

# An introduction to Snowflake - the data cloud



Johan Ludvig Brattås  
Deloitte

# Agenda

---

- A short history
- Overview
- Snowflake as a DB
- Integrations
- Snowpark

# The cloud data warehouse

---

- Initially a response on challenges faced by traditional RDBMS
  - Massively Parallel Processing (MPP)
  - Still a take on EDW
- 





# The cloud data platform

---

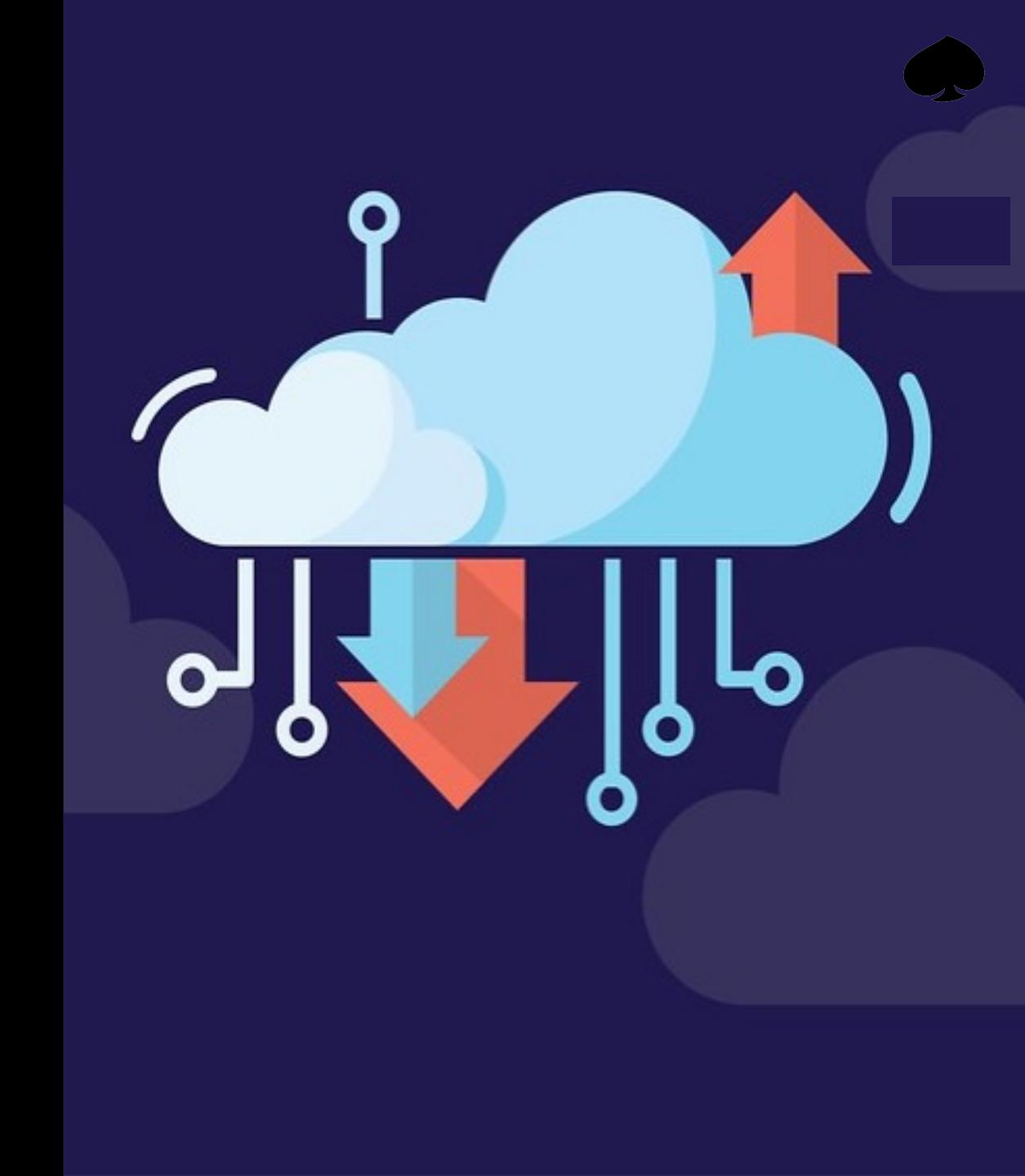
Can data lake functionality and EDW merge somehow?

Suggestions for solving the issues:

