

What is Snowflake?

INTRODUCTION TO SNOWFLAKE

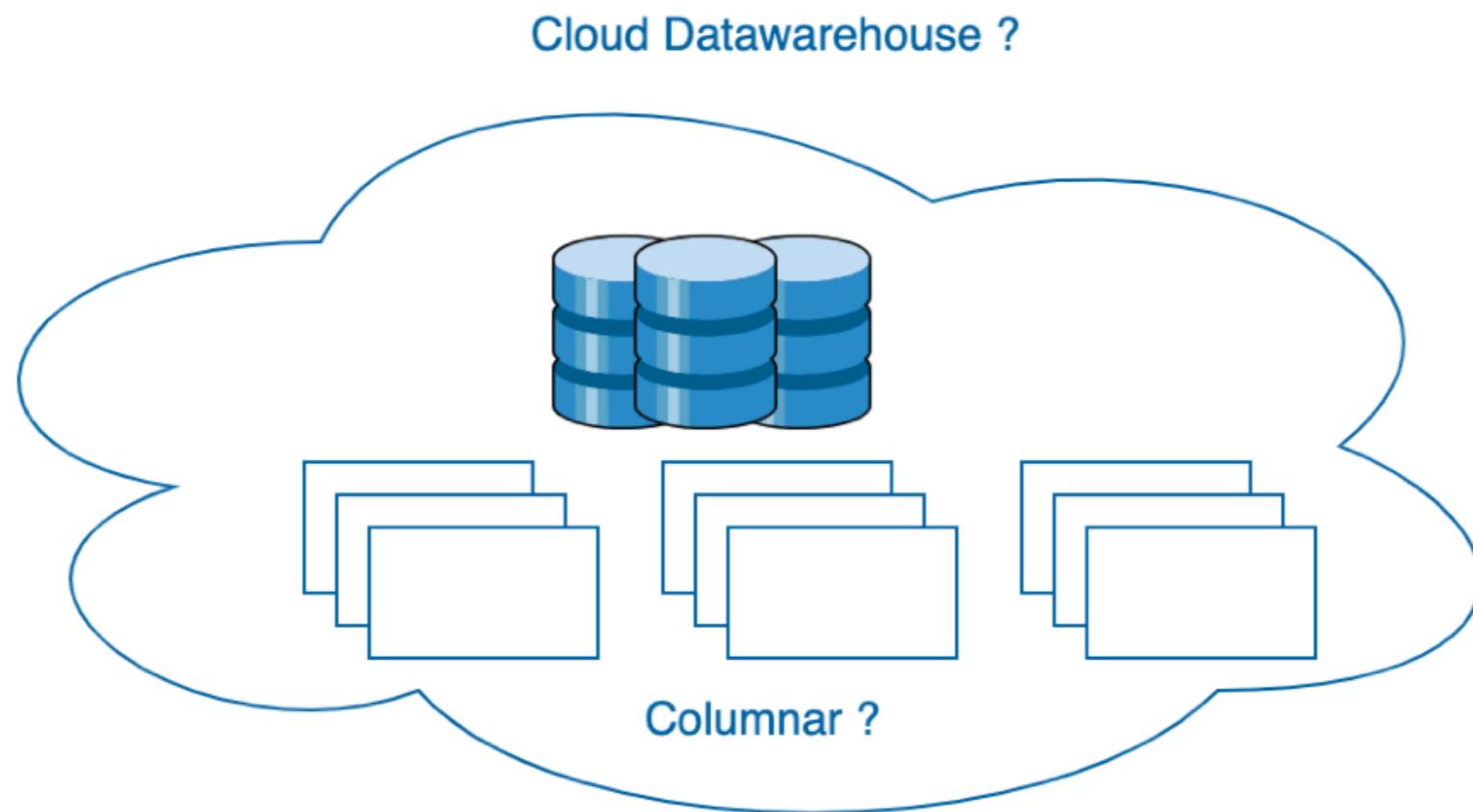


Palak Raina

Senior Data Engineer

What is Snowflake?

