snowflake

الكلية متعددة التخصصات تارودانت
ⵜⴰⵖⵍⵍⴰⵢⵜ ⵜⴰⵣⵜⵓⵏⵜ ⵏ ⵜⵓⵙⴷⴰⵡⵉⵏ
FACULTE POLYDISCIPLINAIRE TAROUDANT

SnowFlake

PRESENTED BY :

Aouni Hamza

Elhamdouni Yassin

SUPERVISED BY :

Mr. Najdi Lotfi

snowflake

PLAN

- Introduction
  - What is Snowflake?
  - Key Concept
  - History
  - Companies Using Snowflake
- SnowFlake architecture
  - Hybrid Architecture:
    - Shared-Disk
    - Shared-Nothing
    - Massively parallel pessing
  - General Architecture of SnowFlake
    - Cloud services layer
    - Compute layer
    - Storage layer
  - Data ingestion reference architecture in Snowflake
- Snowflake’s advantages vs Big Data & RDBMS
  - vs Big Data
  - vs RDBMS
- Practice live
- Conclusion
- Some ressources

snowflake

Introduction

What is Snowflake?

Snowflake is an analytic data warehouse provided as Software-as-Services (SaaS). Snowflake provides data warehouse that is faster, easier to use and more flexible than other traditional data warehouses.

Snowflake data warehouse is not built on existing databases or not on big data software platform as Hadoop.

The Snowflake data warehouse uses a new SQL database engine with unique architecture designed for the cloud.

snowflake

Introduction

Key Concept

Data Warehouse as Cloud Service:

Snowflake data warehouse is a true SaaS offering.

- There is no hardware (virtual or physical) for you to select, install, configure, and manage.
- There is no Software for you to install, configure, and manage.
- Snowflake handles ongoing maintenance, management, and tuning.

Snowflake completely runs on cloud infrastructure. All the components of the Snowflake service run on public cloud infrastructure.

Snowflake uses a virtual compute instance for its compute needs and a storage service for the storage of data. Snowflake can not be run on private cloud infrastructure (on-premises).

Introduction

History

2012 – Company Founded: Snowflake was founded by Benoit Dageville, Thierry Cruanes, and Marcin Zukowski, all of whom were experienced in data architecture and database systems.

Benoit Dageville

Thierry Cruanes

Marcin Zukowski

snowflake

Introduction

History

2012–2014 – Product Development in Stealth Mode: The founders began developing a cloud-native data warehouse from the ground up.

2014 – Beta Testing Begins: Snowflake initiated private beta testing with selected custome.

June 2015 – Public Unveiling of Platform: Snowflake emerged from stealth and publicly launched its cloud data warehousing platform, offering separated compute and storage, and architectural breakthrough at the time.

Post-2015 – Early Market Traction: The product quickly gained adoption, especially among companies seeking scalable, cost-effective, and easy-to-use data warehousing solutions.

snowflake

Introduction

Companies Using Snowflake

Companies

Finance

Western Union :
Leverages it for data-driven decision-making

Technology & Software

Adobe
For real-time marketing analytics and customer data platforms.

Dropbox
Manages analytics and internal data processing.

Insurance

Capital One:
Uses Snowflake to support machine learning and customer analytics.

Media & Entertainment

Sony
Applies Snowflake for content analytics and distribution tracking.

Disney Streaming (Hulu, Disney+)
Uses it for real-time user engagement analysis.

snowflake

Snowflake Architecture

Hybrid Architecture: Shared-Disk & Shared-Nothing

Snowflake architecture is the combination of shared-disk database architecture and shared-nothing database architecture.

- **Shared-disk database architecture**
- **Shared-nothing architecture**
- **Massively parallel processing architecture**

