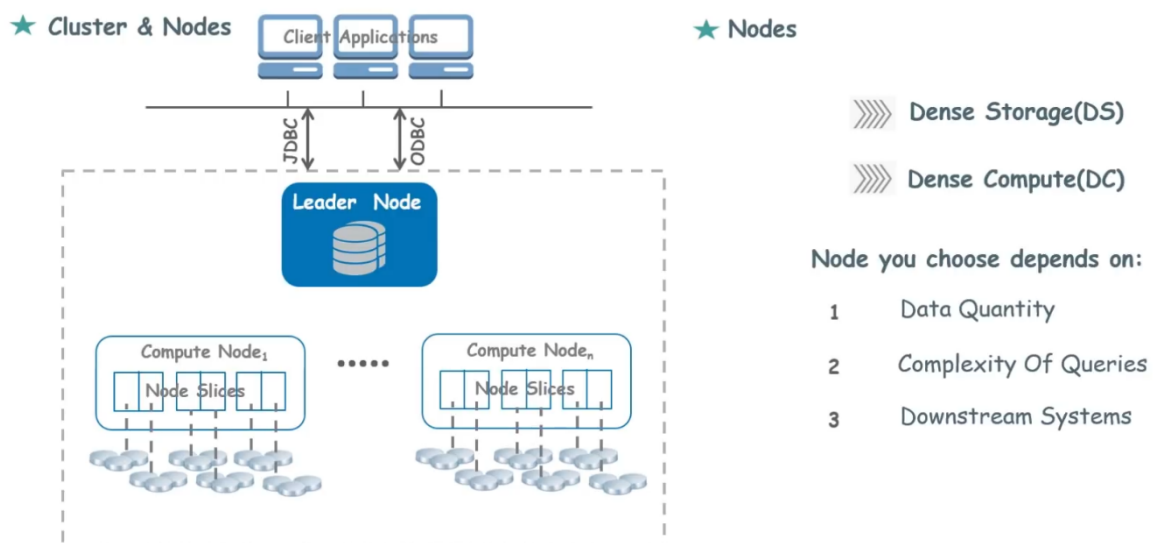


## Red Shift

- A data warehouse is a relational database that is designed for query and analysis rather than for transaction processing.
- It usually contains historical data derived from transaction data.
- To perform analysis, you need a data warehouse not a regular database.
- It is an AWS fully managed petabyte scale data warehouse service in the cloud.
- It gives you fast querying capabilities over structured data using familiar SQL based clients and BI tools
- Queries are distributed and parallelized across multiple physical resources.
- You can easily scale an Amazon Redshift data warehouse up or down with a few clicks.
- Amazon Redshift uses replication and continuous backup to enhance availability and improves data durability.
- Redshift is a SQL based data warehouse used for analytics applications (analytics DB)
- **Example use cases:**
  - Sales reporting, healthcare analytics
  - It is suited for OLAP based use cases
  - Can store huge amount of data (a database) but can not ingest huge amount of data in real time (not like Kinesis do)
- **Redshift can:**
  - Fully recover from a node or component failure.
  - It automatically patches and performs data backup
  - Backup can be stored for a user defined retention period.
  - Is 10 times faster than traditional SQL relational DB
- Redshift has much faster performance than other SQL DBs.
- Data is stored sequentially in columns instead of rows.
- Columnar based DB is ideal for data warehousing and analytics.
- Requires fewer I/O which can greatly enhance performance.
- Redshift automatically selects the compression scheme.

## Key Concepts



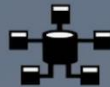
# Amazon RDS vs. Amazon Redshift

The tools have some similarities, but also some crucial differences.

	RDS	REDSHIFT
Data storage limit	64 TB	6 PB (multiple nodes in a cluster, 128 TB per node)
How is data stored?	Locally	Locally, externally (e.g., S3)
Scalability	Limited to master node and its storage size	Spread across multiple nodes in the cluster; much higher scalability than RDS
Support for SQL syntax	Yes	Yes
Design pattern	Online transaction processing (OLTP)	Online analytical processing (OLAP)
Examples of suitable use cases	Serves live requests, such as those coming from a web or mobile application	Asynchronously generates reports out of large amounts of data
Data formats	Accesses data stored internally in the database format	Accesses data stored internally in Redshift's format or data stored externally in JSON, CSV, ORC or Parquet, among others
Supports serverless?	Yes	No