- Logical data warehouse
- Cloud data warehouse
- Virtualization

---

Enter the new cloud data platforms

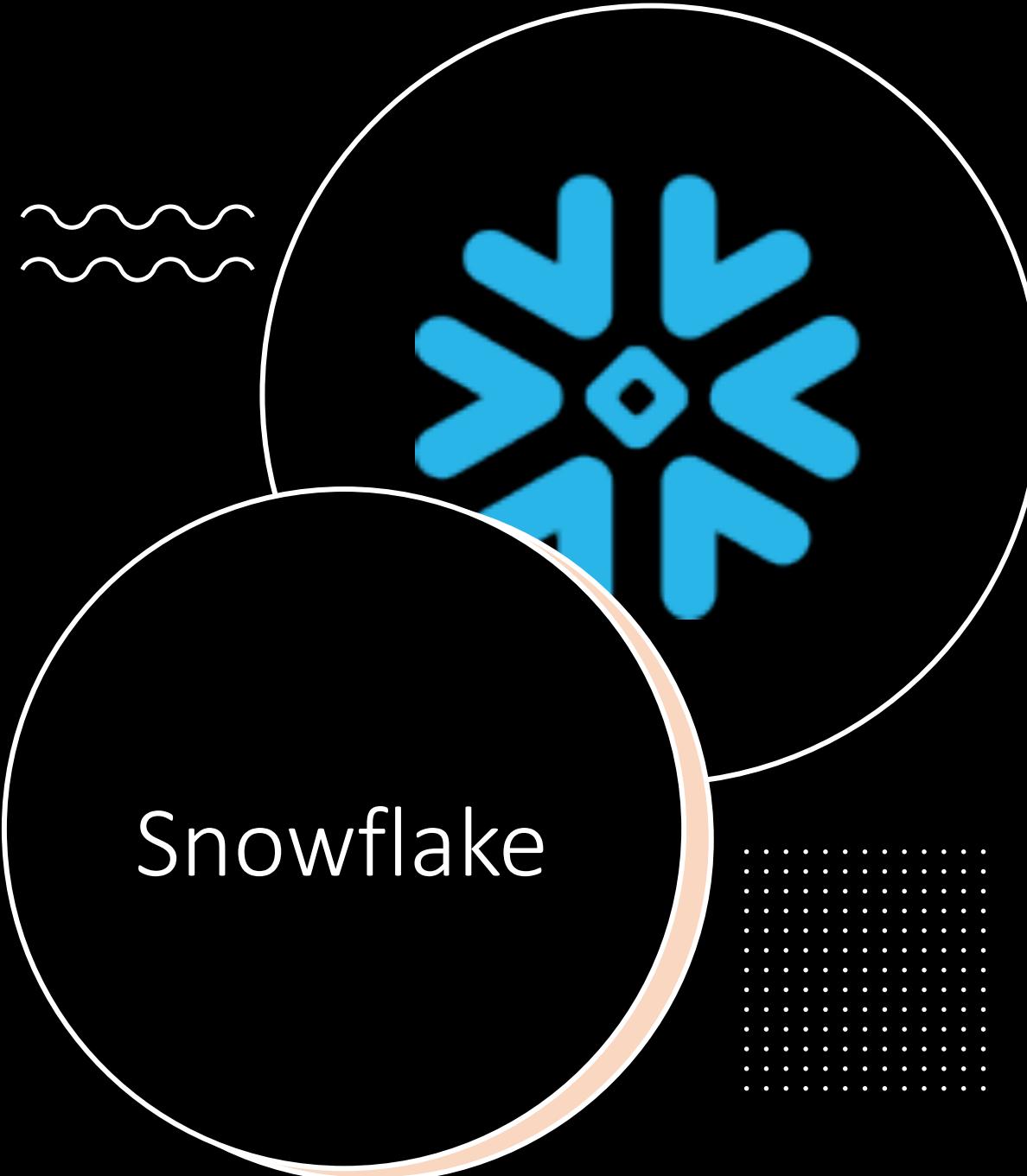


# Definition of a cloud data platform

---

- No longer just your Dad-a-base...
  - Storage supporting diverse data types
  - Compute and tools supporting diverse workloads
  - Tooling for CI/CD, encryption, RBAC etc
  - Data management tools
- 



The graphic features two overlapping circles. The larger circle is black with a white outline and contains the word "Snowflake" in white. The smaller circle is light blue with a white outline and contains the Snowflake logo, which is a blue snowflake icon with a central diamond. There are also three wavy lines above the large circle.

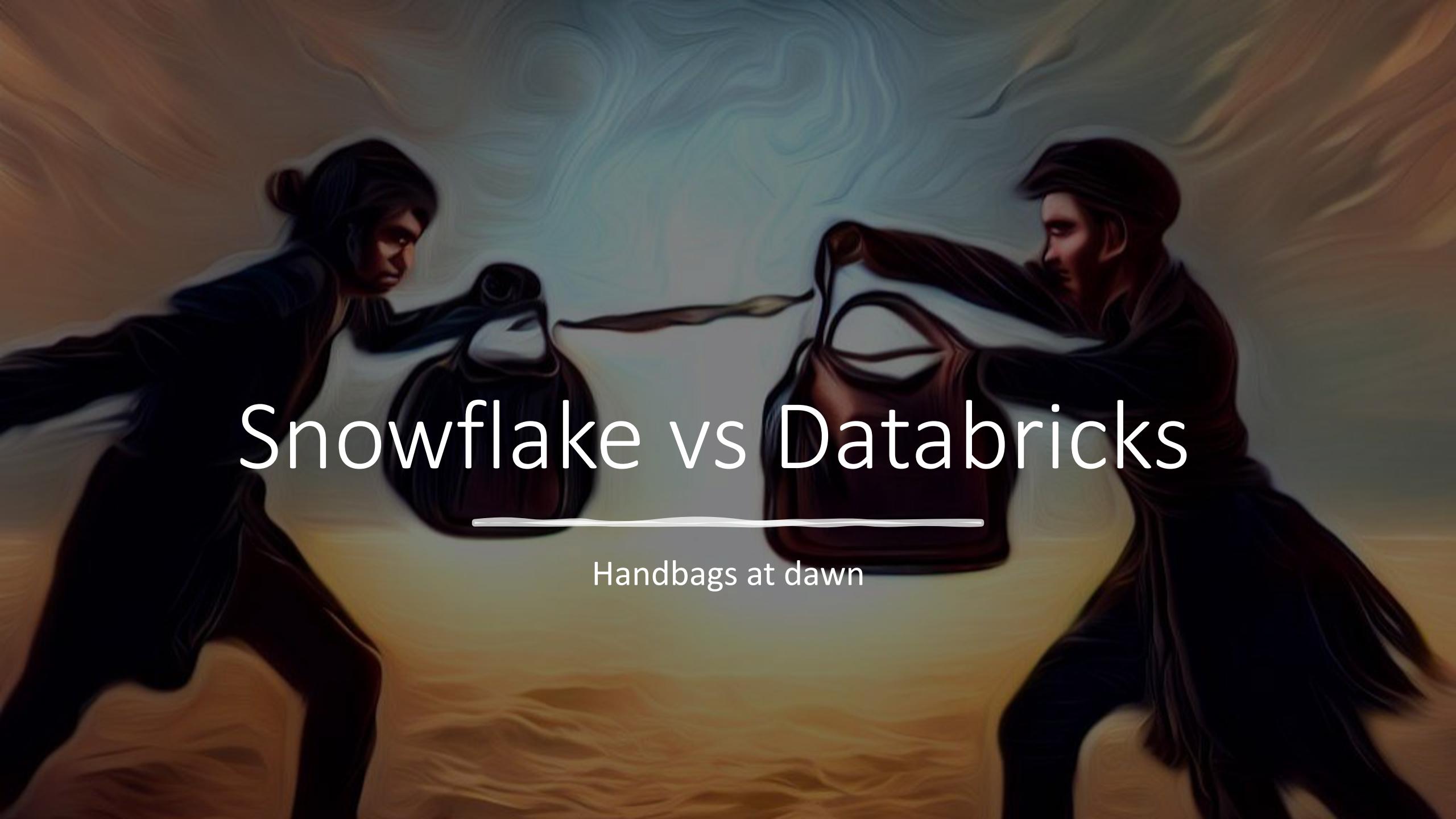
Snowflake

- Established in 2012
- Launched publicly in 2015
- Record IPO in 2020
- Unique architecture with fully separated storage and compute
- Based on ANSI SQL
- Started as a data warehousing service

# Snowflake VS Databricks

---

- Snowflake comes from EDW world
- Databricks from Spark orchestration, data science and python data engineering
- Converge as both have added new features



