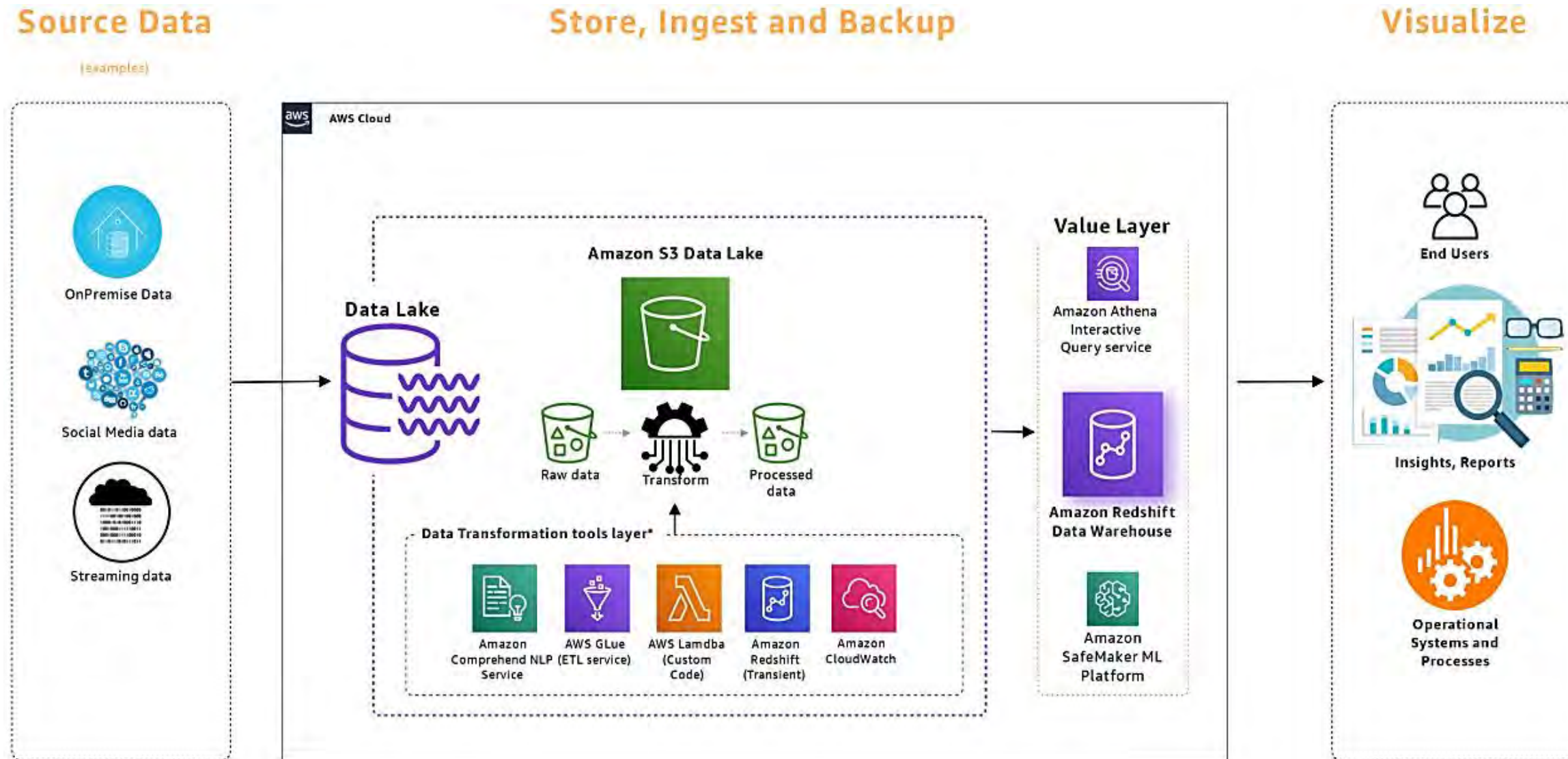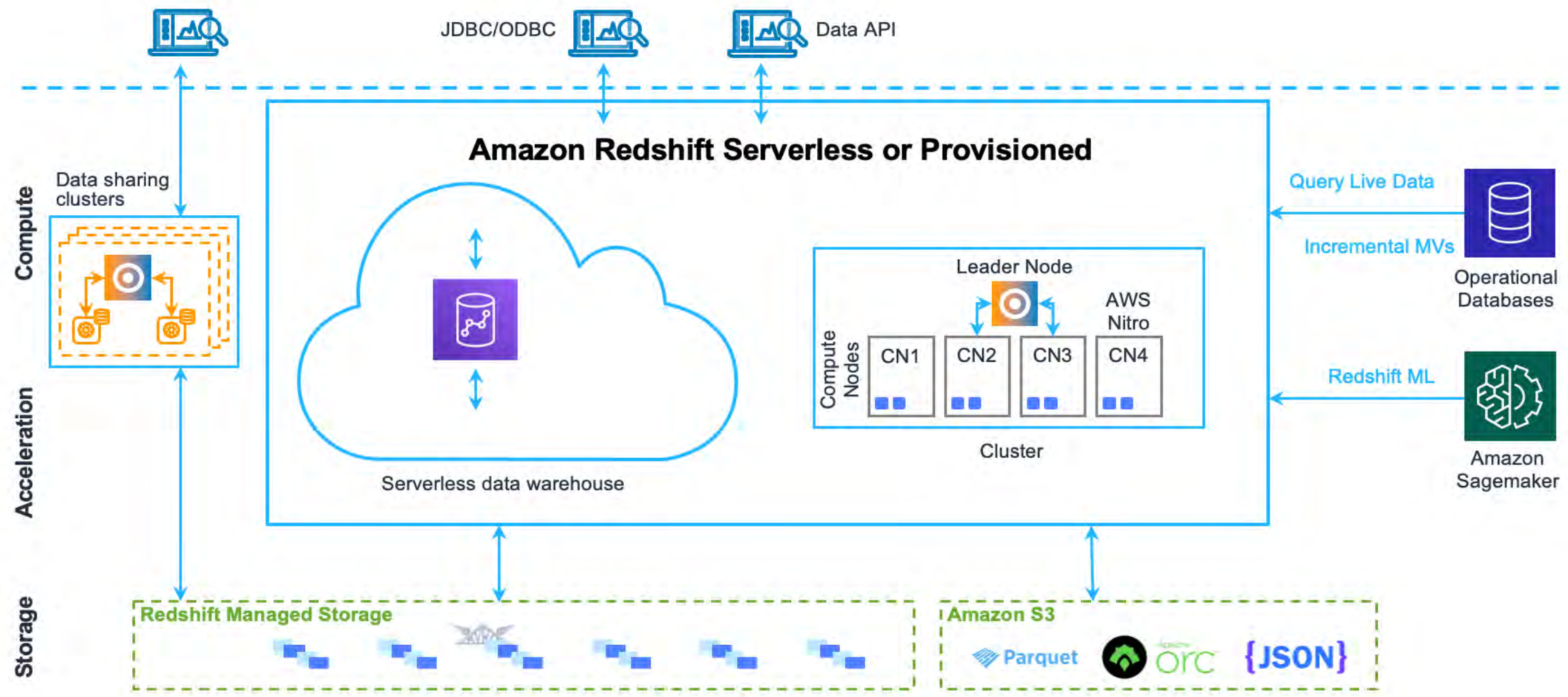# Amazon Redshift