- Cloud data warehouse solution
- Columnar data storage



Cloud data warehouse

Traditional data warehouse



Cloud data warehouse



Advantages of Cloud data warehouses

Scalability 

Accessibility 

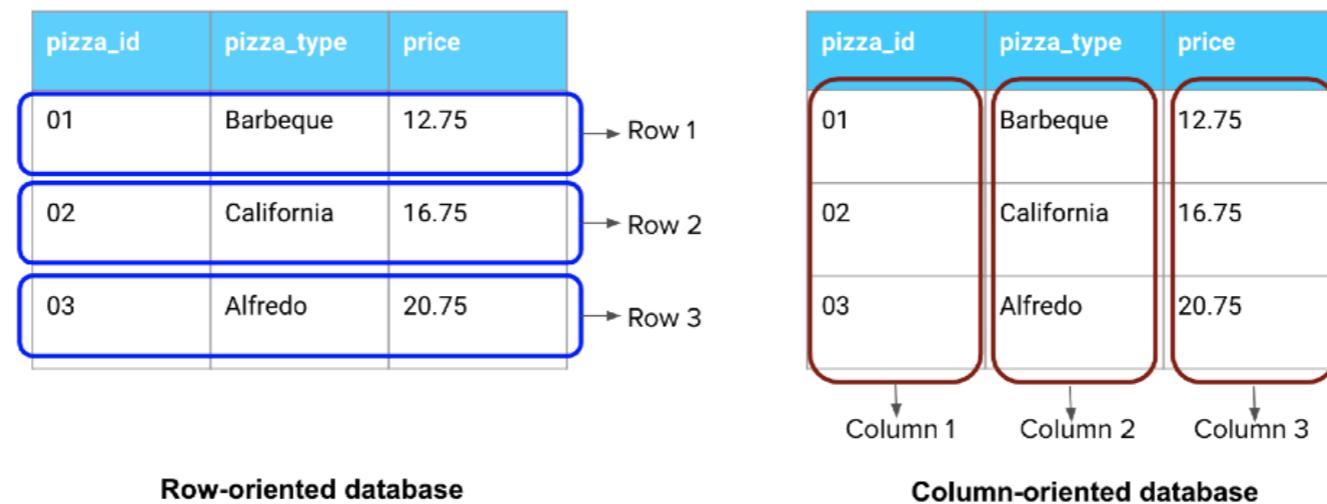
Cost-efficiency 

Lower management effort 

Row vs. Columnar database

pizza_id	pizza_type	price
01	Barbeque	12.75
02	California	16.75
03	Alfredo	20.75

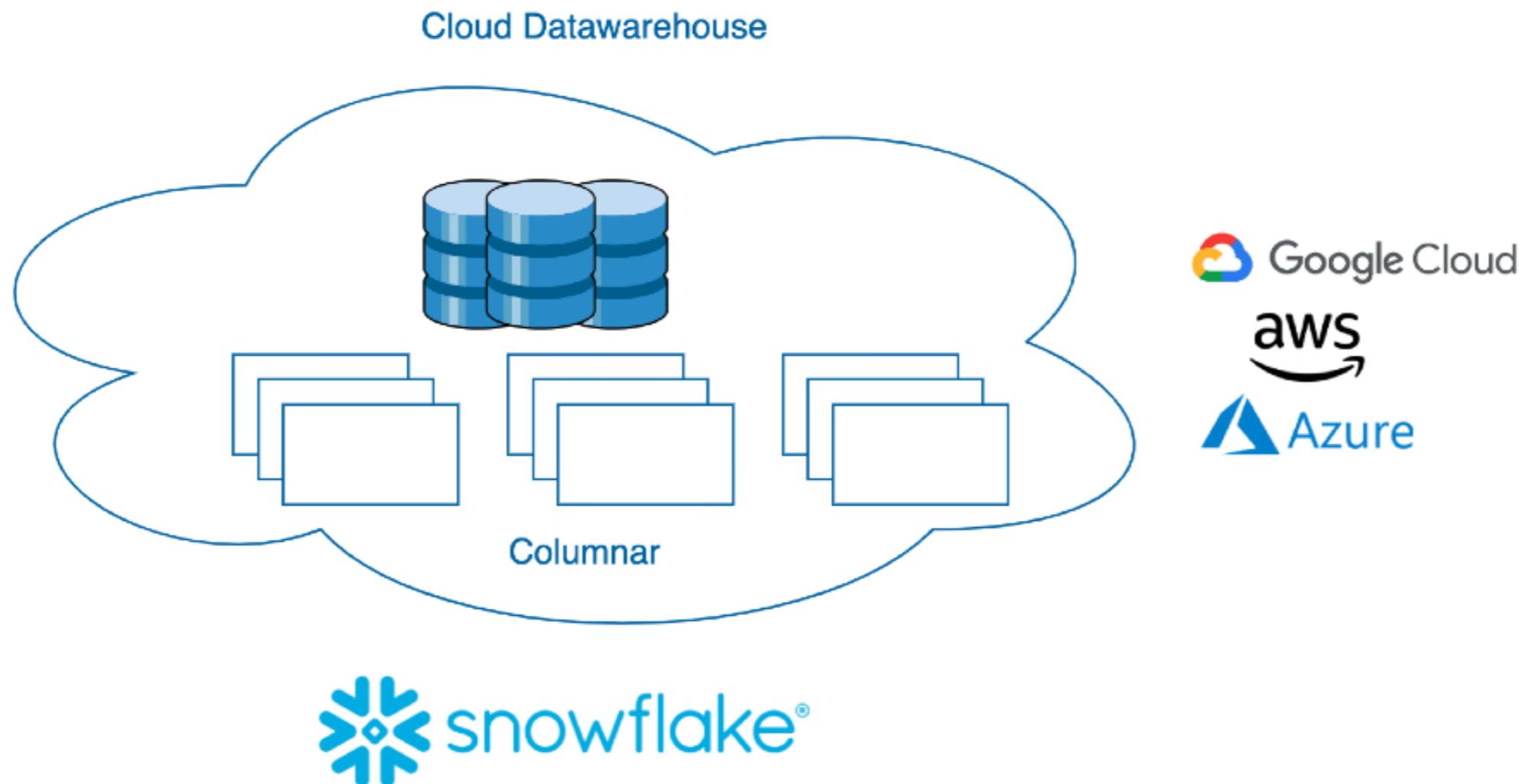
Example: Pizza details table



Row vs. Columnar database

Category	Row	Columnar
Data Organization	By Rows	By Columns
Data Retrieval	Complete records	Relevant columns
Operations	Transactional	Analytical
Example	Postgres, MySQL, Oracle, Microsoft SQL Server	Snowflake, Amazon Redshift, Google BigQuery, Vertica

Snowflake



Snowflake use cases

 Business Intelligence

 Data Science

 Data Ingestion

 Data warehousing

 Data Sharing

Let's practice!

