**Redshift**

**What is Redshift?**

* Amazon Redshift is a fast and powerful, fully managed, petabyte-scale data warehouse service in the cloud. Customers can start small for just $0.25 per hour with no commitments or upfront costs and scale to a petabyte or more for $1,000 per petabyte per year, less than a tenth of most other data warehousing solutions.

**OLAP:**

* OLAP transaction Example:
  + Net profit in large numbers of records.
    - Sum of radios sold in EMEA.
    - Sum of radios sold in Pacific.
    - Unit cost of radio in each region.
    - Sales price of each radio.
    - Sales price – unit cost.
* Data warehousing databases use different type of architecture both from a database perspective and infrastructure layer.

**Redshift configuration:**

* Single node (160Gb)
* Multi-node
  + Leader node: manages client connections and receives queries.
  + Compute node: stores data and performs queries and computations. Up to 128 nodes.

**Redshift – 10 times faster:**

* Columnar data storage: instead of storing data as a series of rows, Amazon Redshift organizes the data by column. Unlike row-based systems, which are ideal for transaction processing, column-based systems are ideal for data warehousing and analytics, where queries often involve aggregates performed over large data sets. Since only the columns involved in the queries are processed and columnar data is stored sequentially on the storage media, column-based systems require far fewer I/Os, greatly improving query performance.
* Advanced compression: columnar data stores can be compressed much more than row-based data stores because similar data is stored sequentially on disk. Amazon Redshift employs multiple compression techniques and can often achieve significant compression relative to traditional relational data stores. In addition, Amazon Redshift doesn’t require indexes or materialized views and so uses less space than traditional relational database systems. When loading data into an empty table, Amazon Redshift automatically samples your data and selects the most appropriate compression scheme.
* Massively Parallel Processing (MPP): Amazon Redshift automatically distributes data and query load across all nodes. Amazon Redshift makes it easy to add nodes to your data warehouse and enables you to maintain fast query performance as your data warehouse grows.

**Redshift pricing:**

* Compute node hours: total number of hours you run across all your compute nodes for the billing period. You are billed for 1 unit per node per hour, so a 3-node data warehouse incur 2,160 instance hours. You will not be charged for leader node hours; only compute nodes will incur charges.
* Backup
* Data transfer (only within a VPC, not outside it)

**Redshift security:**

* Encrypted in transit using SSL.
* Encrypted at rest using AES-256 encryption.
* By default, Redshift takes care of key management.
  + Manage your own key through HSM.
  + AWS Key Management Service.

**Redshift Availability:**

* Currently only available in 1 AZ.
* Can restore snapshots to new AZ’s in the event of an outage.