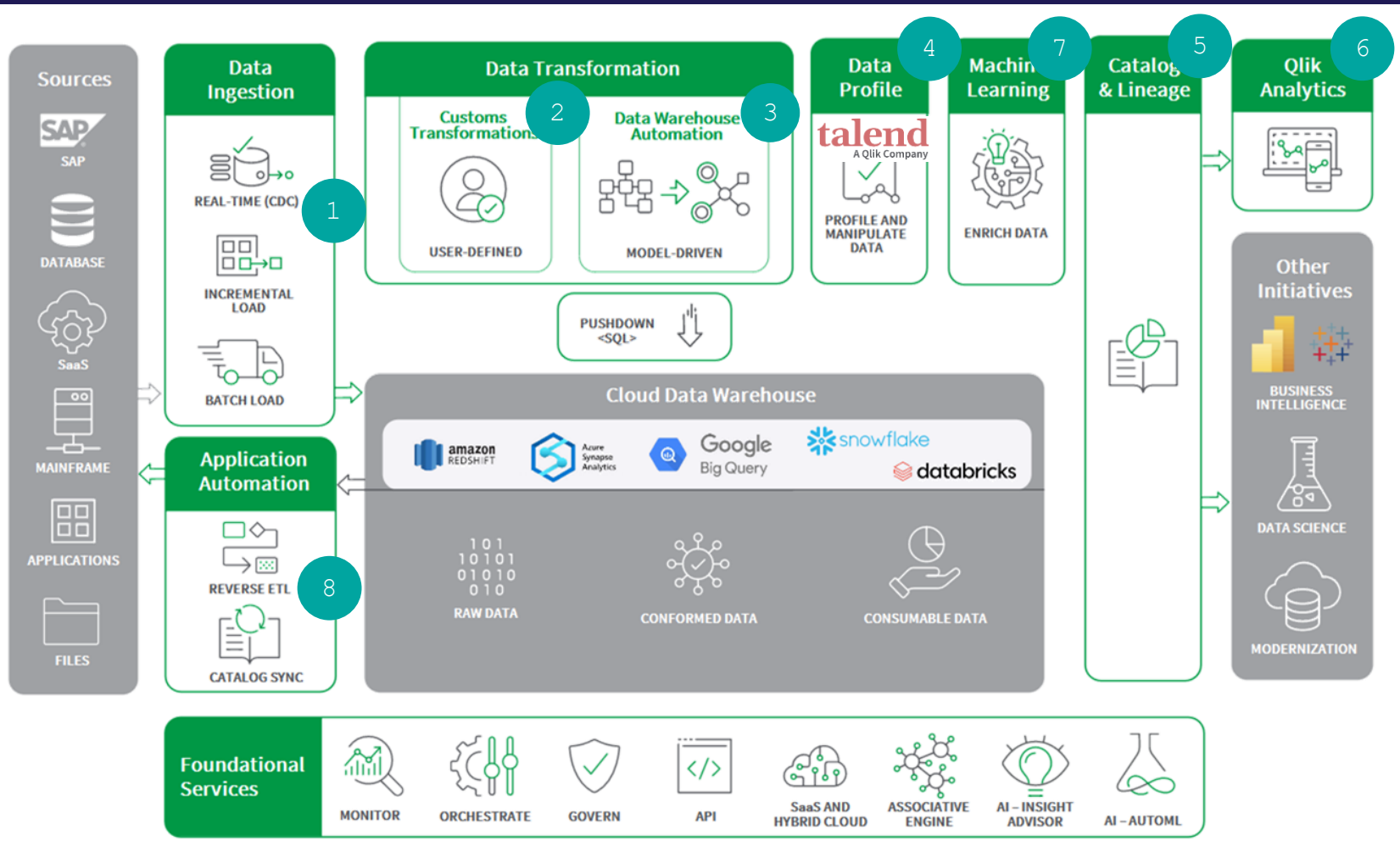
Data Management Track: Snowflake

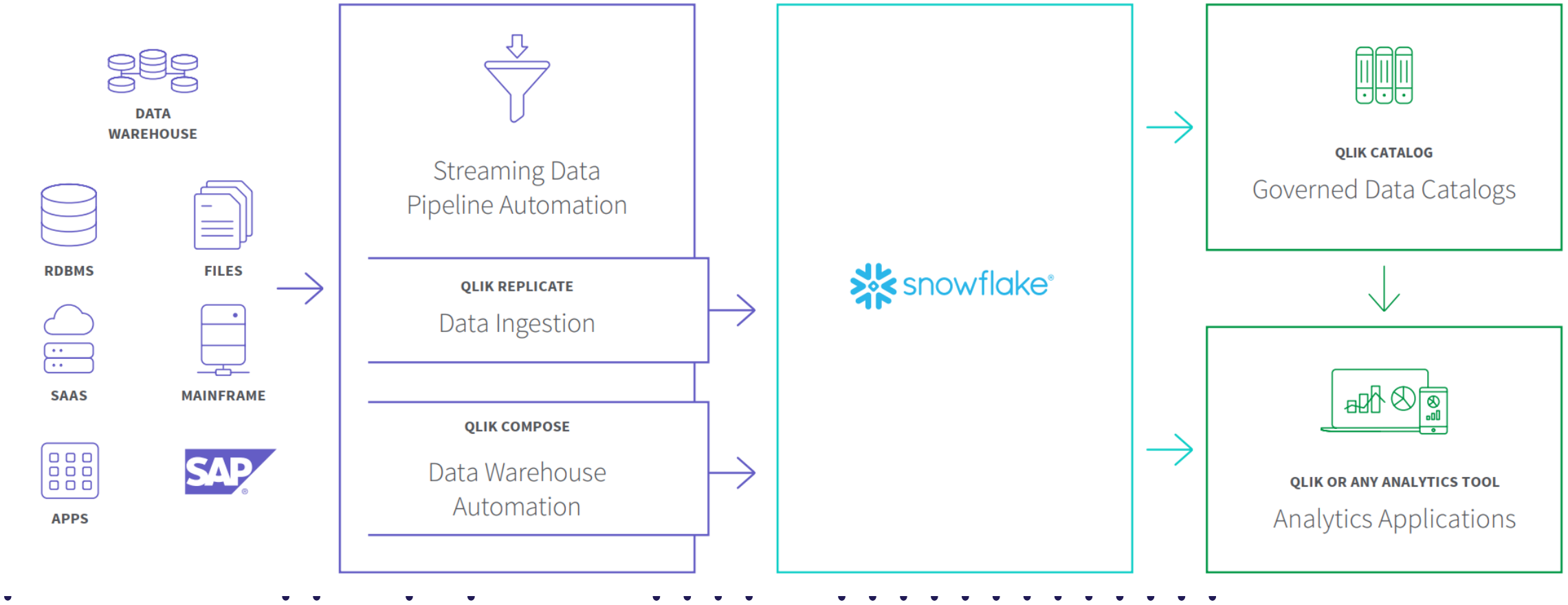
MĀRIS SVILĀNS, Infotrust

Tying the knots: Infotrust and Qlik



1. Real-Time Data Ingestion
2. Custom Transformation
3. Data Warehouse Automation
4. Data Profiling
5. Catalog & Lineage
6. Analytics and Other Initiatives
7. Machine Learning to Enrich Data
8. Reverse ETL