INTRODUCTION TO SNOWFLAKE

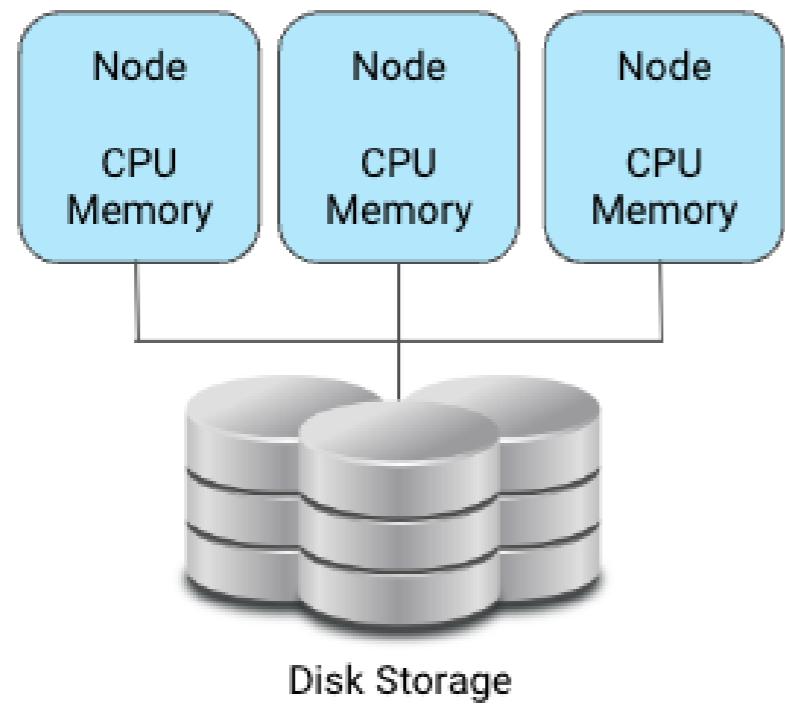
Snowflake Architecture

INTRODUCTION TO SNOWFLAKE

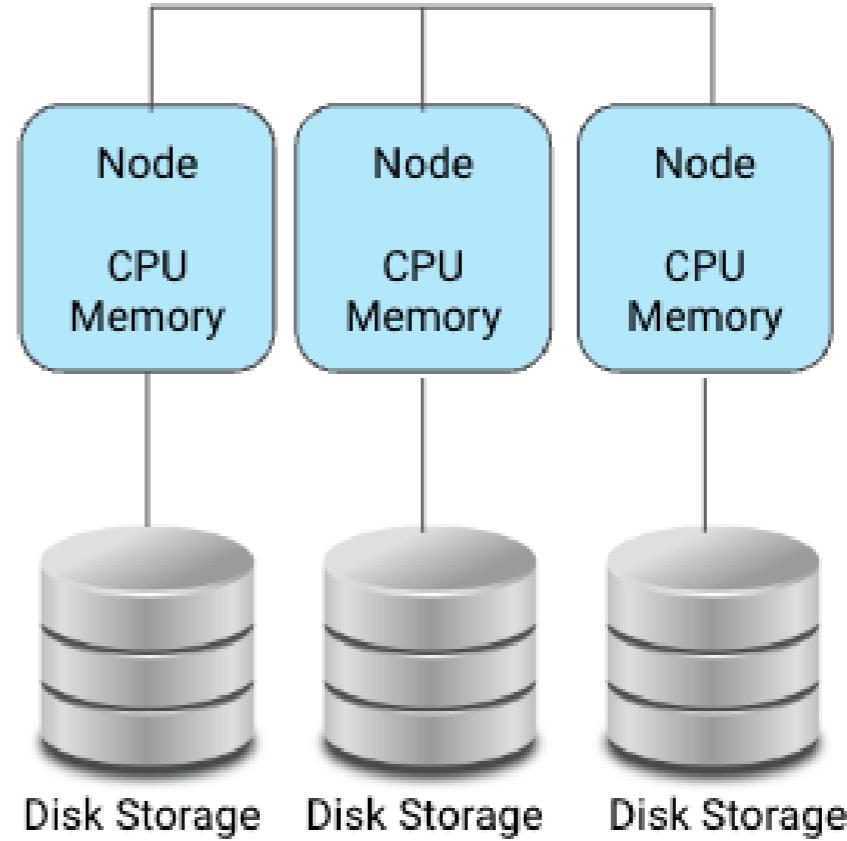


Palak Raina
Senior Data Engineer

Shared-Disk and Shared-Nothing Architecture



