Redshift Cheetsheet -1

- Amazon Redshift is a fast, fully managed data warehouse that makes it simple and cost-effective to analyze all your data using standard SQL and existing Business Intelligence (BI) tools.
- RedShift is a SQL-based data WAREHOUSE used for analytics applications.
- RedShift is an Online Analytics Processing (OLAP) type of DB.
- RedShift is ideal for processing large amounts of data for business intelligence.
- Data can be loaded from S3, EMR, DynamoDB, or multiple data sources on remote hosts
- Redshift can handle petabytes worth of data. Redshift is for Data Warehousing
- RedShift uses replication and continuous backups to enhance availability and improve durability and can automatically recover from component and node failures.
- Redshift can only run in a 1 AZ (Single AZ)

Redshift Cheetsheet -2

- Redshift can run via a single node or multiple-node (cluster)
- A single node is 160 GB in size
- RedShift is 10x faster than a traditional SQL DB.
- A multi-node is comprised of a leader node and multiple compute nodes
- You are bill per hour for each node (excluding leader node in multi-node)
- You are not billed fort he leader node
- You can have up to 128 compute nodes
- Redshift has two kinds of Node Type; Dense Compute and Dense Storage
- Redshift attempts to backup 3 copies of your data, the original, on compute node and on S3
- Similar data is stored on disk sequentially for faster reads

Redshift Cheetsheet -3

- Redshift database can be encrypted via KMS or CloudHSM
- Backup Retention is default to 1 day and can be increased to a maximum of 35 days
- Redshift can asynchronously back up your snapshot to another region delivered to S3
- Redshift uses Massively Parallel Processing (MPP) to distribute queries and data across all loads
- In case of an empty table, when importing Redshift will sample data to create a schema
- RedShift uses columnar data storage;
  - ✓ Data is stored sequentially in columns instead of rows.
  - ✓ Columnar based DB is ideal for data warehousing and analytics.
  - ✓ Requires fewer I/Os which greatly enhances performance.