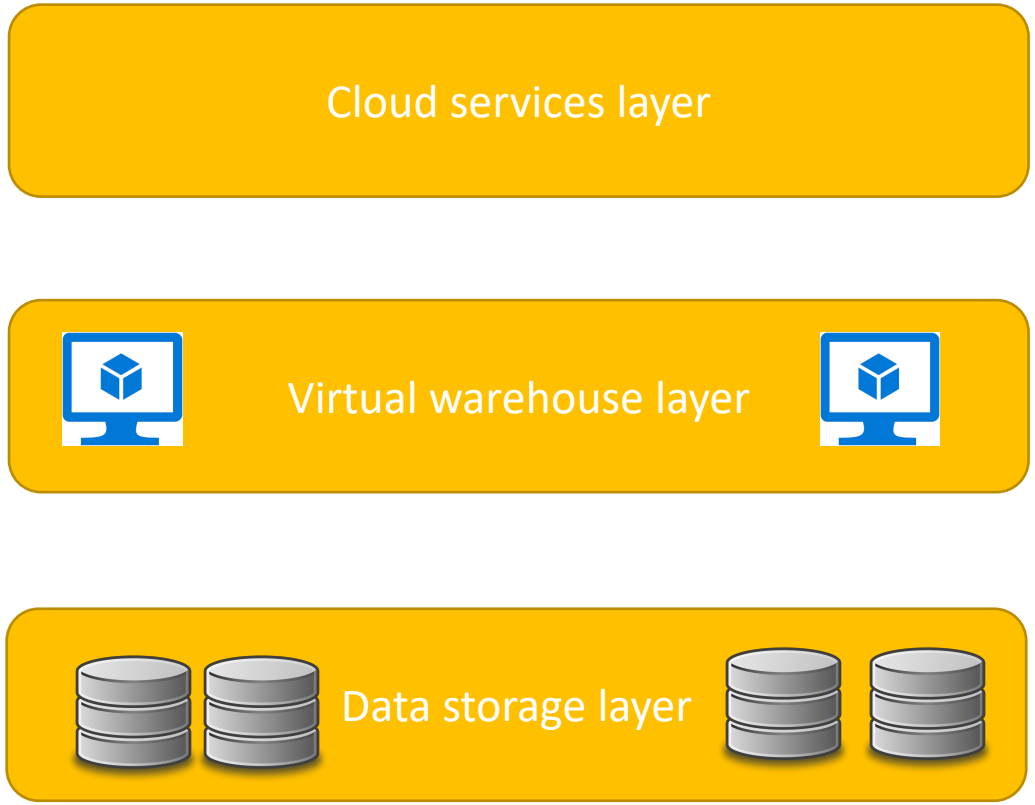
Snowflake fundamentals



Snowflake Architecture

Multi cluster shared data architecture