Introduction to Amazon Redshift

Introduction to Amazon Redshift

- Amazon Redshift is a powerful cloud data warehousing service equipped with the latest technology to store big data and analytics.

- It's a fully managed service that runs complex analytic queries on large data sets... and it's super easy to use!

- Amazon Redshift Serverless lets you access and analyze data without the usual configurations of a provisioned data warehouse.

- Resources are automatically provisioned, and data warehouse capacity is intelligently scaled to deliver fast performance for even the most demanding and unpredictable workloads.
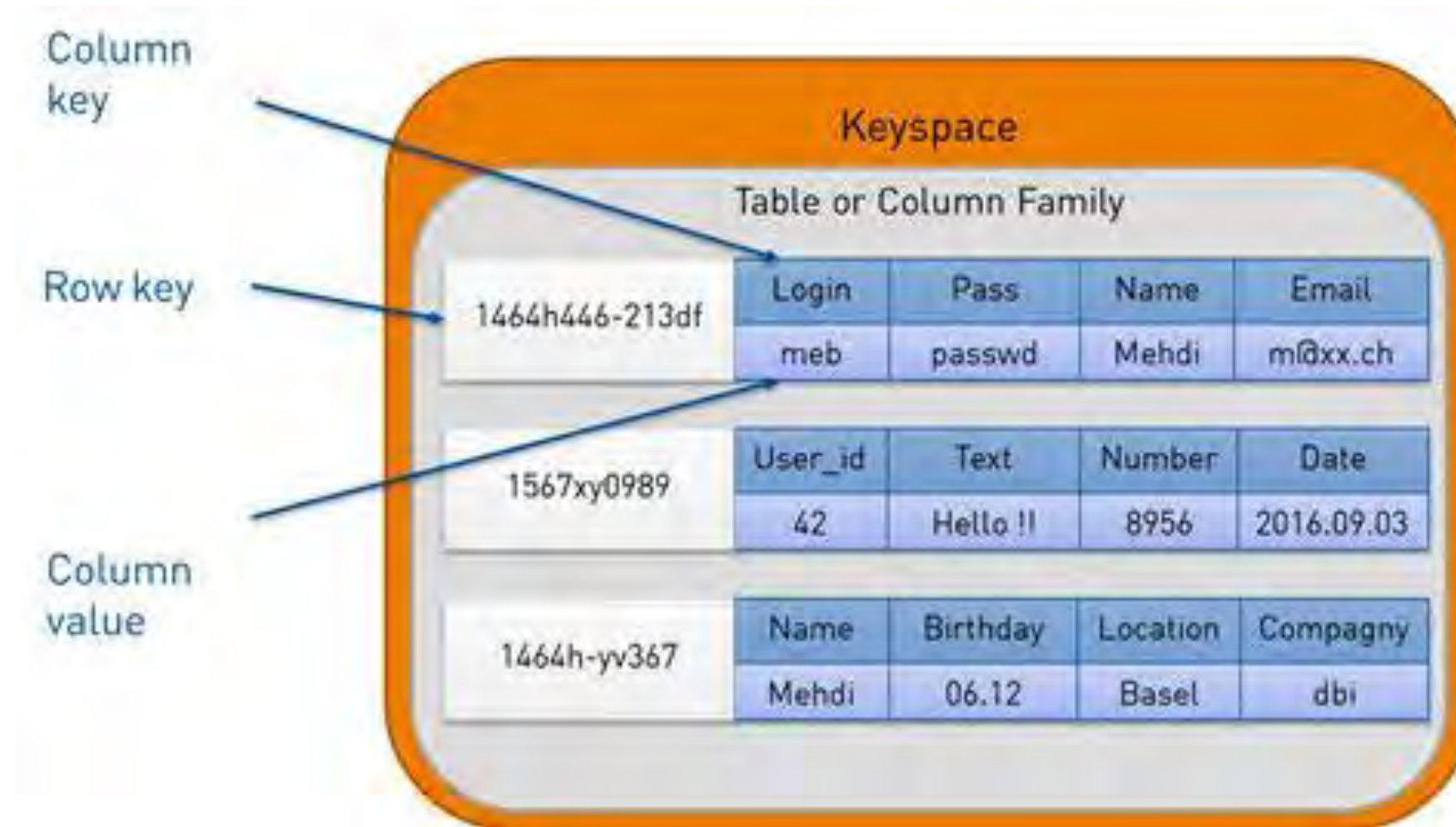
Fully managed data warehouse service