Qlik and Snowflake at a glance



Requirements of a cloud data platform

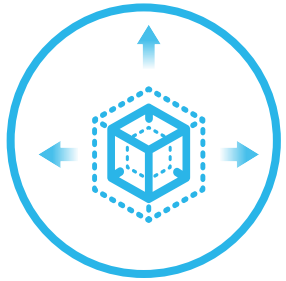
One Platform

One Copy of Data

Many Workloads

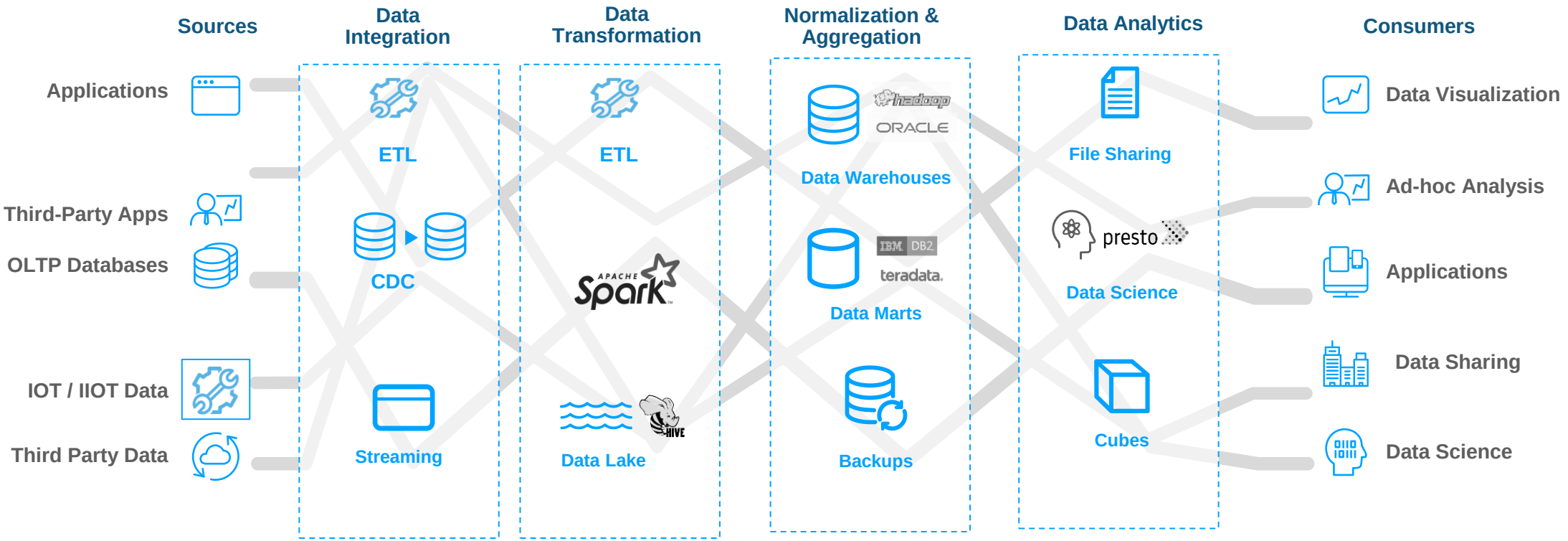
Secure & Governed Access to All Data

Near-zero Maintenance, as a Service



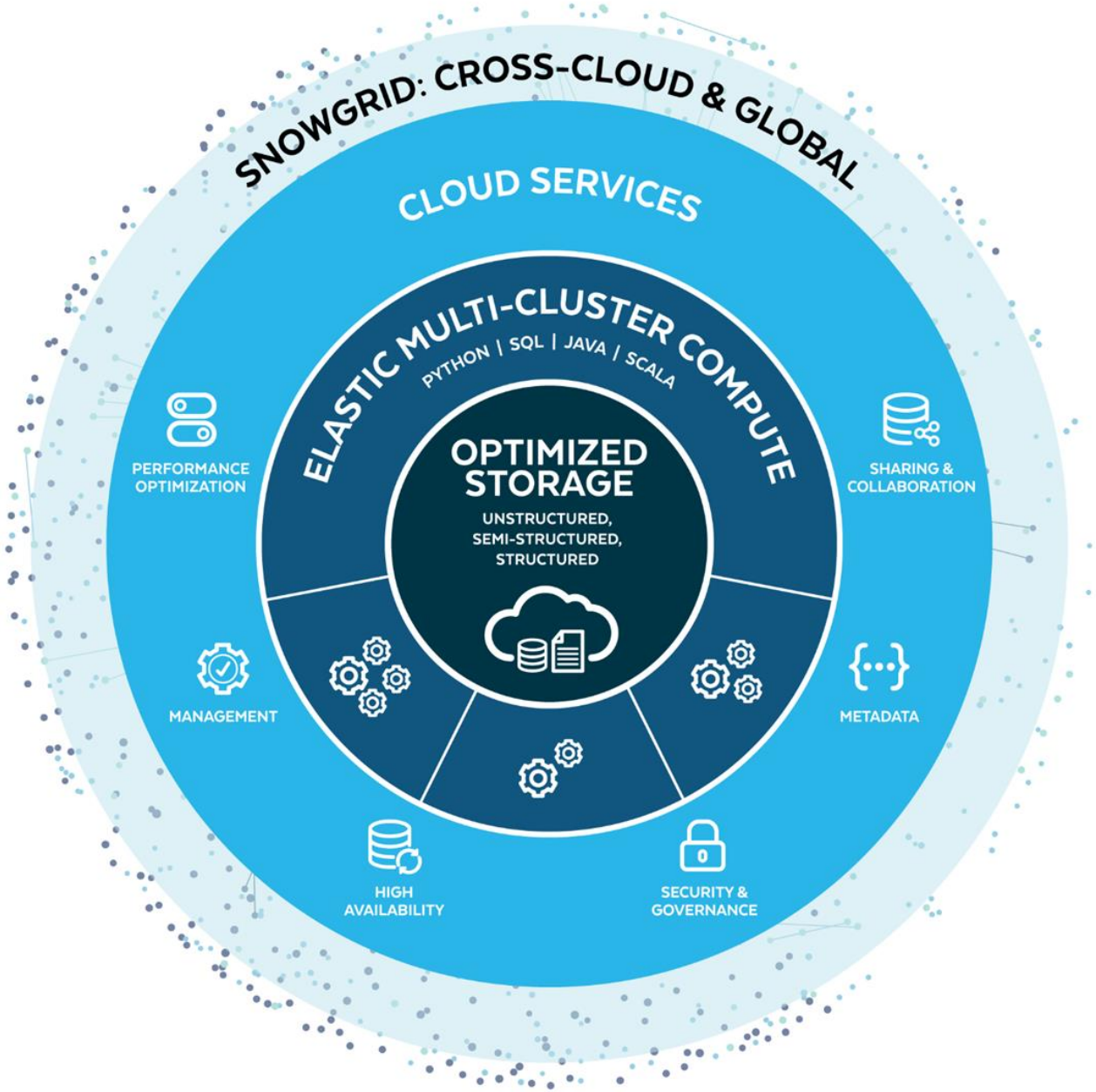
Unlimited Performance and Scale

Challenges with traditional data architecture



Modern data architecture with Snowflake

SNOWFLAKE PLATFORM ARCHITECTURE



ETL/ELT

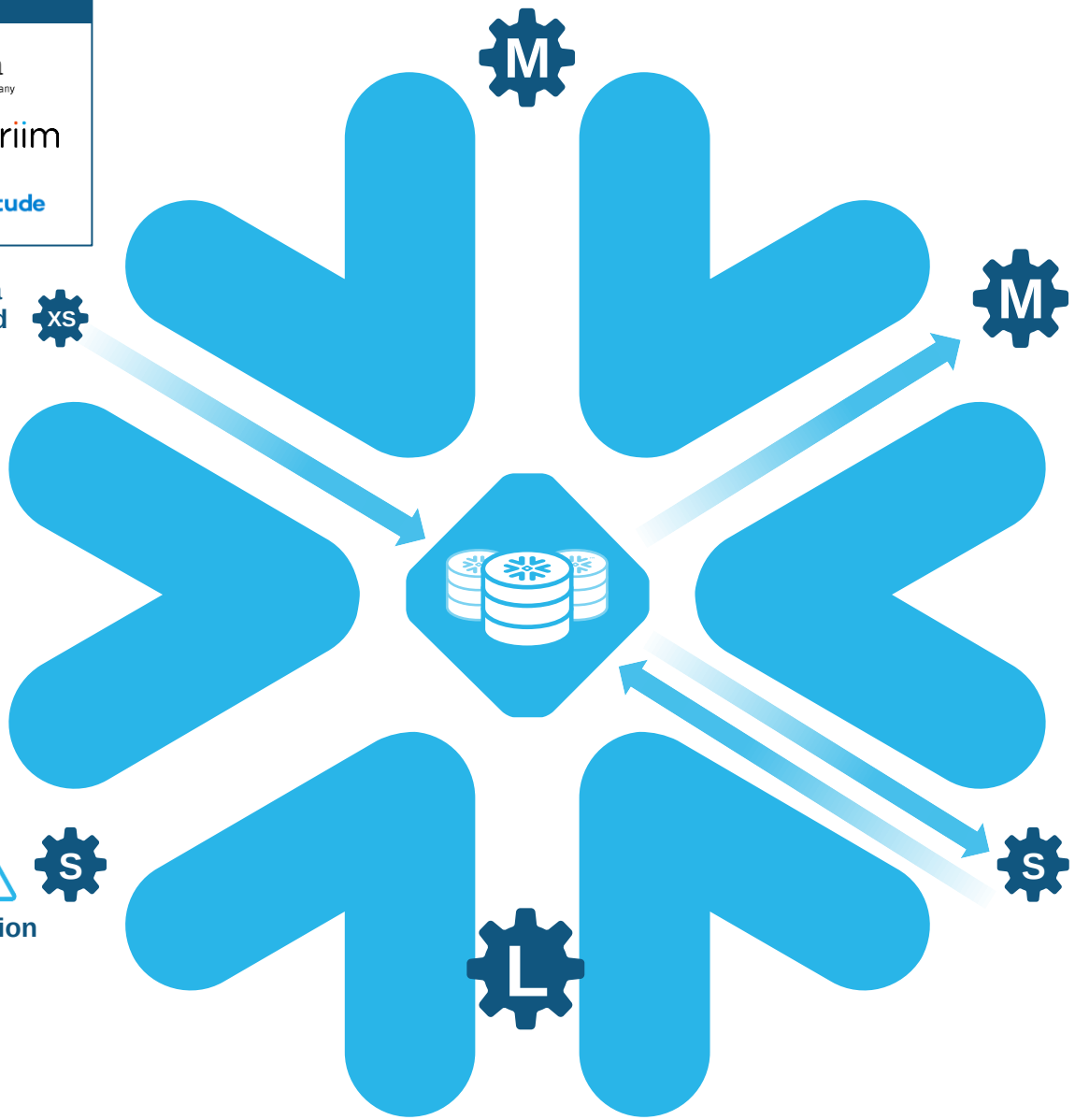
Informatica · talend · Stitch (A Talend Company) · Fivetran · MATILLION · Qlik · striim · Segment · StreamSets · Amplitude