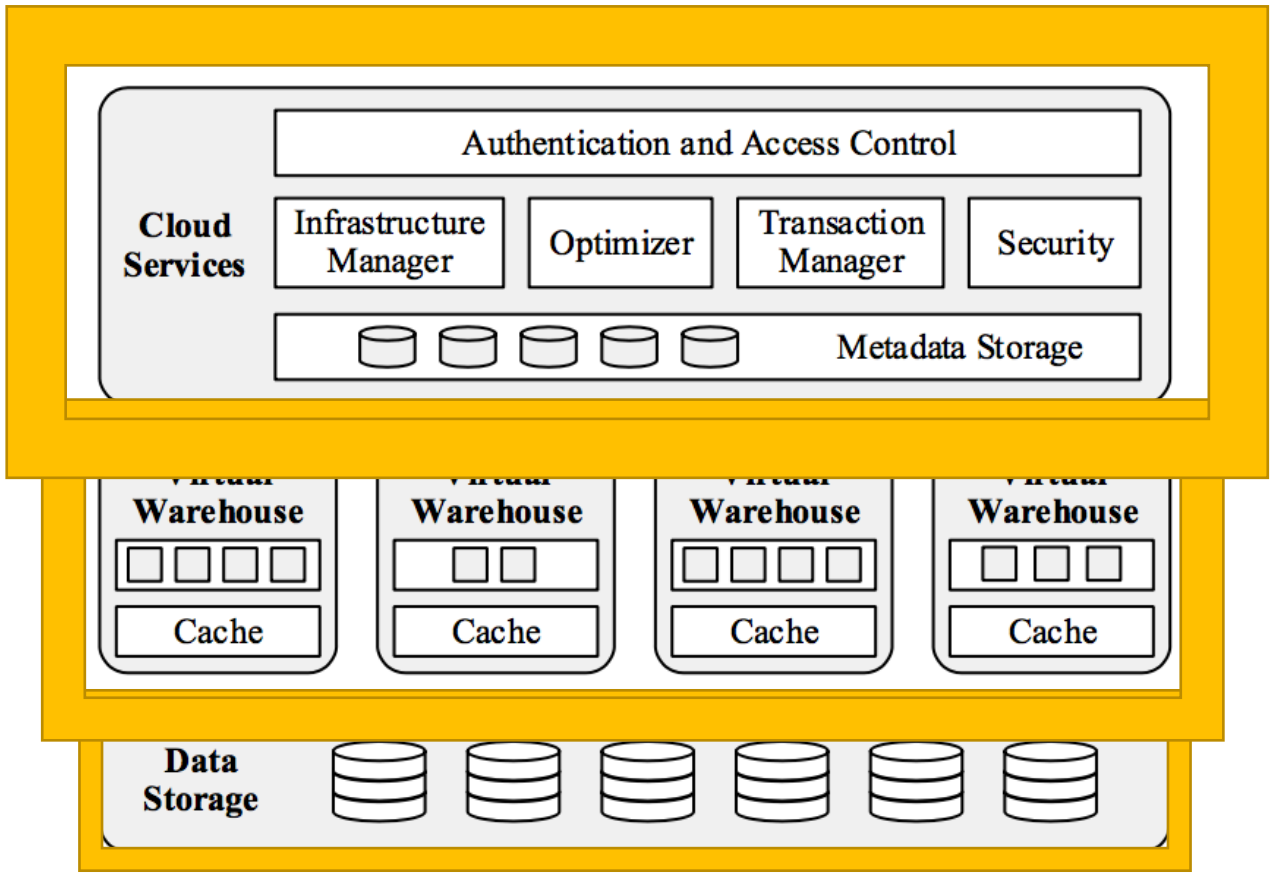


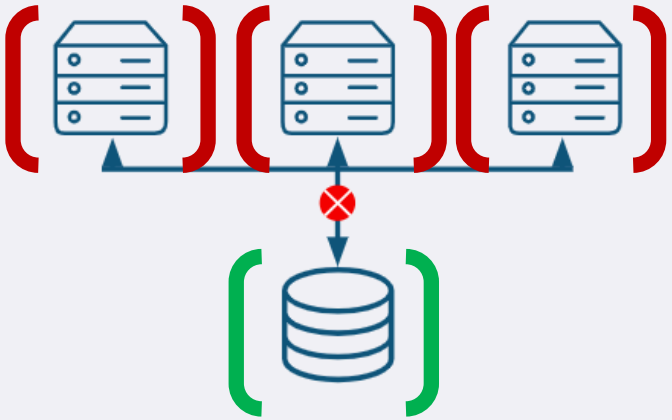
Figure 1: Multi-Cluster, Shared Data Architecture