Shared-Disk Architecture



Shared-Nothing Architecture

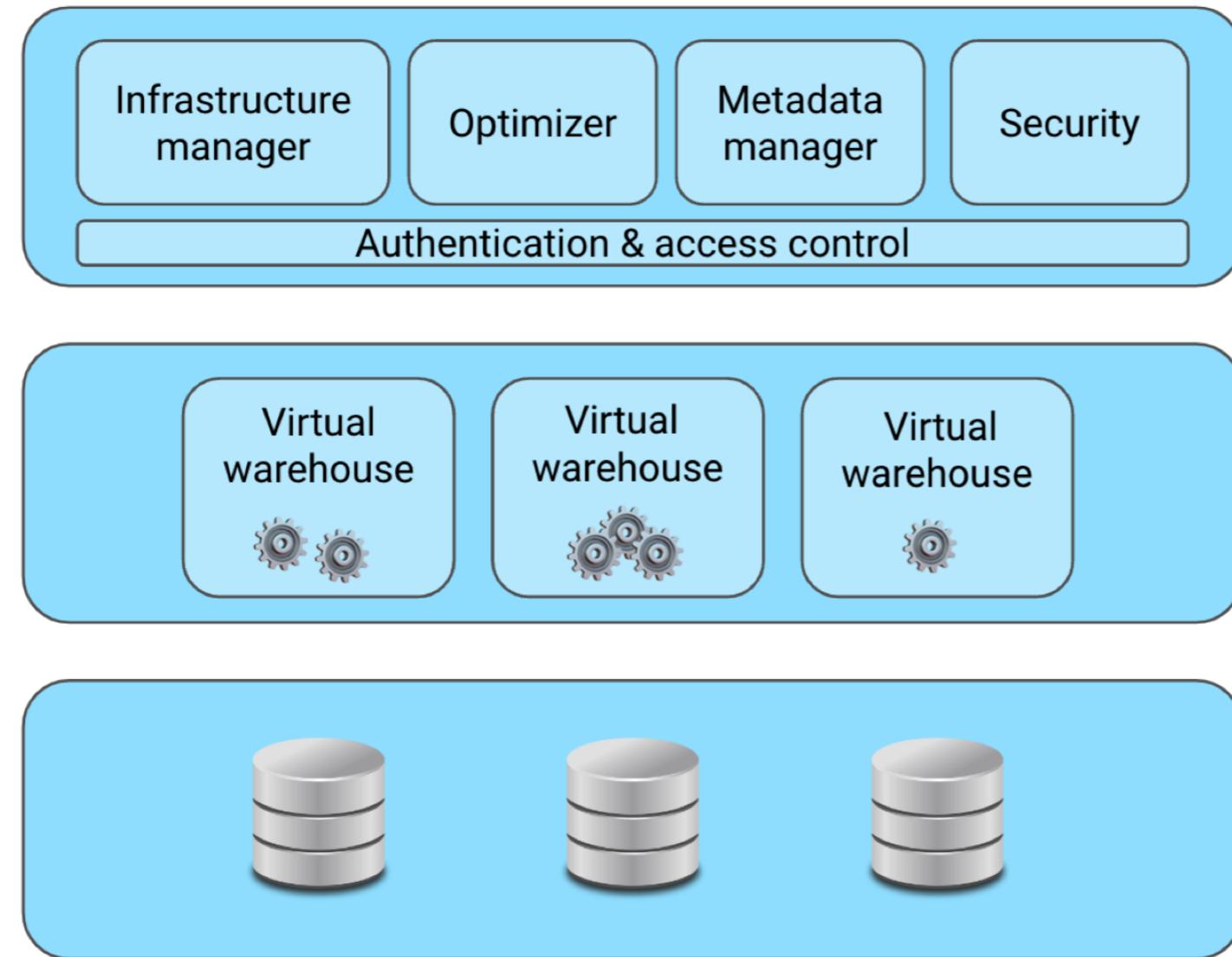
Decoupling Storage & Compute

- Efficient data storage.
- Independent data processing.
- Components operate without interdependence.