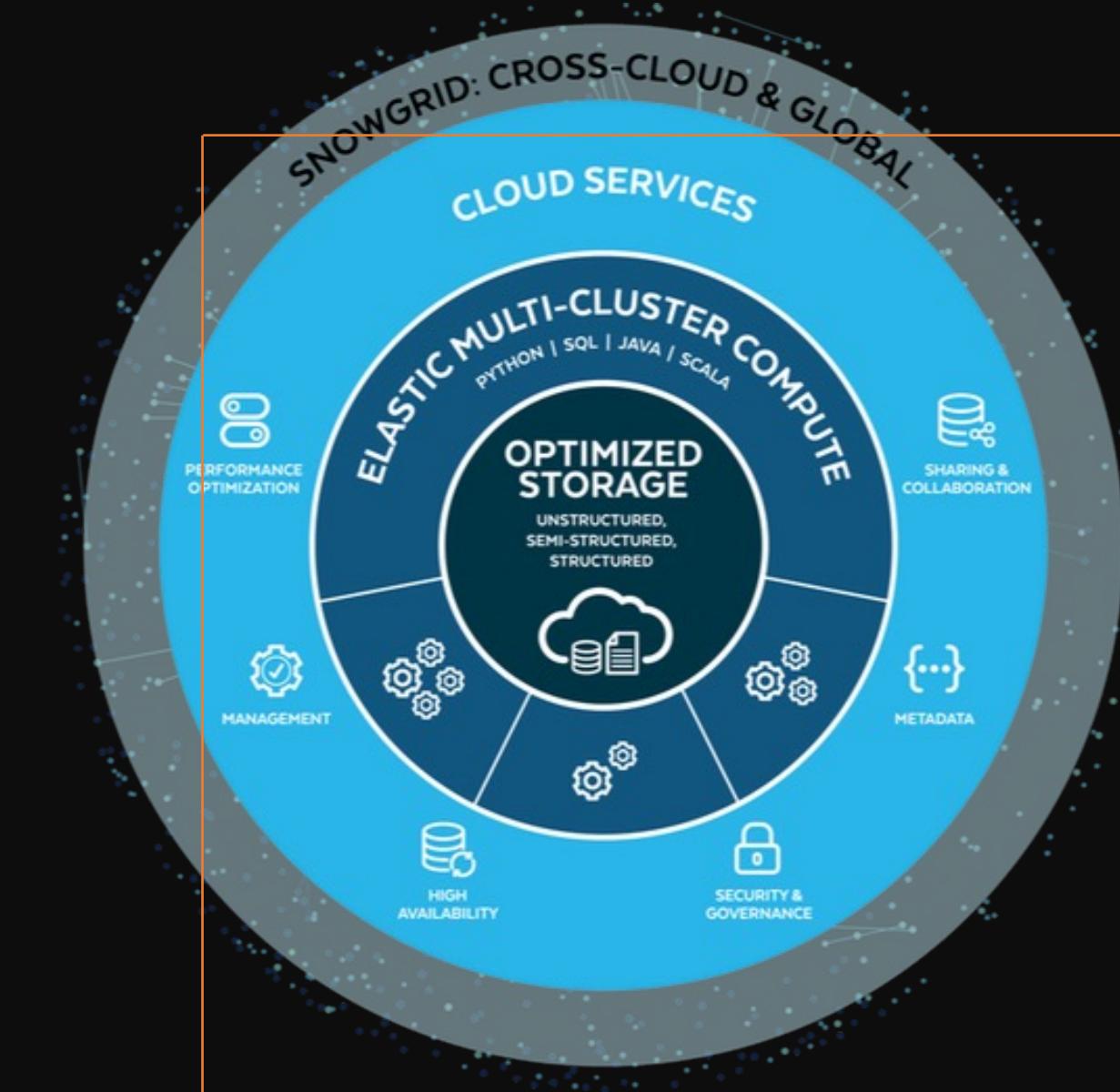
# Snowflake vs Databricks

---

Handbags at dawn

# The Snowflake Architecture

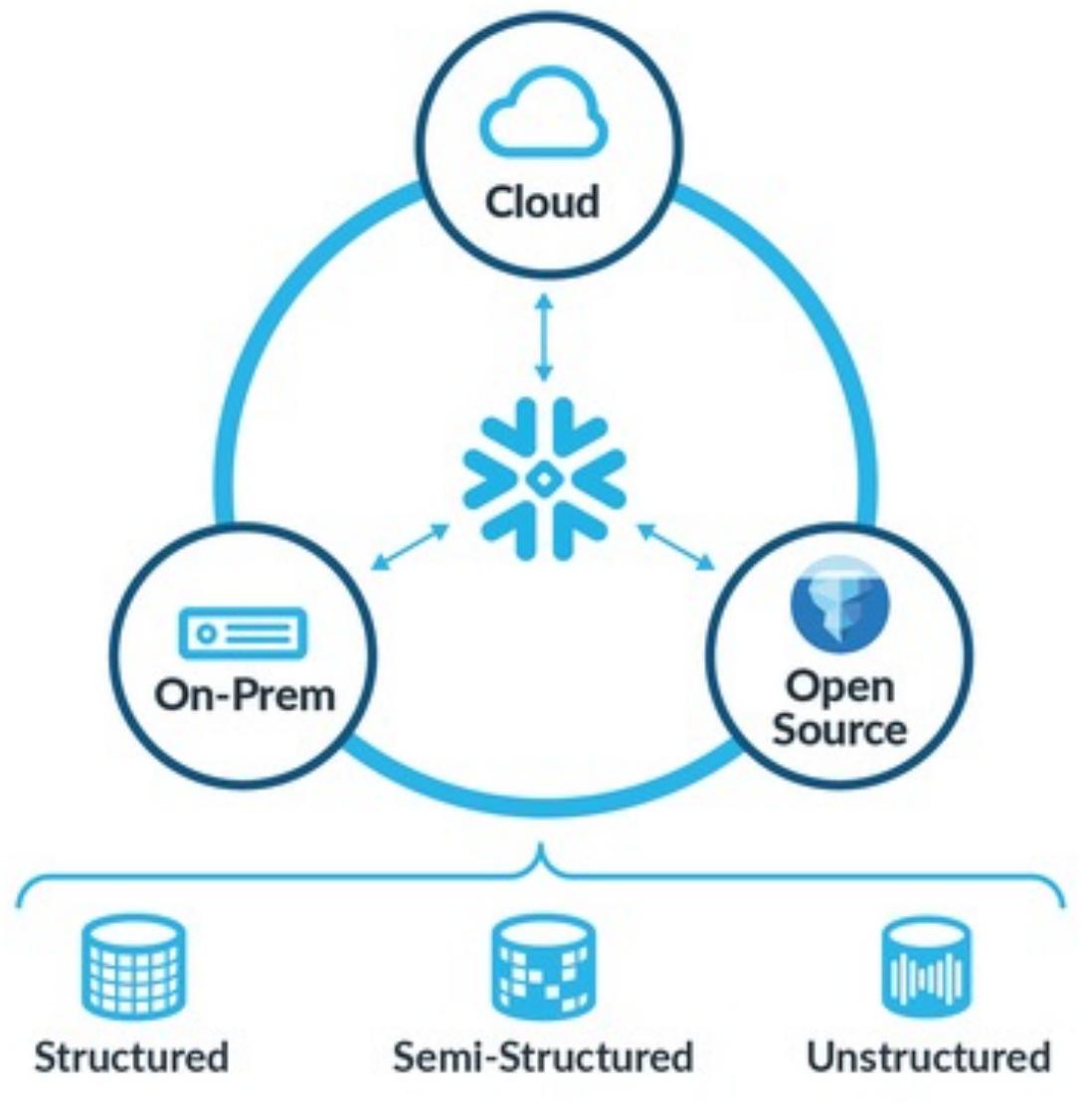
- The core Snowflake platform
  - Storage