This figure introduces the elements of the Amazon Redshift data warehouse architecture
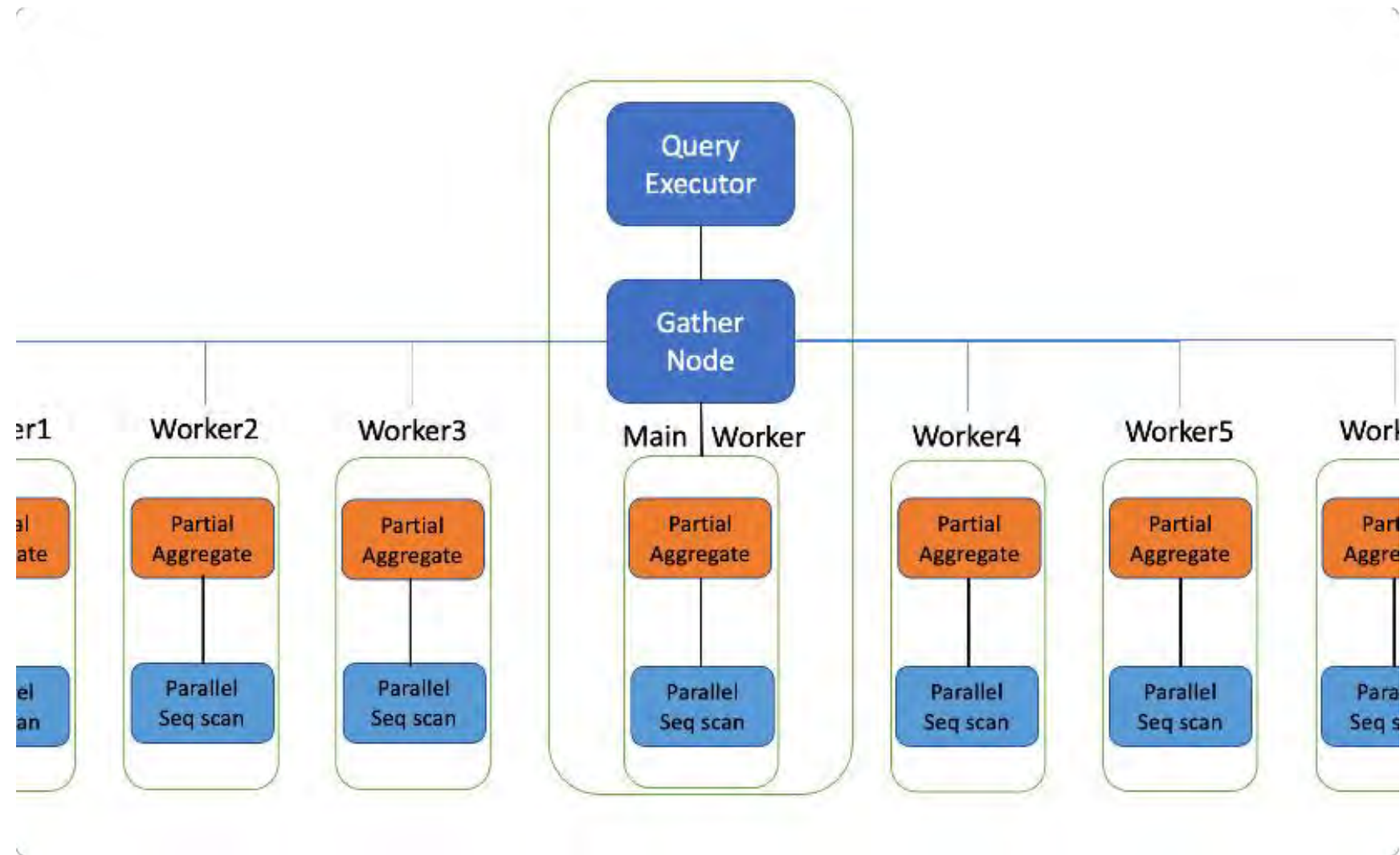
Columnar Storage and Parallel Query Processing