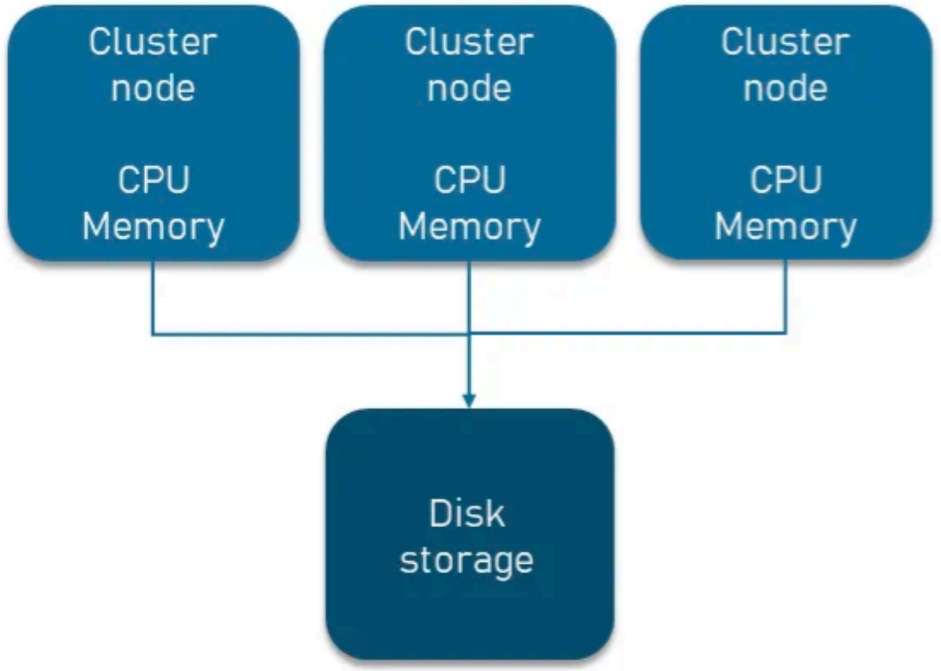
snowflake

Snowflake Architecture

Hybrid Architecture: Shared-Disk & Shared-Nothing

Shared-Disk

In shared-disk architecture, nodes share centralized storage, simplifying management but risking bottlenecks under heavy access.

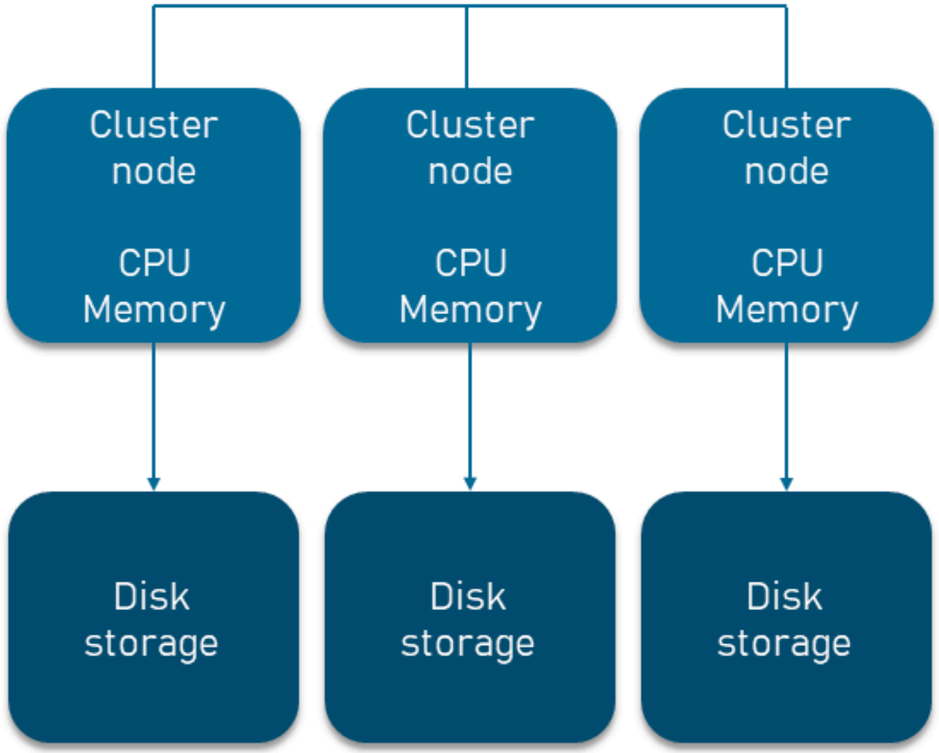


Snowflake Architecture

Hybrid Architecture: Shared-Disk & Shared-Nothing

Shared-Nothing

In a shared-nothing architecture, each node has its own storage and memory, communicating via network to avoid bottlenecks, with efficient data partitioning key to performance.