Snowflake Architecture

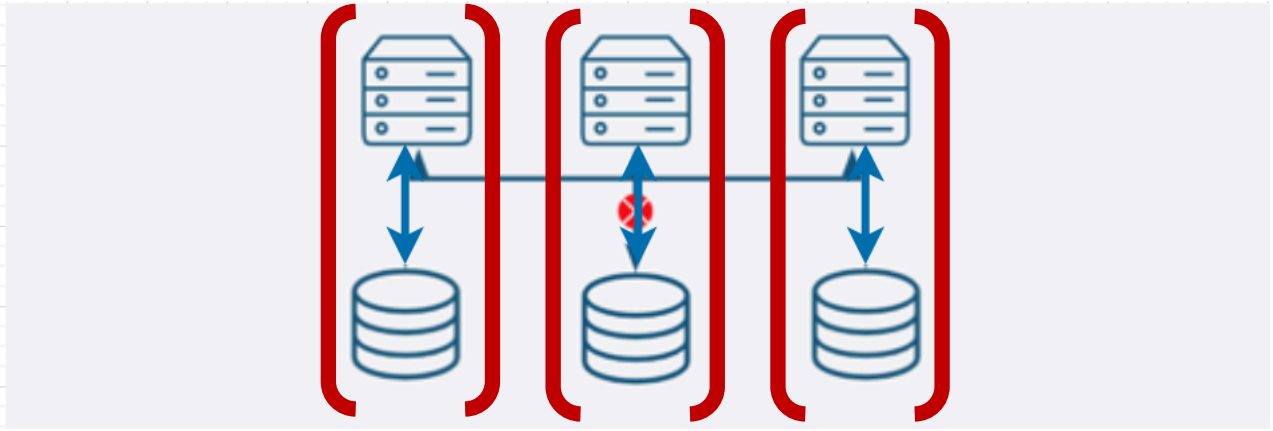
Multi cluster shared data architecture

Shared disk Architecture

- Scalability is limited.
- Hard to maintain data consistency across the cluster.
- Bottle neck of communication with shared disk.



