

# **Amazon Redshift Interview Questions & Answers - Basic Level**

## **What are sort keys and how do they improve performance?**

Sort keys define the order in which rows are stored on disk. They improve query performance by allowing Redshift to skip blocks of data that don't match the filter criteria.

## **What are the different types of sort keys?**

1. Compound Sort Key: Best when queries use a prefix of the sort key.
2. Interleaved Sort Key: Useful when queries filter on any subset of the columns.

## **Explain Redshift Spectrum.**

Redshift Spectrum allows you to run SQL queries directly against data in Amazon S3 without loading it into Redshift. It's useful for querying large volumes of external data.

## **How would you handle data loading in Redshift from S3?**

Use the COPY command to load data in parallel from S3. It's fast and optimized for large volumes of data.

## **What is the purpose of the COPY command?**

The COPY command loads data from S3, DynamoDB, or other sources into Redshift tables efficiently. Best practices include compressing files, using multiple files, and defining column types.

## **How does Redshift handle concurrency and workload management (WLM)?**

WLM lets you define queues with memory and concurrency limits to manage how queries are executed. It ensures that high-priority queries get enough resources.

## **What is vacuuming in Redshift?**

Vacuuming reclaims space and sorts data after updates and deletes. It helps maintain performance and disk efficiency.

## **What is the difference between deep copy and shallow copy?**

## **Amazon Redshift Interview Questions & Answers - Basic Level**

Deep copy creates a new copy of the data, while shallow copy reuses existing metadata pointers. Redshift's CREATE TABLE AS (CTAS) performs a deep copy.

### **How does Redshift handle backups and restores?**

Redshift automatically takes incremental snapshots and stores them in S3. Users can also take manual snapshots and restore them to new clusters.

### **How would you monitor query performance in Redshift?**

Use tools like AWS CloudWatch, STL system tables (e.g., STL\_QUERY), and the Query Performance tab in the Redshift console.

### **How do you design a high-performance Redshift schema for a large dataset?**

Use appropriate dist keys to reduce data movement, sort keys to speed up filtering, avoid unnecessary indexes, and denormalize where it improves performance.

### **What are common performance tuning strategies in Redshift?**

Use compression (encodings), distribution and sort keys, vacuum and analyze regularly, monitor disk usage, and avoid long-running transactions.

### **Explain data distribution skew. How do you resolve it?**

Skew happens when data is unevenly distributed across nodes. Use better dist keys or switch to EVEN/AUTO styles to fix it.

### **How do you optimize Redshift for large-scale analytics workloads?**

Partition data across compute nodes efficiently, use RA3 nodes for scalable storage, apply sort/dist keys, and leverage Spectrum for external data.

### **Compare Redshift with other data warehouses like Snowflake or BigQuery.**

## **Amazon Redshift Interview Questions & Answers - Basic Level**

Redshift: MPP, strong AWS integration. Snowflake: Decoupled compute/storage, auto-scaling. BigQuery: Serverless, pay-per-query model.

### **How would you secure sensitive data in Redshift?**

Use IAM policies, VPC, SSL for data in transit, and encryption (KMS) for data at rest. Use role-based access control for users.

### **What are the key considerations when migrating from an on-premises database to Redshift?**

Data modeling, schema conversion, ETL pipelines, indexing differences, and network security must be evaluated. Use SCT and DMS tools.

### **How does materialized view work in Redshift and when would you use it?**

Materialized views store precomputed results of a query. Use them to speed up complex aggregations and joins that don't change often.

### **Describe the maintenance processes involved in Redshift (vacuum, analyze, etc.).**

VACUUM reclaims storage, ANALYZE updates table stats, DEEP COPY rewrites data for performance. Regular maintenance ensures optimal performance.

### **How does Redshift handle JSON or semi-structured data?**

Use the SUPER data type and PartiQL to query semi-structured JSON data. Redshift Spectrum also supports querying nested JSON files from S3.