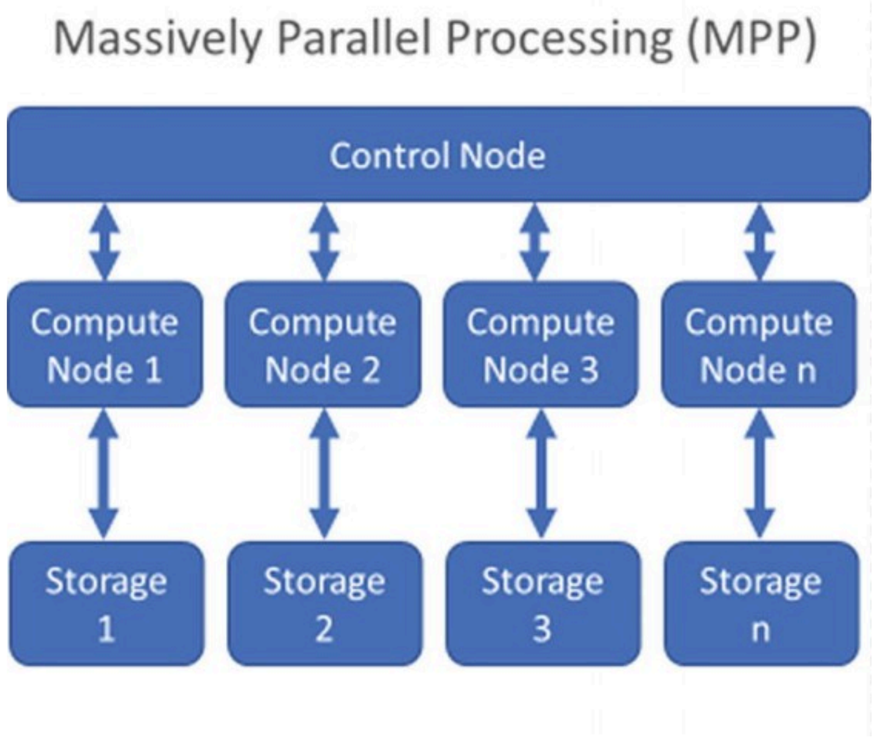
snowflake

Snowflake Architecture

Hybrid Architecture: Shared-Disk & Shared-Nothing

Massively parallel processing

Snowflake processes the queries using MPP (massively parallel processing) compute cluster where each node in the cluster stores a portion of entire data set locally.

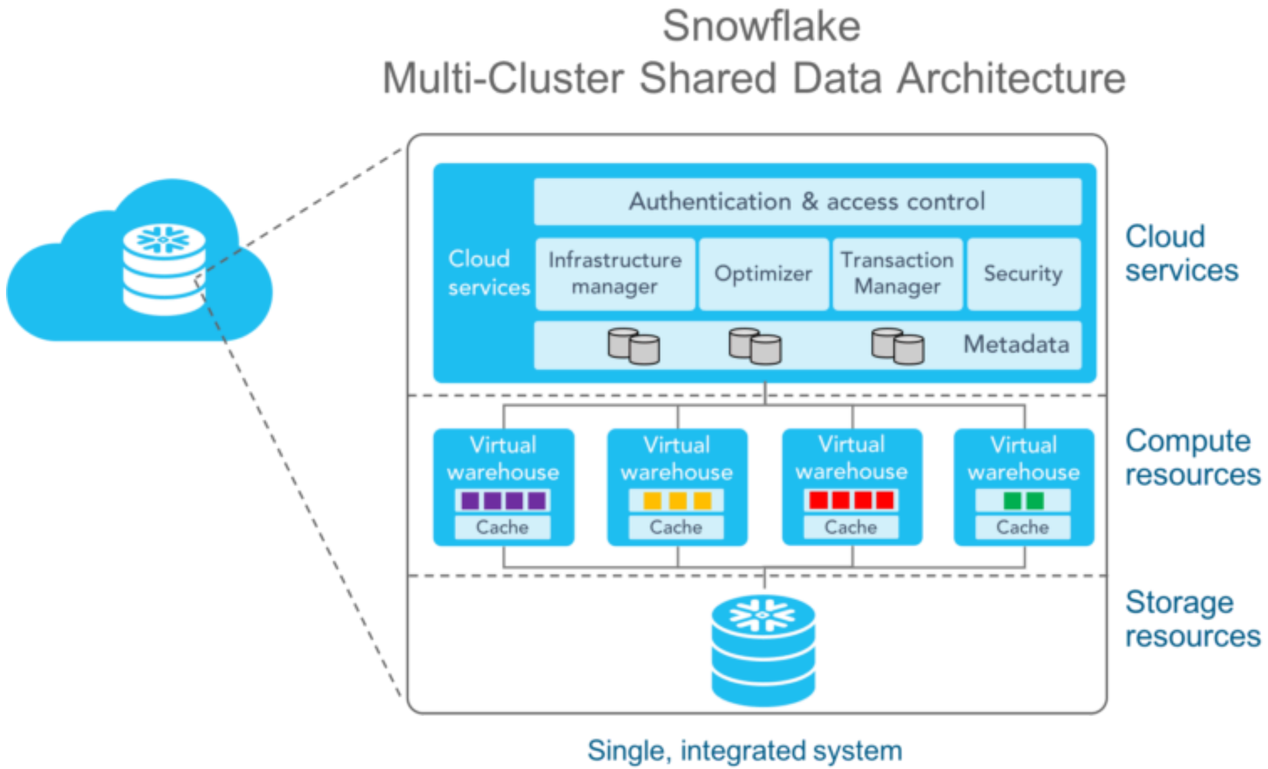


Snowflake Architecture

General Architecture of SnowFlake

Snowflake supports flexible batch and continuous data pipelines in any language, allowing incremental or bulk parsing in various styles. To understand its functionality, explore its three-layer data warehouse architecture:

- **Cloud services layer** : This is the “brain” of Snowflake. It handles query optimization, metadata management, security, and transaction management.
- **Compute layer:** that is, query processing.
- **Database storage:** the central layer of the deepest point of the octahedron

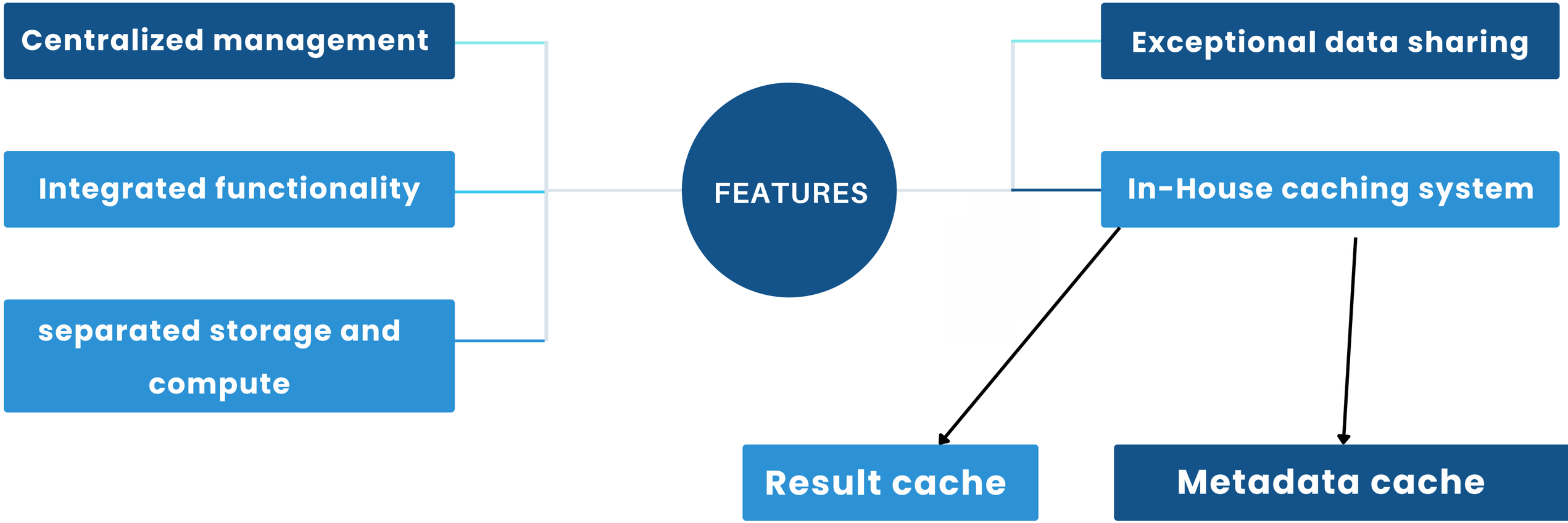


snowflake

Snowflake Architecture

General Architecture of SnowFlake

Cloud services layer



snowflake

Snowflake Architecture

General Architecture of SnowFlake

Compute layer

snowflake

Snowflake Architecture

General Architecture of SnowFlake

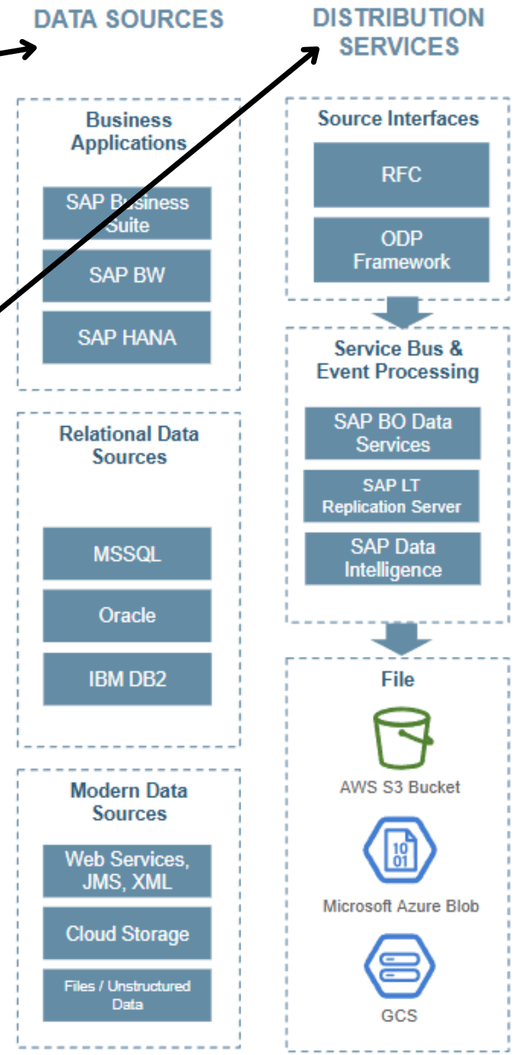