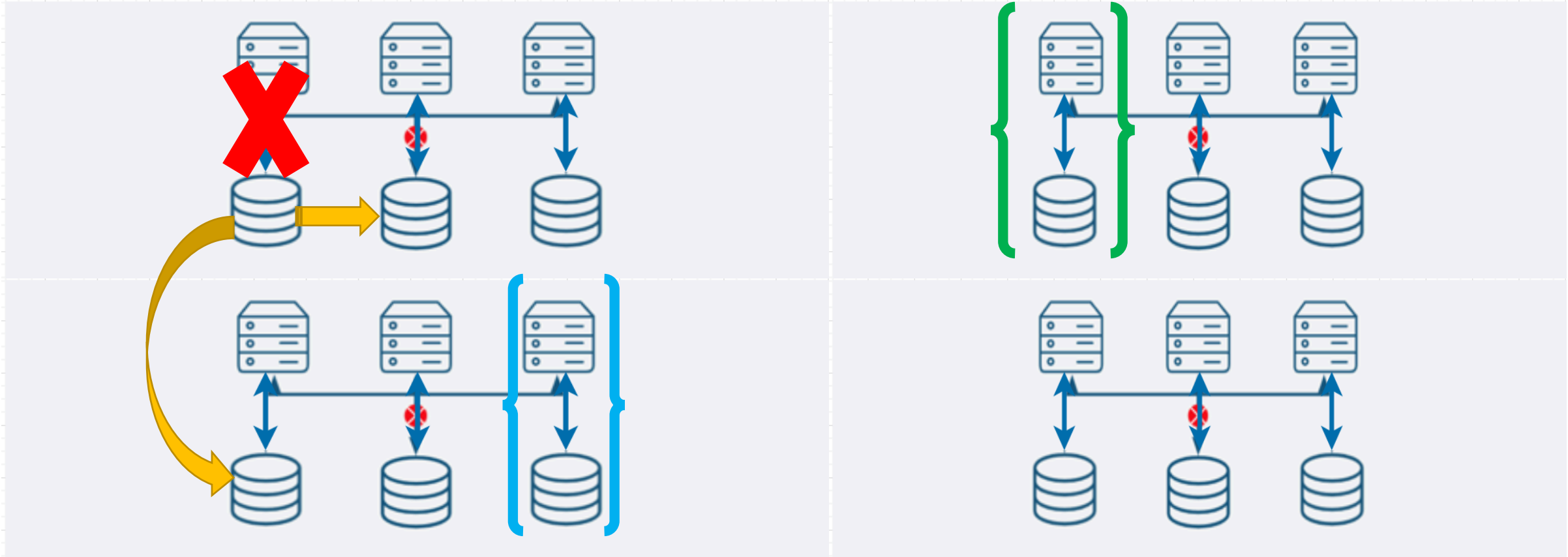
Shared nothing Architecture

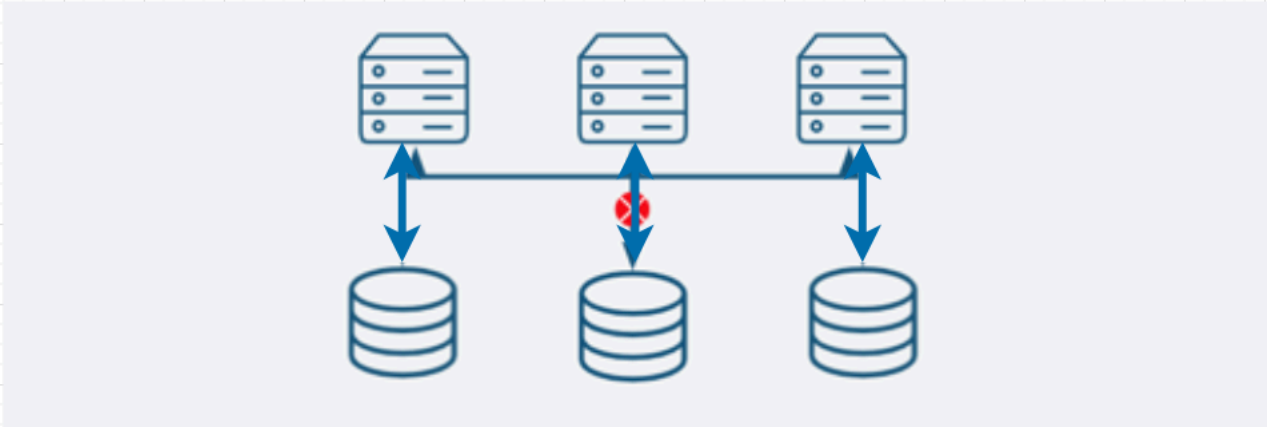


Shared nothing Architecture

It scales processing and compute together.

It moves data storage close to compute.





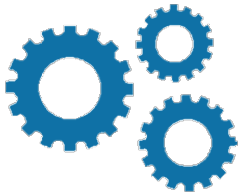
Shared nothing Architecture

Data distributed across the cluster requires shuffling between nodes.

Performance is heavily dependent on how data is distributed across the nodes in the system.

Compute can’t be sized independently of storage.

Heterogeneous workload



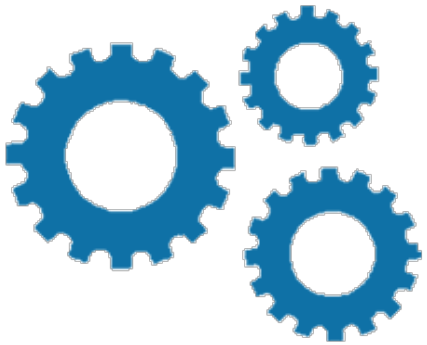
Bulk loading

Low Compute

High I/O

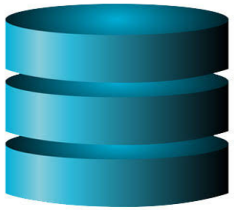
Requires higher I/O bandwidth and light compute.