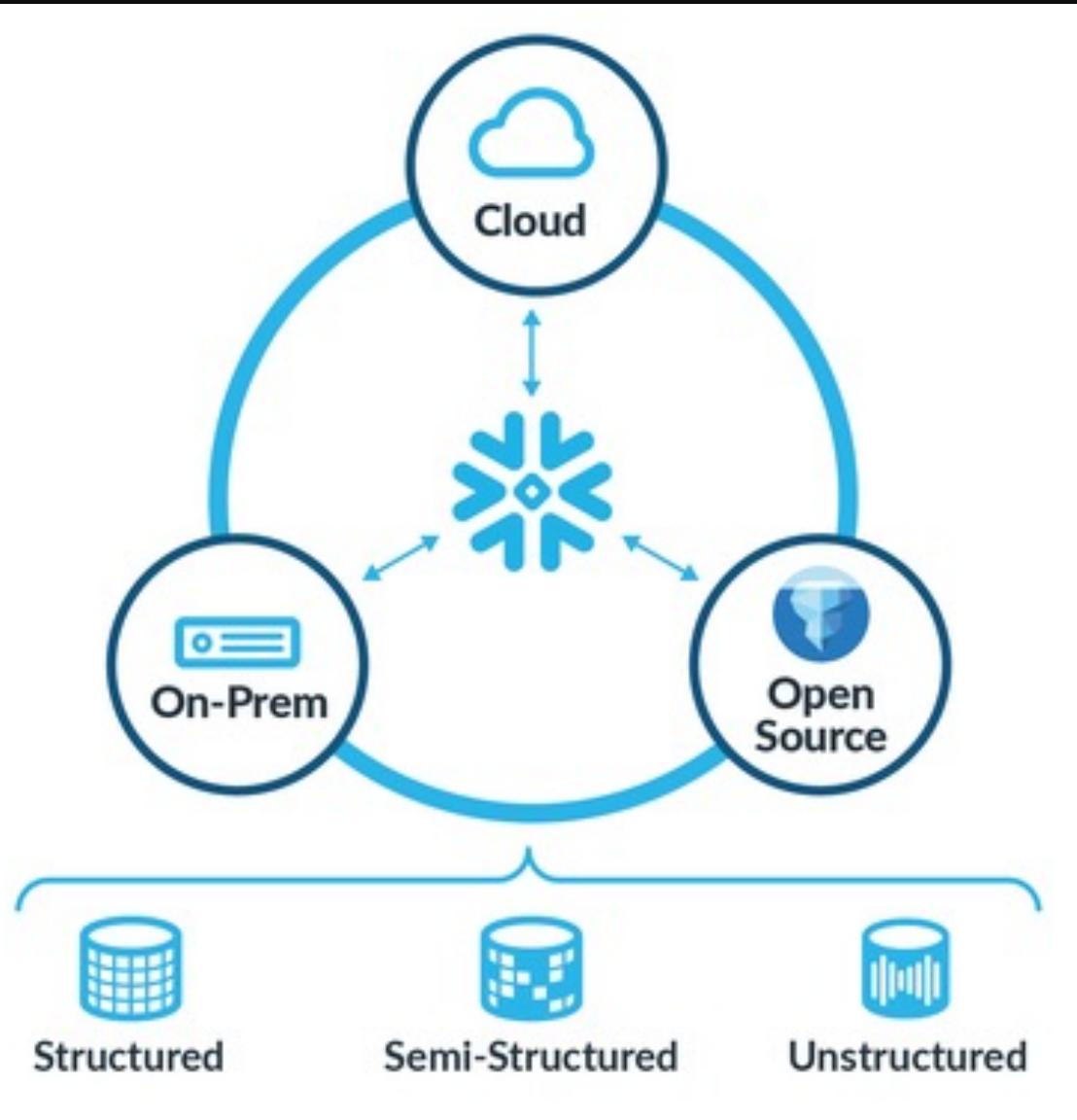


# Storage

---

- Databases for ACID + RDBMS
  - Automated partitioning
  - Time travel
  - Autotuned
- Internal Stage for semi- & unstructured
- External stages to on-prem & cloud





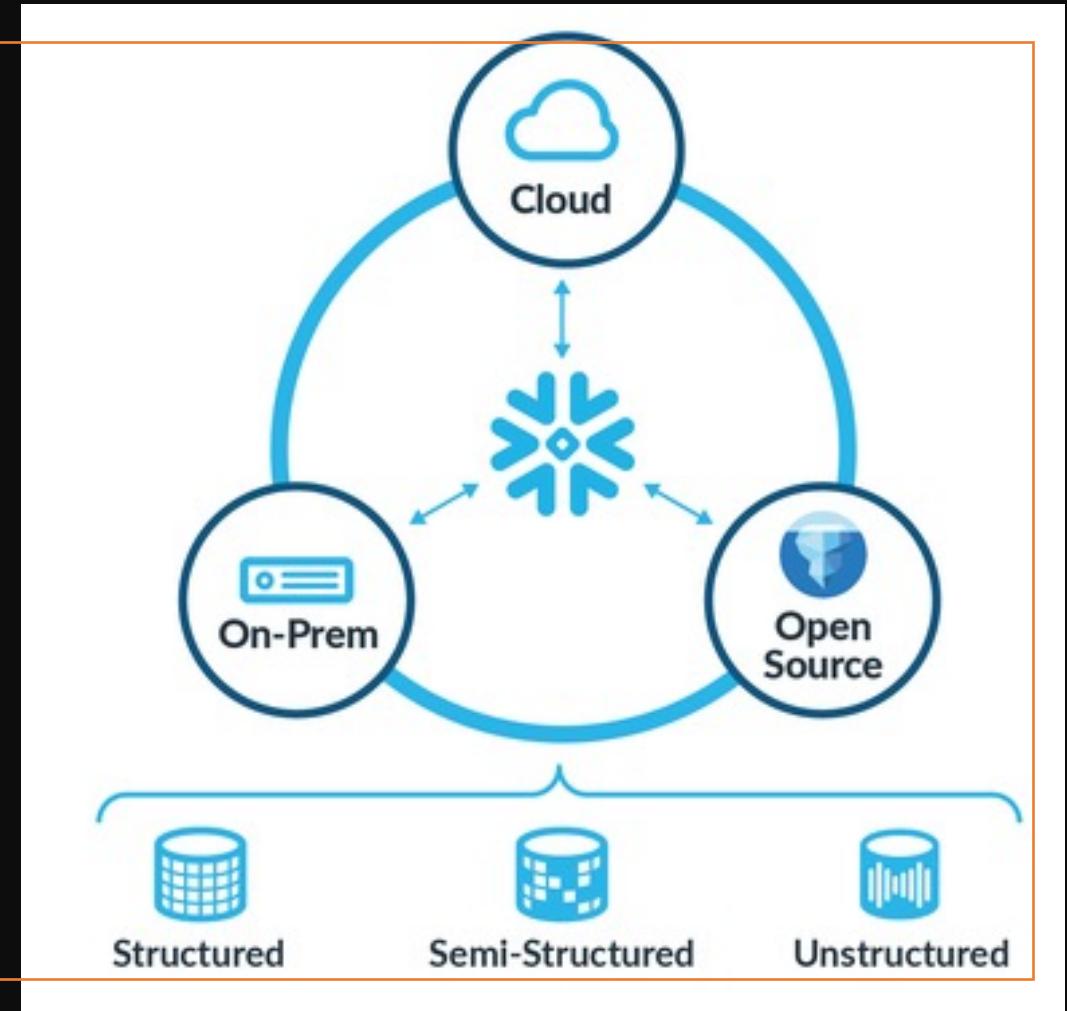
## Storage

---

- Cloud stages support S3, GCS & ADLS
  - On-prem only S3-compatible
  - External stages support
    - JSON/XML/CSV...
    - Avro/Parquet...
    - Apache Iceberg
    - Delta Lake
-