Benefits

- Enhanced scalability.
- Faster data processing and response.
- Cost-effective operations.

Snowflake Architecture



Cloud Services Layer



Compute Layer/
Query Processing Layer

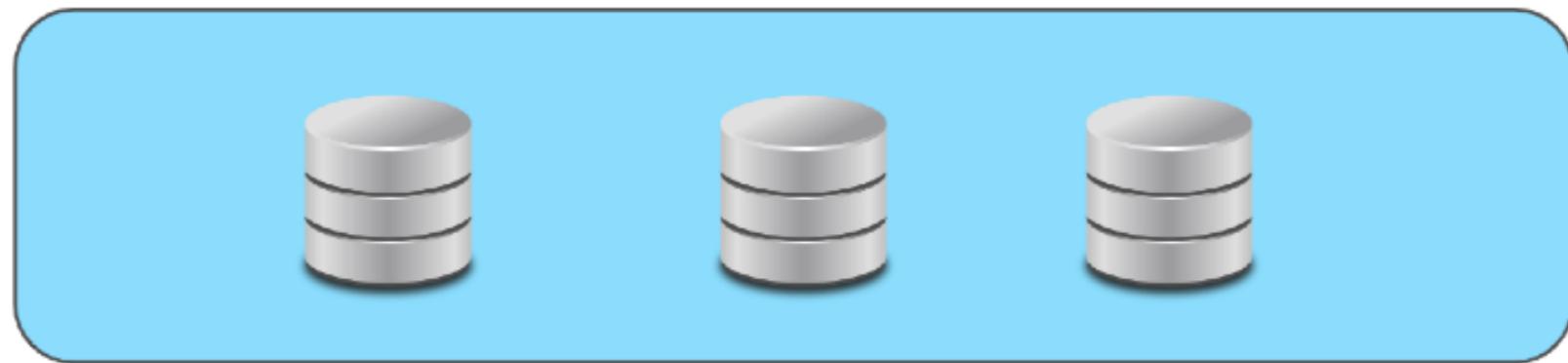


Storage Layer



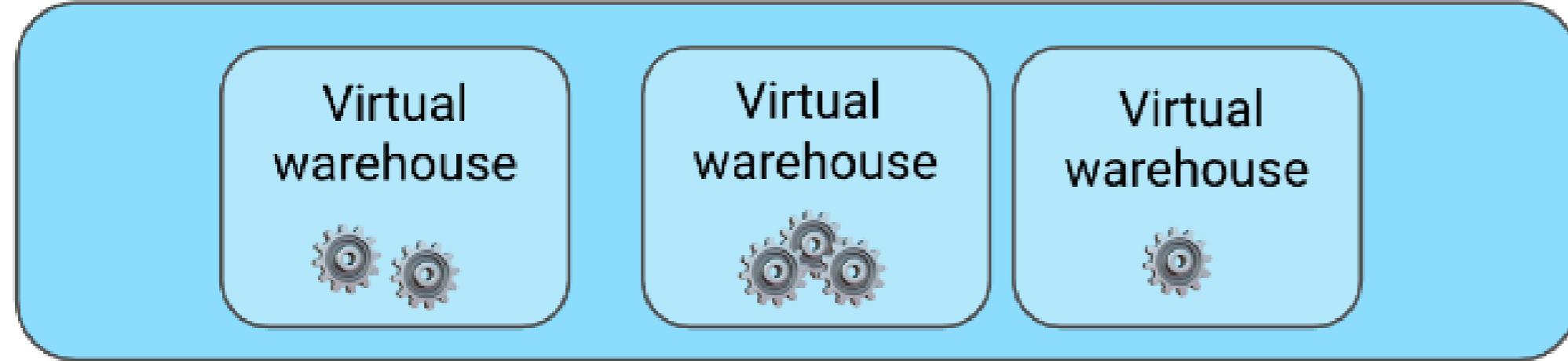
Storage Layer

- Columnar storage
 - Efficient data retrieval
 - Analysis
- Optimized
- Compressed
- Tables, schemas, databases