Heterogeneous workload



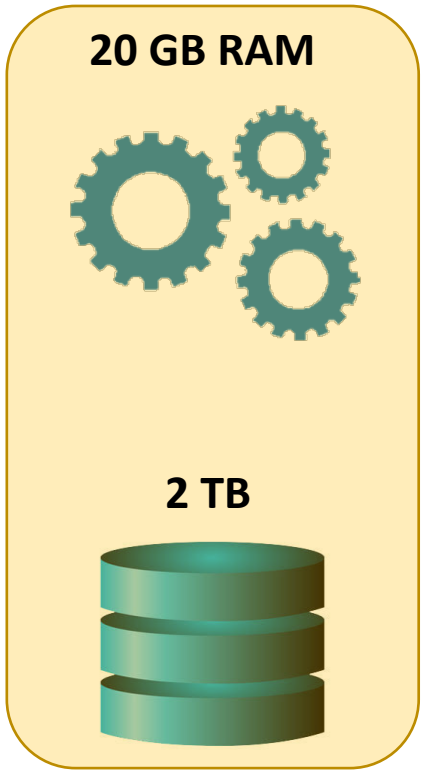
Heavy Compute

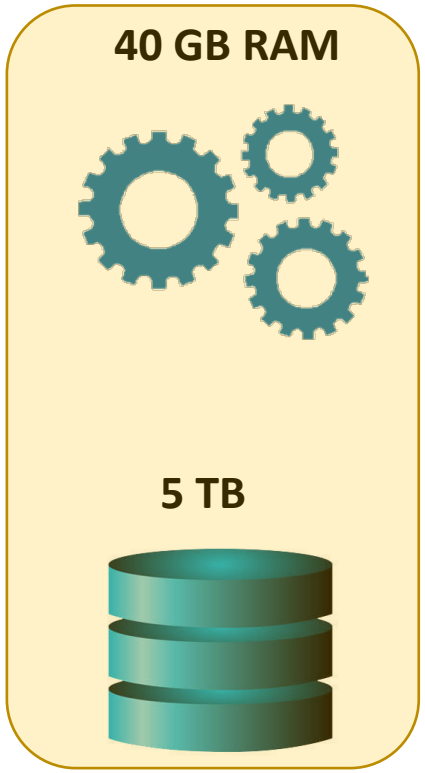
Data processing



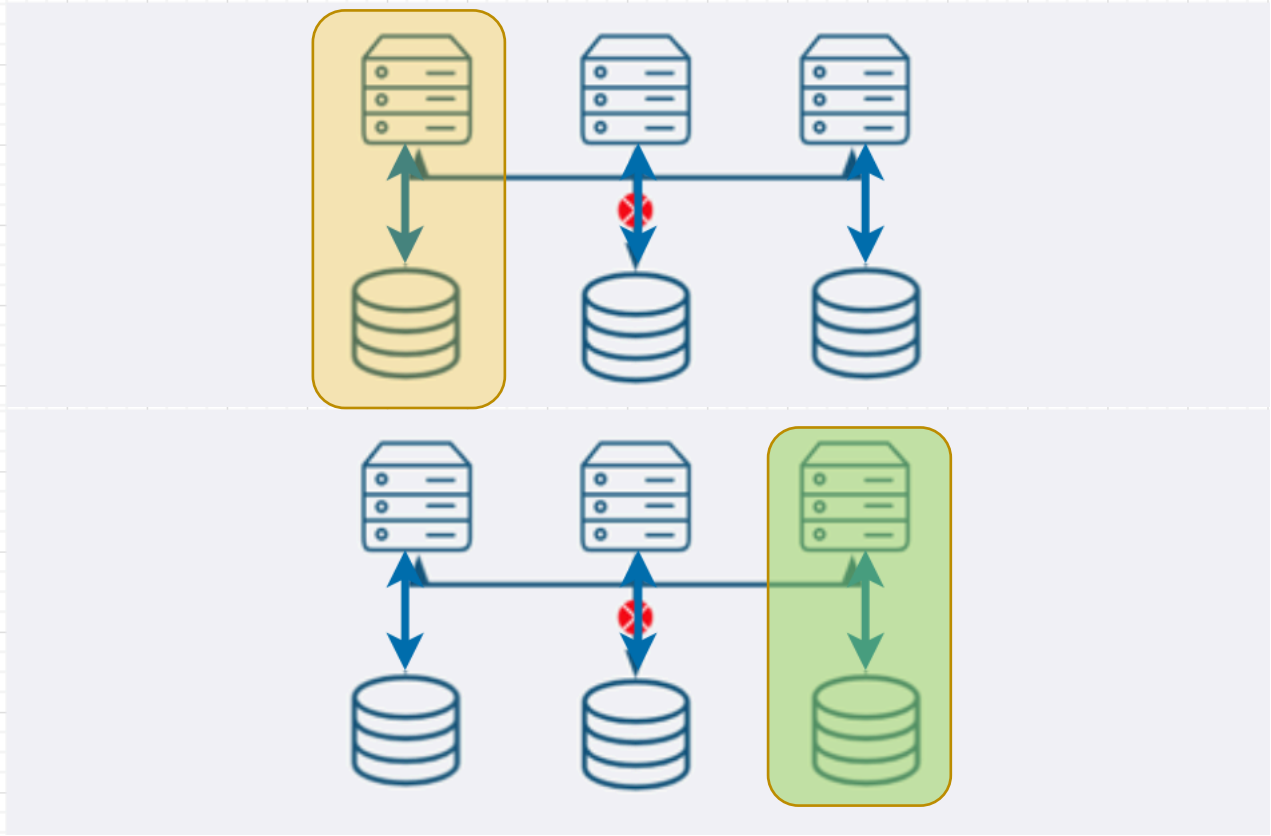
Low I/O

Requires lower I/O bandwidth and heavy compute





Membership changes



Upgrades

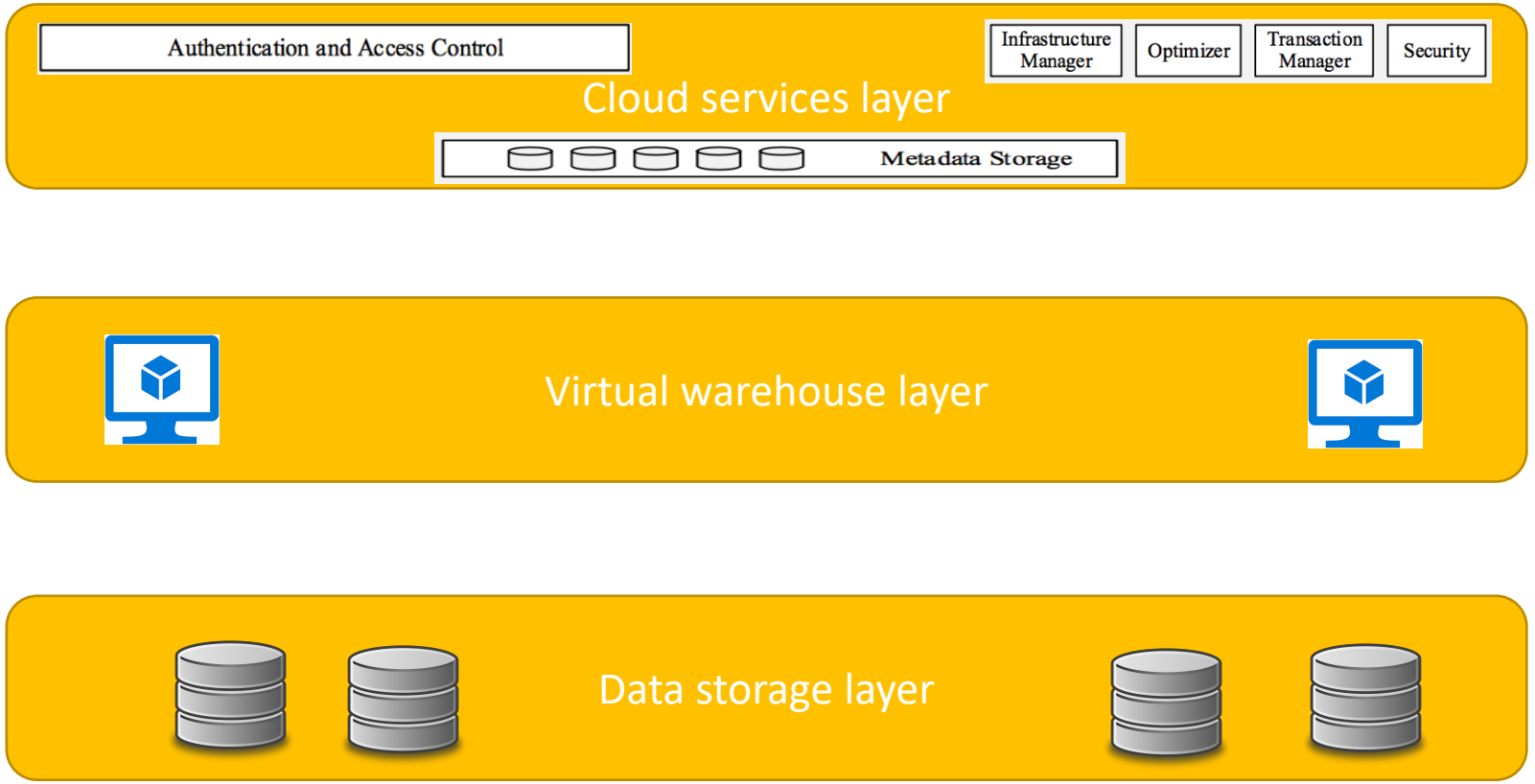
Shared nothing Architecture

Heterogenous Workload and homogenous hardware.

Membership changes.

Problem with software upgrades.

Multi cluster shared data architecture



COST

Snowflake architecture

Impact

ARCHITECTURE DEMO CACHING

- **By the end of this section you will understand how data processing happens under the hood.**
- **You will understand how snowflake architecture layers will interact with each other.**
- **You will understand how caching works in snowflake.**