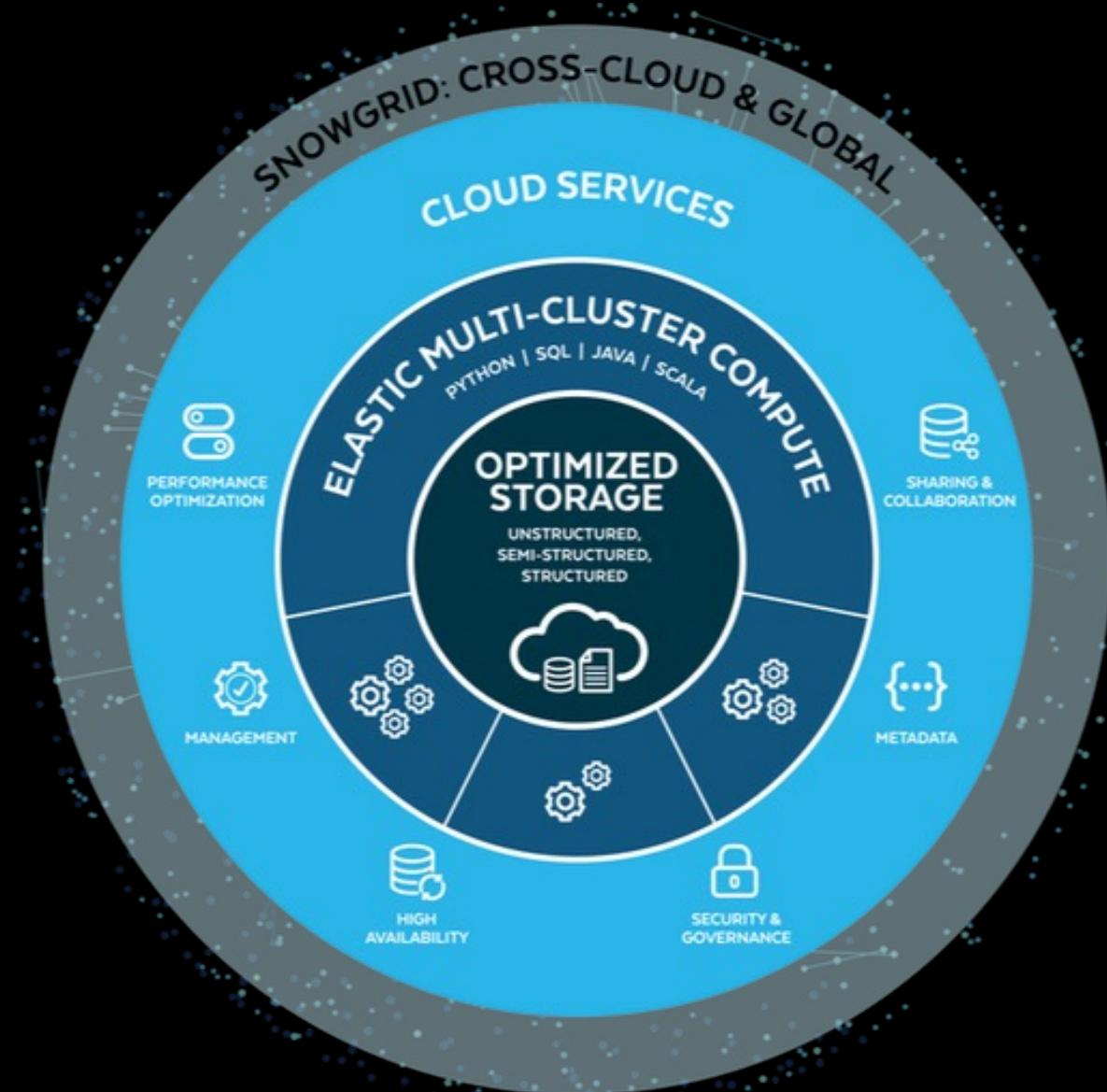
# Storage

- Create External Tables
- Build materialized views on semi-structured data



# The Snowflake Architecture

- The core Snowflake platform
  - Storage
  - Compute

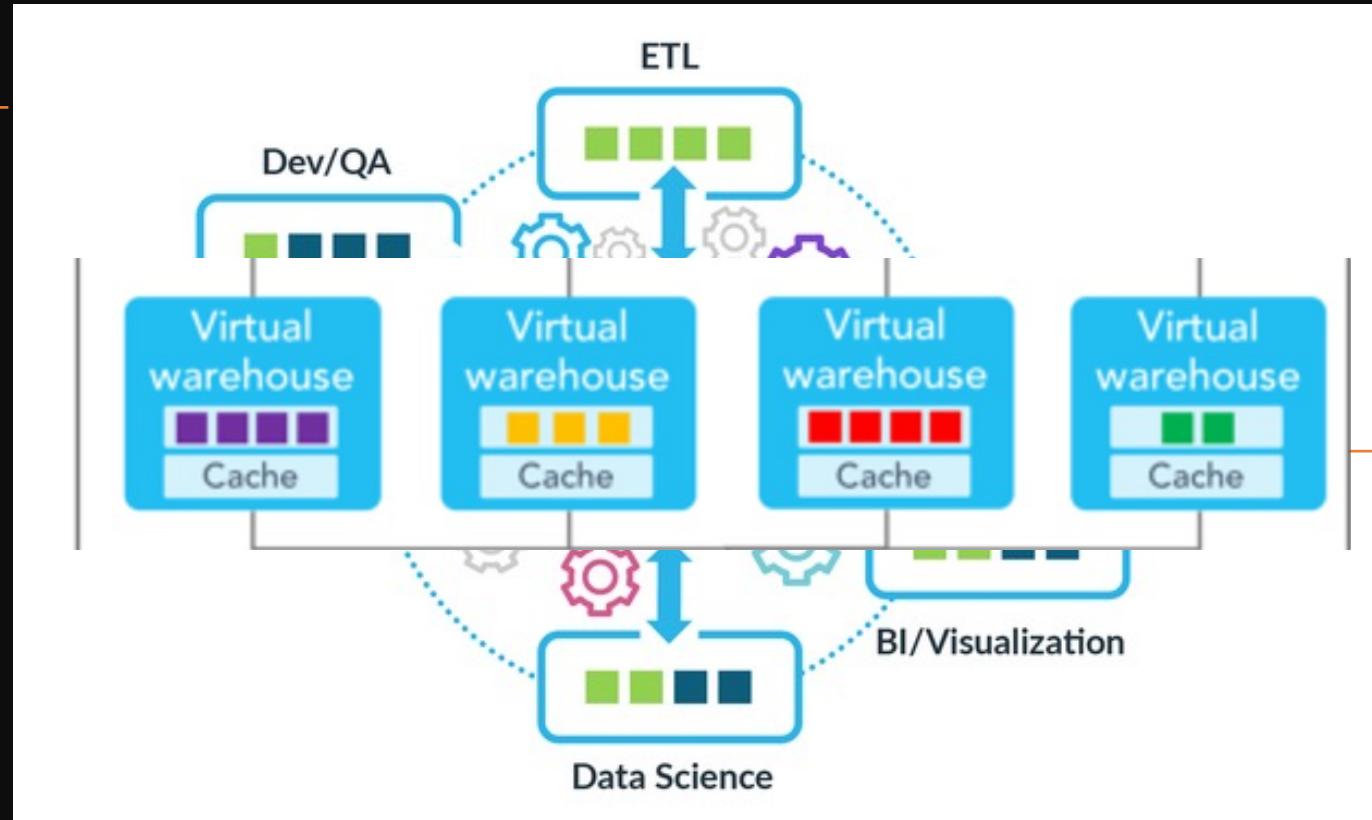


# Compute

- Called warehouses
- Elastic
  - From XS -> 6XL
- 2 types
  - Normal
  - Snowpark (memory) optimized
- Auto-pause + instant restart

