CS 764: Topics in Database Management Systems

Lecture 19: Snowflake

Xiangyao Yu

11/9/2020

Announcements

Reviews for the remaining lectures

- You can skip up to three reviews

Guest lecture on Wednesday

- Guest lecture: 13:00–14:00 Wednesday
- Round-table discussion: 14:00–15:00 Wednesday
- **Use the lecture zoom link for both the lecture and the discussion**

Today's Paper

The Snowflake Elastic Data Warehouse

Benoit Dageville, Thierry Cruanes, Marcin Zukowski, Vadim Antonov, Artin Avanes,
Jon Bock, Jonathan Claybaugh, Daniel Engovatov, Martin Hentschel,
Jiansheng Huang, Allison W. Lee, Ashish Motivala, Abdul Q. Munir, Steven Pelley,
Peter Povinec, Greg Rahn, Spyridon Triantafyllis, Philipp Unterbrunner

Snowflake Computing

ABSTRACT

We live in the golden age of distributed computing. Public cloud platforms now offer virtually unlimited compute and storage resources on demand. At the same time, the Software-as-a-Service (SaaS) model brings enterprise-class systems to users who previously could not afford such systems due to their cost and complexity. Alas, traditional data warehousing systems are struggling to fit into this new environment. For one thing, they have been designed for

Keywords

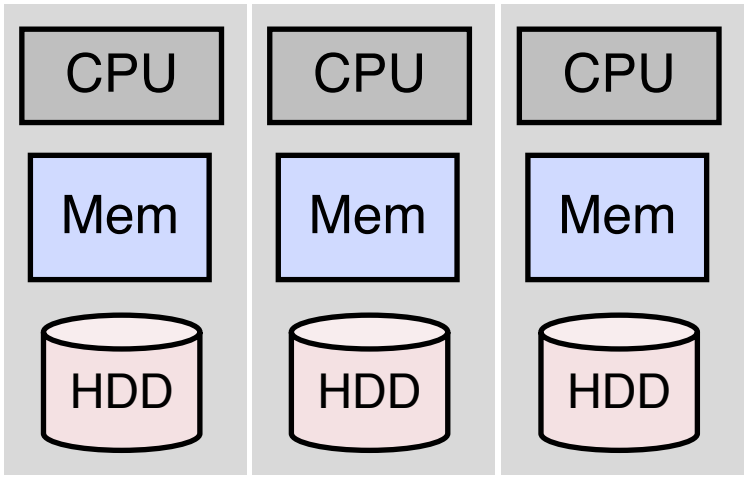
Data warehousing, database as a service, multi-cluster shared data architecture

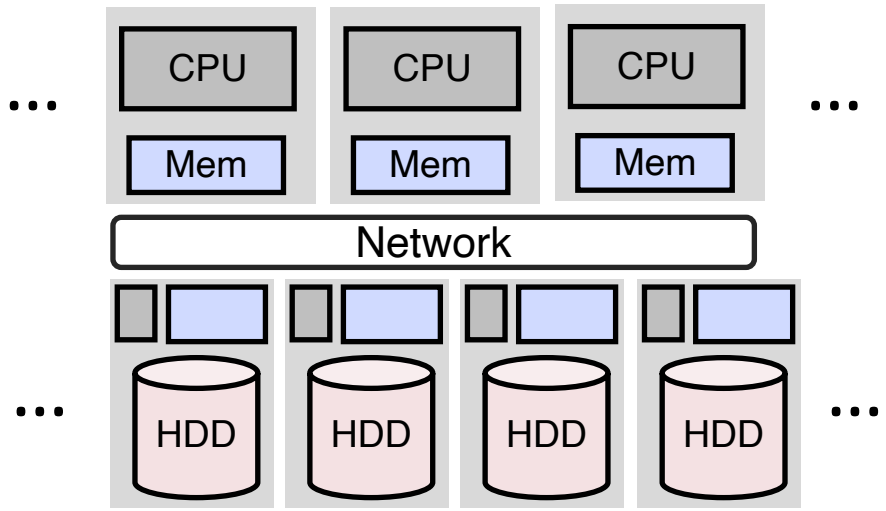
1. INTRODUCTION

The advent of the cloud marks a move away from software delivery and execution on local servers, and toward shared data centers and software-as-a-service solutions hosted by

SIGMOD 2016

On-Premises vs. Cloud





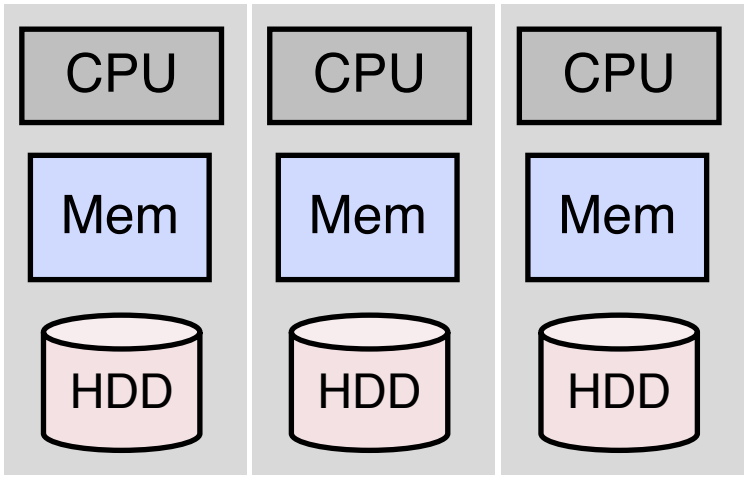
On-premises

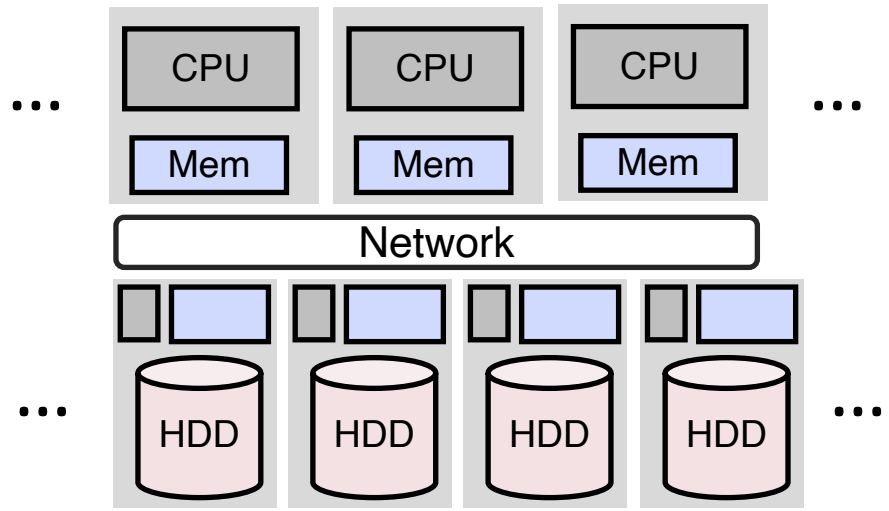
- Fixed and limited hardware resources

Cloud

- Virtually infinite computation & storage, Pay-as-you-go price model

On-Premises vs. Cloud





On-premises

- Fixed and limited hardware resources
- Shared-nothing architecture

Cloud

- Virtually infinite computation & storage, Pay-as-you-go price model
- Disaggregation architecture

Database in Cloud

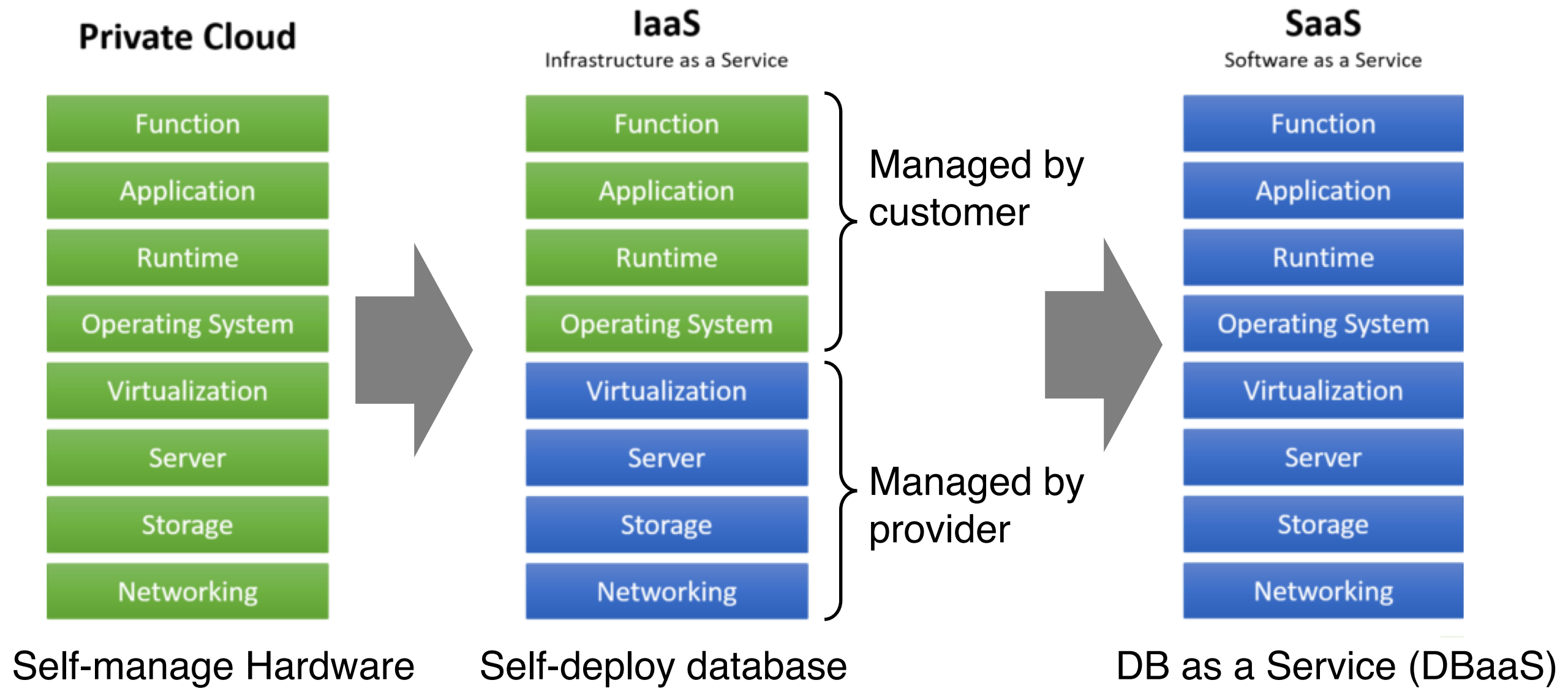
- Cloud databases can be **cheaper**
  - Economies of scale

- Cloud databases can be **better**
  - Simpler administration, development, and deployment
  - Unlimited storage and computation
  - Pay-as-you-go
  - Geo-distribution