Marketing Analytics/Reporting/BI

SIGMA · ThoughtSpot · SISENSE · MODE · Power BI · Periscope Data (by Sisense) · Qlik · looker · tableau

Structured
Semi-Structured
Unstructured

Data Load



Data Transformation

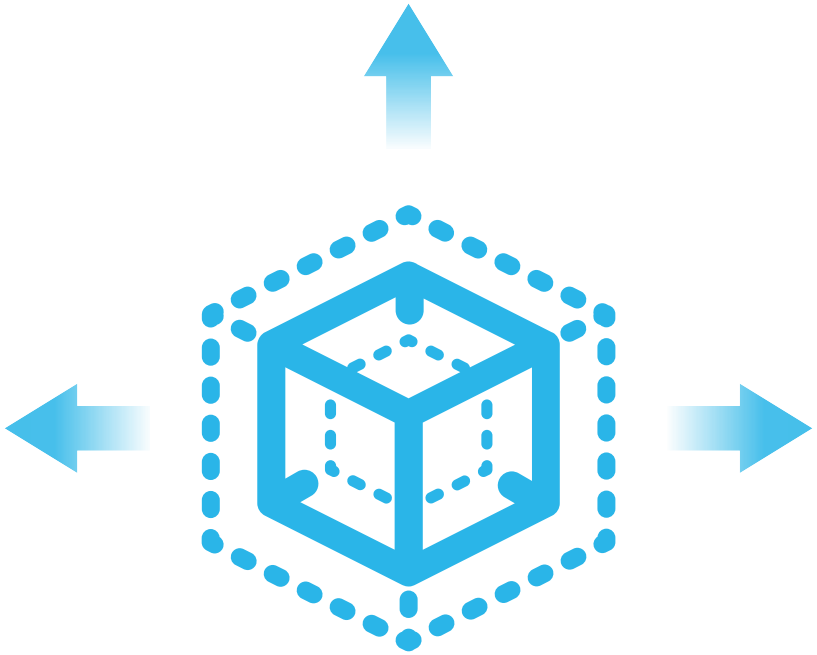
Intelligent Infrastructure:

- Logical Model
- Security
- Query Planning & Optimization
- Transactional Control

Data Science

alteryx · DataRobot · data iku · Amazon SageMaker · DOMINO · H2O.ai

3 Dimensions of Scaling



ACROSS

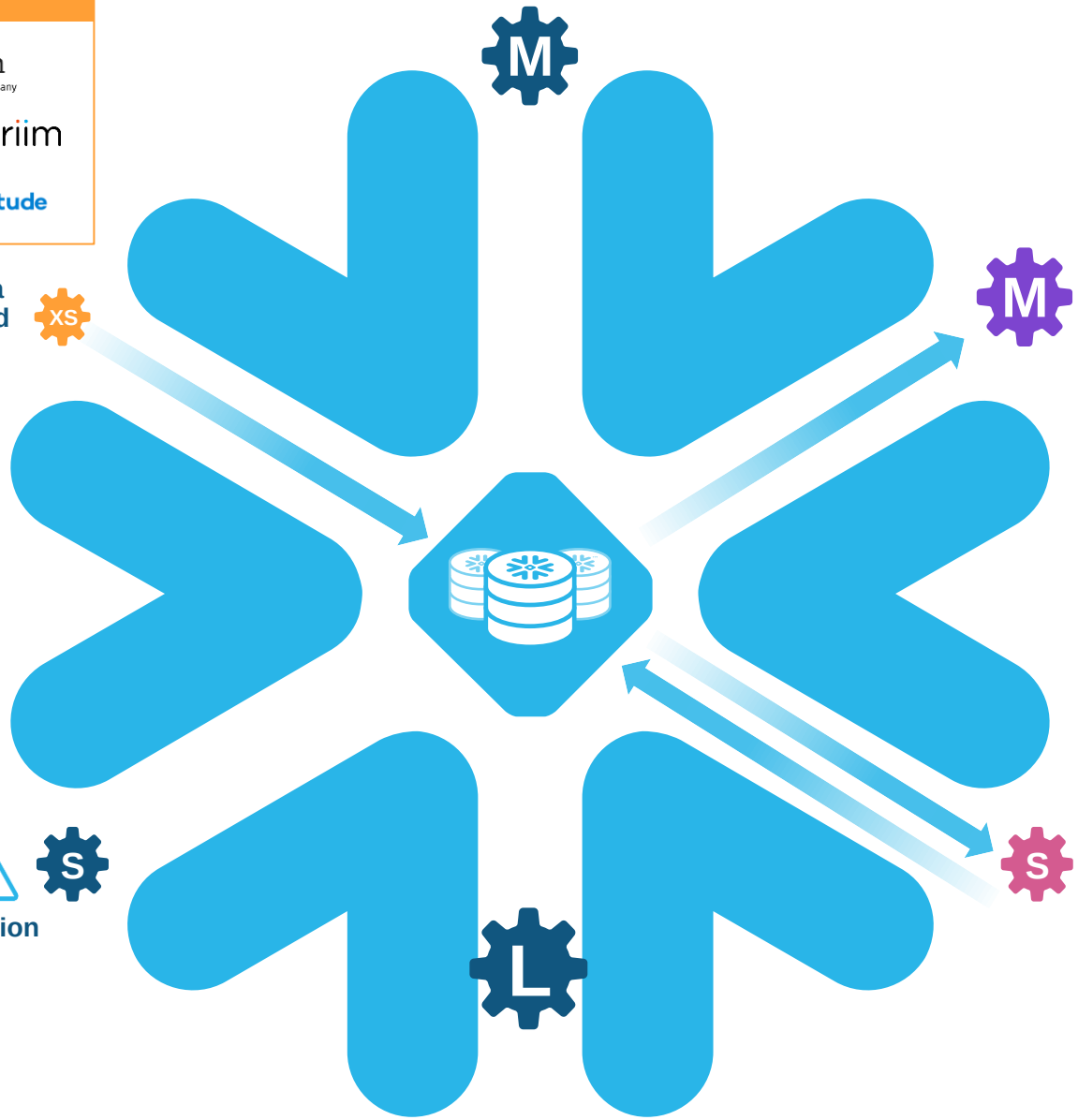
- Many competing workloads
- Resource contention
- Isolate on separate warehouses

ETL/ELT

Informatica · talend · Stitch (A Talend Company) · Fivetran · MATILLION · Qlik · striim · Segment · StreamSets · Amplitude

Structured
Semi-Structured
Unstructured

Data Load



Data Transformation

Marketing Analytics/Reporting/BI

SIGMA · ThoughtSpot · SISENSE · MODE · Power BI · Periscope Data by Sisense · Qlik · looker · tableau

Data Science

alteryx · DataRobot · data iku · Amazon SageMaker · DOMINO · H_2O.ai

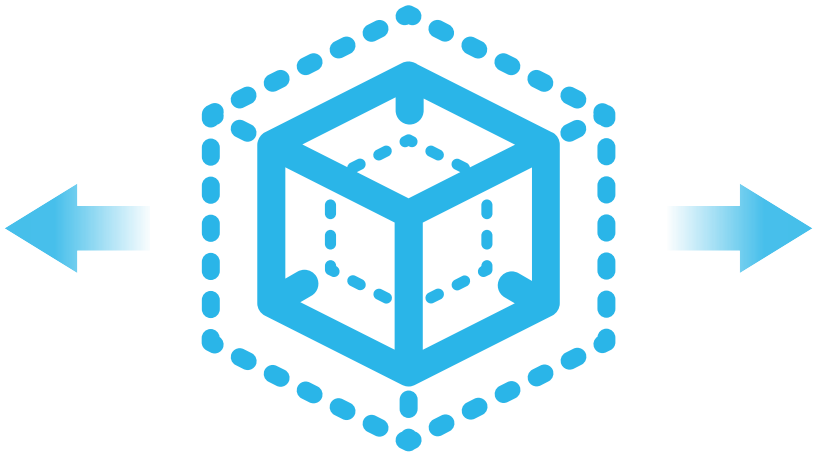
Intelligent Infrastructure:

- Logical Model
- Security
- Query Planning & Optimization
- Transactional Control

3 Dimensions of Scaling

UP

- Single query performance
- More data, more complex queries
- Add more servers to the cluster



ACROSS

- Many competing workloads
- Resource contention
- Isolate on separate warehouses

