

## \* Snowflake Documentation

→ Analytic cloud, Data Warehouse provided as a service.

→ Faster, Easier to use & Flexible.

→ Client just loads & querying the data & snowflake will take care of the rest.

→ Clean, Easy to navigate interface.

→ Menu Bar ⇒

① Databases

② Warehouse

③ Worksheet

④ History

⑤ Help menu

⑥ User menu. → Login, Roles Security (Public)

## \* Help menu

→ Access to snowflake documentation

→ Can download software.

Through command line client & other interface

## \* Database ⇒ ANSI Std SQL

SQL → To store, maintain & query structure & semi-structured data.

## Snowflake

⇒ All data in Snowflake is stored in databases.

\* Databases ⇒ Logical grouping of objects ⇒  
Tables & Views ⇒ Organized into Schemas.

Create, Clone, Drop, Transfer ownership to different user.

⇒ select any row ⇒ Panel opens  
→ Display security details for the object.  
+ View & Manage objects

\* Warehouse ⇒ Cluster of servers.  
→ Provides compute resources  
→ Execute queries  
→ Load data  
→ Perform other DML operations

→ Create, Drop, Resume, Suspend Warehouse  
→ ownership, Configure

\* NIKE → Data Warehouse implementation  
Data migration.

\* Bagmane Tech Park → Indiranagar  
not later than 31st sept. Captive centre.  
CV Roman Nagar

\* Teradata → Conversion → Big Data into cloud.

→ Snowflake → system → Cloud Based Data  
for Big data

Hosting → AWS

\* Quad 1, Quad 2 → Commercial data  
Legacy system → Teradata  
into cloud → Snowflake.

\* Challenges ⇒ Data Duplication.  
Lot of data

\* Role → BA? Database, SQL, Oracle  
Teradata

\* Data Analyst → Insights → model → Imp<sup>^</sup>  
Biz Decisions

Skill Gap

Database, Datawarehouse, Big Data, AWS ⇒ Read the  
Amazon S3, EC2, Agile Story Points ⇒ Docs.  
Client Disc. **weschool** IT Engs → soft implementation  
Welingkar Education

## SQL Worksheet

→ Interactive ANSI SQL interface for running queries & performing DDL & DML ops

DML → Select, Insert, Update, Delete.

→ Used to store, Modify, retrieve, delete & Update data in database.

→ Creating & using multiple warehouse to separate resources & intensive tasks.

↳ Data Loading & queries → Isolating queries from others.

→ Resize warehouse at any time to improve performance. , Load script file.

→ Results can be exported to a delimited text file.

## \* Monitor Queries

→ History page → Displays all SQL commands executed over last 14 days.  
↓  
from Client slw. ⇒ Filter

Query ID ⇒ To Examine the details of the query →  
→ view query results (last 24 hrs)

docsnowflake.net → Learning material

## \* Architecture & Key Concepts

→ Snowflake's Cloud data Warehouse

↳

↳ New SQL Data Warehouse build for the cloud.

↳ Snowflake runs on Amazon Web Services cloud (AWS)

↳ Delivered as a Data Warehouse as a service.

## \* Snowflake

- Software Configuration
- Resource Management.
- Availability
- Authentication.
- Data Protection
- Tuning & Optimization.

## \* Traditional DW (Shared Disk Archi)

- Multiple nodes on a single cluster to access data shared on a single storage system.
- Shared Nothing Architecture
- Shared portion of data in each node in each cluster in the Data Warehouse.

\* Snowflakes → Combines all the benefits of this ~~architecture~~ in a innovative new design takes full advantage of the cloud.

⇒ Snowflake

→ Multi Cluster - Shared data Architecture.

→ Consist of 3 separate layers.

- ↳
- ① Data Storage
  - ② Compute
  - ③ services layer → Manages all.

→ Each Layer scale independently & includes Built in Redundancy.

⇒ Organizes data into Logical Database. Containing one or more Schemas. → Skeleton structure  
↳ Logical view of DB

Defines

\* Schema → How the data is organized.  
→ Their Relations & associated.  
→ Formulates all the constraints.  
→ It defines tables, views & integrity constraints.

DBMS Data types ① VARCHAR ② NUMBER  
③ BOOLEAN ④ TIMESTAMP.

Non RDBMS → Semi-structured ⇒ Variant

→ Stores all data in

→ Amazon S3 cloud storage as

**weschool**  
S.P. MANDAL'S  
Welingkar Education

## \* Snowflake

→ Data is converted into an Optimized Columnar Compressed Encrypted format & stores the data in Snowflake private S3 buckets.