Columnar Storage

Redshift uses a columnar storage technique that provides high-speed performance and efficiency by only selecting the columns it needs to perform the query.
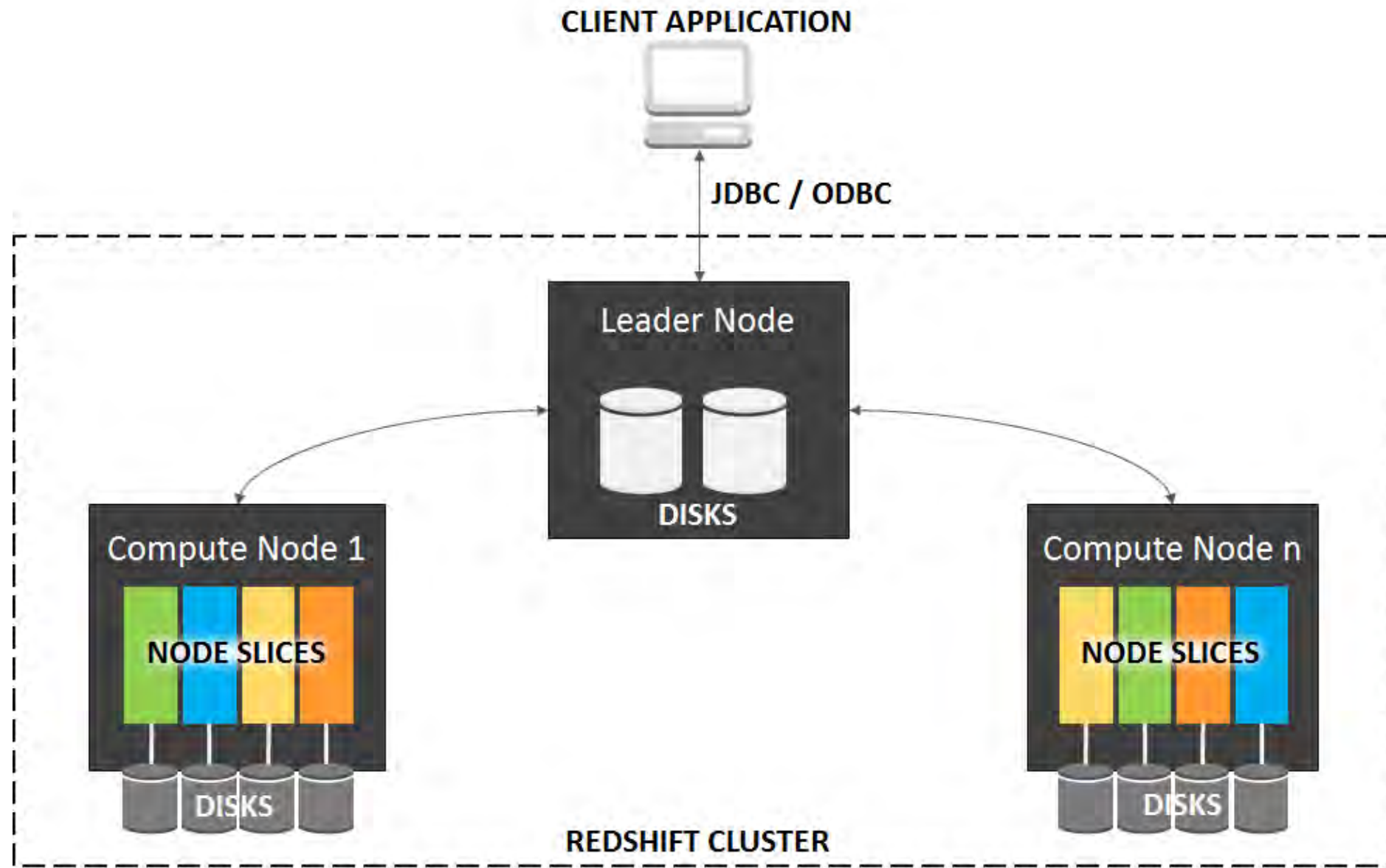
iam**neo**   www.iamneo.ai

Columnar Storage and Parallel Query Processing

Parallel Query Processing

Redshift utilizes parallel processing to execute complex queries involving multiple nodes to quickly provide results in seconds.

iam**neo**   www.iamneo.ai

Creating and Configuring Amazon Redshift Clusters

Creating and Configuring Amazon Redshift Clusters

Some of the essential concepts and terminologies that you to keep in mind when working with Amazon Redshift:

➤ **Clusters**
- **Leader node**
- **Compute node**

➤ **Node slices**

➤ **Databases**

Creating a Cluster

1 **Step-by-Step Guide**

Creating an Amazon Redshift cluster can be done with just a few clicks. By following the AWS Management Console's step-by-step guide, you can have your cluster up and running in no time!
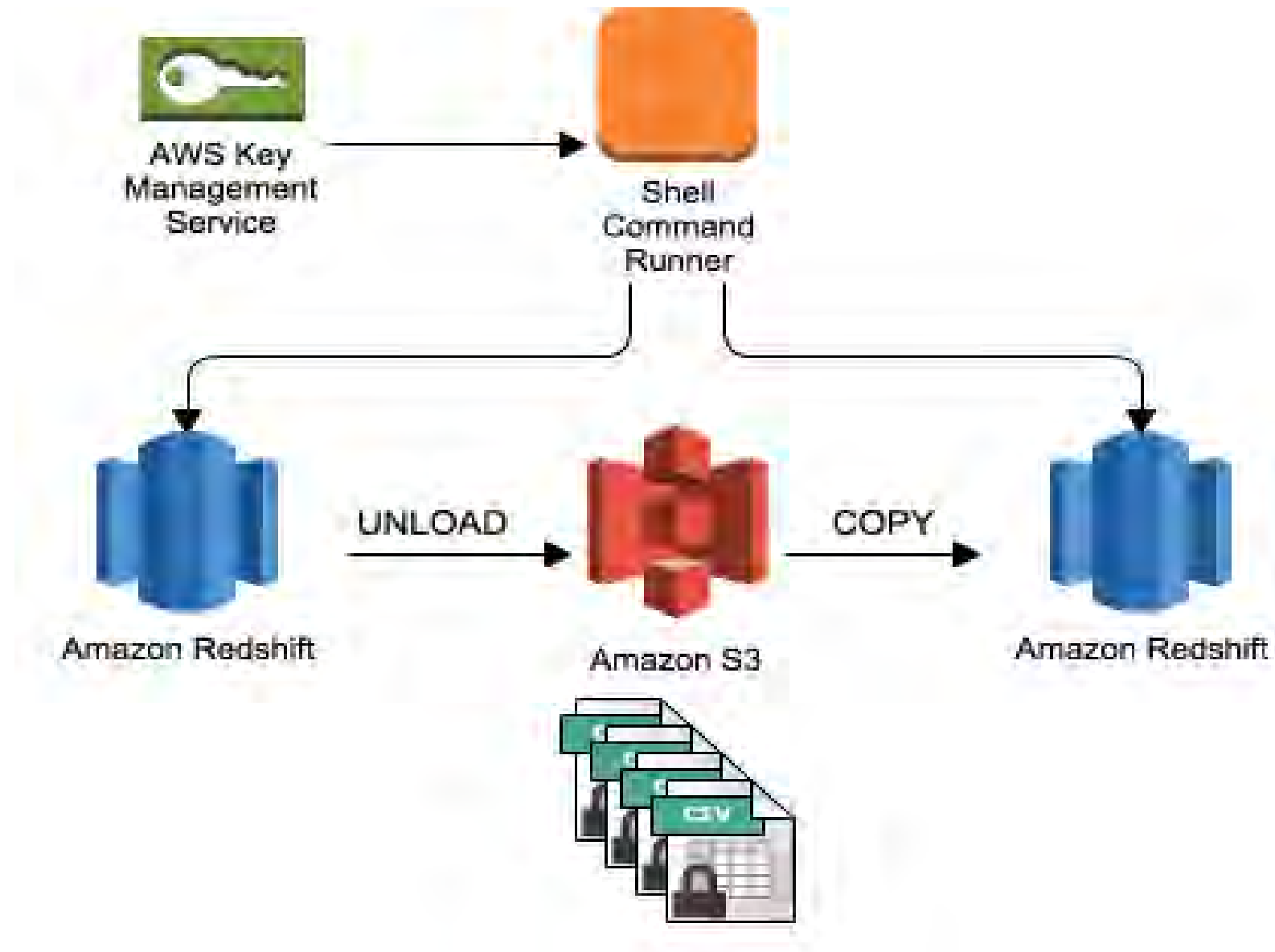
2 **Cli or SDK**