Storage layer

Snowflake Architecture

Data ingestion reference architecture in snowflake

Data ingested from various sources

Data from these sources is collected and processed through Source Interfaces



Snowflake Architecture

Data ingestion reference architecture in snowflake

Conformed data is consumed through Consumer Interfaces

Data is ingested into the Snowflake Data Cloud using interfaces like JDBC, ODBC, and Copy into,

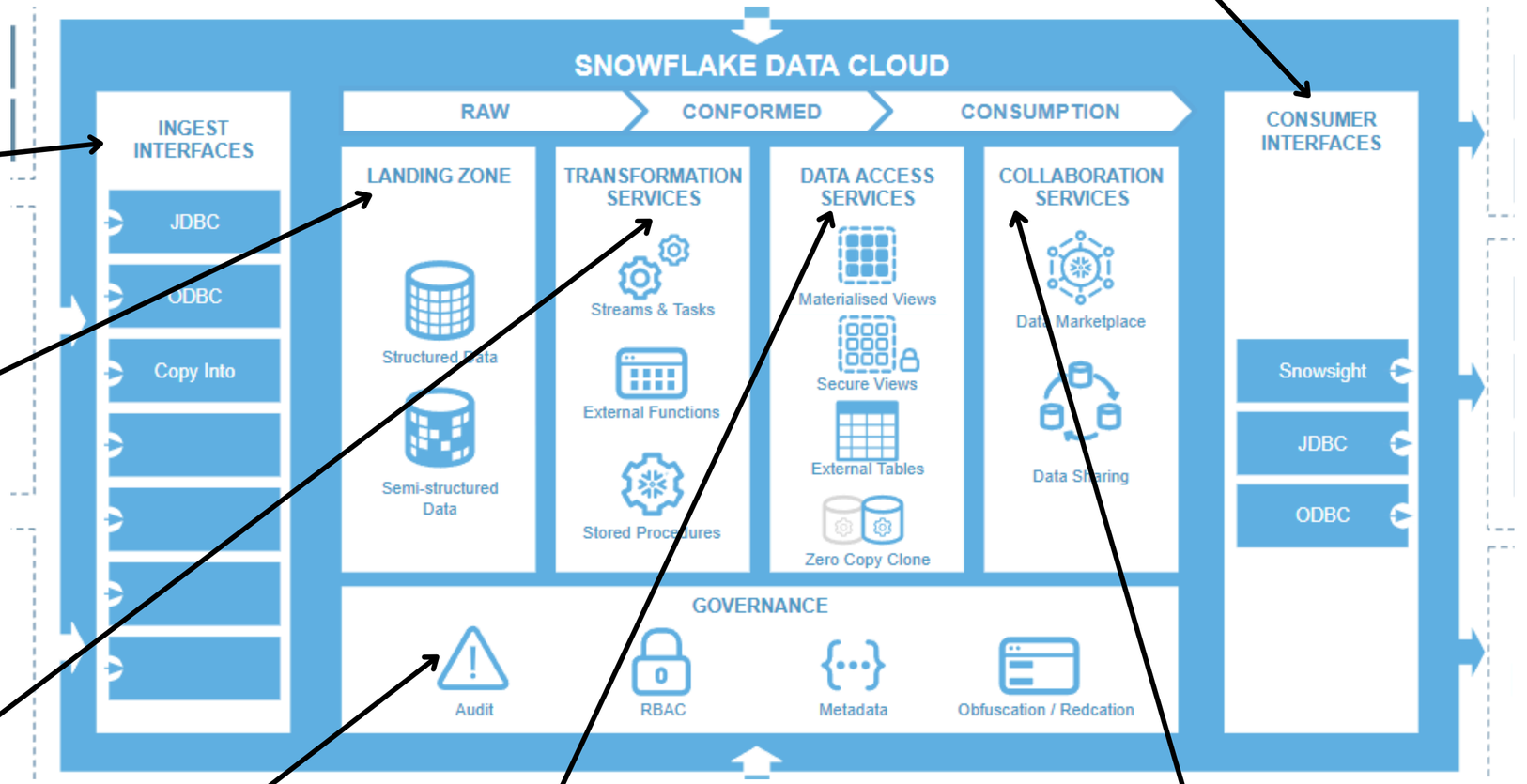
Once ingested, data is stored in the Landing Zone as S or SS data, serving as the initial storage area within Snowflake