## \* Compute Layer

→ Queries are Executed in Snowflake using resources from Amazon EC2 Service.

→ Allows to create multiple independent Compute cluster called virtual warehouses access the same data Storage layer without contention or performance degradation snowflake's

→ Processing queries → as soon as it is provisioned scaled up & down without any downtime or disruption.

→ Loading & querying Concurrently.

## \* Service Layer → Co-ordinate & Manages the Entire System

→ Manage Metadata, performs query compilation



→ Clients → JDBC & ODBC drivers

→ Command line client & web based user interface  
↳ SQL Commands.

→ Snowflakes eliminates all knobs & tuning parameters.

\* Snowflakes **USERS**

↳ Create DB, schemas, tables, views

↳ Create virtual warehouse.

↳ Load & query data

↳ **Snowflake** → Handles the rest

\* Lifecycle of a query

→ Queries send to the Services layer using any supported client or interface

↳ Validates that user is sending their query if Authorized to access the data & perform

Service layer →

VW → Executes query

**weschool**  
Welingkar Education

resolve & returned to the client.

Data Recovery, Compression

# \*] Introduction to Virtual Warehouse

\* Warehouse ⇒ One or more clusters of  
↓  
Computes resources.

→ To execute queries & perform other  
DML operations.

→ Data Loading & Data Analysis.

→ Maintain Development & testing Environment  
Separate from production.

→ Size ⇒ S, M, <sup>Large</sup> ~~Complex~~ ⇒ Determines no. of

→ Double the amt of resources available  
Servers in each cluster  
in the Warehouse.

⇒ Warehouse can be created and managed  
Using either the Web interface or SQL

> Creating Warehouse Using the User interface,  
(Name, Size, max no. of clusters) ⇒ Return <sup>Aut</sup>

> Once created ⇒ starts provisioning servers.

→ queries → queued  
↳ completed when resources become available.

⇒ Resizing of Warehouse Resizing & Multi Cluster Warehouse.

→ Queries are not running fast enough or Data is loading slow ⇒ Increase the size.

\* Multi Cluster Warehouse

→ Adapts to fluctuating no. of users & queries

⇒ Warehouse → Performance benefits for slow running queries & Data loading.

→ Size & no. of clusters

→ How long it runs

→ Warehouse consumes credits for each hour or fraction of an hour.

\* Database & Querying

⇒ Logical Containers

↳ organize your data into schemas & tables.

→ Query across databases in a single query.

# Databases, Schemas & Tables

## Feature of Snowflake

⇒ Zero Copy Cloning of DB, Schemas & tables

\* Zero Copy Clone ⇒ only uses storage if you change or add data

⇒ Sample DB ⇒ Read only, no changes  
→ Query ✓

→ Contains multiple DB each stored in a separate schema.

Worksheet ⇒ For building & Executing queries  
→ Performing all other SQL operations  
ID & status of each query.

→ Performance & behaviour of query ⇒ Click Profile to  
→ Graphical Representation.

Amazon EC2  $\Rightarrow$  Amazon Elastic Compute Cloud  
Amazon S3  $\rightarrow$  Amazon Simple Storage Service

AWS  $\rightarrow$  Rent virtual computers on which to run their own Computer App<sup>n</sup>.

EC2  $\rightarrow$  Encourages Scalable deployment of App<sup>n</sup> by providing a Web Service.

$\hookrightarrow$  Cloud Computing Web service offered by AWS

AMI Amazon Machine Image.

EC2  $\rightarrow$  Web service that provides secure, resizable compute capacity in the cloud.

- $\rightarrow$  Designed to make web scale Cloud computing easier for Developers
- $\rightarrow$  Quick scale capacity  $\rightarrow$  New server instances
- $\rightarrow$  Failure resistant App<sup>n</sup> by Dev in <sup>mining</sup> AWS
- $\rightarrow$  Isolate them from Common failure scenario.

CSV <sup>structured</sup> → Data Loading

\* Unstructured → JSON, AVRO & Parquet <sup>quet</sup>

\* upload ⇒ Staging the files

⇒ Load the data (Internal & External locations)

→ Provides convenient secure internal staging location for each table.

Amazon S3 → Externally staged.

Need to supply bucket URL & access credentials

→ Running virtual warehouse

- WH Extracts data from each file & insert it as rows into the table.