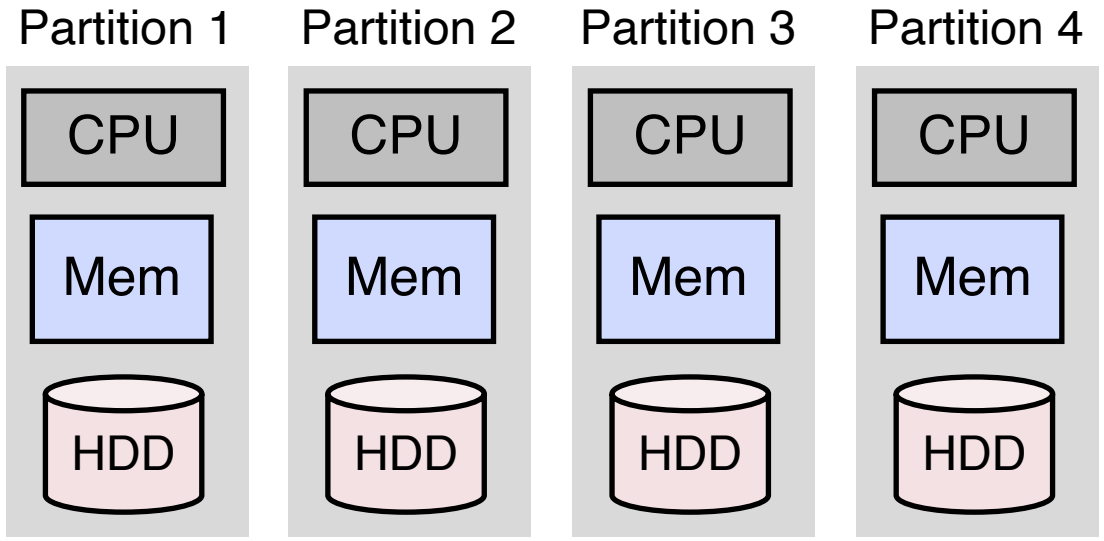
Amazon Athena

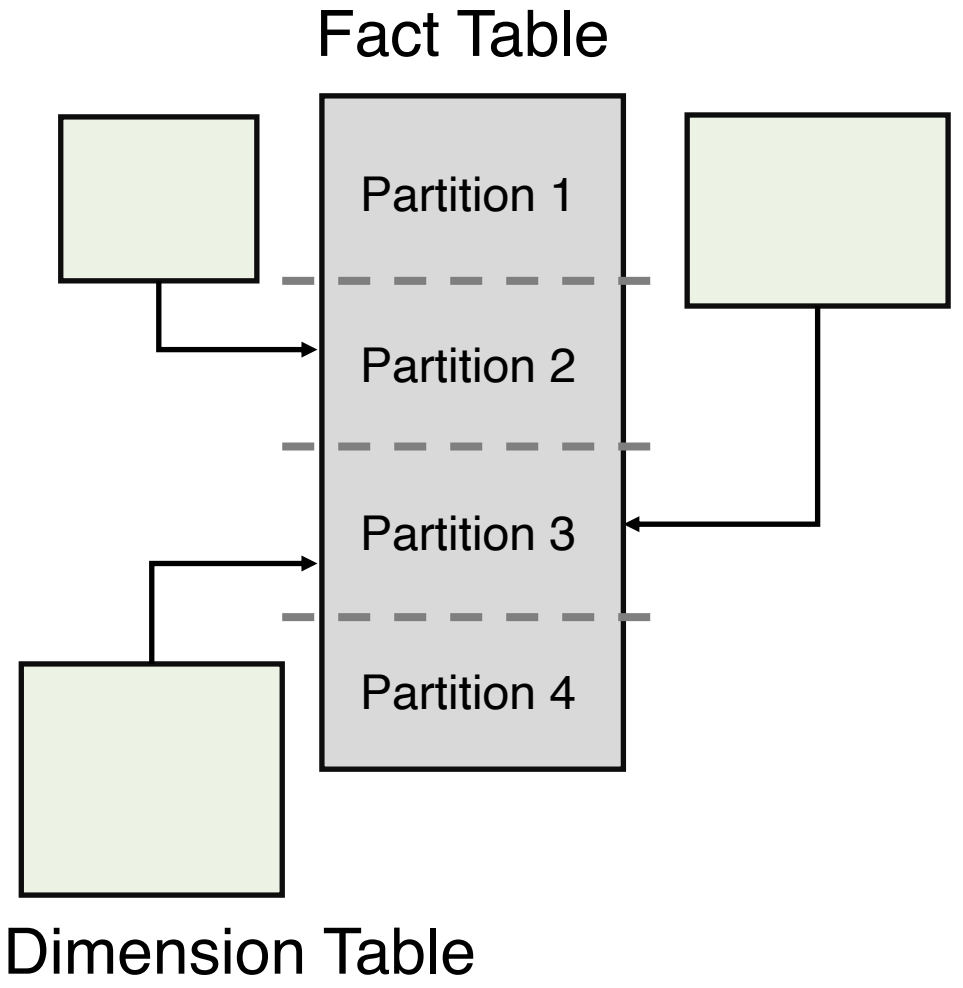
Cloud
Spanner

Cloud Computing Architectures



Shared Nothing – Advantages

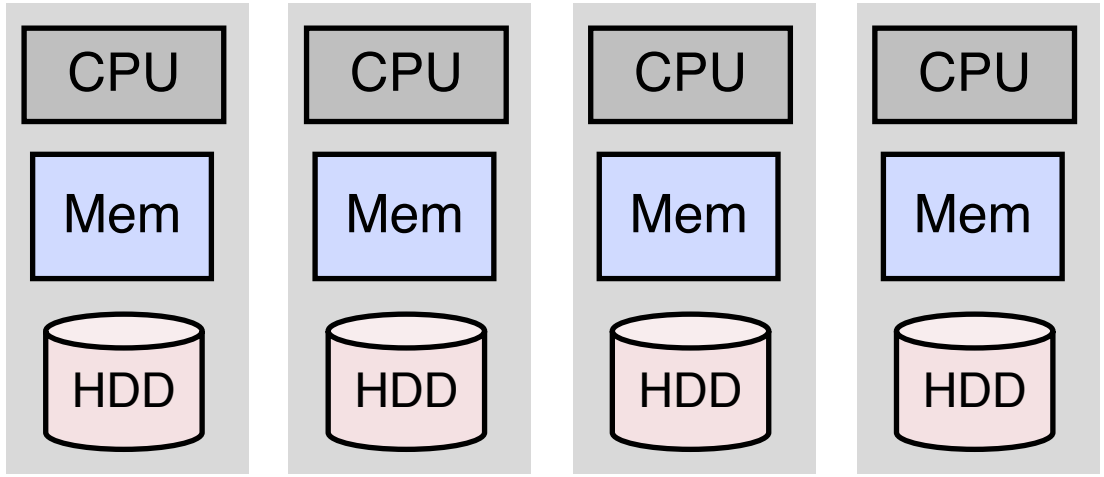


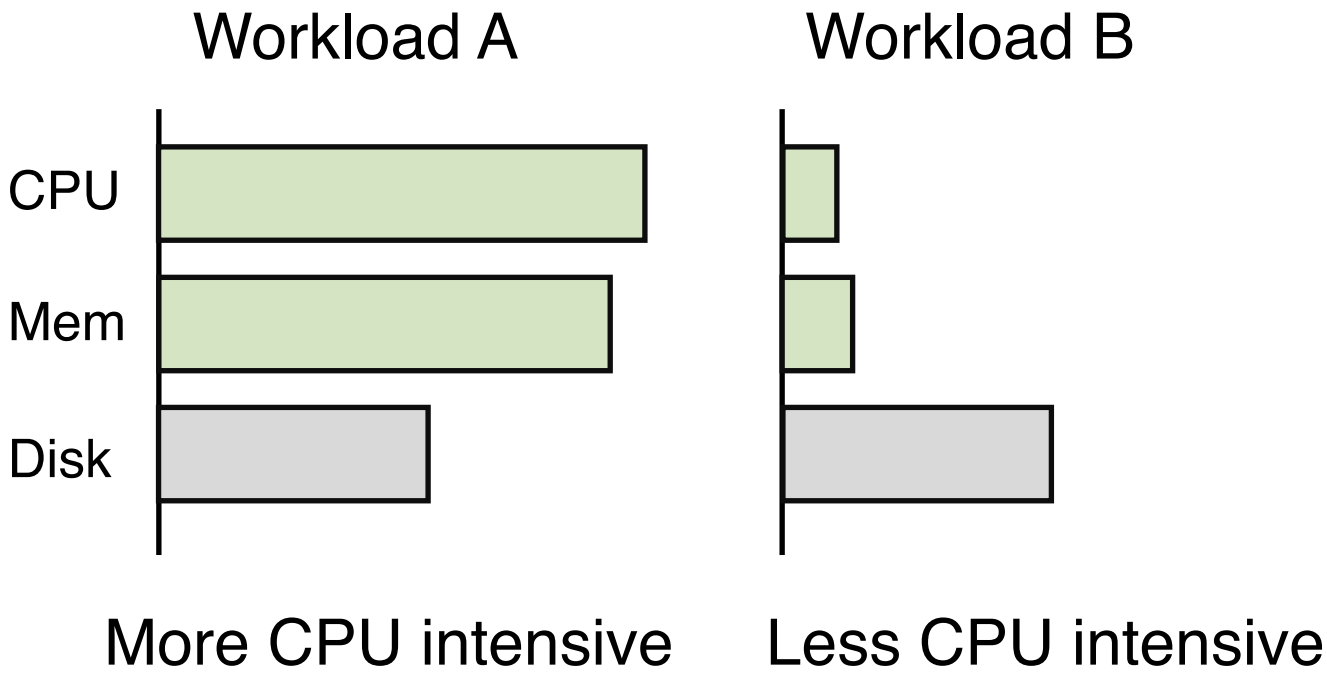


Scalability: horizontal scaling

- Scales well for star-schema queries

Shared Nothing – Disadvantages

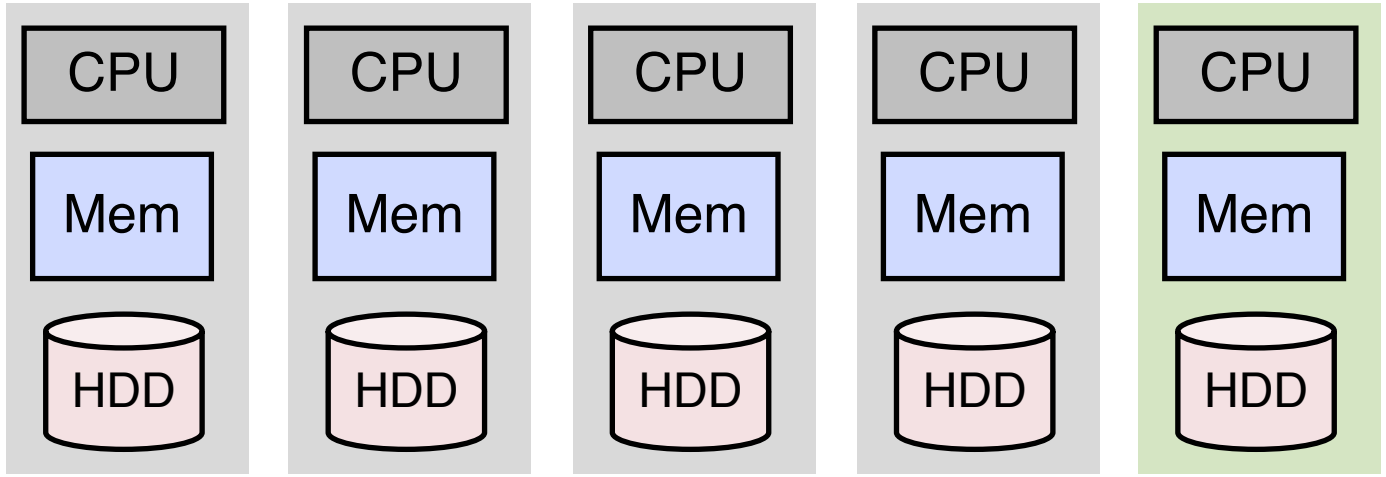




Heterogeneous workload

- Static resource provisioning cannot adjust to heterogeneous workloads