Data is transformed converting raw data into conformed data ready for

Specific governance functions include Audit

Transformed data is made accessible, enabling efficient querying and utilization

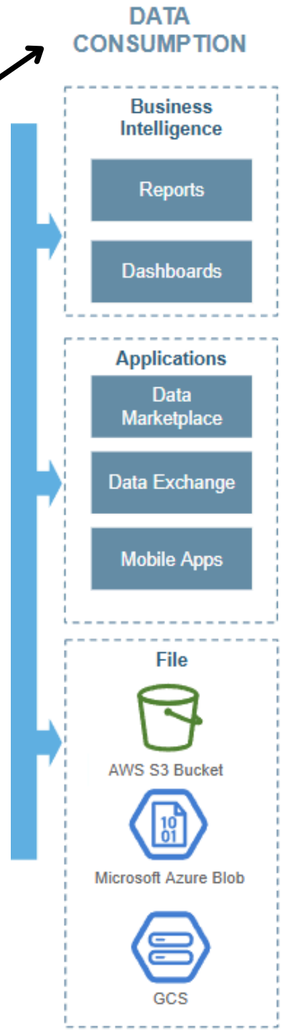
allowing data to be shared and collaborated on within the ecosystem.



Snowflake Architecture

Data ingestion reference architecture in snowflake

This layer focuses on delivering processed data to end-users and applications in an accessible and actionable manner



Snowflake advantages vs Big Data & RDBMS

Traditional Big Data ecosystems: Hadoop

Feature	Snowflake	Hadoop & RDBMS
Scalability	Elastic cloud scaling, instant compute & storage separation	Complex cluster management (Hadoop), limited vertical scaling (RDBMS)
Performance	Blazing-fast queries with auto-optimization	Batch-heavy, slower for ad-hoc (Hadoop), constrained by schema (RDBMS)
Ease of Use	Zero setup, fully managed, SQL-native	Steep learning curve (Hadoop), rigid schema design (RDBMS)
Concurrency	Unlimited users, no contention	Limited by nodes (Hadoop), locking issues (RDBMS)

snowflake

Snowflake’s advantages vs Big Data & RDBMS

On-Premises RDBMS: PostgreSQL

SNOWFLAKE’S ADVANTAGES OVER POSTGRESQL

SCALABILITY:

Snowflake handles massive analytics with isolated compute clusters.

MULTI-CLOUD SUPPORT:

Snowflake integrates seamlessly across AWS, Azure, and GCS, unlike PostgreSQL's single-server dependency.