ETL/ELT

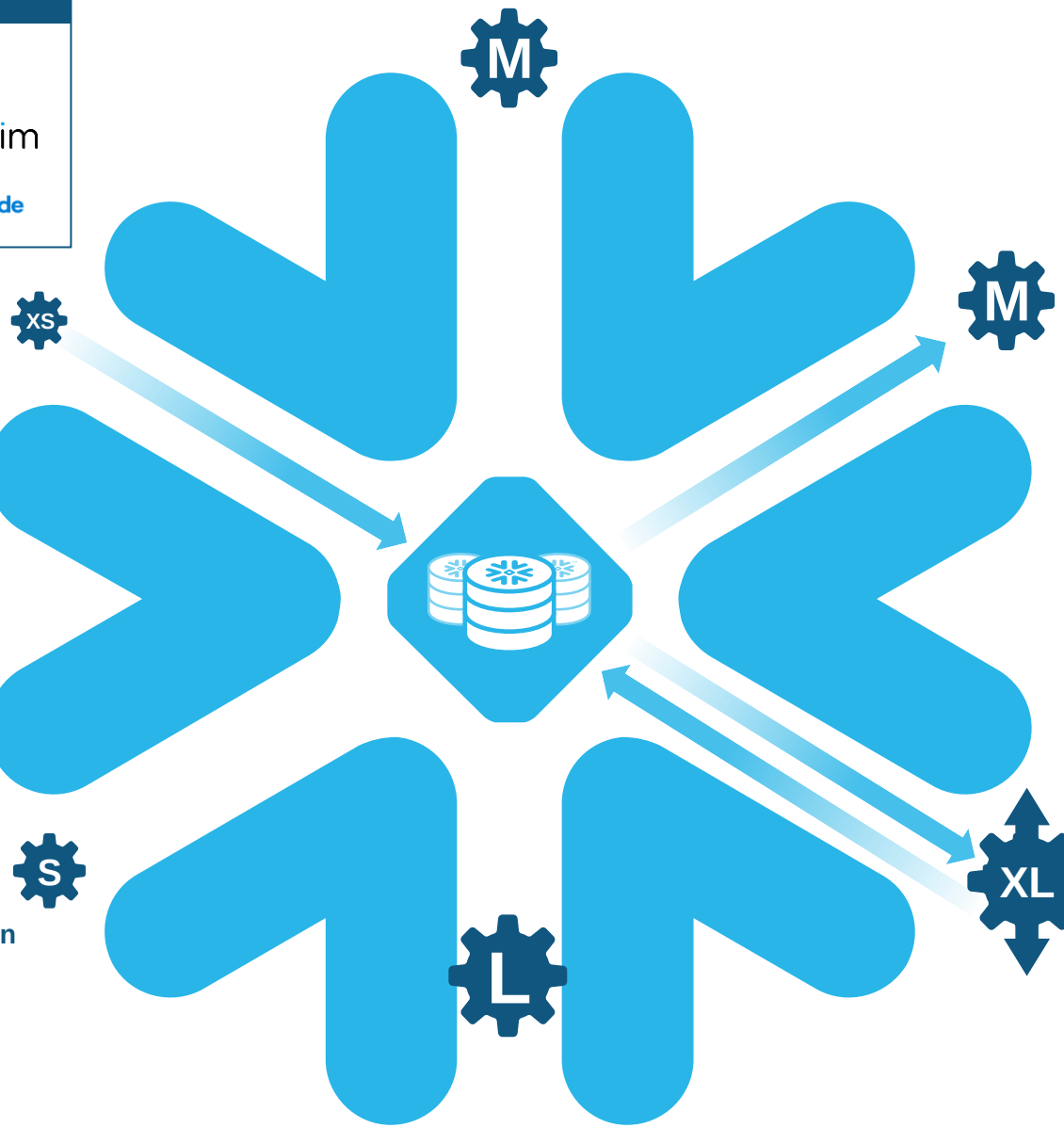
Informatica · talend · Stitch (A Talend Company) · Fivetran · MATILLION · Qlik · striim · Segment · StreamSets · Amplitude

Marketing Analytics/Reporting/BI

SIGMA · ThoughtSpot · SISENSE · MODE · Power BI · Periscope Data (by Sisense) · Qlik · looker · tableau

Structured
Semi-Structured
Unstructured

Data Load



Data Transformation

Intelligent Infrastructure:

- Logical Model
- Security
- Query Planning & Optimization
- Transactional Control

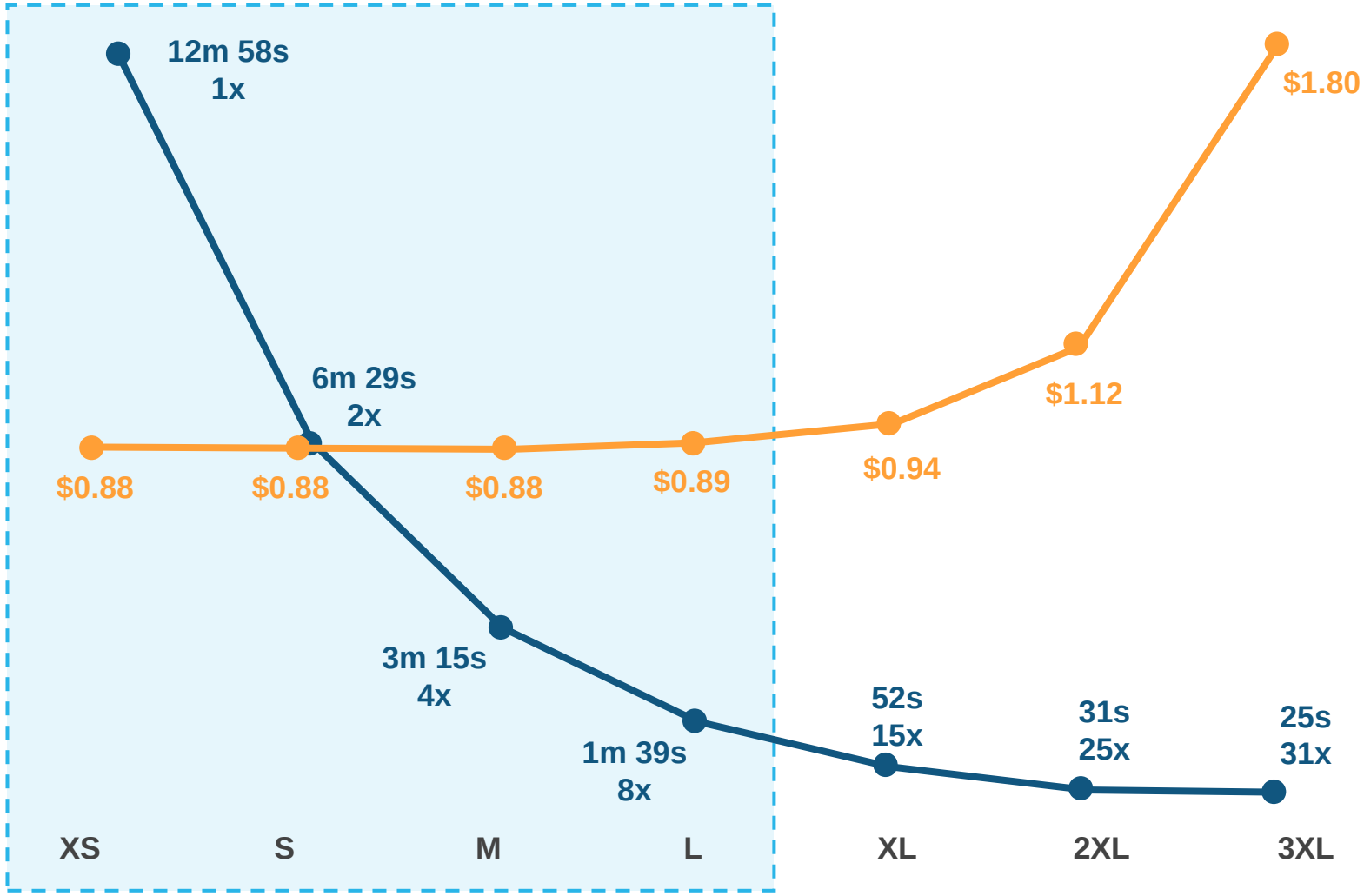
Data Science

alteryx · DataRobot · data iku · Amazon SageMaker · DOMINO · H2O.ai

Scale Up – Loading 1BN Records

- Doubling the number of servers halves the run time
- But you pay per-server, per second of compute
- So you get your answer

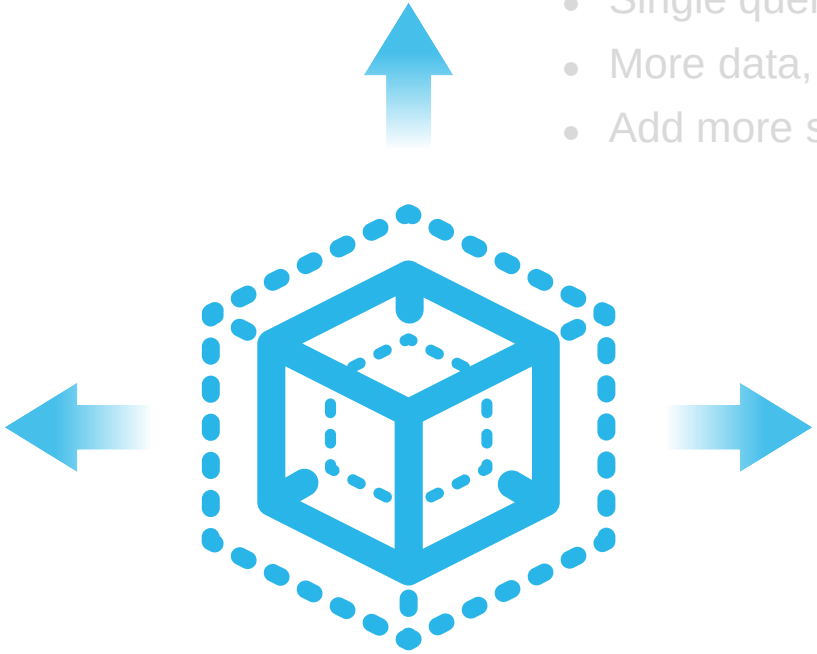
8X FASTER FOR THE SAME COST



3 Dimensions of Scaling

UP

- Single query performance
- More data, more complex queries
- Add more servers to the cluster