Shared Nothing – Disadvantages

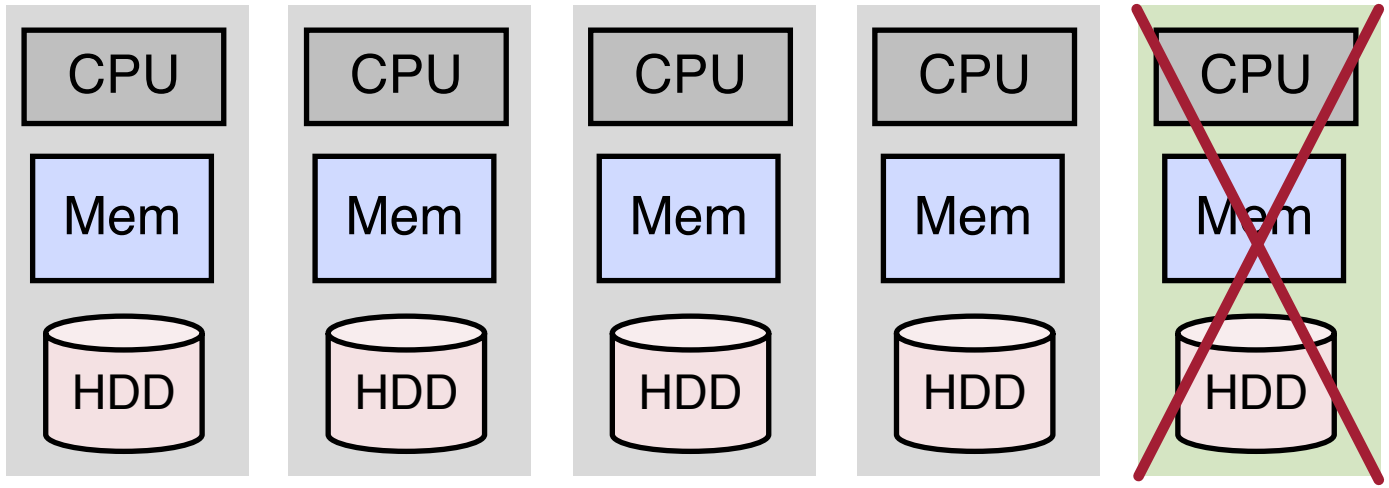


Heterogeneous workload

Membership changes

- Add a node: data redistribution

Shared Nothing – Disadvantages

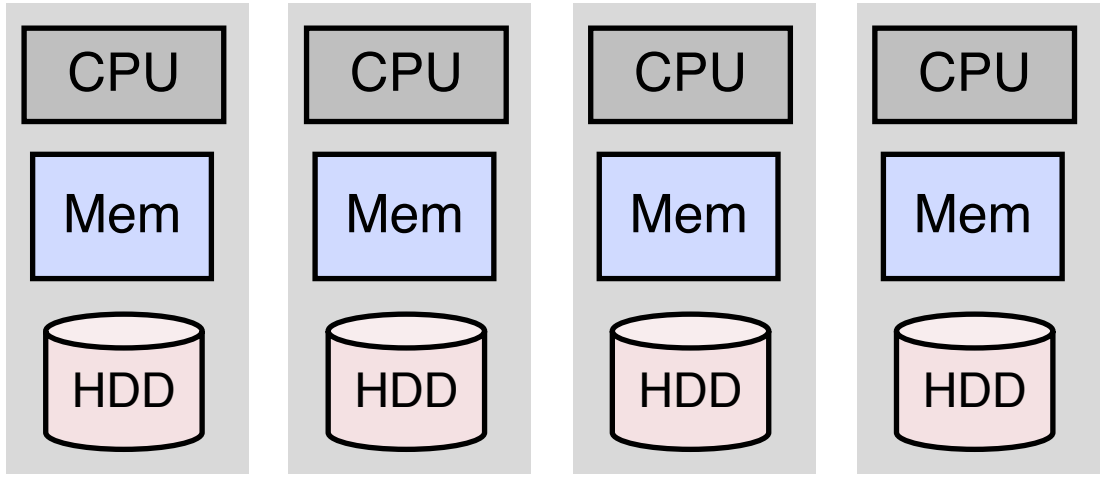


Heterogeneous workload

Membership changes

- Add a node: data redistribution
- Delete a node: similar to the fault tolerance problem

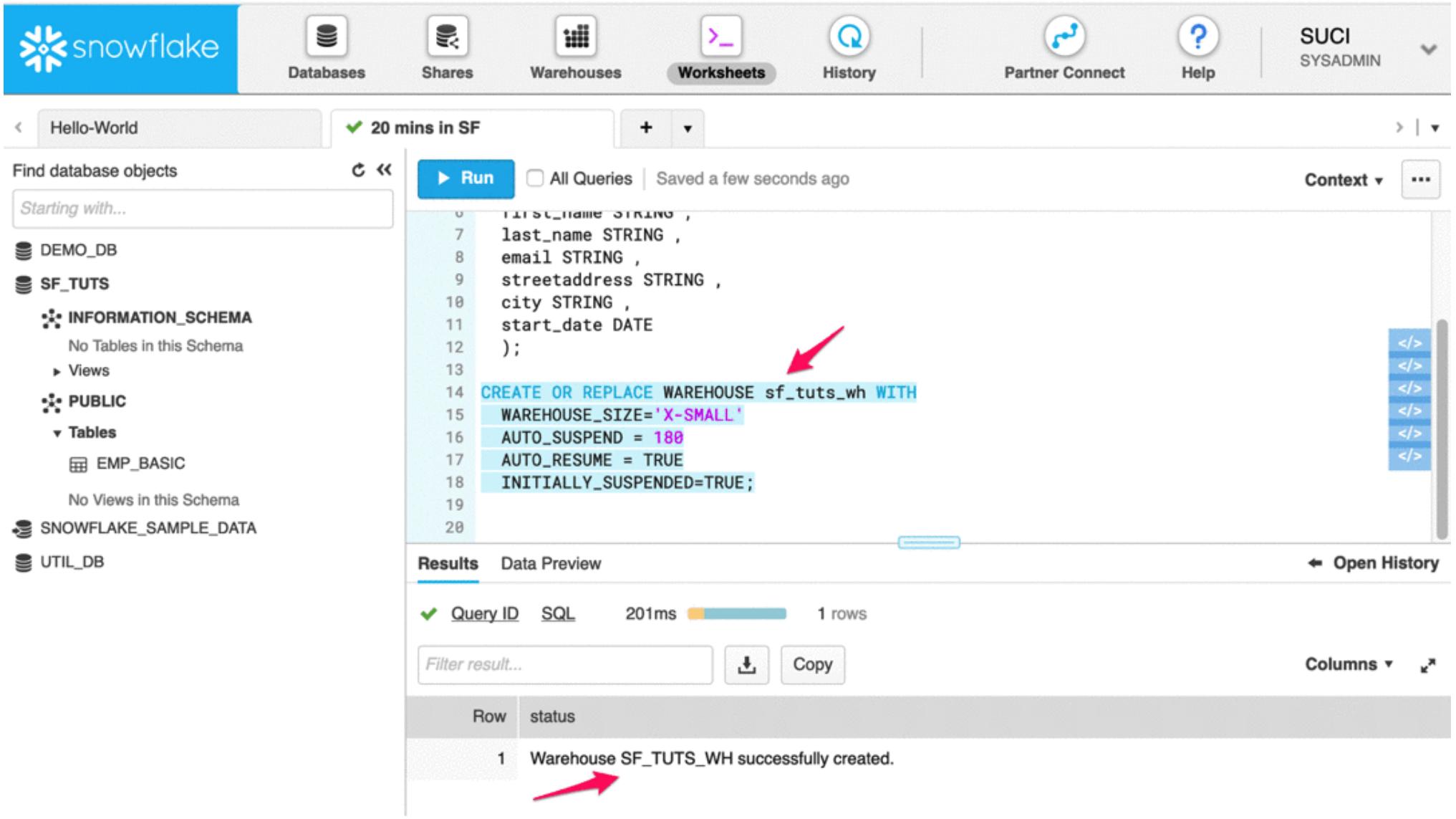
Shared Nothing – Disadvantages



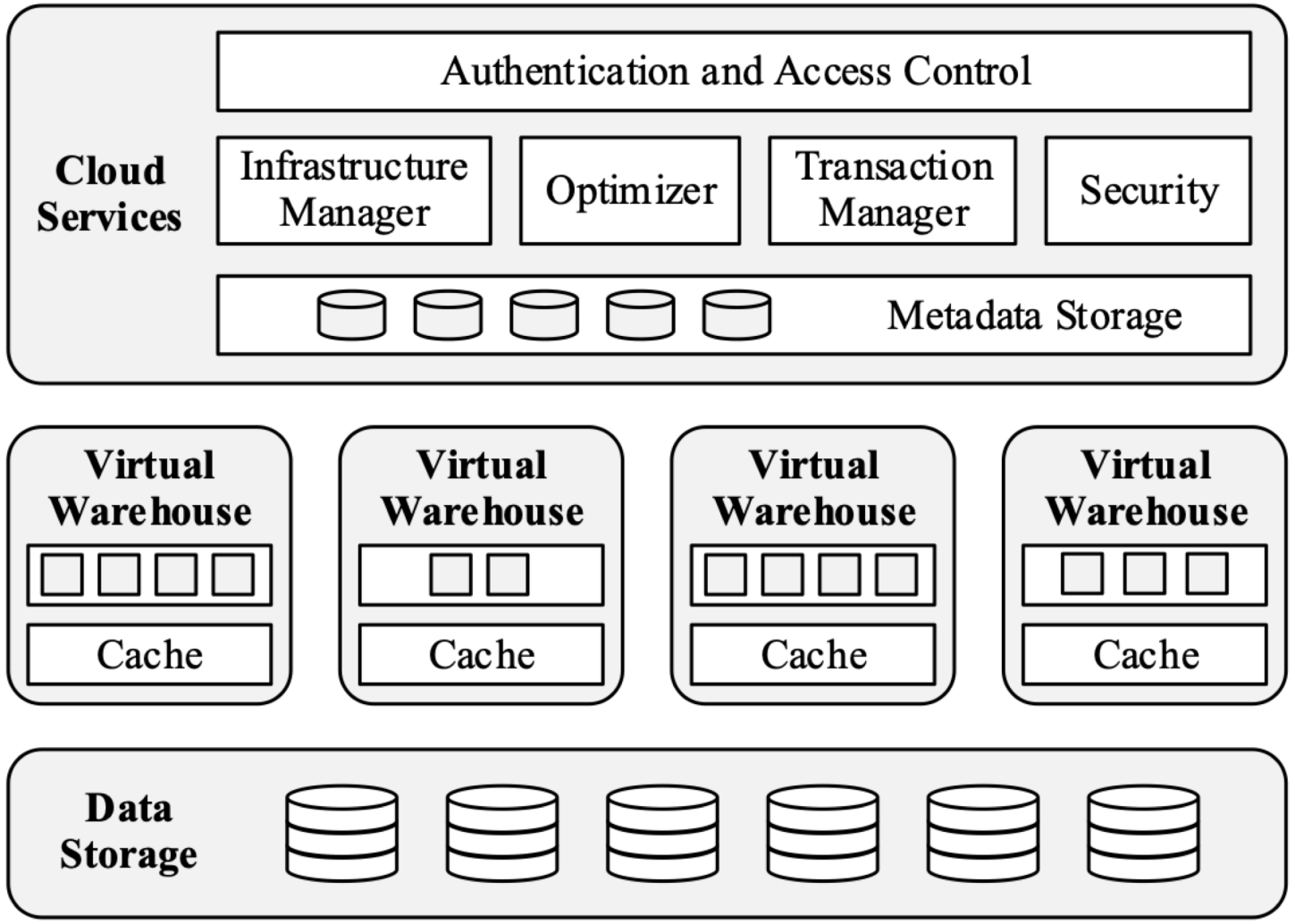
Heterogeneous workload
Membership changes
Online upgrade