For more complex production environments, the AWS Management Console offers a Command Line Interface (CLI), as well as Software Development Kits (SDKs), for scripting Amazon Redshift service tasks.
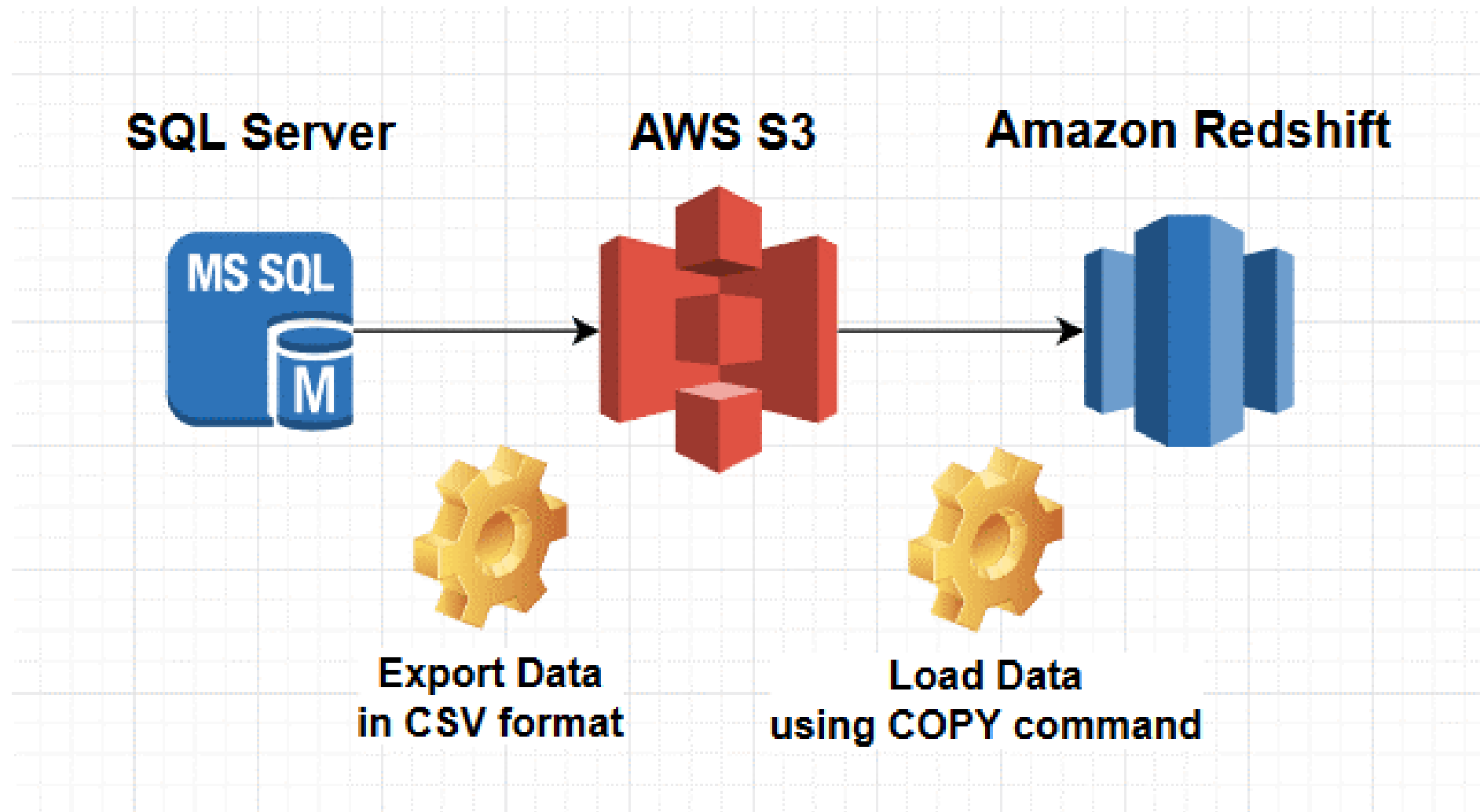
Configuring cluster node types

Node size	vCPU	RAM (GiB)	Managed storage quota per node
ra3.xlplus	4	32	32 TB
ra3.4xlarge	12	96	128 TB
ra3.16xlarge	48	384	128 TB

