Amazon Redshift Interview Questions & Answers - Basic Level

What is Amazon Redshift?

Amazon Redshift is a fully managed, petabyte-scale data warehouse service in the cloud. It allows you to run complex queries and perform analytics on large datasets using standard SQL.

What are the key components of Redshift architecture?

Leader Node: Manages query planning and coordination.

Compute Nodes: Store data and execute queries.

Client Applications: Send SQL queries to the leader node.

Columnar Storage: Improves performance and compression.

How is Redshift different from traditional RDBMS systems?

Columnar Storage: More efficient for analytics.

Massively Parallel Processing (MPP): Distributes workloads across nodes.

Optimized for Read-heavy Workloads: Great for OLAP, not OLTP.

Cloud-native: Managed service with easy scaling.

What is a Redshift Cluster?

A Redshift cluster is a set of nodes (1 leader, 1+ compute nodes) that together perform data storage and processing tasks. The cluster is the fundamental unit of compute and storage.

What are nodes in Redshift?

Leader Node: Parses queries and creates execution plans.

Compute Nodes: Perform the actual data operations (join, filter, aggregate).

How does Redshift store data internally?

Data is stored in a columnar format, which allows better compression and faster reads for analytical queries.

What is columnar storage and how does it help?

Amazon Redshift Interview Questions & Answers - Basic Level

Columnar storage stores data by column instead of row. This reduces I/O for analytical queries, improves compression, and speeds up aggregations and filtering.

What are the different node types in Redshift?

Dense Storage (DS2): More storage, less performance.

Dense Compute (DC2): More CPU and RAM, less storage.

RA3: Modern node type that decouples storage and compute for better scalability.

What is a distribution key?

A distribution key defines how data is distributed across compute nodes. A good dist key minimizes data movement during joins and improves performance.

What are the different distribution styles?

KEY: Based on the value of one column.

EVEN: Evenly distributes rows (default).

ALL: Copies the entire table to all nodes (used for small dimension tables).

AUTO: Redshift decides the best distribution style.