- Similar to membership change but affect all nodes

Web User Interface (Serverless)



Multi-Cluster Shared-Data Architecture



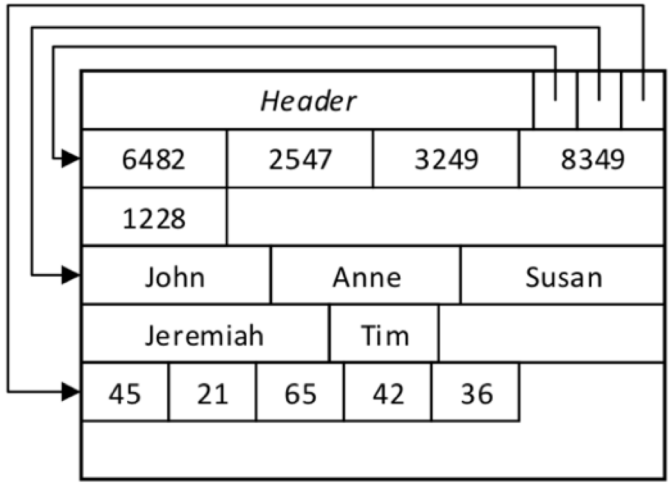
Control layer

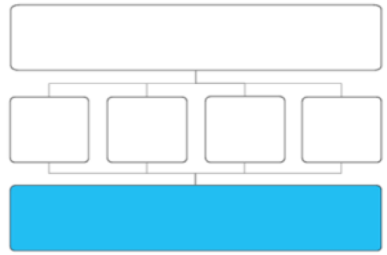
Compute layer

Storage layer

Architecture – Storage

Data format: PAX





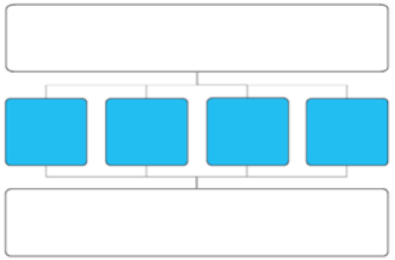
Data horizontally partitioned into immutable files (~16MB)

- An update = remove and add an entire file
- Queries download file headers and columns they are interested in

Intermediate data spilling to S3

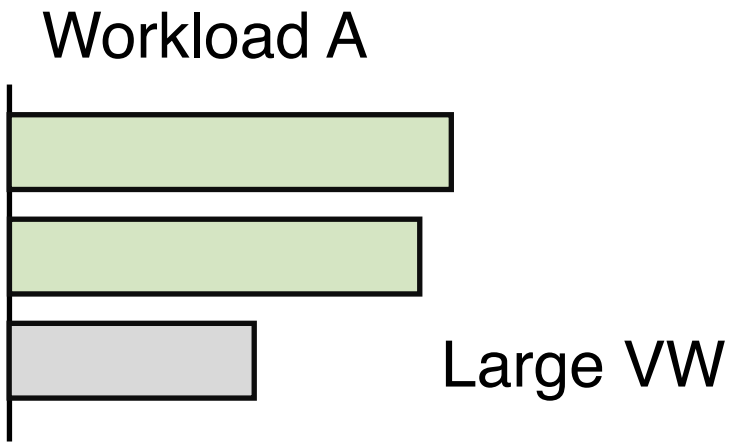
Architecture – Virtual Warehouse

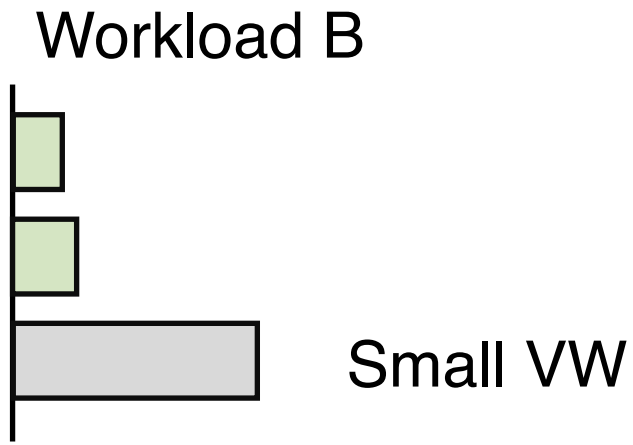
T-Shirt sizes: XS to 4XL



Elasticity and Isolation

- Created, destroyed, or resized at any point (may shutdown all VWs)
- User may create multiple VWs for multiple queries

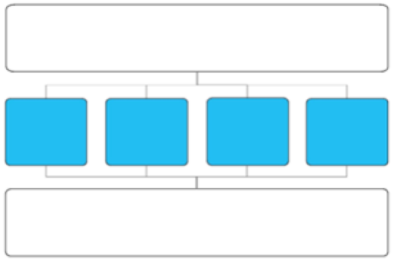


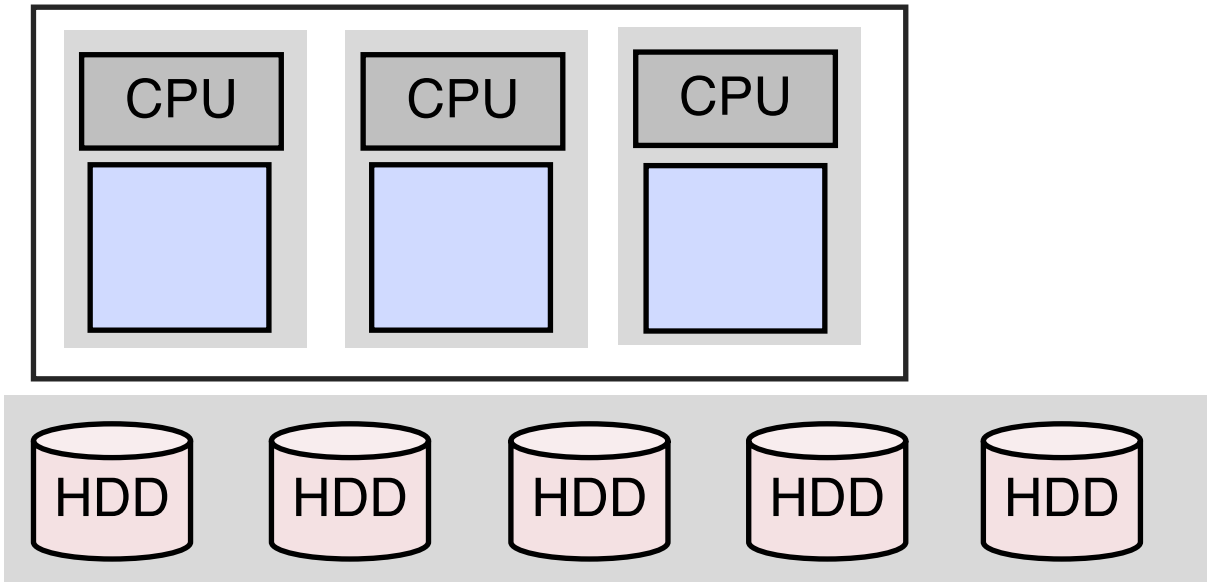


Architecture – Virtual Warehouse

Local caching

- S3 data can be cached in local memory or disk

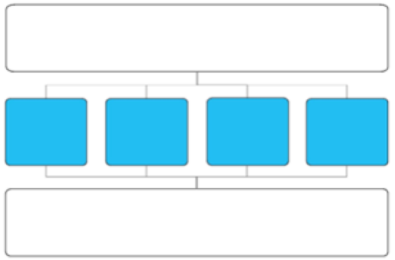




Architecture – Virtual Warehouse

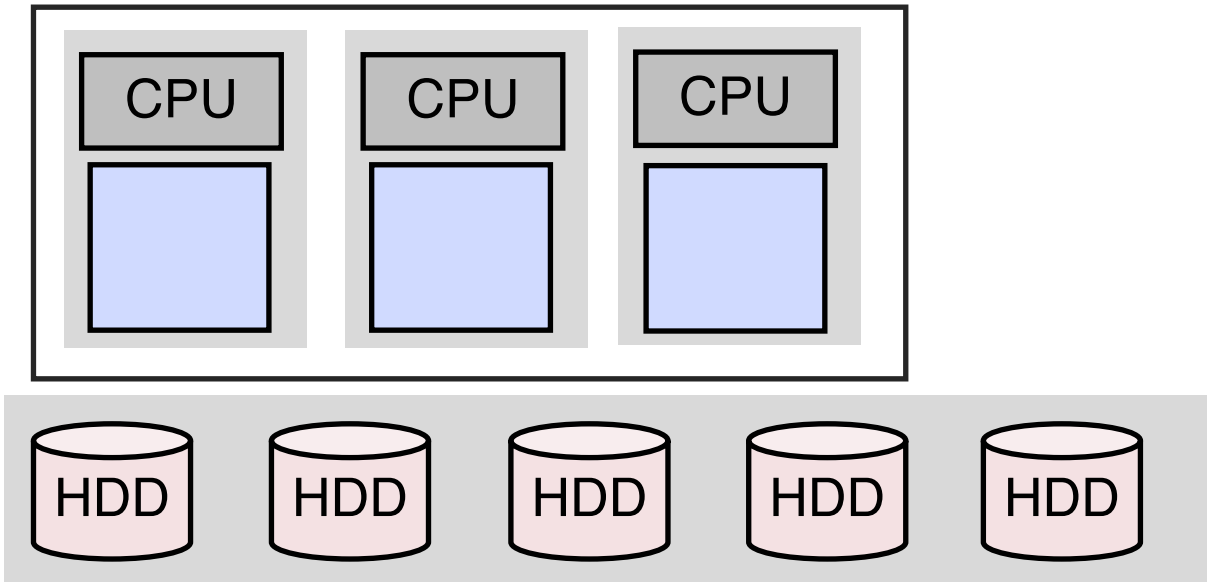
Local caching

- S3 data can be cached in local memory or disk



Consistent hashing

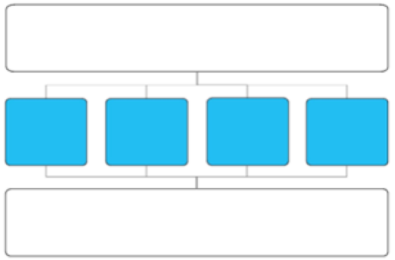
- When the hash table (n keys and m slots) is resized, only n/m keys need to be remapped