Configuring cluster Security

Loading data in Redshift using COPY commands

The **Redshift COPY Command** is a very powerful and flexible interface to load data to Redshift from other sources. That said, it does have its share of limitations, specifically when it comes to enforcing data types and handling duplicate rows.
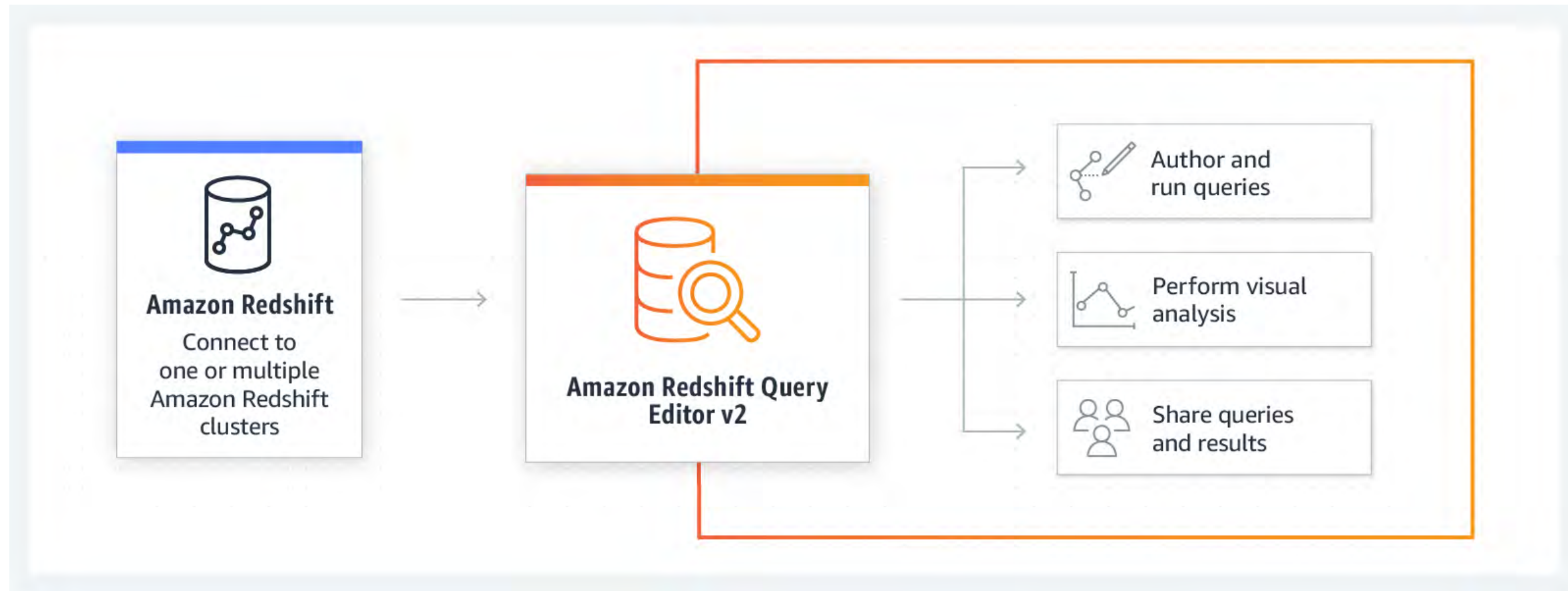
Loading and managing data in Redshift using UNLOAD commands

Methods to Unload Data from Amazon Redshift to S3

- Method 1: Unload Data from Amazon Redshift to S3 using the UNLOAD command

- Method 2: Unload Data from Amazon Redshift to S3 in Amazon Parquet Format
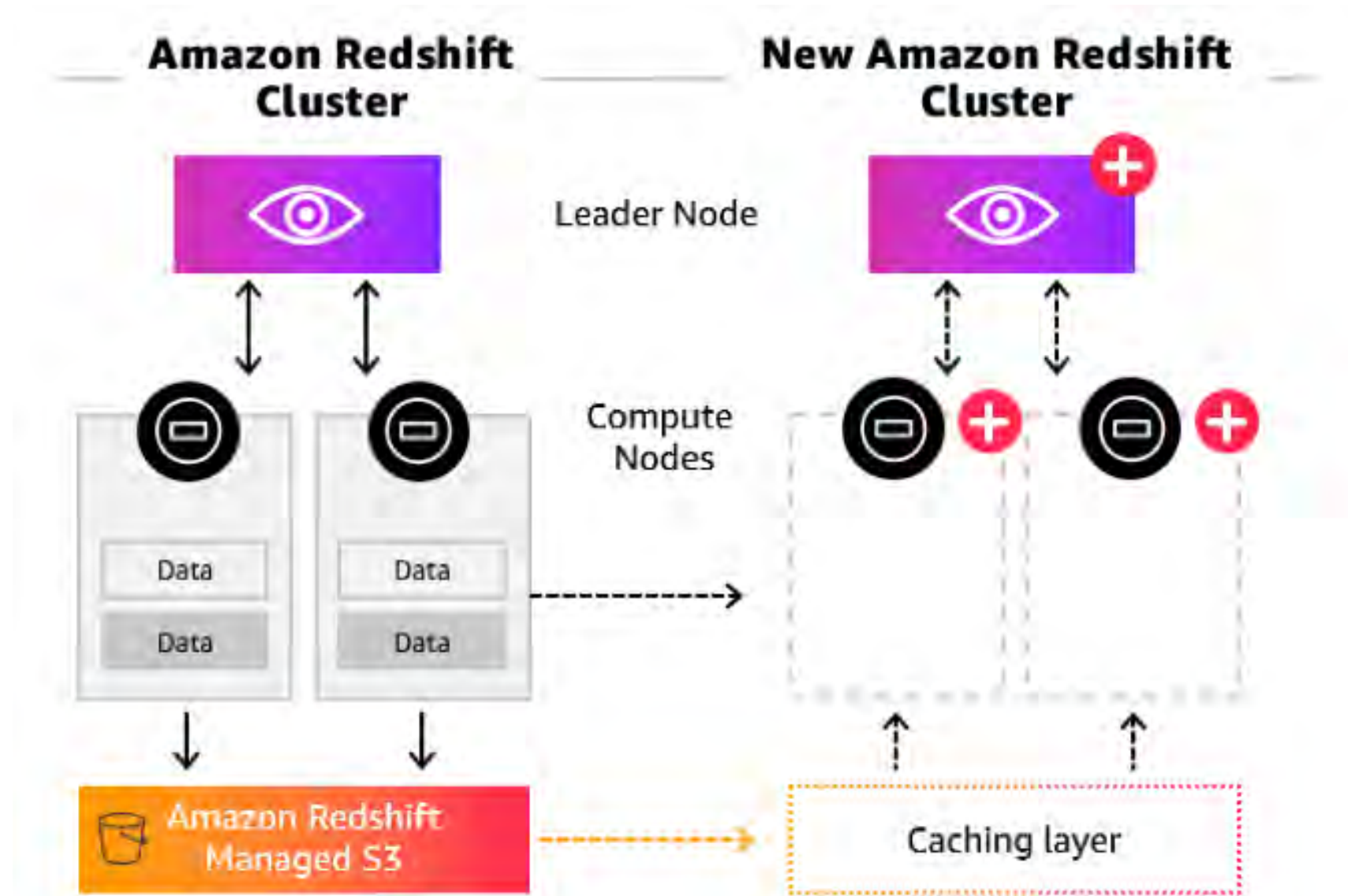
Querying in Amazon Redshift