Loading by web interface or SQL

\* SQL

demo\_db database

→ Internal staging loc<sup>r</sup> for table

Datatype ⇒ String

@ ⇒ Stage the files

% ⇒ Internal staging locations for the specified table.

\* USE Warehouse: Dest<sup>r</sup> → Copy into ~~@~~ Contacts  
Locn<sup>r</sup> of Staged Files → from @ % Contacts  
Data files to ~~loc~~ Pattern  
nos 1 to 4 → on - Error

\* Snowflake

\* Powerful ops for error handling

Query) Select \* from Contacts limit 10;

Combining 2 phase in single operation.

loading ⇒ web interface or SQL  
SQL → Large files<sup>no. of</sup>

# \*] Amazon Web Service (AWS)

- On-demand access to highly durable storage
- Low cost compute
- High performances databases
- Tools to manage resources
- Pay for use
- Database driven applications.
- Compliance, secure, Resize
- Can create scalable applications of any size & shape.

## \*] AWS

- Global Cloud platform.
- Allows to host & manages servers in the internet
- Used by around 80% of fortune 500 companies to host their infrastructure.
- IaaS → provide fair service as a service
  - ↳ Manage the backup & the power supply (Recovery)
- Platform as a Service (PaaS) → Java, Ruby, PHP
- SaaS → Email ⇒ SES,
  - Queuing services ⇒ SQS



## \* Cloud storage platform

→ Lot of storage options. ⇒ EBS & S3  
(Elastic Block Store)

## \* AWS & Hosting provider

→ Gives a lot of services.

→ Users can run your App<sup>n</sup> on cloud.

## \* Billing → Per hour Billing

→ Every instance or activity ⇒ Micro billing

on EC2 →

→ Transparent

## \* S3 Bucket ⇒ Charged on a per GB basis.

→ Micro-billing.

⇒ Sign up process is Easy.

⇒ Launch up servers & big machines without buying hardware.

→ Simple billing Dashboard ⇒ Reports

→ Services → stable (7-8 years) ⇒ 80% 4 major outages.

→ Trusted vendor - Advisor

→ Region specific.

→ Not more than 2-3 hours

\* EC2 → Elastic Compute Cloud  
→ Base servers.

→ This service will give you a machine which you can launch and you can run SW on those.

→ We can get small or big machines based on our requirements.

\* Virtual Private Cloud (VPC)

→ Amazon will not allow full control of cloud.

→ Will give you chunks of their cloud ⇒ VPC

→ VPC allows you to create networks in the cloud and then run your servers in those networks

\* Simple Storage Cloud (S3)

⇒ Opportunity to upload & share files.

⇒ Files Storage & sharing service

RDS → Relational Database Service

\* RDS ⇒ Allows you to run and manages databases on the cloud  
→ All major flavors of databases right from SQL Server to Oracle, MySQL, PostgreSQL

→ Launch ⇒ Aurora  
⇒ High performance Database

\* Route 53 ⇒ DNS  
→ Manage DNS Service.  
→ Where we can point our dns to Amazon & they <sup>(AWS)</sup> take care of the stuff.

→ Global DNS Service  
→ Scalable DNS Service  
→ Scales according to demand.

\* Elastic Load Balancing (ELB)

→ ELB is a service gives you the opportunity to load balanced incoming traffic to multiple machines.

→ This way we can scale up your web applications to any number of users.

→ Auto scaling ⇒ Adds capacity on the fly to Elastic Load balancers  
our website is never down due to load

# \* Pricing

- Region specific pricing
- Virginia ⇒ Cheapest region. { Good hold & headquarters }
- Oregon
- Team specific pricing → Cheaper rather than on-demand basis
- Reserved instances cheaper than on-demand ones
  - ↳ discount ⇒ 20-60% ⇒ sign for 3 years term.

# \* Spot Resources

- Spot instances ⇒ Bidding Market base
- ⇒ Bid for a price
- ⇒ terminated if someone bid higher.
- ↳ Ad-hoc tasks. ⇒ Not critical

AWS ⇒ Multiple availability zones per region.

→ Massive Data Centres ⇒ (300,000 - 5,00,000) Servers

# \* Future of AWS (Scaleup → Cost benefits)

- 64 Services currently ⇒ <sup>Across</sup> (IaaS, PaaS, SaaS)
- Focusing on ML, SaaS products

user can take control of the service that user want to utilize. (we to upload) → Reducing cost. Etc