Architecture – Virtual Warehouse

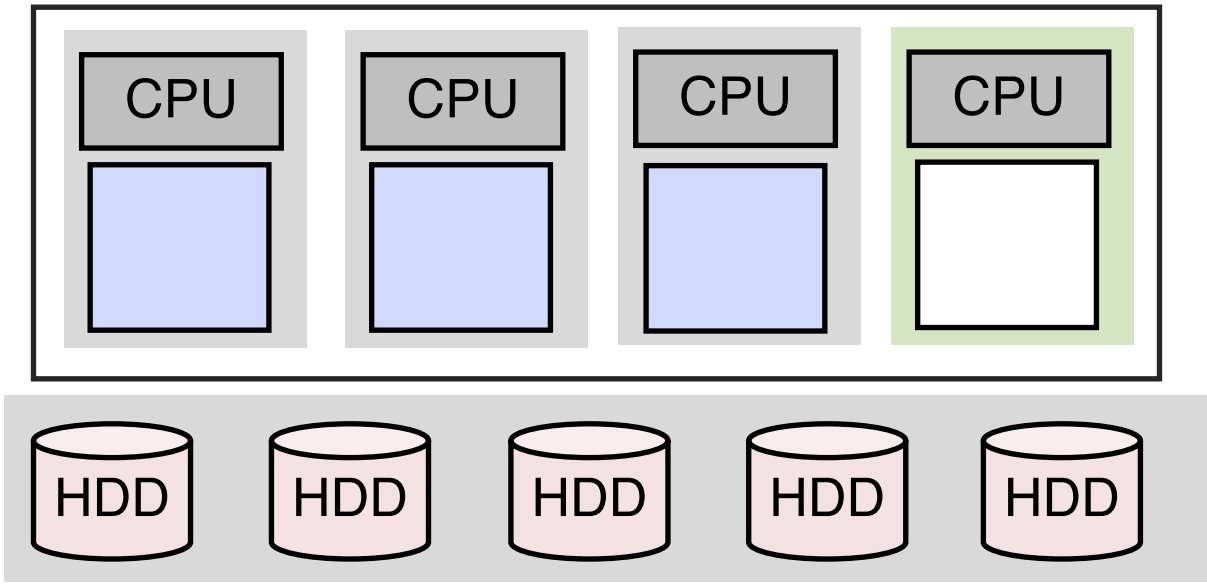
Local caching

- S3 data can be cached in local memory or disk



Consistent hashing

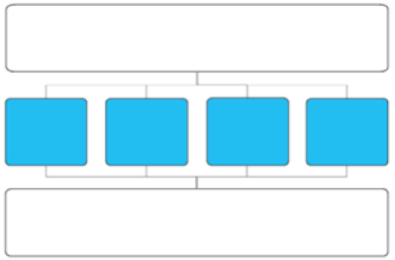
- When the hash table (n keys and m slots) is resized, only n/m keys need to be remapped



Architecture – Virtual Warehouse

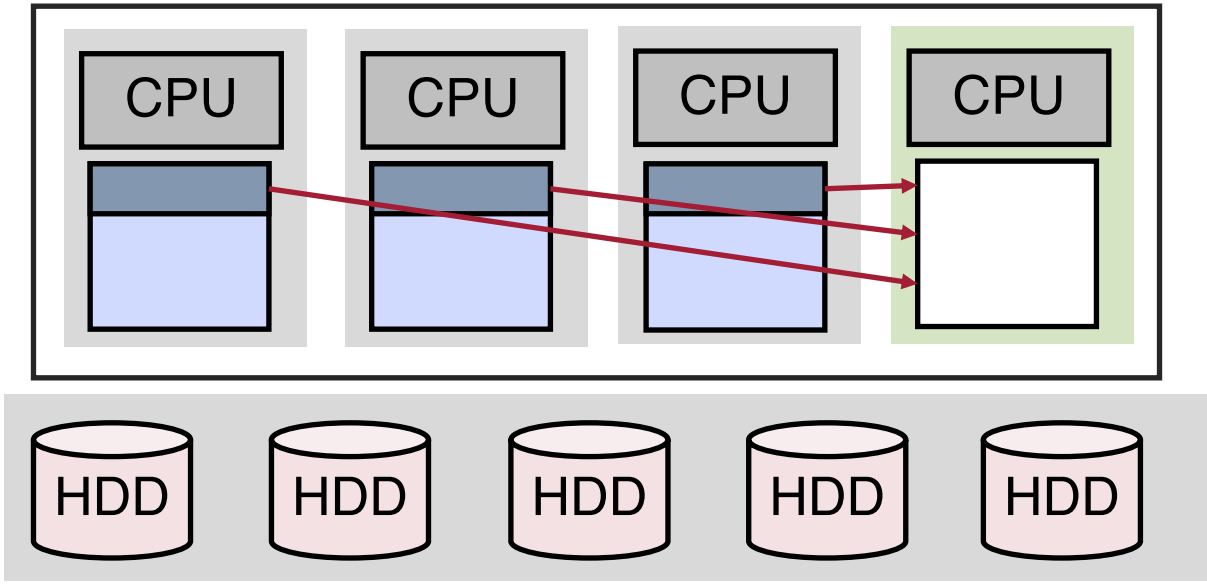
Local caching

- S3 data can be cached in local memory or disk



Consistent hashing

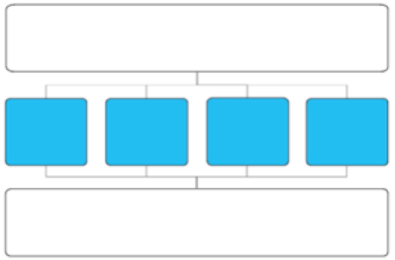
- When the hash table (n keys and m slots) is resized, only n/m keys need to be remapped



Architecture – Virtual Warehouse

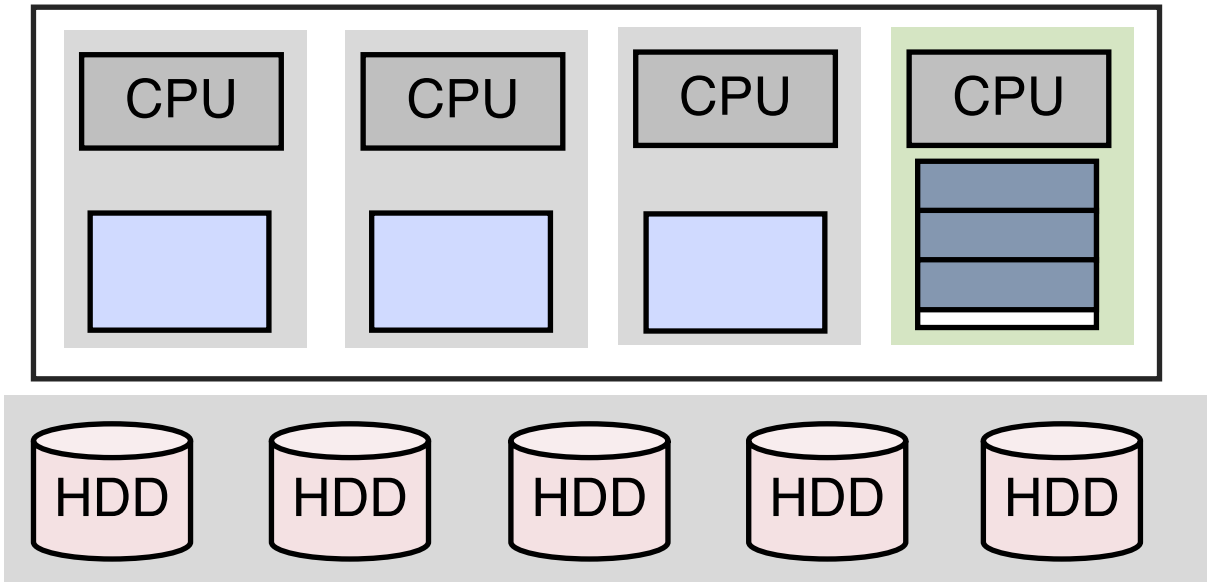
Local caching

- S3 data can be cached in local memory or disk



Consistent hashing

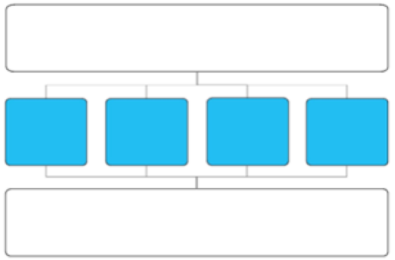
- When the hash table (n keys and m slots) is resized, only n/m keys need to be remapped



Architecture – Virtual Warehouse

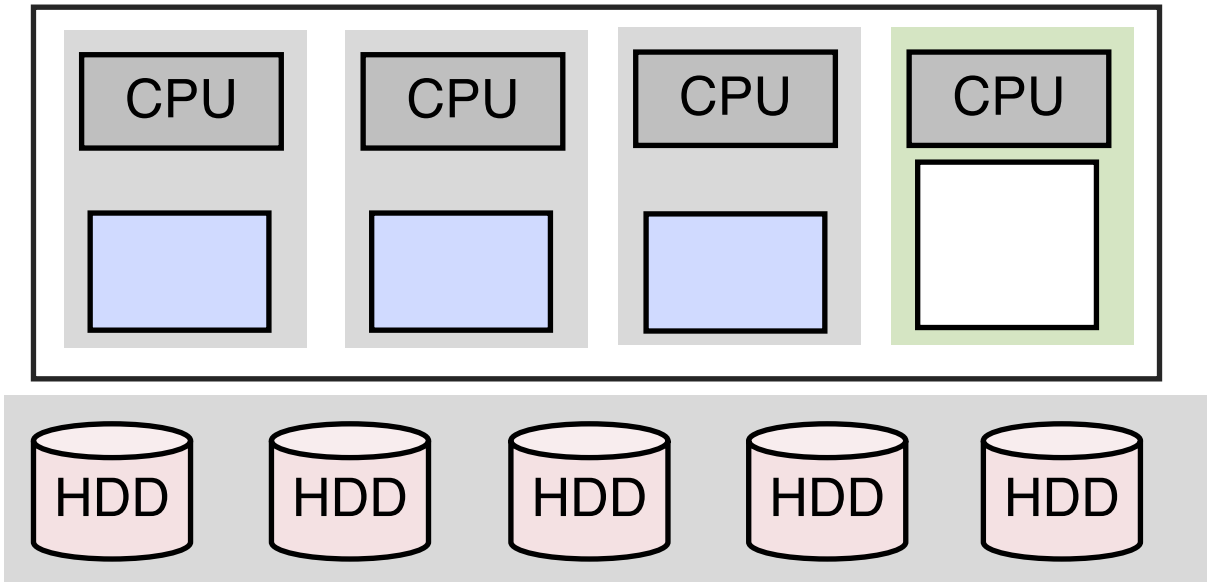
Local caching

- S3 data can be cached in local memory or disk



Consistent hashing

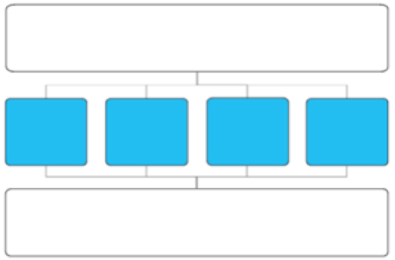
- When the hash table (n keys and m slots) is resized, only n/m keys need to be remapped
- When a VW is resized, no data shuffle required; rely on LRU to replace cache content



Architecture – Virtual Warehouse

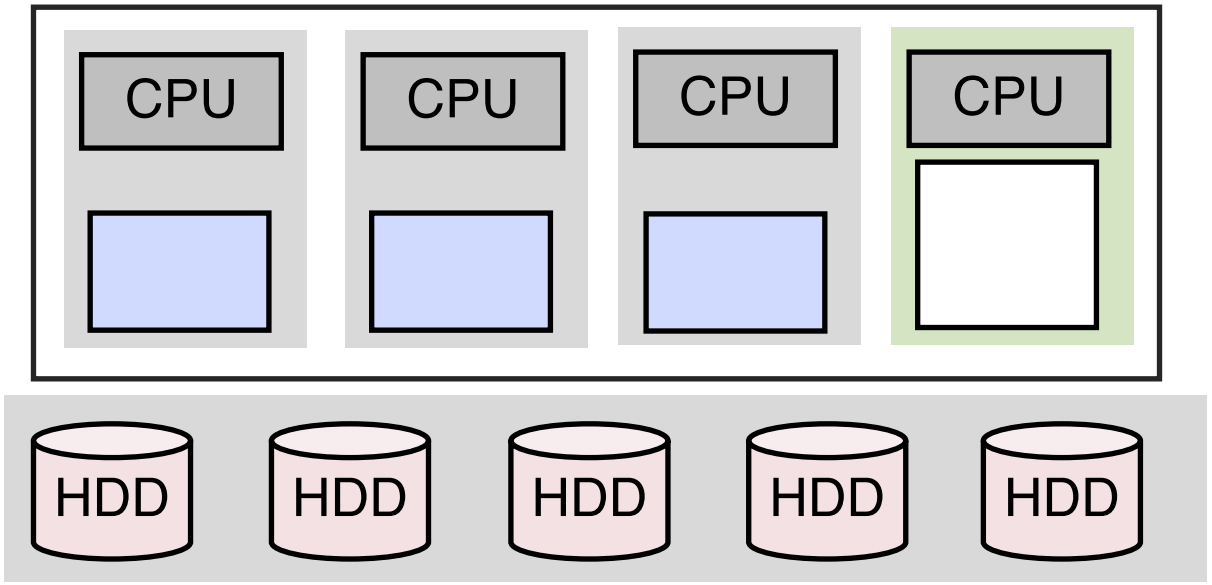
Local caching

- S3 data can be cached in local memory or disk



Consistent hashing

- When the hash table (n keys and m slots) is resized, only n/m keys need to be remapped
- When a VW is resized, no data shuffle required; rely on LRU to replace cache content

