Glue ETL's

1. Create the Glue ETL job using pyspark.

2. Read the data from product table

3. Convert into parquet format.

4. Script Path

5. Temporary / cash.

ETL operation

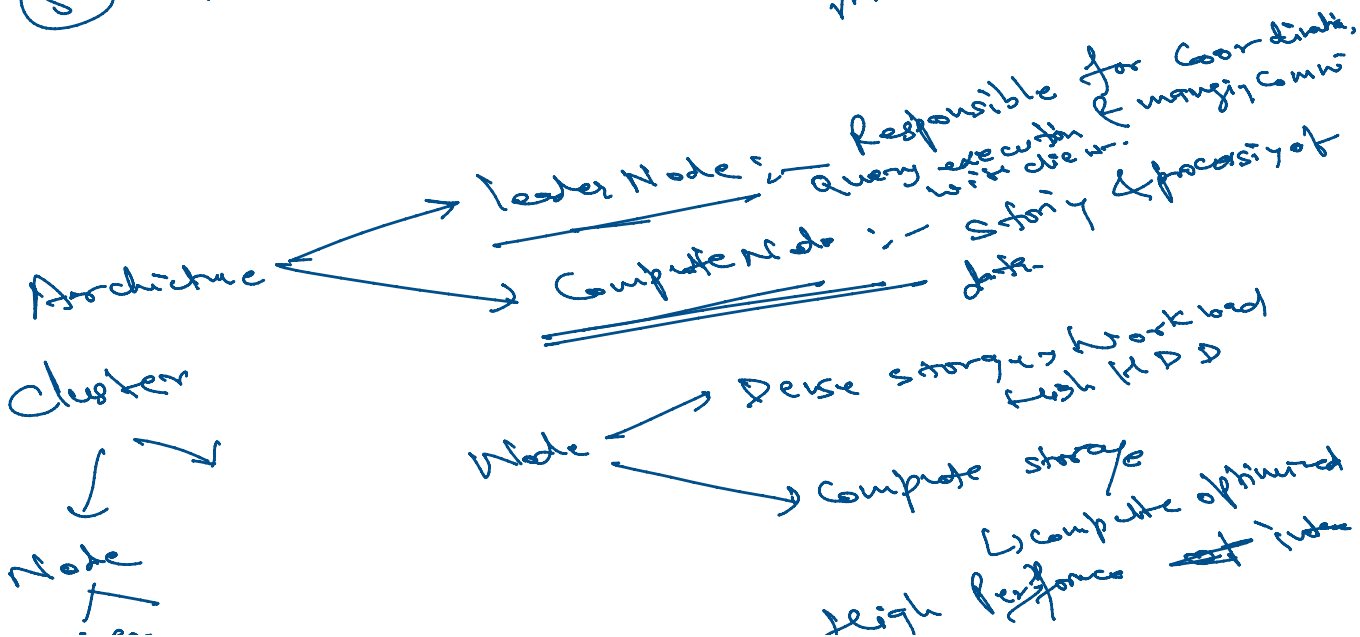1. Apply Mapping → Apply transformation to the dynamic frame.

2. Resolve choice.
   └ Handle column type conflicts that arises when data from different or format is combined

Use SQL Query in the glue.

Benefit of Redshift

1. Scalability → Petabyte

2. Simplicity

3. Cost effectiveness

4. Column data storage

5. Query optimization ── 1. Compress the data
   2. Store in columnar format
   3. Zone map & Query optimization Method.

Architecture ── Leader Node :- Responsible for Coordinate, Query execution & managing communi with cluster.

Compute Node :- Storing & processing of data.

Cluster

↓ ↓

Node

Node ── Dense storage → Workload with HDD

└ Compute storage
  └ compute optimized
    High Performance instance

Note
↑
slicer
128MB

High Performce set
Workload.