ACROSS

- Many competing workloads
- Resource contention
- Isolate on separate warehouses

Out

- More users
- More queries simultaneously
- Spin up more clusters

ETL/ELT

Informatica · talend · Stitch (A Talend Company) · Fivetran · MATILLION · Qlik · striim · Segment · StreamSets · Amplitude

Marketing Analytics/Reporting/BI

SIGMA · ThoughtSpot · SISENSE · MODE · Power BI · Periscope Data by Sisense · Qlik · looker · tableau

Structured
Semi-Structured
Unstructured

Data Load

Multi-cluster

Data Transformation

Data Science

alteryx · DataRobot · data iku · Amazon SageMaker · DOMINO · H2O.ai

Intelligent Infrastructure:

- Logical Model
- Security
- Query Planning & Optimization
- Transactional Control

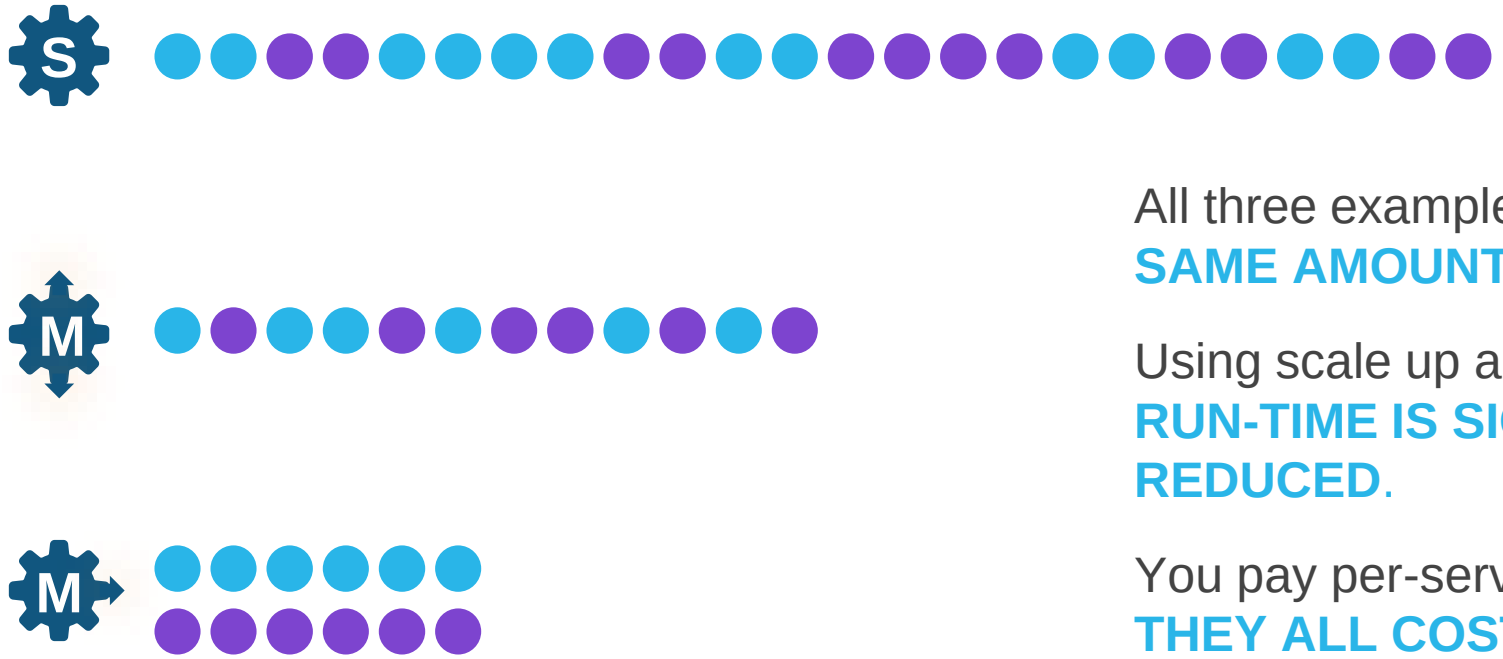
Scale Out: Multi-cluster Warehouses

Scale for Concurrency

Scale for Performance

	1	2	3	4	5	6	7	8	9	10
4XL	128	256	384	512	649	768	896	1024	1152	1280
3XL	64	128	192	256	320	384	448	512	576	640
2XL	32	64	96	128	160	192	224	256	288	320
XL	16	32	48	64	80	96	112	128	144	160
L	8	16	24	32	40	48	56	64	72	80
M	4	8	12	16	20	24	28	32	36	40
S	2	4	6	8	10	12	14	16	18	20
XS	1	2	3	4	5	6	7	8	9	10

All Together - Scale, Elasticity, Cost



All three examples contain the **SAME AMOUNT OF WORK**.

Using scale up and scale out, total **RUN-TIME IS SIGNIFICANTLY REDUCED**.

You pay per-server, per-second so **THEY ALL COST THE SAME**.

ETL/ELT

Informatica · talend · Stitch (A Talend Company) · Fivetran · MATILLION · Qlik · striim · Segment · StreamSets · Amplitude

Marketing Analytics/Reporting/BI

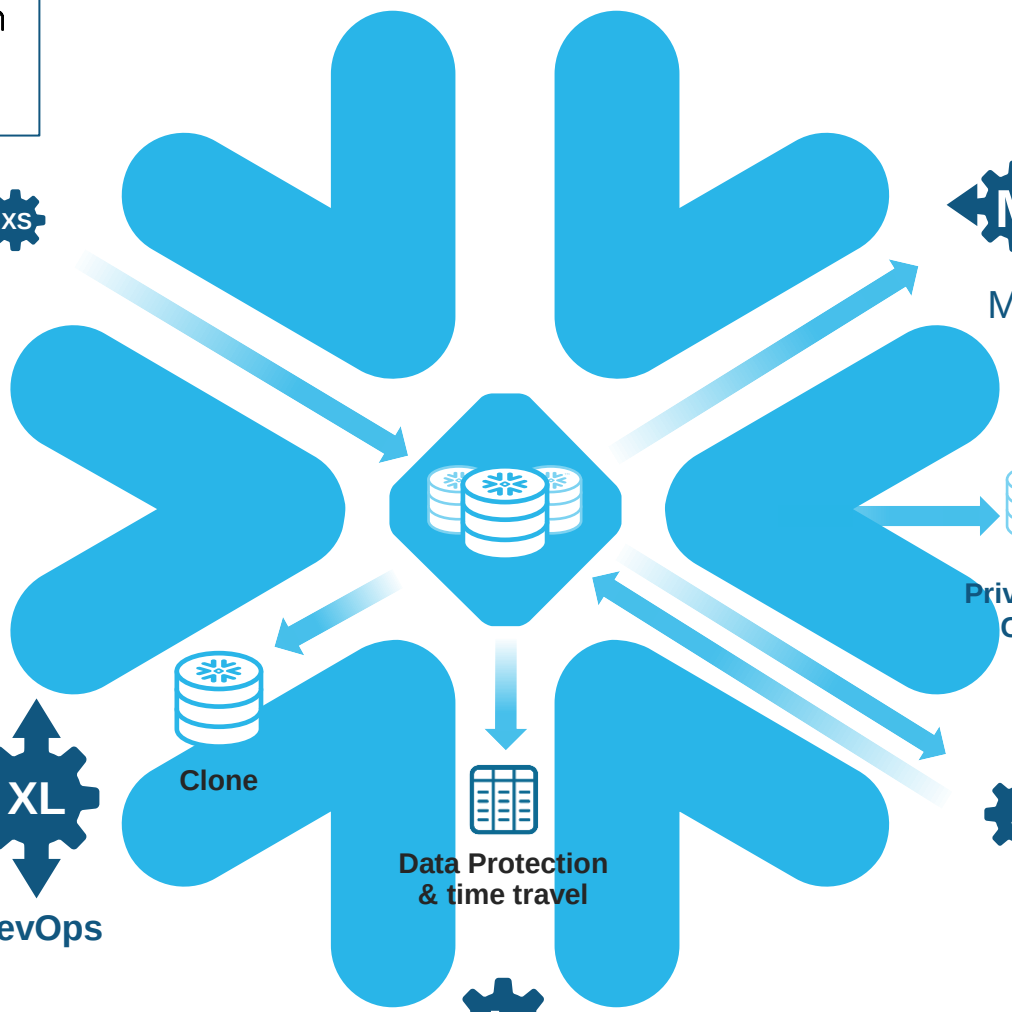
SIGMA · ThoughtSpot · SISENSE · MODE · Power BI · Periscope Data by Sisense · Qlik · looker · +ableau