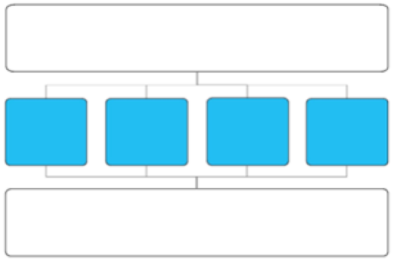


File stealing to tolerate skew

Architecture – Virtual Warehouse

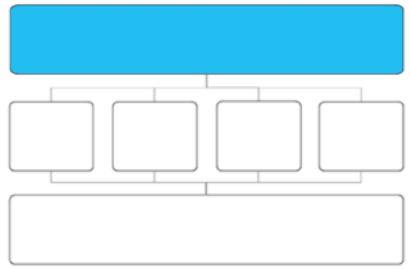
Execution engine

- Columnar: SIMD, compression
- Vectorized: process a group of elements at a time
- Push-based



Architecture – Cloud Services

Multi-tenant layer shared across multiple users



Query optimization

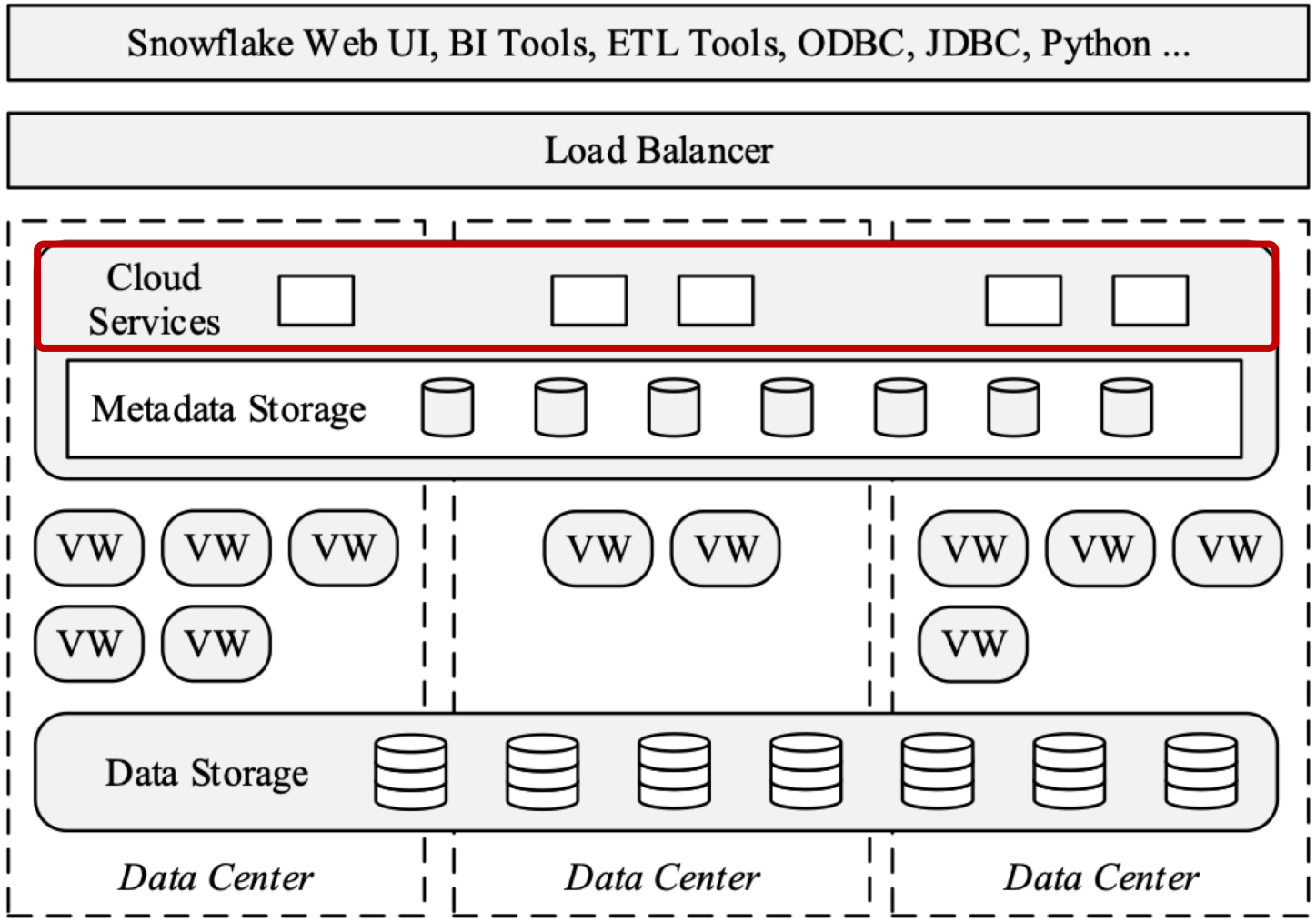
Concurrency control

- Isolation: snapshot isolation (SI)
- S3 data is immutable, update entire files with MVCC
- Versioned snapshots used for time traveling

Pruning

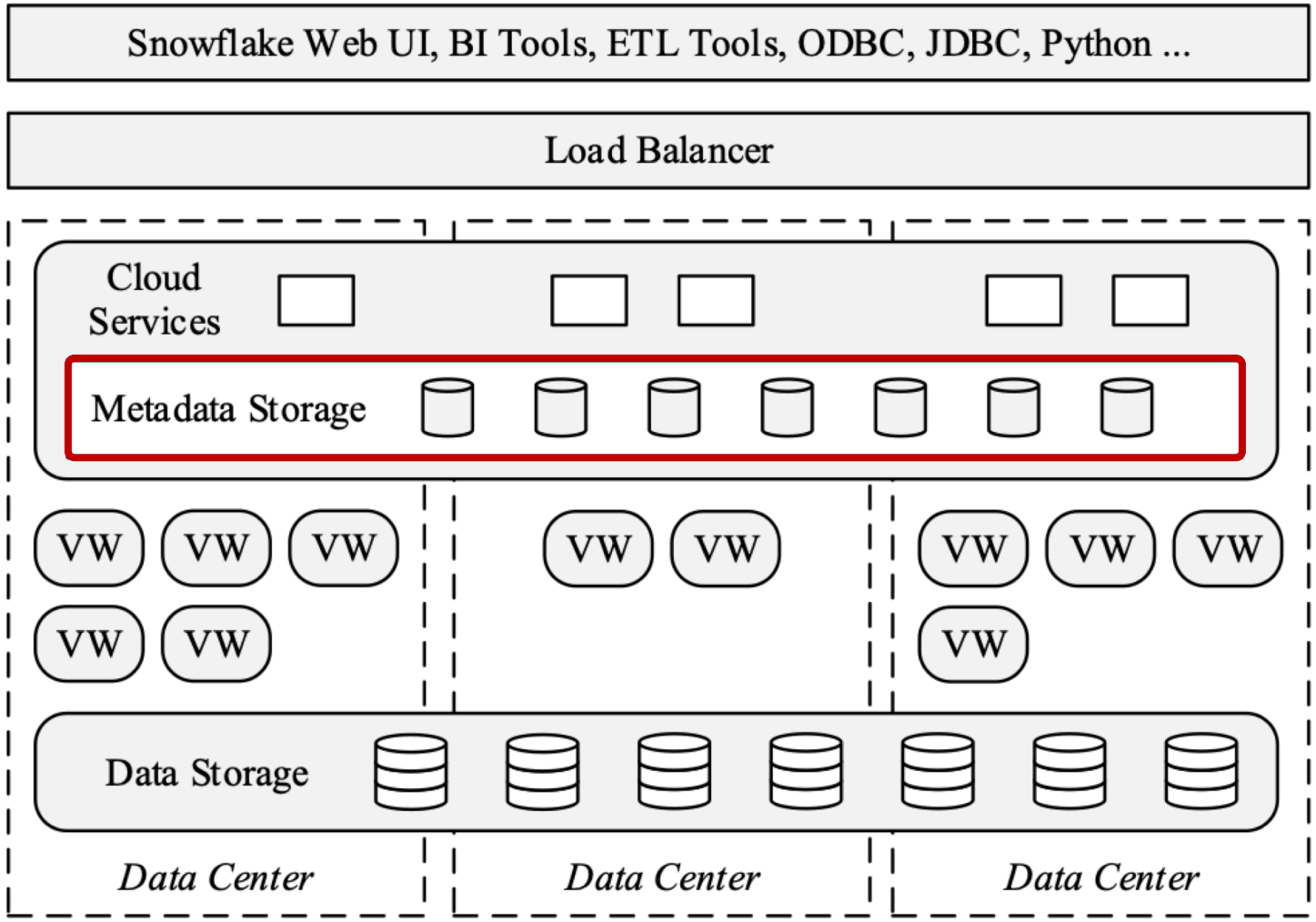
- Snowflake has no index (same as some other data warehousing systems)
- Min-max based pruning: store min and max values for a data block

