Day - 9 (24th May)

24 May 2024 10:08

1. Data Migrate from On-prem to RDS
   - ↓ EC2 instance
   - ↳ Serverless Service.

EC2 instance → Config., Set environmental Variable.

Database Setup.

↓ Dummy data.

Migration of Database in EC2 instance to RDS Database.

1. Migration Using DMS. → Amazon Database migration service.

```
mysqldump -u root -p ec2db >ec2db.sql
mysql -h <replace-rds-end-point-here>-P 3306 -u rdsuser -p rdsdb <ec2db.sql
mysql -h <replace-rds-end-point-here>-P 3306 -u rdsuser -p
USE rdsdb
SELECT *FROM table1;
```

AWS Redshift → Data ware house.

↓

Centralized data repository.

↳ Query & Analysis

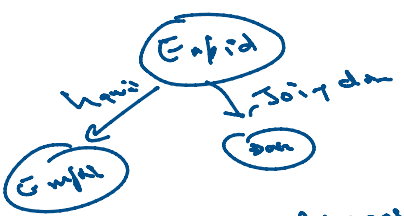
Data Modelling:- Blueprint or Conceptual representation of data.

Process Design & organize your data Model.

1. Conceptual DM → Initial gathering phase & high level view of data
2. Logical DM → structure of data - entities, attribute & table
3. Physical DM → Actual create storage & push the data or retrieve the data.

Data Modelling technique:-

- ER Model:- → Entity Relationship
  - Entity & Attribute Relation.
- Multi Dimensional Model :-
  - Data in format of Facts &



Multi Dimension

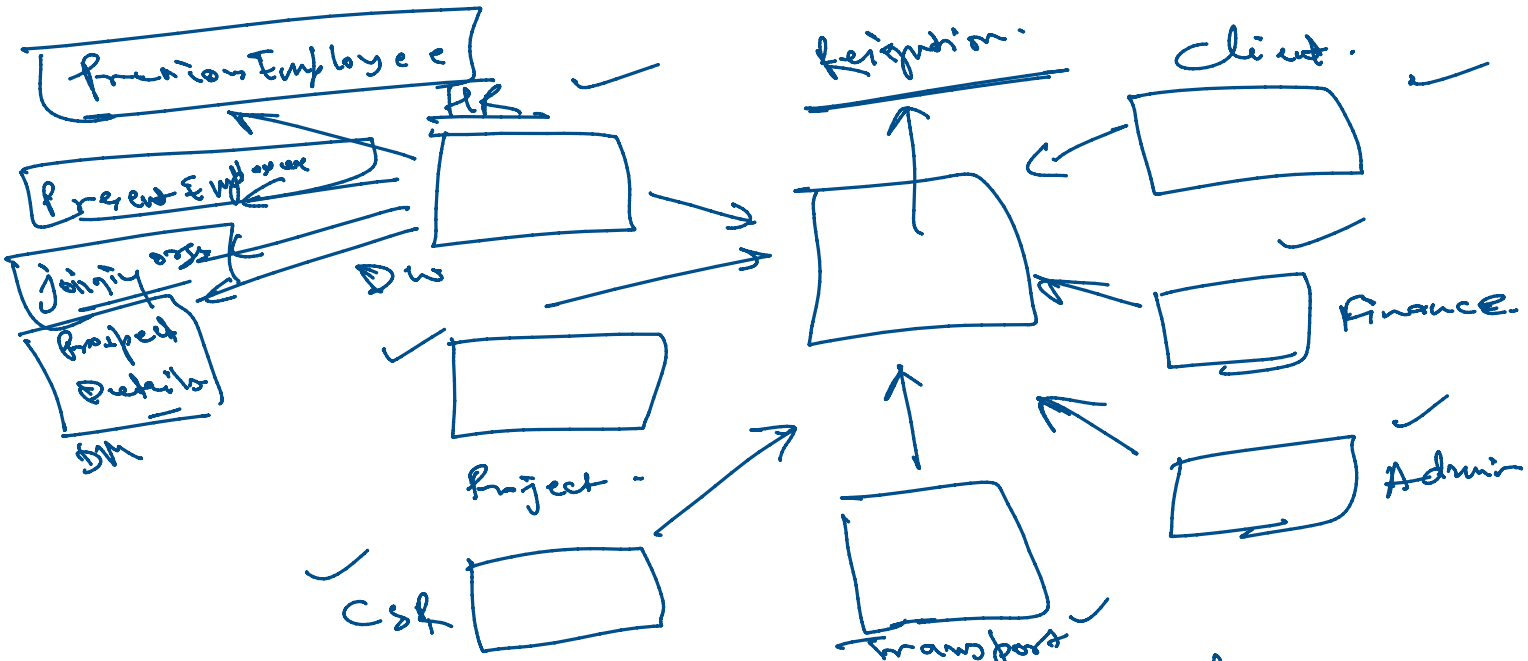
Measurable events or transaction → Data in format of Facts & Dimension