Amazon Redshift
Connect to one or multiple Amazon Redshift clusters

Amazon Redshift Query Editor v2

Author and run queries

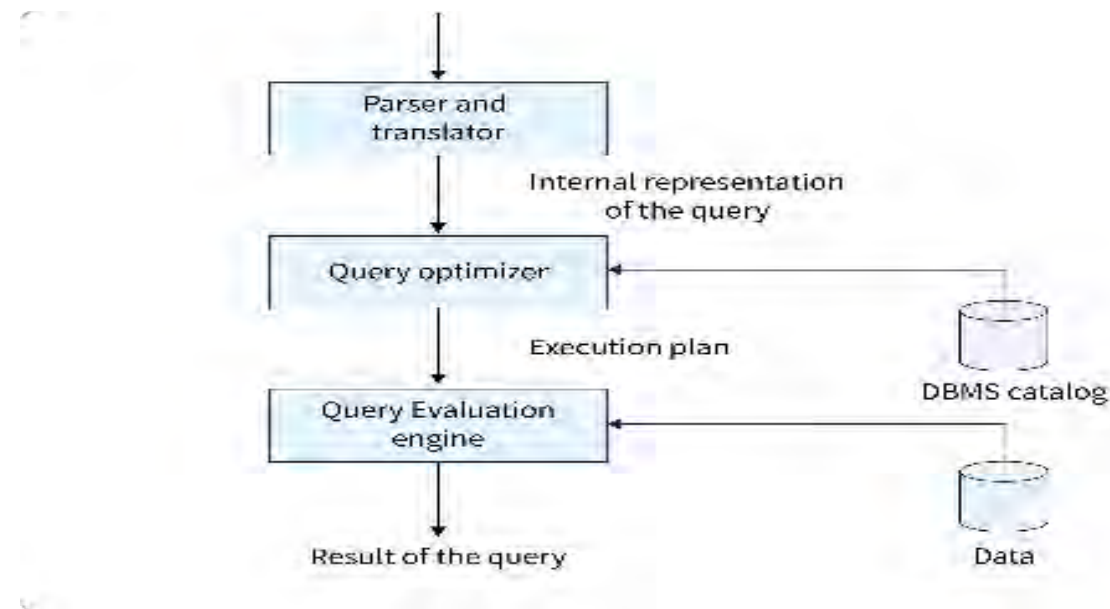Perform visual analysis

Share queries and results

Amazon Redshift Query Editor v2.0 is a web-based analyst workbench for you to explore, share, and collaborate on data with your teams in SQL through a common interface.

Optimizing Performance in Amazon Redshift

The following diagram is an architectural illustration of how Automatic table optimization works:
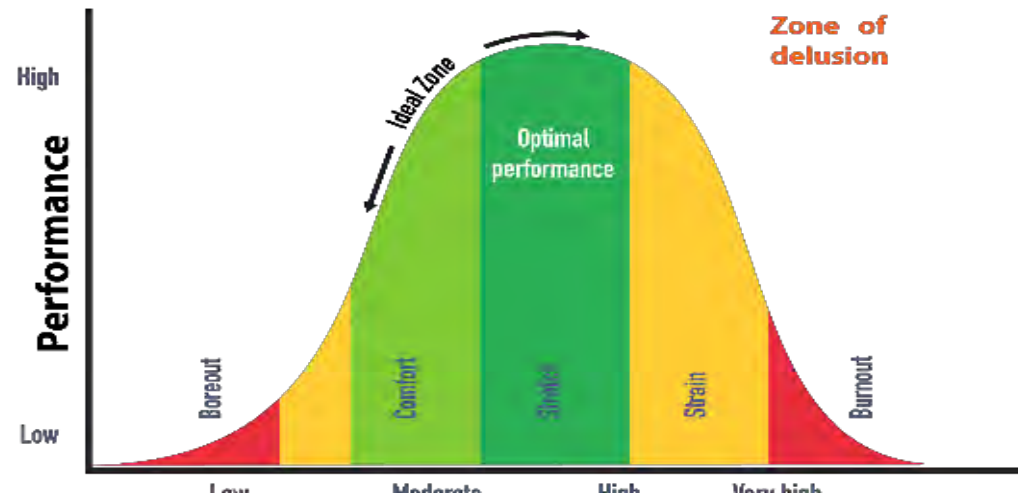