High Availability and Fault Tolerance



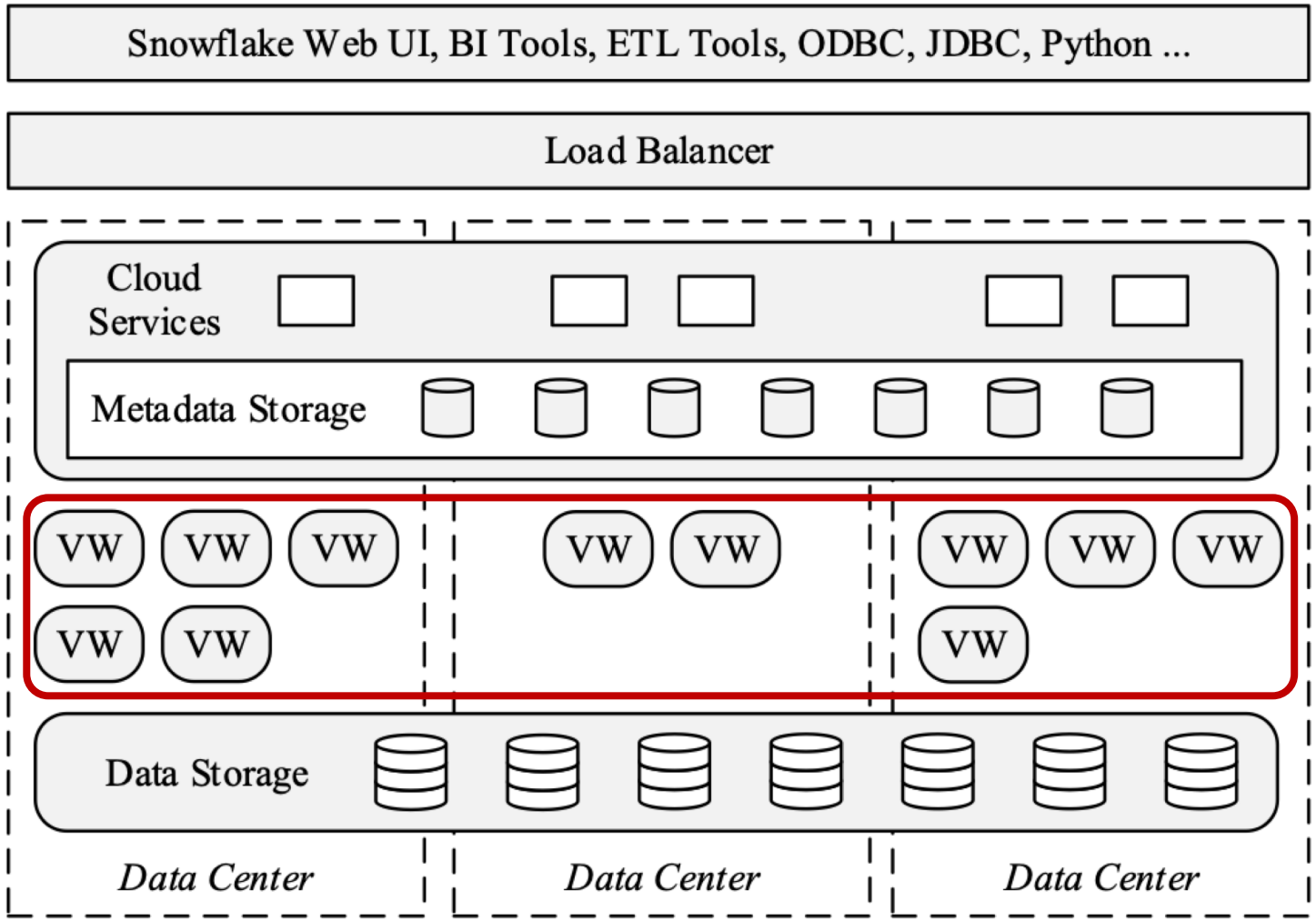
Stateless services

High Availability and Fault Tolerance



Replicated metadata

High Availability and Fault Tolerance



One node failure in VW

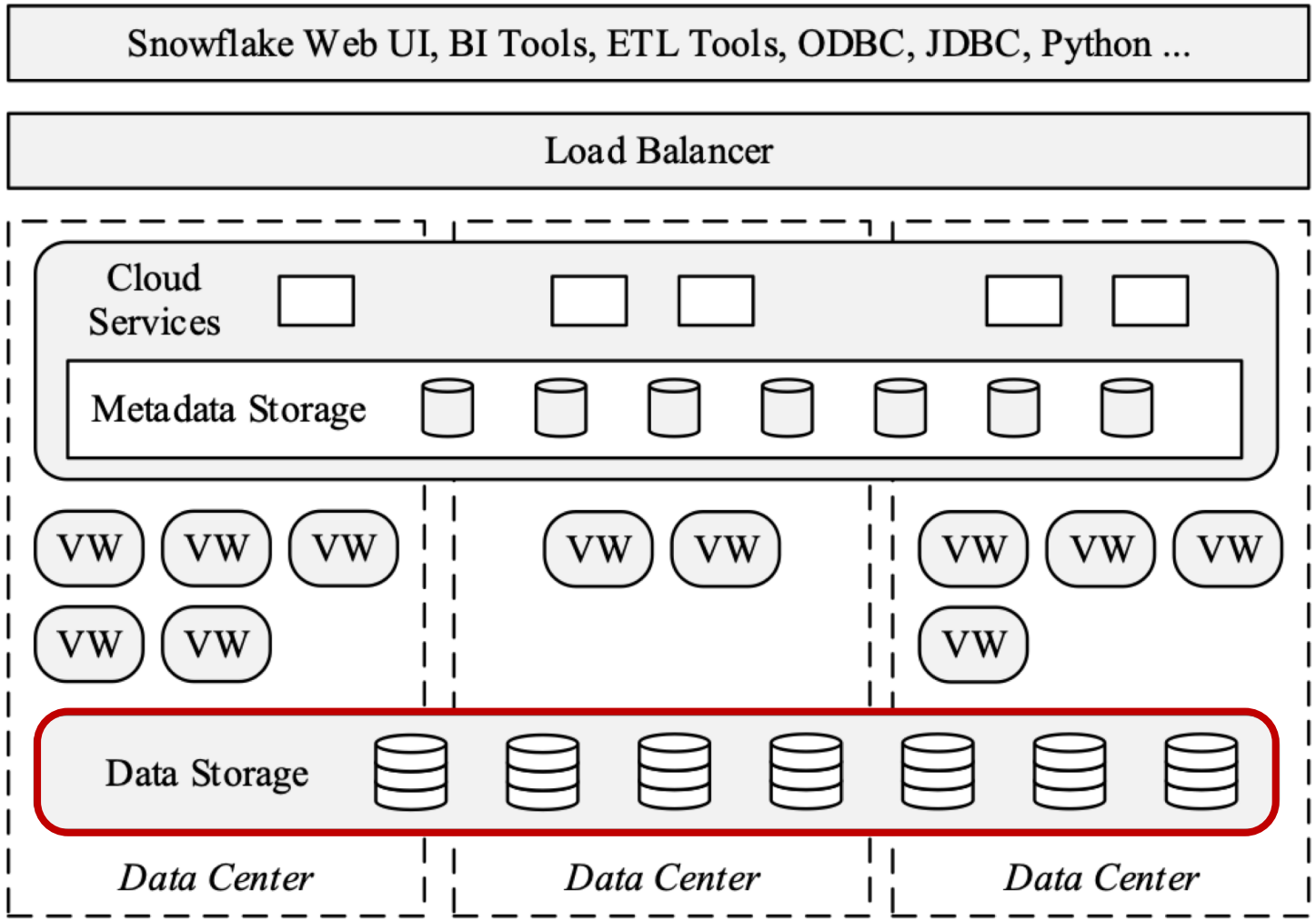
- Re-execute with failed node immediately replaced
- Re-execute with reduced number of nodes