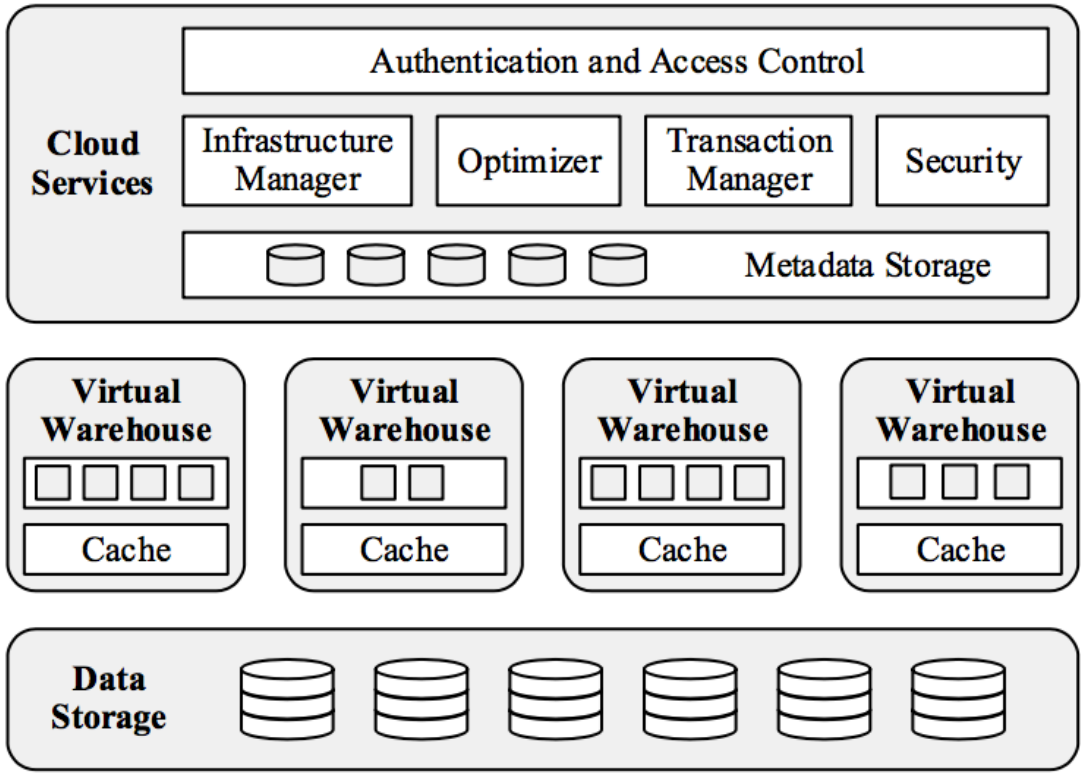
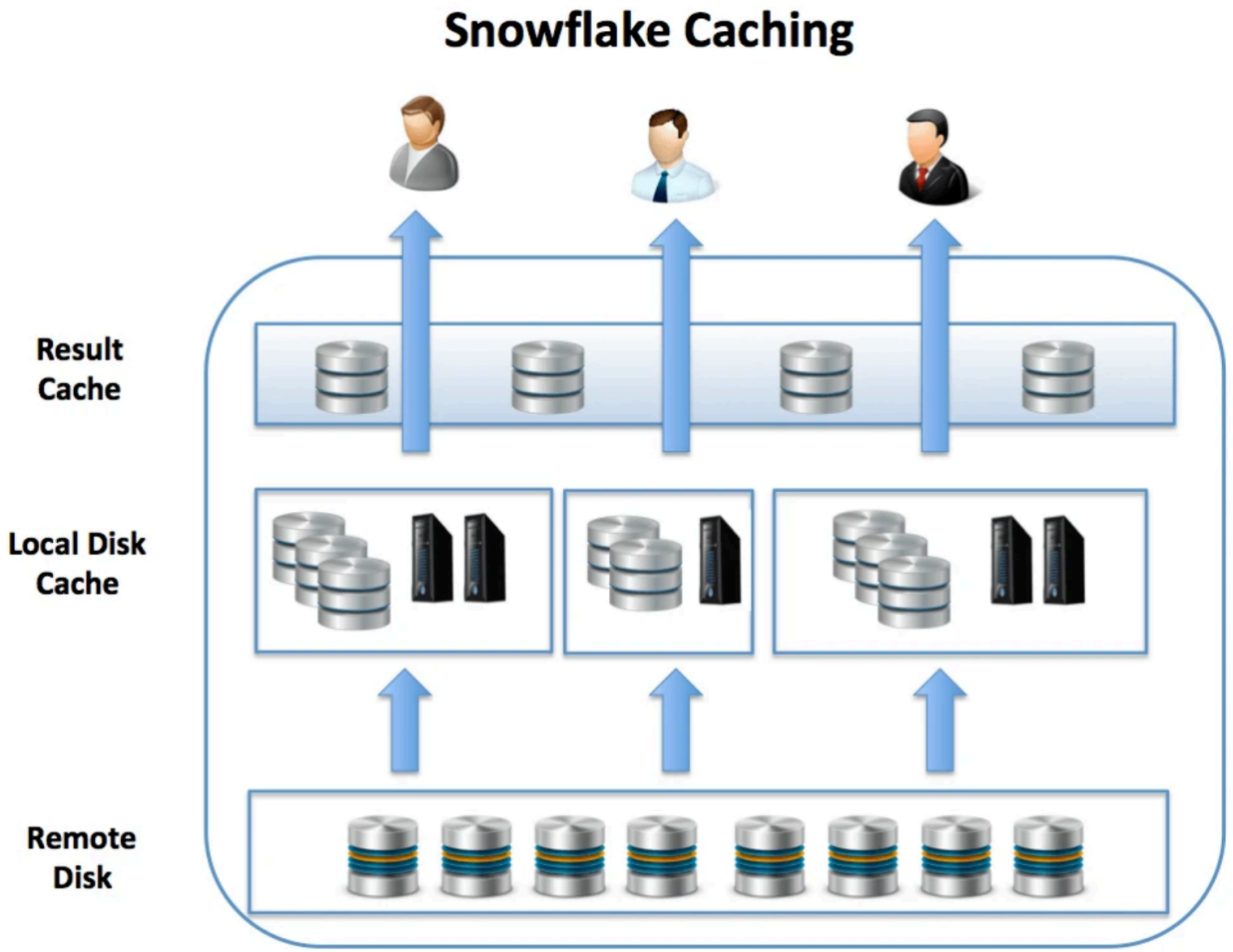


Figure 1: Multi-Cluster, Shared Data Architecture

Architecture



Caching in snowflake.