iamneo

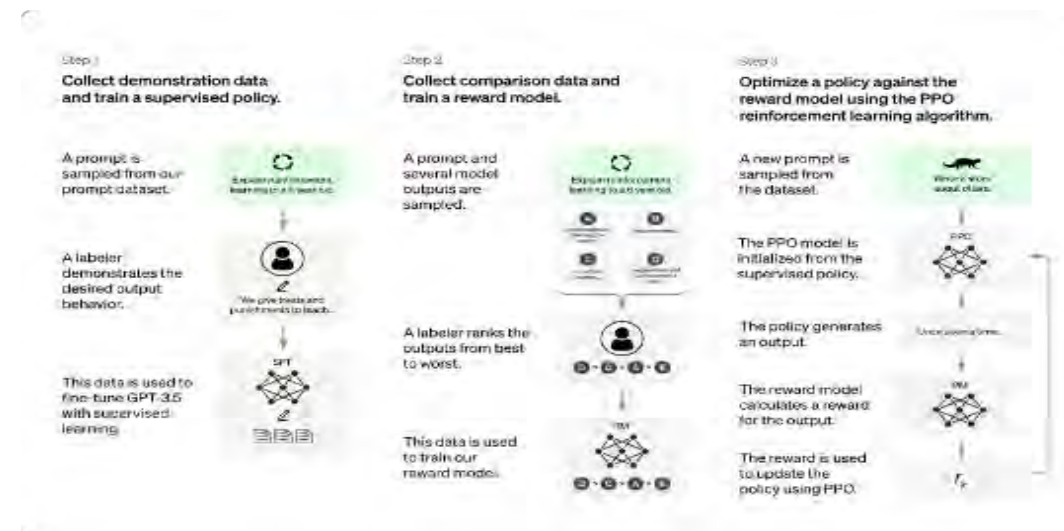www.iamneo.ai

Query optimizer

- The Amazon Redshift query run engine incorporates a query optimizer that is MPP-aware and also takes advantage of the columnar-oriented data storage.

- The Amazon Redshift query optimizer implements significant enhancements and extensions for processing complex analytic queries that often include multi-table joins, subqueries, and aggregation.
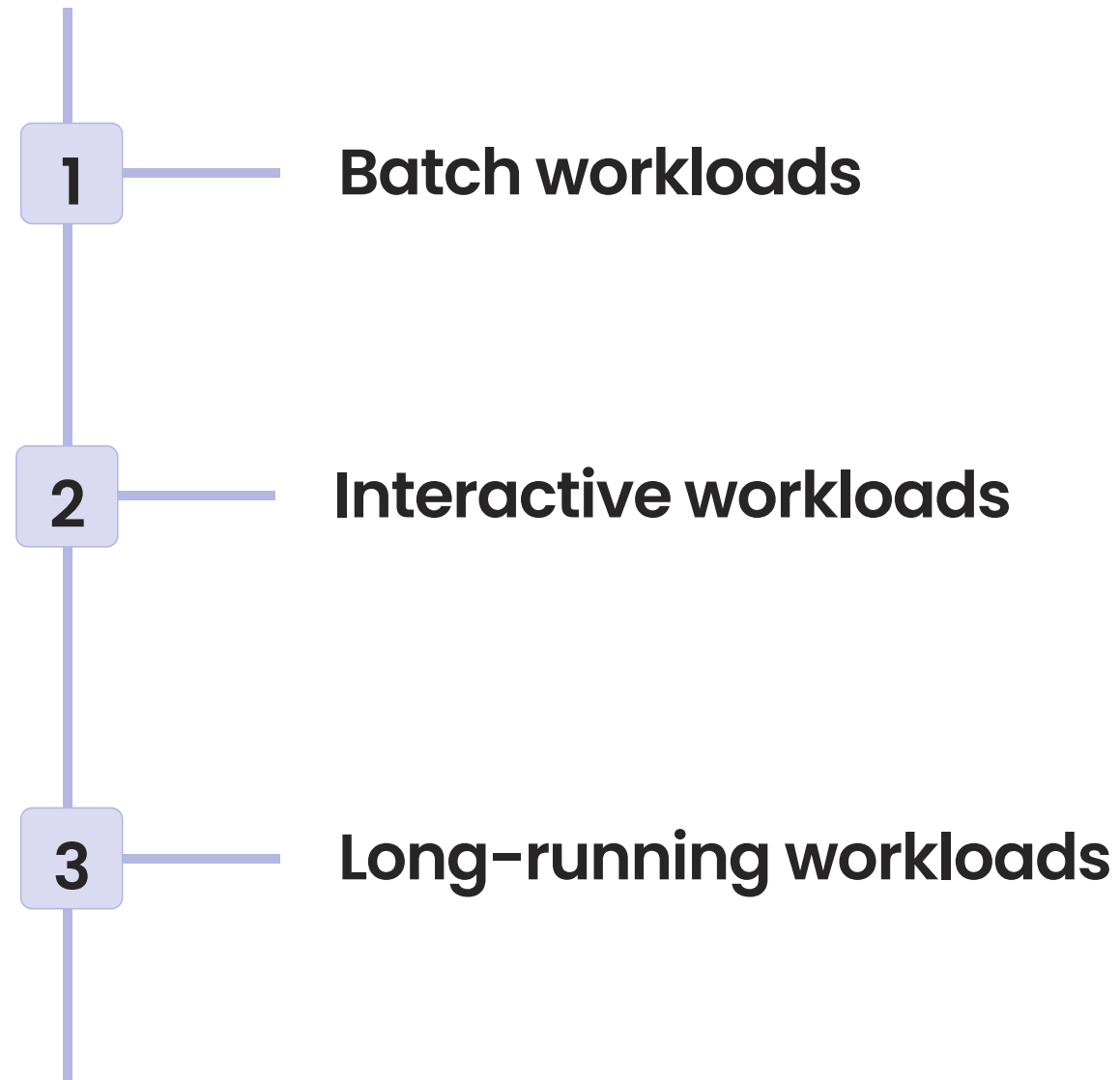
Utilizing Redshift query tuning performance

- Use Redshift performance tuning features to optimize performance, such as WLM, concurrency scaling, and workload management.

- Learn how to use advanced techniques for fine-tuning Redshift performance, such as workload management, parallel query execution, and optimizing disk utilization.

iamneo

www.iamneo.ai

Types of workloads in Redshift

1 — Batch workloads

2 — Interactive workloads

3 — Long-running workloads

Workload management when dealing with large datasets

1   **Granularity of queries**

2   **Query queuing**

3   **Workload management via WLM**

Scaling concurrency in Redshift

Maximize cluster size

Use appropriate distribution keys

Be mindful of query complexity

Avoid hotspots in your data

Benefits of effective workload management and concurrency scaling in Redshift

➢ **Improved query performance**

➢ **Reduced costs**

➢ **Improved data quality**