Whole AZ failure

- Re-execute by re-provisioning a new VW

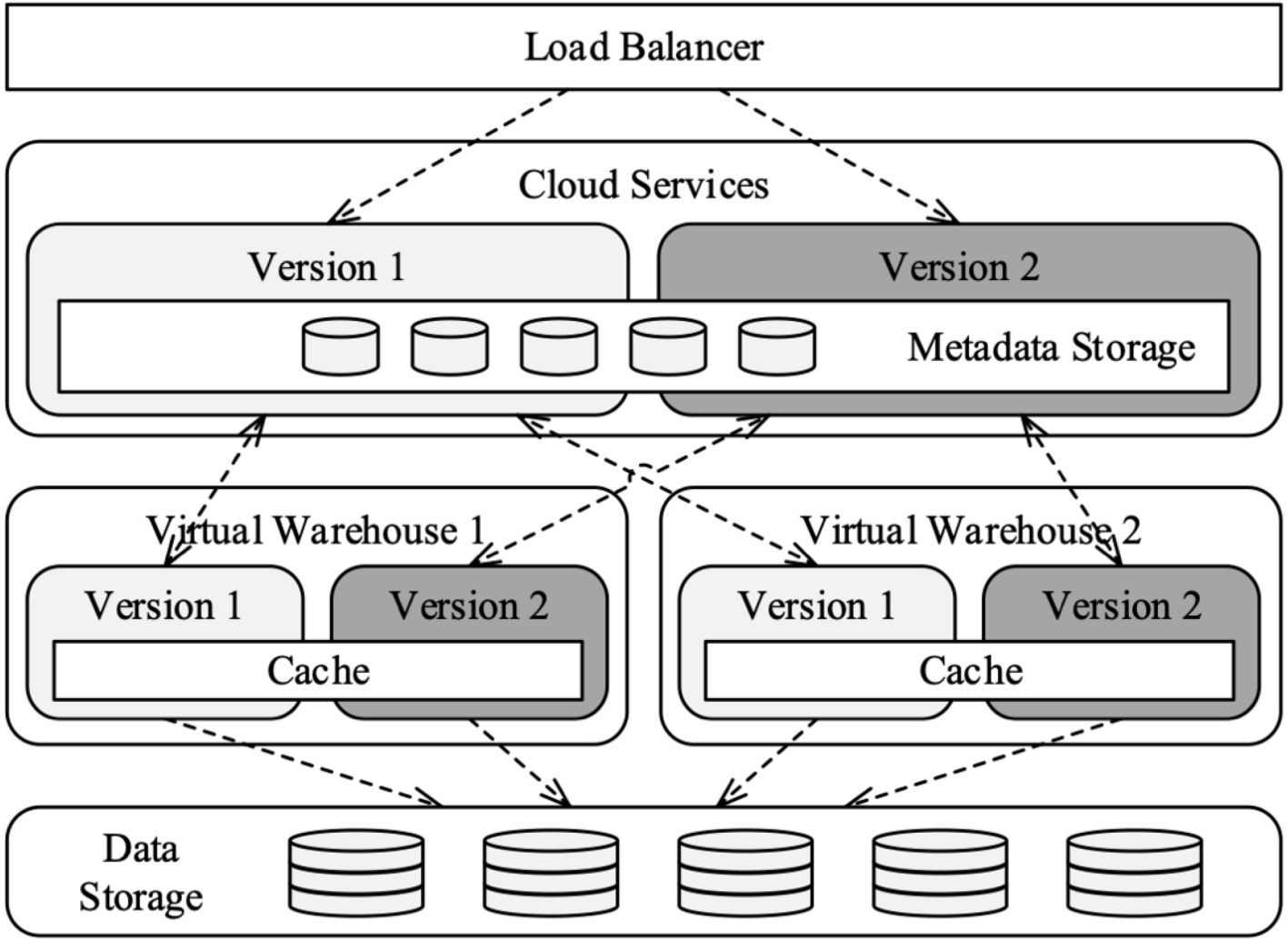
Hot-standby nodes

High Availability and Fault Tolerance



S3 is highly available and durable

Online Upgrade



Deploy new versions of services and VWs

Semi-Structured Data

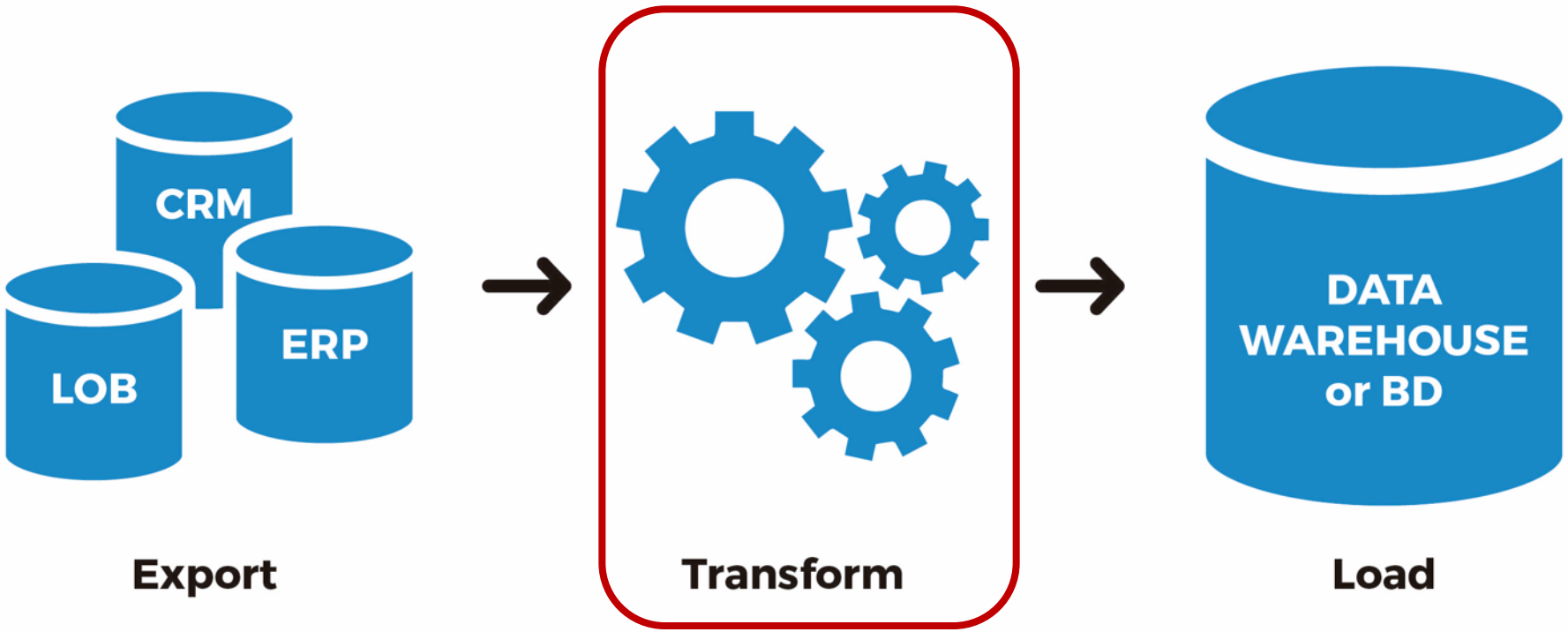
Extensible Markup Language (**XML**)

```xml
<?xml version="1.0" encoding="UTF-8"?>
<customers>
    <customer>
        <customer_id>1</customer_id>
        <first_name>John</first_name>
        <last_name>Doe</last_name>
        <email>john.doe@example.com</email>
    </customer>
    <customer>
        <customer_id>2</customer_id>
        <first_name>Sam</first_name>
        <last_name>Smith</last_name>
        <email>sam.smith@example.com</email>
    </customer>
    <customer>
        <customer_id>3</customer_id>
        <first_name>Jane</first_name>
        <last_name>Doe</last_name>
        <email>jane.doe@example.com</email>
    </customer>
</customers>
```

JavaScript Object Notation(**JSON**)

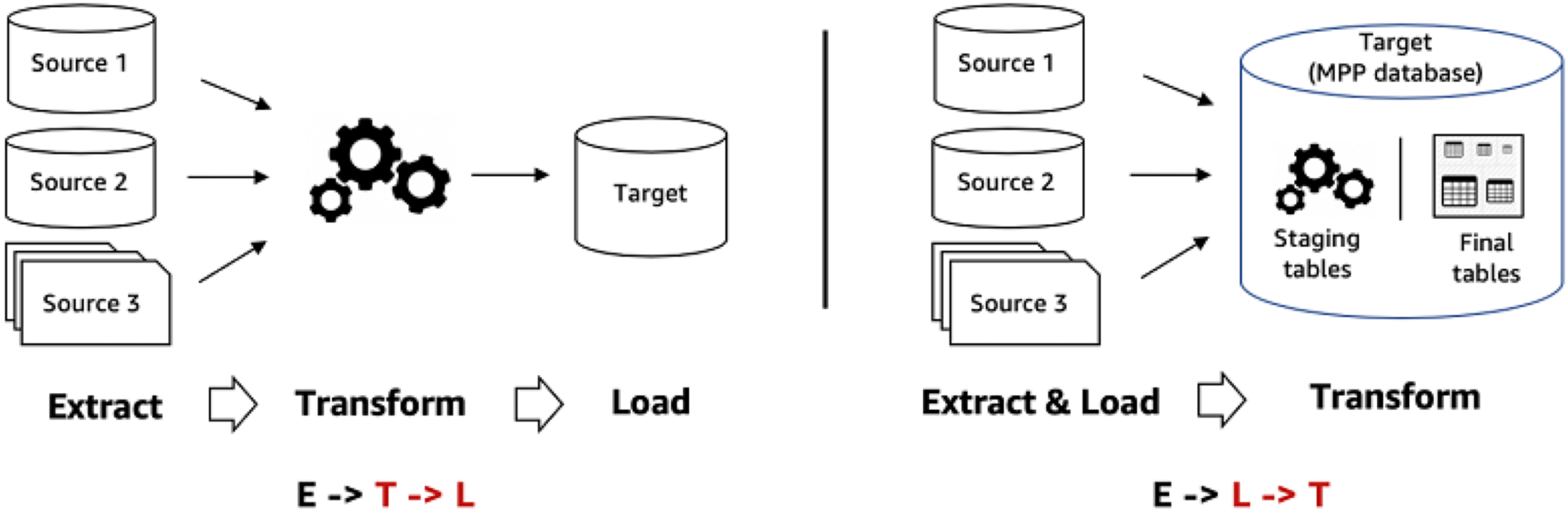
```json
{
    "orders": [
        {
            "orderno": "748745375",
            "date": "June 30, 2088 1:54:23 AM",
            "trackingno": "TN0039291",
            "custid": "11045",
            "customer": [
                {
                    "custid": "11045",
                    "fname": "Sue",
                    "lname": "Hatfield",
                    "address": "1409 Silver Street",
                    "city": "Ashland",
                    "state": "NE",
                    "zip": "68003"
                }
            ]
        }
    ]
}
```

Extract-Transform-Load (ETL)



Transform (e.g., converting to column format) adds latency to the system

ETL vs. ELT



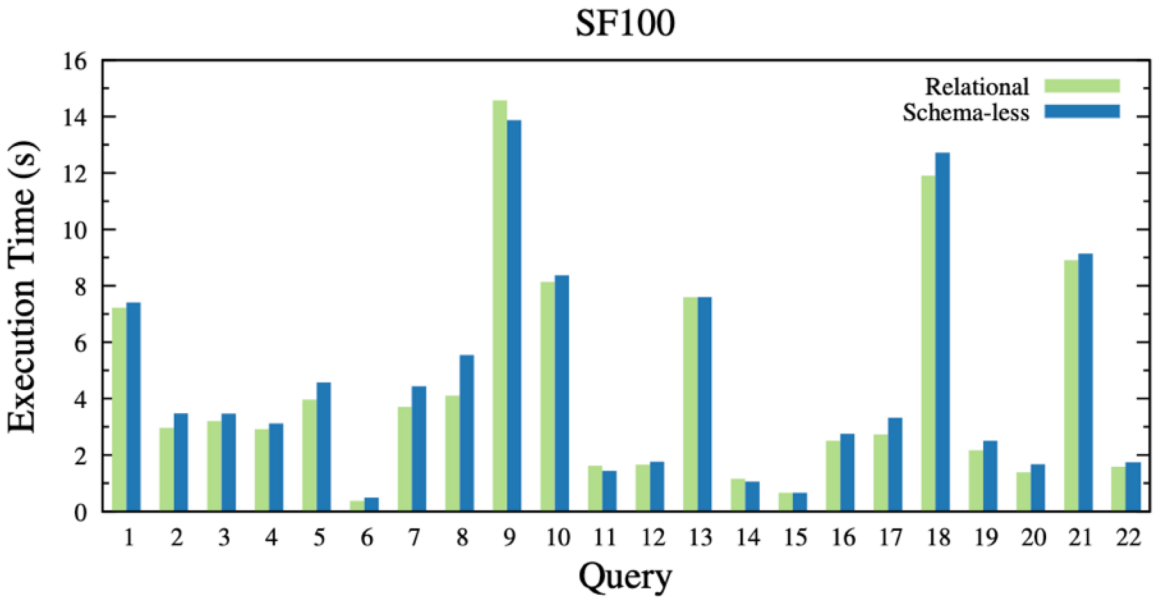
Picture from https://aws.amazon.com/blogs/big-data/etl-and-elt-design-patterns-for-lake-house-architecture-using-amazon-redshift-part-1/

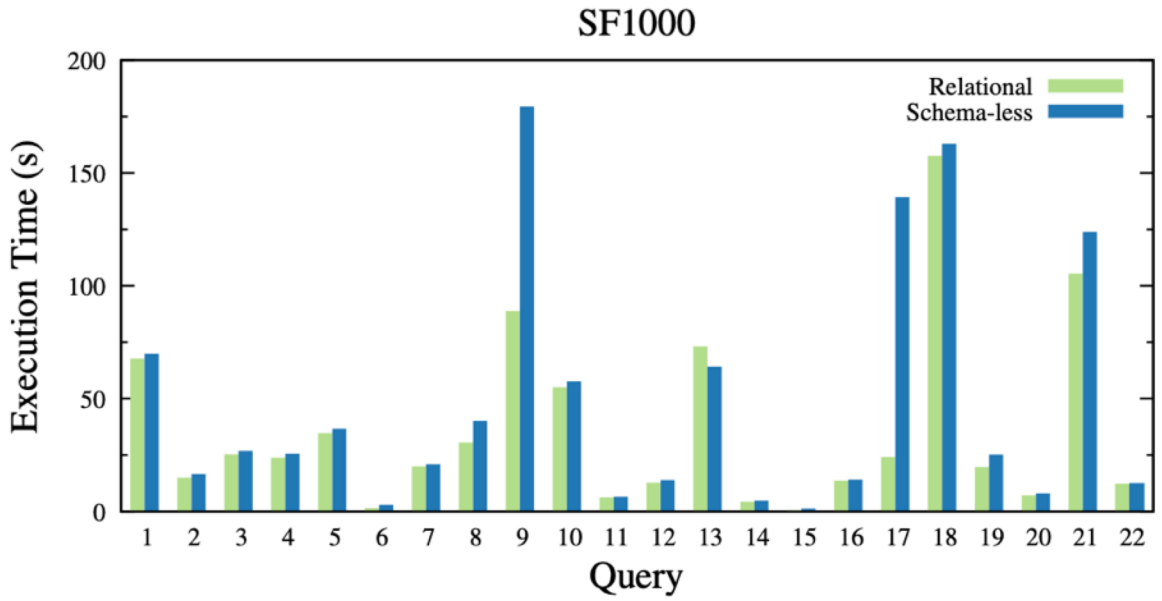
Optimization for Semi-Structured Data

Automatic type inference

Hybrid columnar format

- Frequently paths are detected, projected out, and stored in separate columns in table file (typed and compressed)
- Collect metadata on these columns for optimization (e.g., pruning)





Other Cloud Data Warehousing Systems [1]

Database-as-a-Service (DBaaS)

- Redshift
- Redshift Spectrum
- Athena

Query engines

- Presto
- Apache Hive

Cloud agnostic OLTP DBMS

- Vertica

[1] Tan, Junjay, et al. "Choosing a cloud DBMS: architectures and tradeoffs." VLDB 2019

Summary

Storage-disaggregation vs. shared-nothing

Virtual warehouse

- Control layer
- Compute layer
- Storage layer

ELT for semi-structured data

Snowflake – Q/A

Is security an important aspect of database?

AWS Redshift vs. AWS Aurora vs. Snowflake?

Would S3 supporting update/edit change the design of Snowflake?

Concurrency control other than MVCC in Snowflake?

Updates are inefficient?

Next Lecture

Next lecture is a guest lecture from Snowflake

Submit a review for the guest lecture

- **DDL: Friday 11:59pm**
- Note that you can use one of your three chances to skip this review

Before next Monday (Nov. 16), submit review for

- Xiangyao Yu, et al. PushdownDB: Accelerating a DBMS using S3 Computation, ICDE 2020