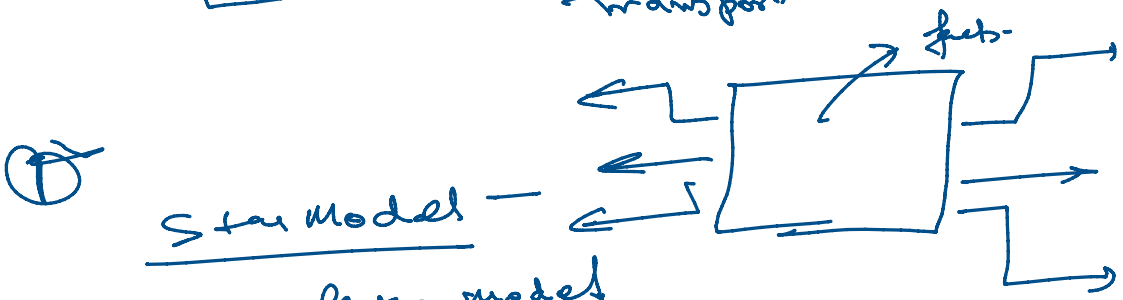
Model → Fact → Measure, Sales, Transaction, Quantity

Dimension → Business / Attribute of event

Two Basic Dimensional Model technique.

1. Star Model
2. Snowflake Model





Star Model

Snowflake model

Data Redundancy :- Huge data → lots duplicate.

→ Commercial → Amazon Redshift, SQL Server, IBM Netezza

Datawarehouse → Commercial → Amazon Redshift, Microsoft SQL Server, IBM Netezza, Oracle Exadata.

→ Open Source → Apache Hive, Apache Cassandra, Postgresql

↓

Datamart → subset of Datawarehouse.

Amazon Redshift

1. Scalability → Petabytes.
2. Pay-as-you-go
3. Performance.
4. Columnar dataset.

Query optimization. → Zone map. Query optimization method.

Architecture

→ Leader Node →

→ Compute Node →

Select * from table1;

Calculate the cost ← LN → Coordinate b/w User & Redshift Query execution.

Cluster → Node → Slices

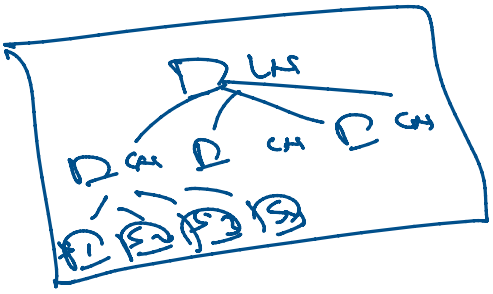
CN CN CN CN

Node

→ Dense storage → High HDD

→ Compute storage → High performance. High Workload.

Size of each slice: 128MB



cluster.