## What is Redshift?

Data Warehouse



Distributed Workloads



Supports Petabytes of Data



Optimized for large queries



## Why use Redshift?

Elastic Scaling



Managed - Almost zero maintenance



Optimized Query Performance



Support Thousands of Users with a Single Cluster



Flexible Pricing Model



Integrates well with other AWS Services

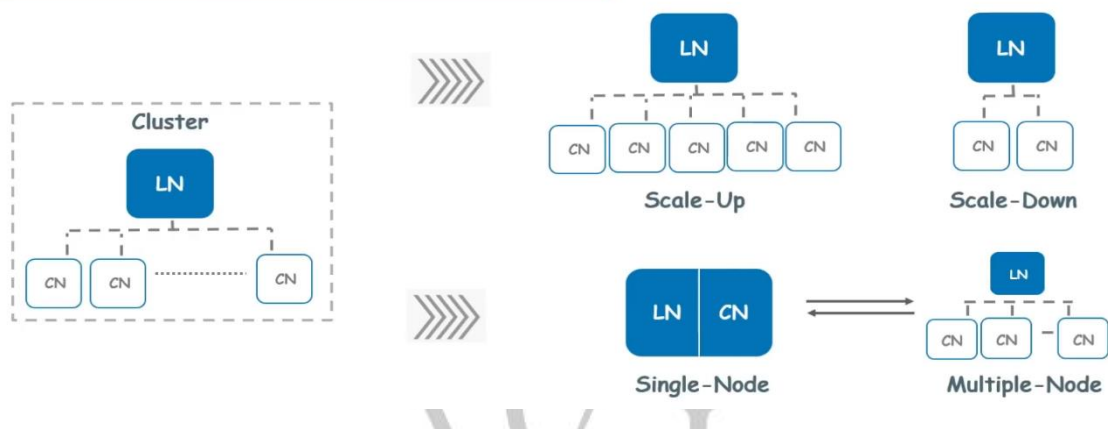


## Why Use Amazon Redshift

### 1 Easy to set-up, deploy & manage



### 2 Scales quickly to meet your needs



### 3 10x better & faster performance

#### ★ Columnar Data Store

Row Storage			Column Storage		
SSN	NAME	AGE	SSN	NAME	AGE
107135024	Jenson	25	107135024	Jenson	25
382634557	Sam	27	382634557	Sam	27

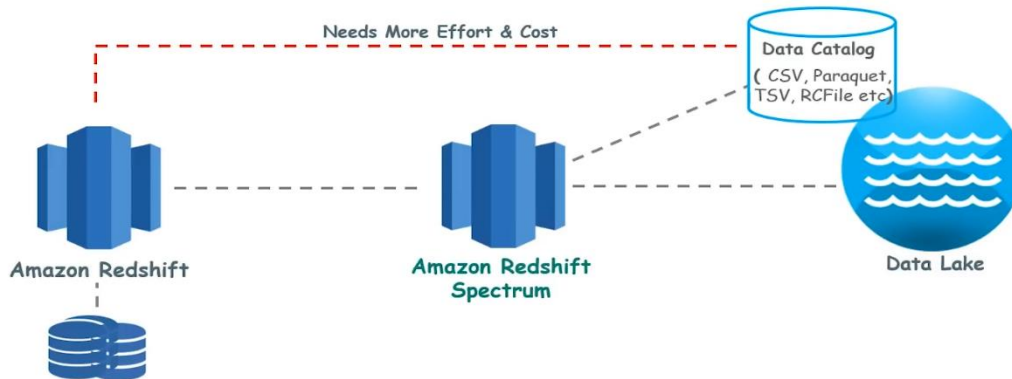
107135024   Jenson   25	382634557   Sam   25	382634557   107135024   ....	Jenson   Sam   .....
Block 1	Block 2	Block 1	Block 2

### 4 Cost - Effective



5

*Allows to query from data lake*

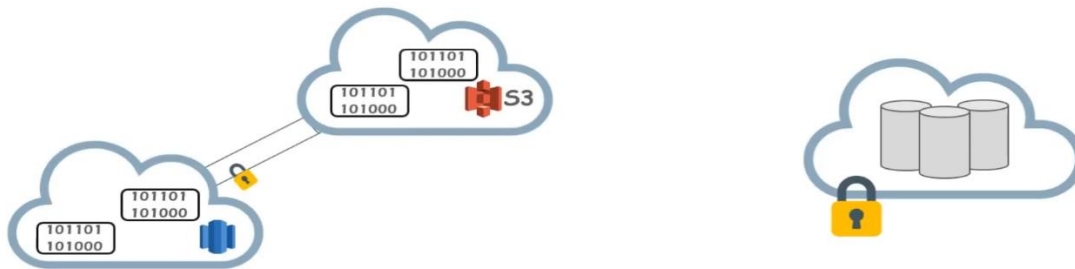


6

*Data is secure in Redshift*

★ Backup & Recovery

★ Encryption



## Cost & Pain Points

Instance Based  
OnDemand vs Reserved

Redshift Spectrum

Large Clusters can be very expensive

Advanced Setup can be daunting