Structured
Semi-Structured
Unstructured

Data Load

XS

Multi-cluster



Your Business Ecosystem

Privacy-Preserving Collaboration

Snowflake Marketplace

Your Teams & Internal Users

Data Transformation

DevOps

Clone

Data Protection & time travel

Finance/DBAs

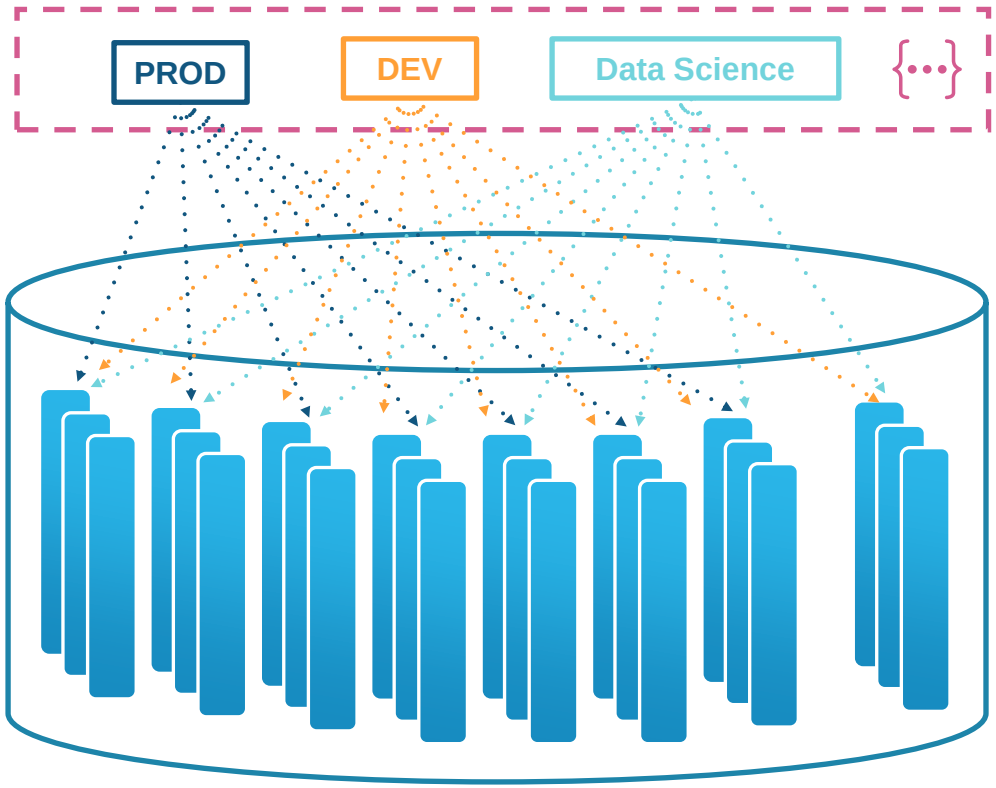
Intelligent Infrastructure:

- Logical Model
- Security
- Query Planning & Optimization
- Transactional Control

Data Science

alteryx · DataRobot · data iku · Amazon SageMaker · DOMINO · H2O.ai

Zero-Copy Cloning



The Metadata layer keeps track of every micro-partition file in every customer database.

Creating a DEV environment usually means copying the PROD database

Limited to subset of full Prod

Up to 2x storage requirement

Periodic refreshes

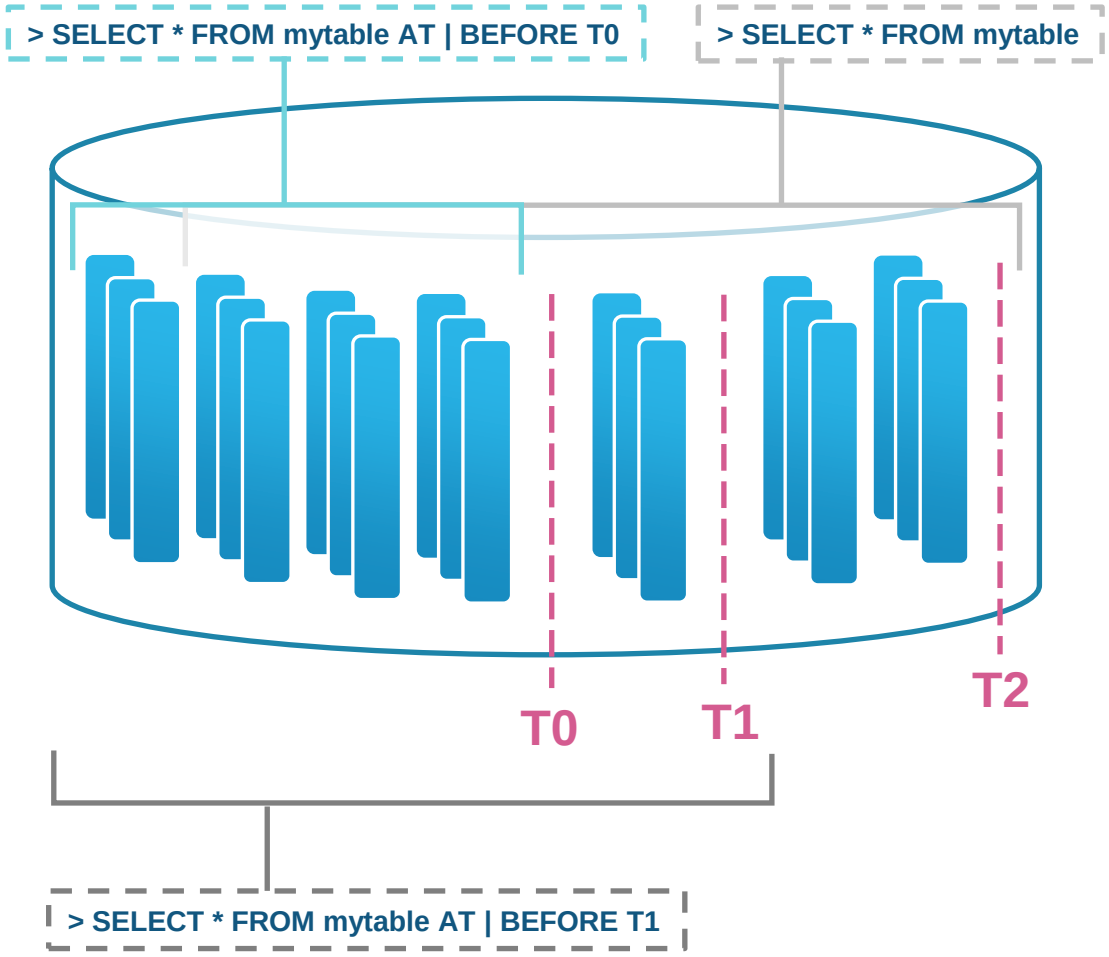
Snowflake Zero-Copy Clones

Simply “point” to the same files

Consumes zero additional storage

Changes to either DB are isolated

Time Travel



T0 – Initial state of database

T1 – `update myTable set colX = Y where…`

T2 – ELT job loads new data

Previous versions of data automatically retained

AT | BEFORE [timestamp | statement | offset]

CLONE AT | BEFORE to recreate a prior version

UNDROP recovers from accidental deletion

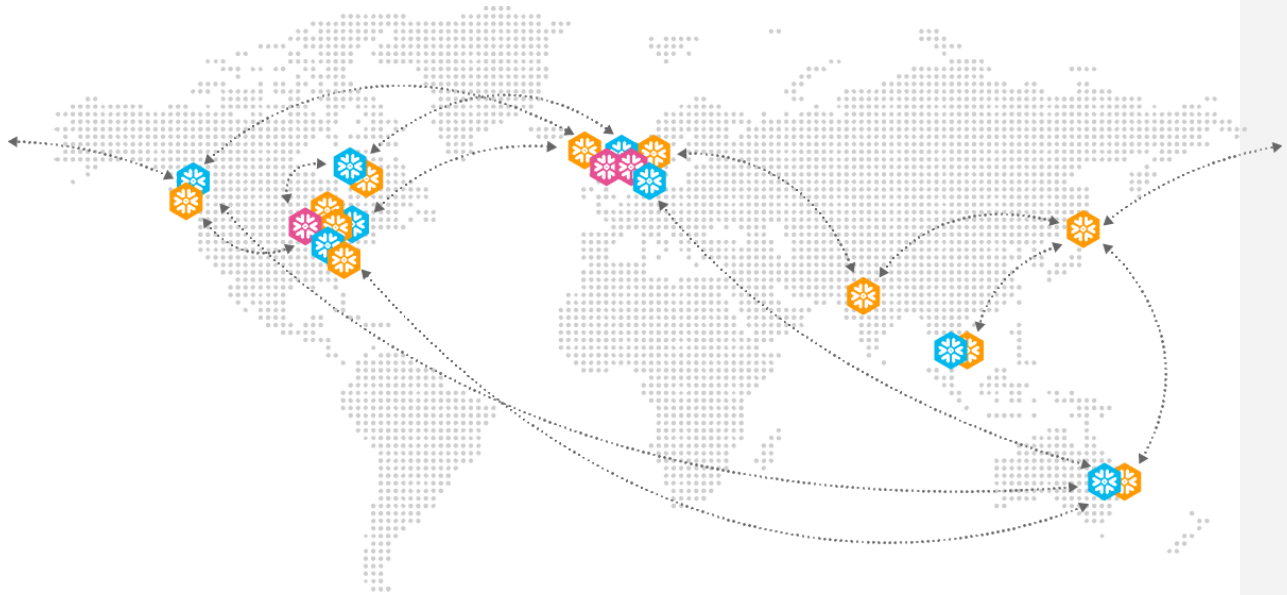
Accessed via SQL extensions

AT | BEFORE [timestamp | statement | offset]