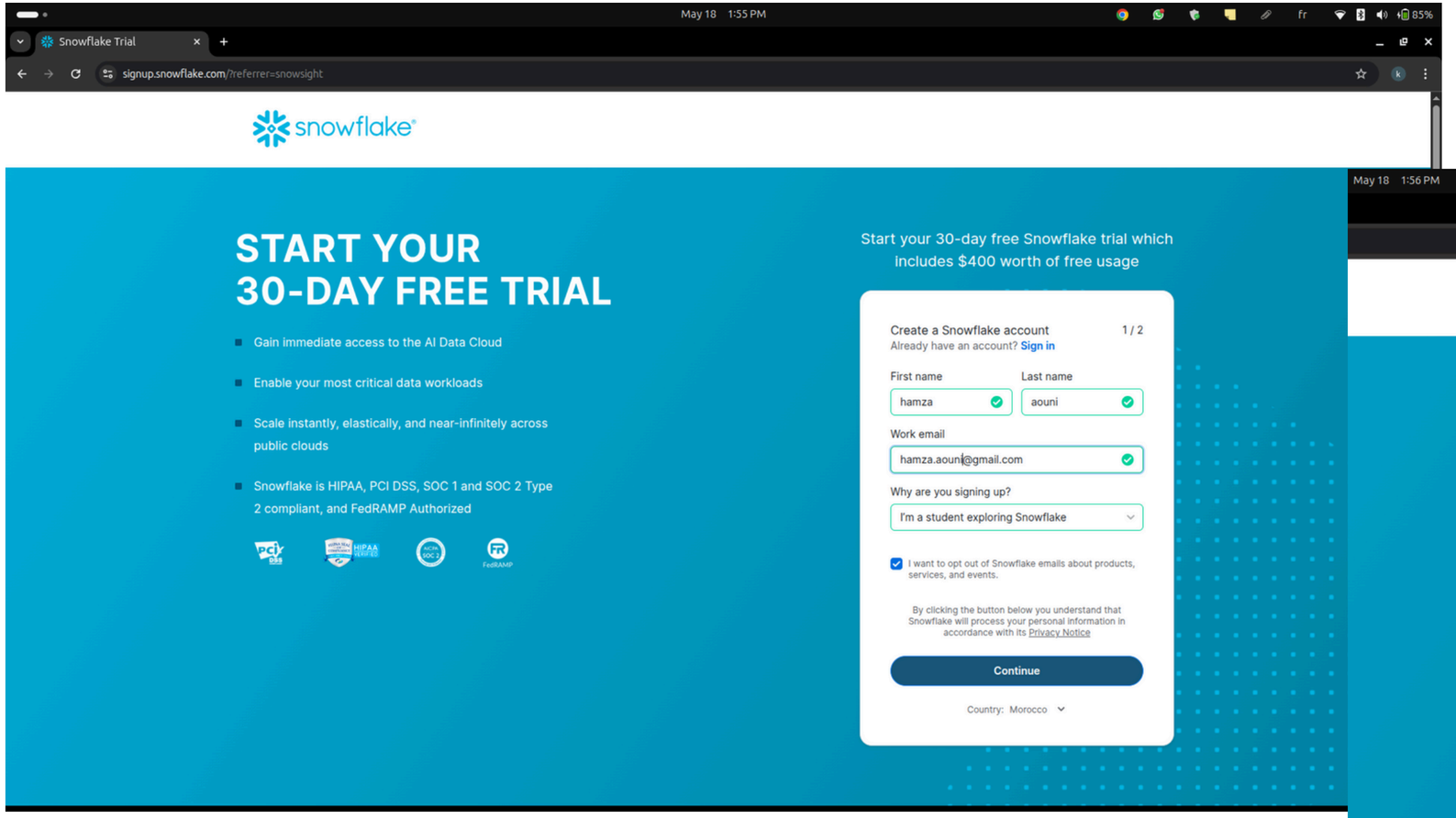
DESIGN:

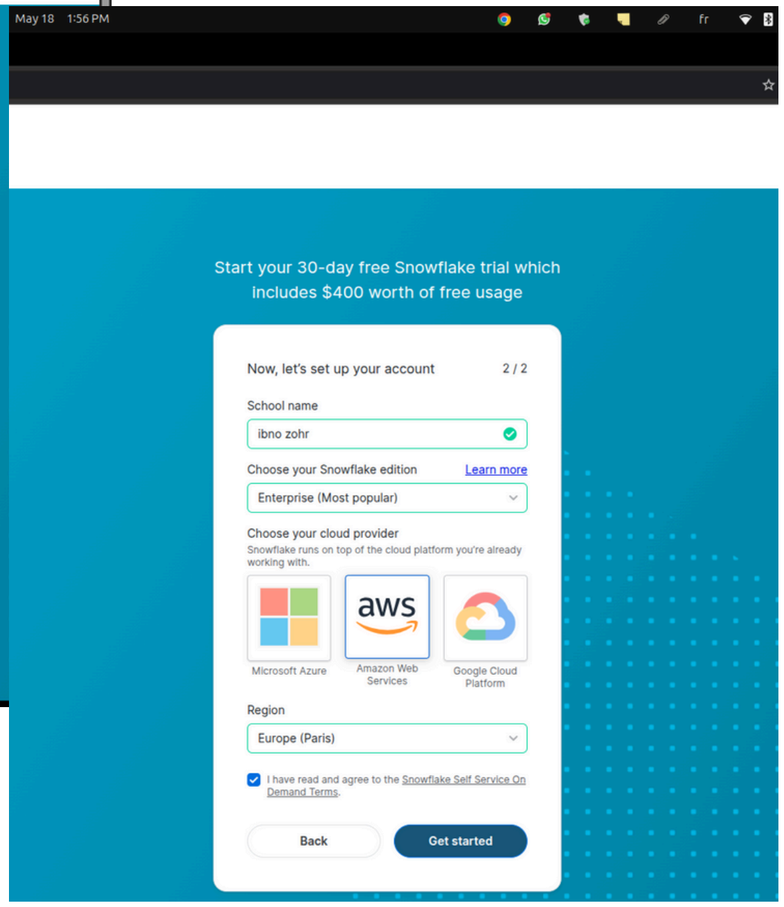
Built for analytics (Streams & Tasks), unlike PostgreSQL's transactional focus.