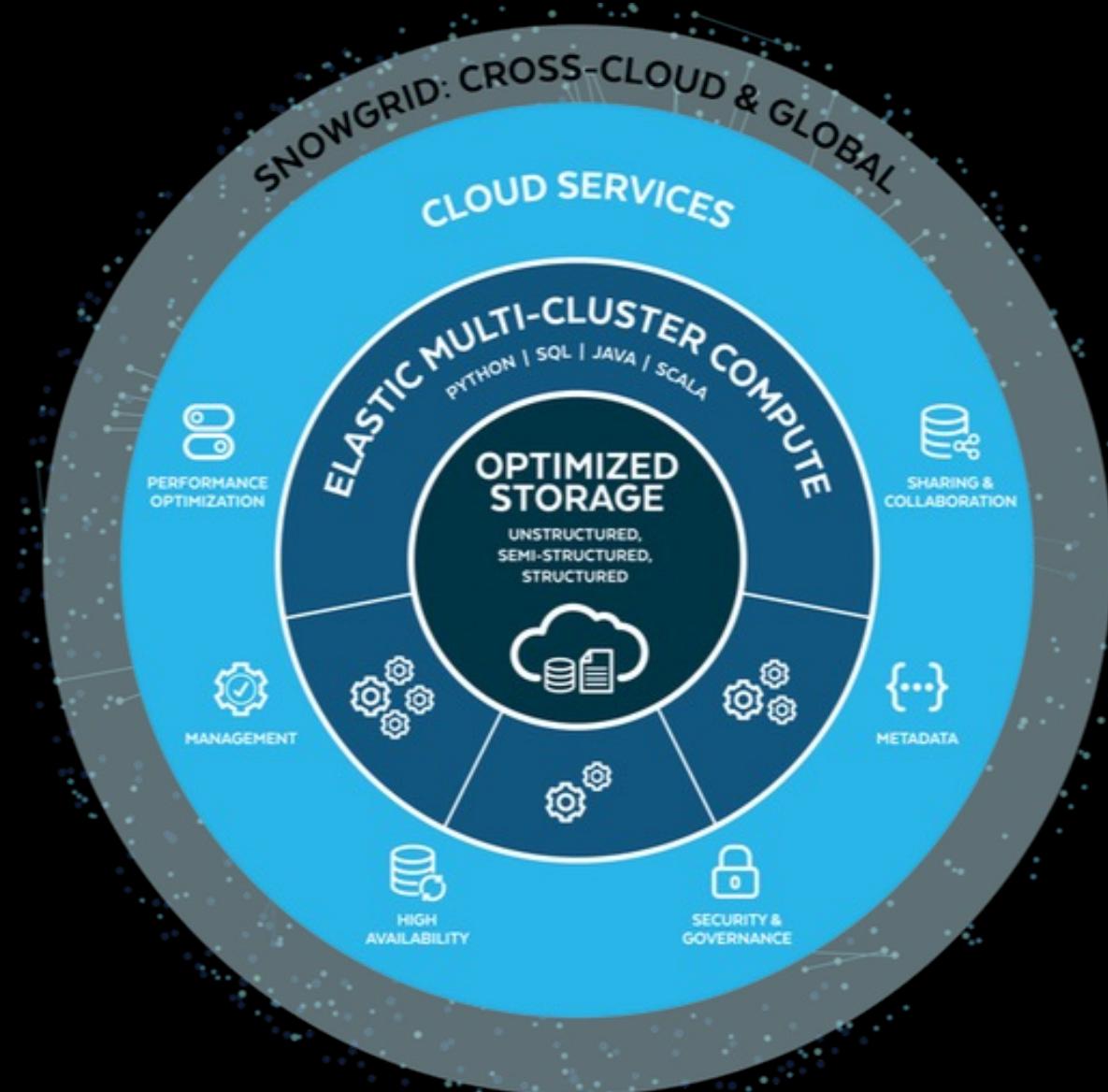
# Compute

- Consists of CPU & RAM
- Cache
- Separate warehouses per usecase
- Be mindful of auto-pause = cache emptied
- Plan your usecase usage patterns



# The Snowflake Architecture

- The core Snowflake platform
  - Storage
  - Compute
  - Cloud Services



# Cloud Services

---

- The central administration and control layer
- 4 pillars
  - Maintenance & tuning
  - Administration
  - Networking & Encryption
  - Resource Manager



# Cloud Services – 4 pillars

---

- Maintenance & tuning
- Administration
- Networking & Encryption
- Resource Manager



# Cloud Services – 4 pillars

---

- Maintenance & tuning
- Common meta-data repository
- Snowflake is “DBA-free”
  - Auto-tuning of queries
  - Auto-partitioning
  - Auto-indexing/”Indexfree”