CLONE AT | BEFORE to recreate a prior version

UNDROP recovers from accidental deletion

SNOWGRID



Maintain global business continuity

Eliminate disruptions, deliver better experiences, and comply with changing regulations through unique cross-cloud, cross-region connectivity.

Share data with no ETL or silos

Remove the barriers to data, regardless of cloud, region, workload, or organizational domains. Get instant access and distribution through a single copy of data.

Cross-cloud governance controls

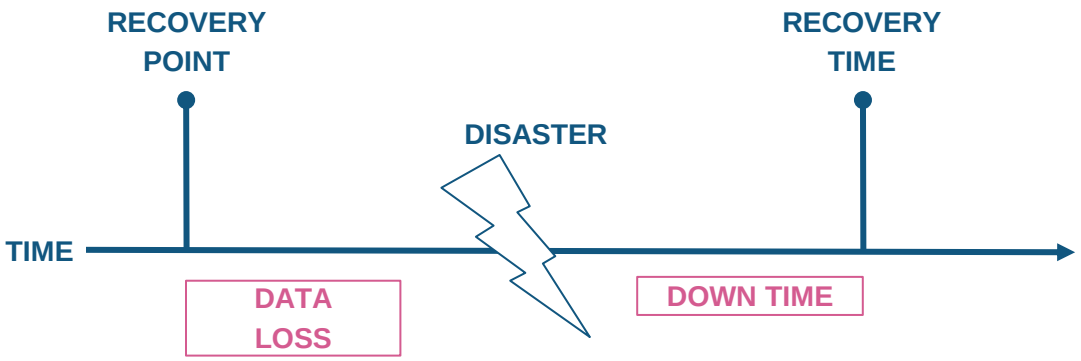
Simplify governance at scale with flexible policies that follow the data for consistent enforcement across users and workloads.

Tap into the extended ecosystem

Enrich insights with a network of third-party data. Discover and run new functions for extended workflows.

WHAT DO WE MEAN BY BUSINESS CONTINUITY?

1. Identify critical applications
2. Analyze business impact
3. Create plan & objectives



SNOWFLAKE ENABLES BUSINESS CONTINUITY

Failure	Mitigation
Customer Error	**Snowflake Features** Time Travel Fail-safe
Single Instance Failure	**Snowflake Built-in Redundancy** Triple-redundancy for critical services Automatic retries for failed parts of a query
Zone Failure	**Snowflake Built-in Redundancy** Using Availability Zones on AWS, Azure, GCP Using Availability Sets on Azure
Region Failure	**Snowflake Features** Cross-Region Database Replication Cross-Region Database Failover
Multi-Region Failure	**Snowflake Features** Cross-Cloud Database Replication Cross-Cloud Database Failover

DATABASE REPLICATION & FAILOVER

1. **Cross-Cloud & Cross-Region Replication**
   Business Continuity & Disaster Recovery
   Secure Data Sharing across regions/clouds
   Data Portability for Account Migrations

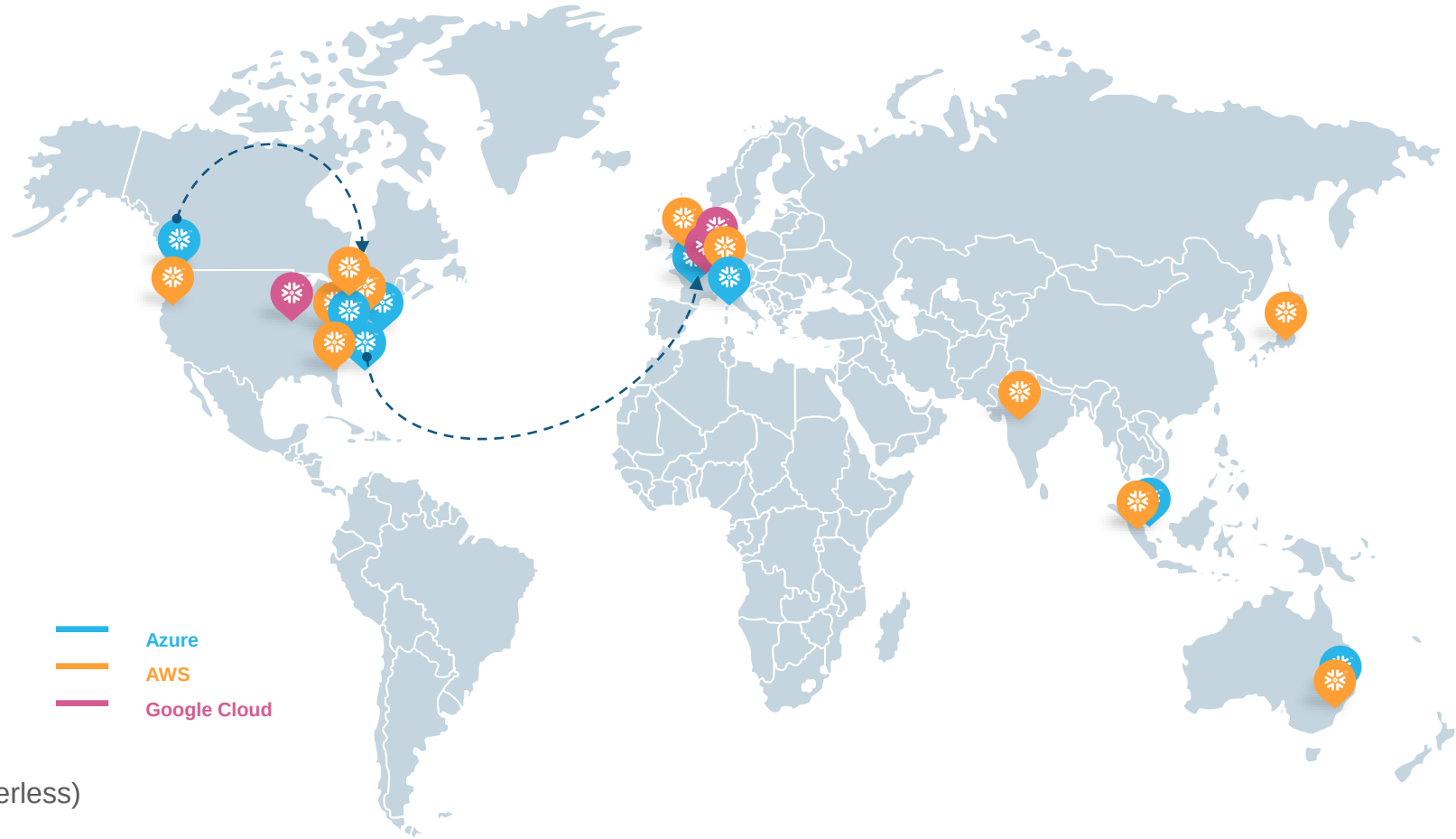
2. **Zero Performance Impact on Primary**
   Asynchronous Replication

3. **Reduced Data Loss**
   Incremental Refreshes

4. **Instant Recovery**
   Read: Readable Secondary Databases
   Write: Database Failover

5. **Secure**
   Data Encrypted at-rest & in-transit
   Tri-secret secure compatible

6. **Cost Effective**
   Replication Costs: Data Transfer & Compute (serverless)
   Control which databases to replicate



CLIENT REDIRECT

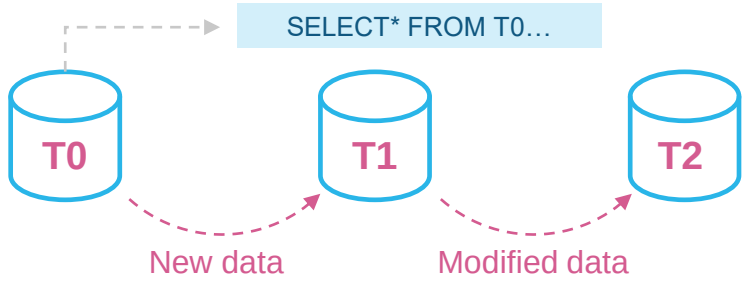
Redirect client connections to the region and cloud
le server-side command