Compute Layer

- Query execution
- Virtual warehouses



Compute Layer/
Query Processing Layer

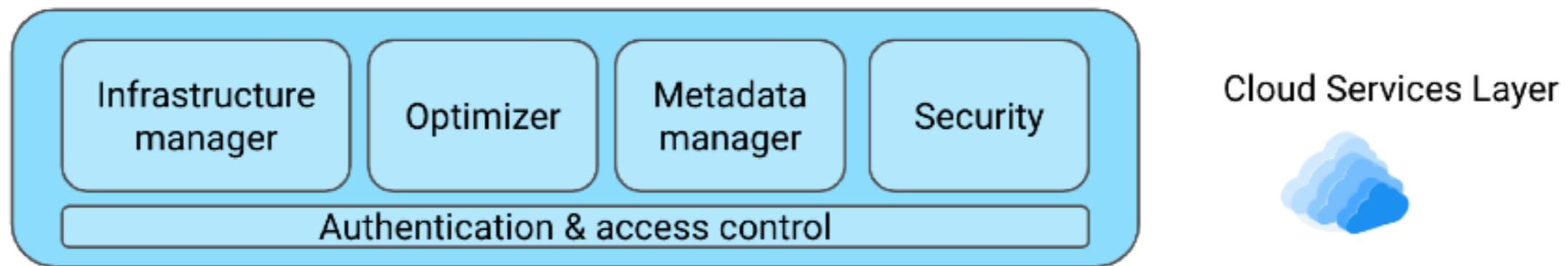


Compute Layer - virtual warehouse

- Scalability
- Performance
- Cost-effectiveness
- Different sizes: XS, S, M, L, XL

Cloud Services Layer

- Infrastructure management
- Query Optimization
- Authentication
- Access control
- Security



Let's practice!

INTRODUCTION TO SNOWFLAKE

Snowflake Competitors and why use Snowflake

INTRODUCTION TO SNOWFLAKE



Palak Raina
Senior Data Engineer

Snowflake Competitors



Google
Big Query



databricks



amazon
REDSHIFT



PostgreSQL

Comparison: Architecture, Scalability & Management

Feature	Snowflake	Google BigQuery	Amazon Redshift	Databricks	Postgres
Architecture	Decoupled storage and compute	Serverless, Decoupled storage and compute	Depends on the cluster you chose	Decoupled storage and compute	Relational database management system (RDBMS): Traditional monolithic architecture
Scalability	Faster autoscaling	Automatic Scaling	Scaling with some delay	Automatic Scaling	Manual
Management	Low			Medium	

Security & Data Support

Security

- Access Controls
- Encryption

Unstructured Data Support - Databricks

- Text, images, and audio

Semi-Structured Data Support

- JSON
- Avro
- Parquet
- CSV

Integration & Pricing

Feature	Snowflake	Google BigQuery	Amazon Redshift	Databricks	Postgres
Integrations with Cloud Providers	Multi-cloud	Multi-cloud through connections	Single Cloud	Multi-cloud	Can be deployed on multi-cloud
Pricing	Compute (per second) and storage billed separately	Compute(per slot) and storage billed separately. Options: On-demand or Capacity pricing	Compute (per hour) and storage billed separately	Compute (per second/per minute) and storage billed separately	Open-source (no license fees), varies with hosting provider

What makes Snowflake Unique?

- Unique architecture - decoupling compute, storage
- Secure data sharing
- High performance
- Multi-Cloud Support
- Pricing

Why use Snowflake?

- Large-scale data sets
- Multi-Cloud Support
- Pricing
- Ease of use

SnowflakeSQL vs. PostgreSQL

- American National Standards Institute(ANSI) SQL

Example:

```
select * from orders
```

Let's practice!

INTRODUCTION TO SNOWFLAKE