# Cloud Services – 4 pillars

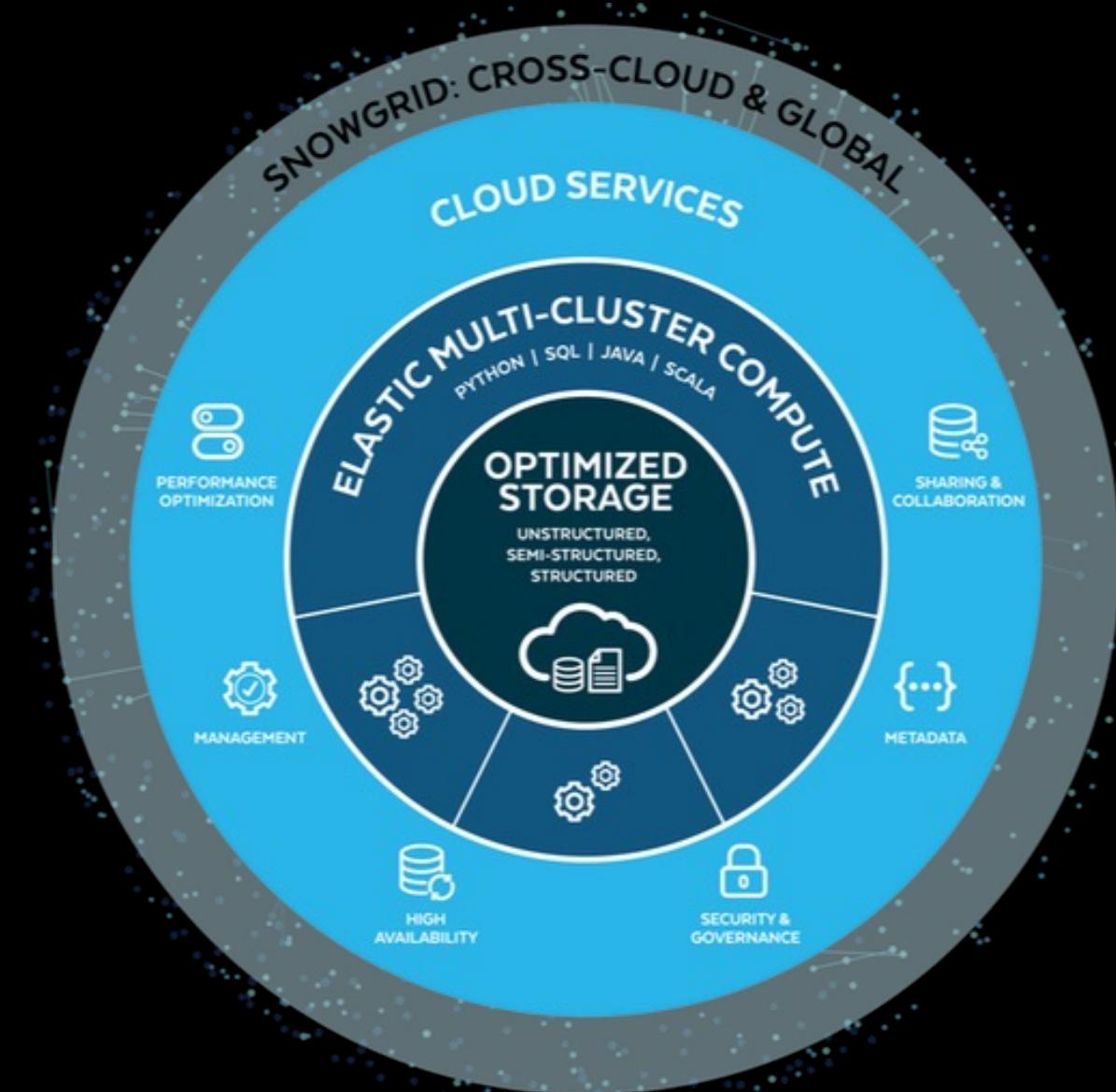
---

- Administration
    - Transaction manager
    - Security/RBAC
    - Authentication & Authorization
  - Networking & Encryption
    - Intra-cluster
    - Cloud connectivity
  - Resource Manager
    - Cluster management
- 



# The Snowflake Architecture

- The core Snowflake platform
  - Storage
  - Compute
  - Cloud Services
  - Snowgrid