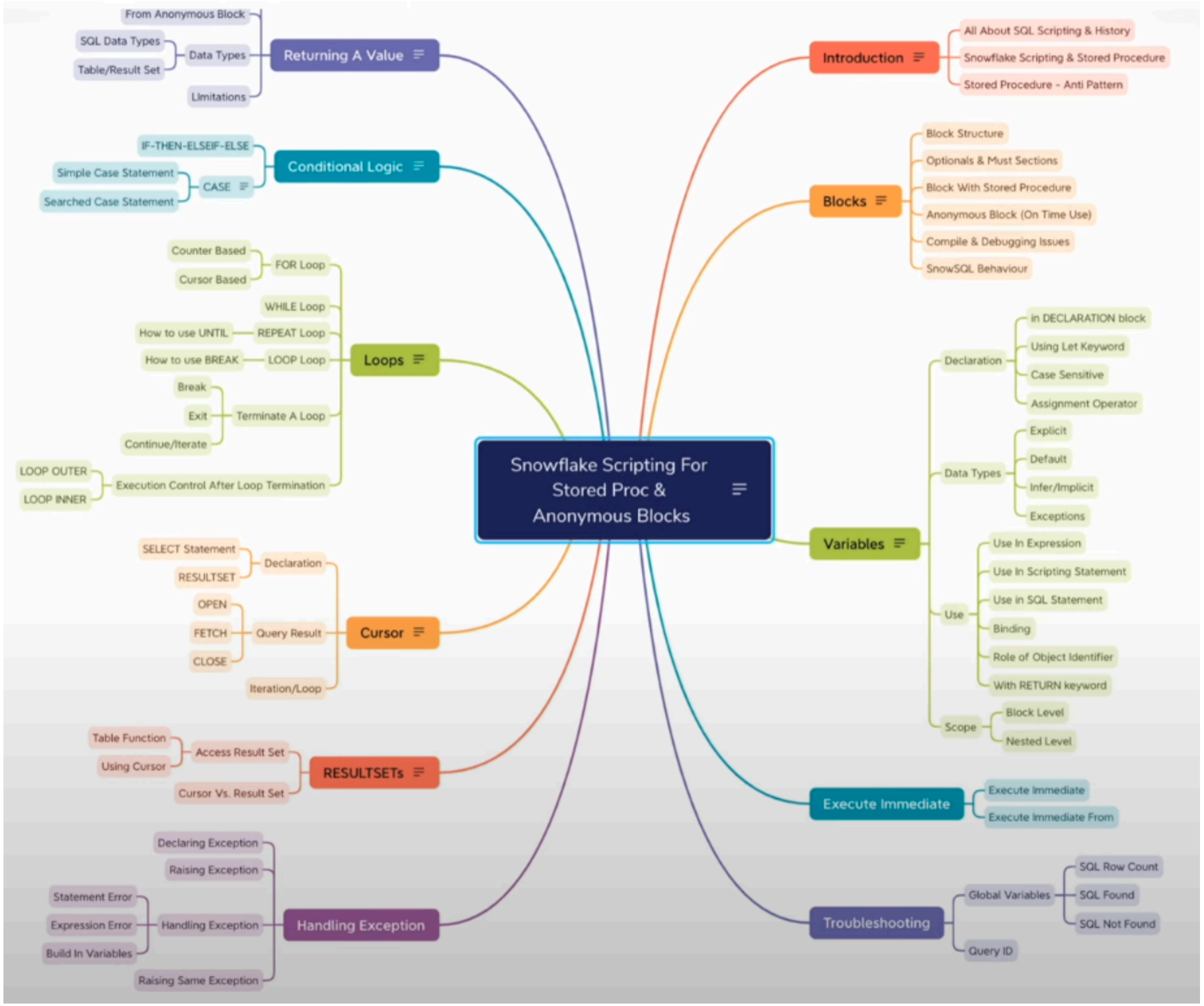
snowflake

↳ **Practice live**

Snowflake Free Trial Edition







"If you're skilled in SQL and PL/SQL, working with Snowflake becomes pretty straightforward"

snowflake

CONCLUSION

"Snowflake's serverless architecture and robust security make it a top choice for modern data-driven businesses."

Some Ressources

- Customers CSV with 100 records

- How To Load Data Into Snowflake
- SnowFlake course
- SnowFlake Architectuer

- SnowFlake Getting Started
- Snowflake Architecture
- softweb solutions
- SnowSQL - developers

snowflake

Thank you very much!