BENEFITS

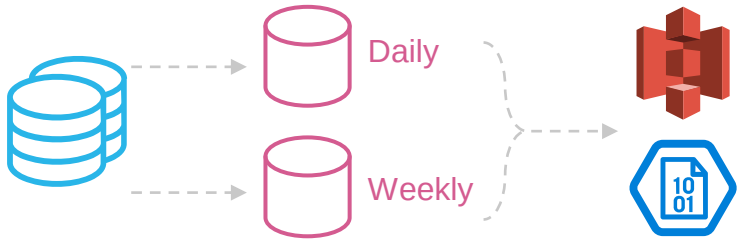
- **New Connection URL**
  Can be failed over across regions and clouds
- **Supports all clients**
  SnowSQL, Python, JDBC, ODBC, Go, Node.js, .NET, Snowflake UI
- **Redirects instantly**
  Connections redirect within 30-45 seconds
- **Supports PrivateLink**
  - Private connections routed via customer's DNS
  - Customer creates & updates CNAME

COMPREHENSIVE DATA PROTECTION

Protection against infrastructure failures
All data transparently & synchronously replicated
3+ ways across independent infrastructure



Protection against corruption & user errors
“Time travel” feature enables instant roll-back to
any point in time during chosen retention window



Long-term data protection
Zero-copy clones + optional export to cloud
object storage enable user-managed data copies

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

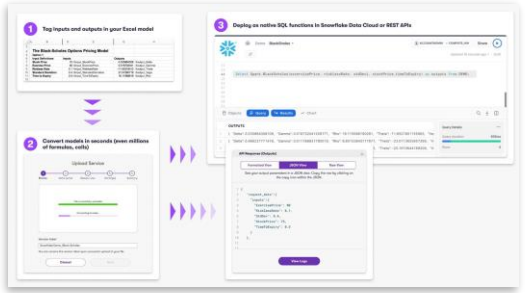
Recovery

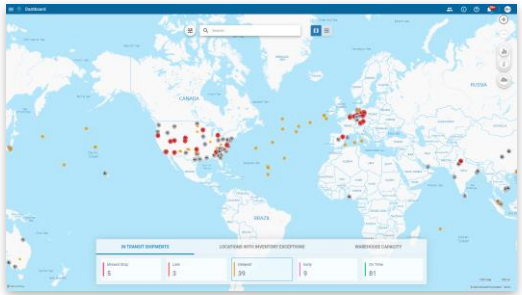
We give you options to ensure your data can be recovered in case of an accident or worse

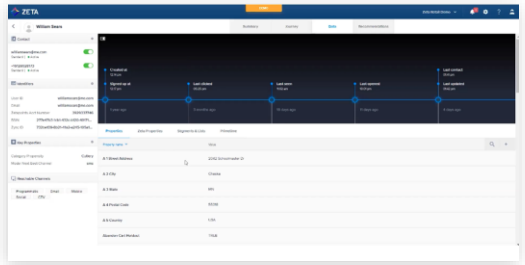
[Snowflake Security Product Documentation]

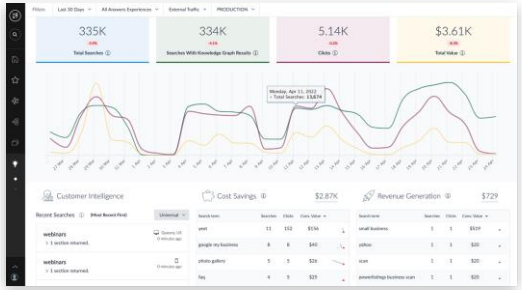
Building Apps in the Data Cloud

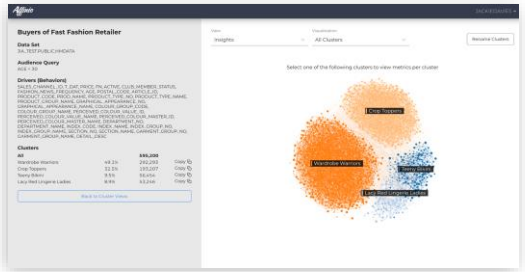
Programmability to put your data to work

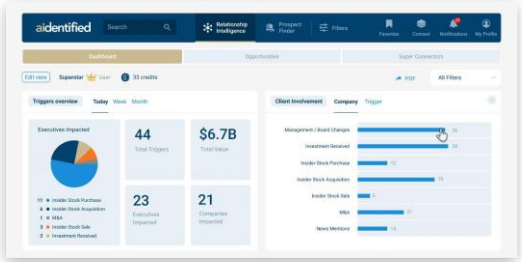












Build your way, but faster

Code directly in Python and Java with Snowpark; securely work with your favorite libraries; and rapidly prototype live applications with Streamlit.

Support dynamic demand

Easily scale to support growing usage without the SRE burden through the full power of Snowflake's platform.

Deliver better experiences

Unlock new ways to experience data through apps that run natively in you and your customer's Snowflake accounts.

SNOWFLAKE SUPPORTED REGIONS

Available on customer's cloud & region of choice

Generally Available

US West (Oregon)
US East (Ohio)
US East (N. Virginia)
Canada Central (Montreal)
US East (Commercial Gov - N Virginia)
US Gov West 1
EU (Ireland)
Europe (London)
EU (Frankfurt)
EU (Stockholm)
Asia Pacific (Tokyo)
AWS Pacific (Mumbai)
Asia Pacific (Singapore)
Asia Pacific (Sydney)
Asia Pacific (Seoul)

Generally Available

West US 2 (Washington)
Central US (Iowa)
East US 2 (Virginia)
Canada Central (Toronto)
US Government (Virginia)
North Europe (Ireland)
West Europe (Netherlands)
Switzerland North (Zurich)
Southeast Asia (Singapore)
Australia East (New South Wales)
Japan East (Tokyo)

Generally Available

US Central 1 (Iowa)
Europe West 2 (London)
Europe West 4 (Netherlands)

2022 Worldwide BDA Software Revenue by Vendor (Top 10)

Vendor	2021 Revenue ($M)	2022 Revenue ($M)	2022 Share (%)	2021-22 Growth (%)
Microsoft	$13,211.1	$15,836.9	15.0%	19.9%
Oracle	$8,603.4	$9,039.6	8.6%	5.1%
SAP	$7,286.7	$7,313.2	6.9%	0.4%
Amazon Web Services	$5,022.6	$7,016.2	6.7%	39.7%
Salesforce	$5,121.2	$5,829.6	5.5%	13.8%
IBM	$4,139.9	$4,324.8	4.1%	4.5%
SAS	$3,090.4	$3,296.7	3.1%	6.7%
Google	$2,099.8	$3,259.4	3.1%	55.2%
Adobe	$1,904.7	$2,173.2	2.1%	14.1%
Snowflake	**$1,016.3**	**$1,763.6**	**1.7%**	**73.5%**
Rest of Market	$39,560.0	$45,399.6	43.1%	14.8%
Total Market	**$91,056.0**	**$105,252.9**	**100.0%**	**15.6%**

Source: *IDC's Worldwide Semiannual Big Data and Analytics Software Tracker*, May 2023

snowflake®

AS OF APRIL 30, 2023 - Q1 FY24

PRODUCT REVENUE [1]

Q1 FY24

$590.1M

+ 50% YoY Growth

NET REVENUE RETENTION RATE [2]

151%

TOTAL CUSTOMERS [2]

8,167

+ 29% YoY Growth

$1M CUSTOMERS [2]

373

+ 80% YoY Growth
Customers with Trailing 12-Month Product Revenue Greater than $1M

FORBES GLOBAL 2000 CUSTOMERS [2]

590

+ 15% YoY Growth

SNOWFLAKE MARKETPLACE LISTINGS [3]

1,894

Total Listings
+ 3% QoQ Growth

CUSTOMER SATISFACTION

DRESNER CUSTOMER SATISFACTION SCORE [4]

100%

Of Customers Recommend Snowflake for Sixth Consecutive Year

NET PROMOTER SCORE (NPS) [5]

72

Most Customers Would Recommend Snowflake to a Friend or Colleague

SNOWFLAKE DATA CLOUD

1. For the three months ended April 30, 2023. 2. As of April 30, 2023. Please see our Q1 FY24 earnings press release for definitions of net revenue retention rate, customers with trailing 12-month product revenue greater than $1 million (which definition includes a description of our total customer count), and Forbes Global 2000 customers. 3. As of April 30, 2023. Each live dataset, package of datasets, or data service published by a data provider as a single product offering on Snowflake Marketplace is counted as a unique listing. A listing may be available in one or more regions where Snowflake Marketplace is available. 4. Dresner Advisory Services: 2023 Wisdom of Crowds® Analytical Data Infrastructure (ADI) Market Study, January 2023. 5. As of June 2022. If a customer fails to (i) respond to each required question in the survey or (ii) submit a complete set of responses by the end of the survey period, we consider that customer's survey incomplete. Starting with our NPS as of June 2022, we exclude incomplete survey responses from the calculation.

PROVEN BY THOUSANDS OF CUSTOMERS

iNFOTRUST

Questions?