# Snowgrid

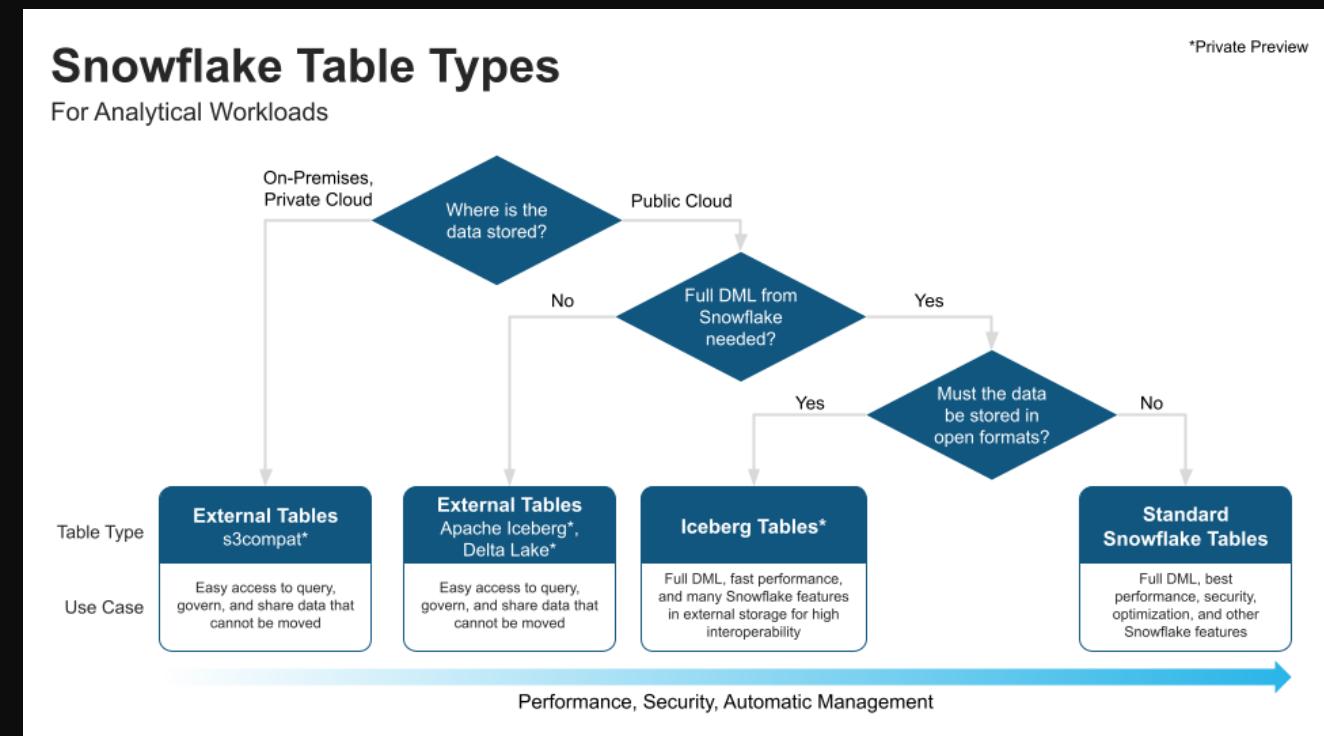
---

- Snowgrid
  - Global Snowflake internal network
- Cloud Agnostic



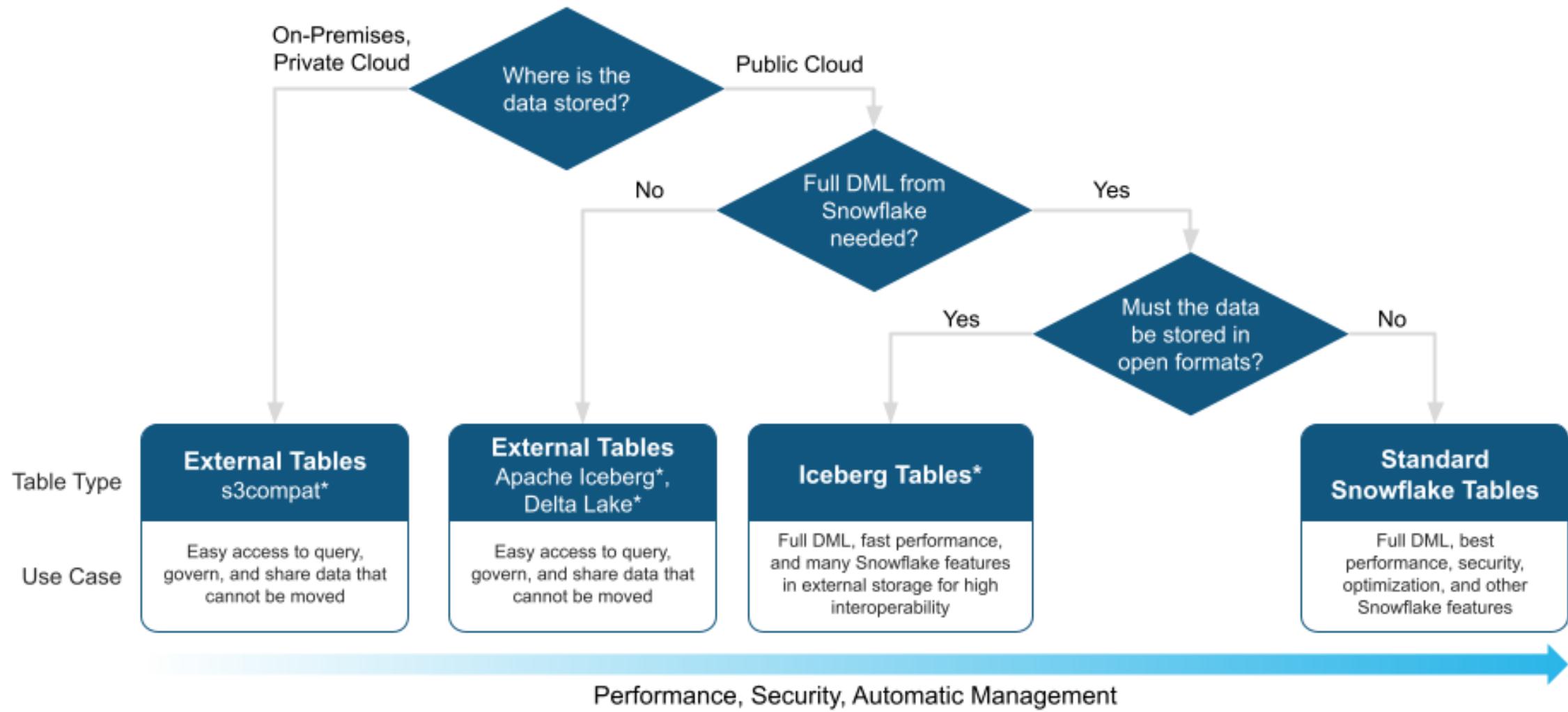
# Integrations

- Integration
- Stages
- External Tables
- Dynamic Tables
- Snowpipes
- Unistore



# Snowflake Table Types

For Analytical Workloads



# The Snowflake Eco-system

---

- Snowpark
- Streamlit



# Snowpark

- Expands Snowflake from traditional RDBMS
- Python – offers traditional dataframe APIs
- Also ML modelling and operations APIs
- Can run inside warehouses
- Can run on containers (Snowpark Container Services)
- Now also notebook experience



# Streamlit

- Company acquired by Snowflake 2022
- Build interactive apps with Python that runs on Snowflake
- Web apps, widgets – with unique URLs that can be shared
- Still in public preview



# But what about Gen AI?

- Cortex (private preview)
- Fully managed service for building and running AI applications
- Llama + Snowflake LLMs
- Vector functionality





# The Snowflake Marketplace

- From the consumer
  - Search, discover and sample datasets globally
  - Access datasets – some free, some commercial
  - No need to run ETL processes to fetch data
  - Directly start querying the data inside own account
  - Can combine internal and marketplace data

# The Snowflake Marketplace

- From the producer
  - Share data with users outside your organization
  - This done through listings
  - Listings can be global or limited to select users/organizations
  - Datasets can be a one-off, an update or stream.
  - No special development needed
  - Listings can be private, free or paid



# So what about cost?

---

- Pricing tiers
- Pricing models





# Pricing tiers

Standard



Enterprise



Business Critical



Virtual Private Snowflake



Your standard service – single cluster, multi-warehouse  
1 day time travel  
data marketplace and exchange

Standard+ Multi-cluster warehouse  
90 day time travel

Annual rekey of encrypted data  
Materialized views, Dynamic Datamasking, Search Optimization

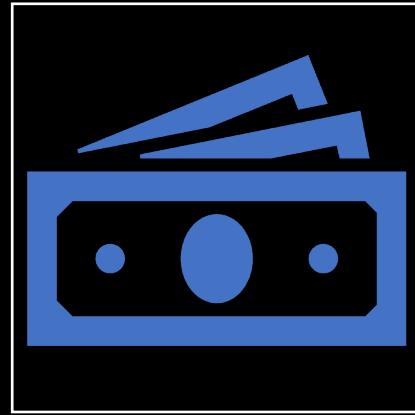
Enterprise+ HIPAA support, PCI Compliance  
Tri-Secret Secure w/CMK  
AWS Private Link support

DB Failover and Fallback, AWS API Gw Private Endpoint support

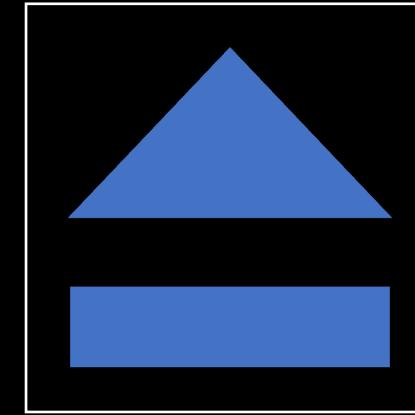
Business Critical+

Customer Dedicated Virtual Servers w/encryption key in memory  
Customer Dedicated Metadata store

# Pricing



Pay as you go



Pre-purchased



# Pricing model

Snowflake Credit – the base of cost in Snowflake pricing.  
Real price depends on pricing tier (\$2.6/\$3.9/\$5.2)

Separate pricing for storage, compute and Cloud Services.

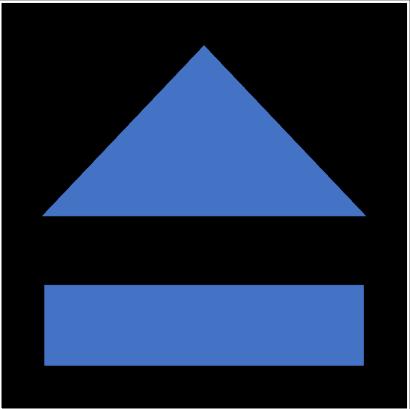
Storage - \$40 per Terabyte per Month

Compute starts at 1 credit for XS Warehouse -> 512 Credits for 6XL  
Snowpark Optimized starts at 6 for M Warehouse -> 786 Credis for 6XL

Cloud Services based on workload – but up to 10% of daily credits included free

Additional prices for data transfer to other cloud providers or within same provider but different country or continent

Prices shown are guidance only – based on Snowflake list prices for Azure West-Europe pr November 2023



## Pre-purchased

Capacity Storage is at \$23 per Terabyte / Month

Compute & Cloud Service price depends on size of commitment + time of purchase ++

Prices shown are guidance only – based on Snowflake list prices for Azure West-Europe pr November 2023





# Johan Ludvig Brattås

## Director, Deloitte

 /johanludvig

 @intoleranse

 jbrattas@deloitte.com



GitHub

### Chronic volunteer

Co-organizer – DataSaturday Oslo  
President – MDPUG Oslo  
Frequent volunteer in general

### When not geeking out over new tech

Teaching coeliacs how to bake gluten free  
Baking  
Hiking  
Gardening

# Session evaluation

Your feedback is important to us

**Evaluate this session at:**

[www.PASSDataCommunitySummit.com/evaluation](http://www.PASSDataCommunitySummit.com/evaluation)

