

Crash Course on Modern Data-Warehousing using

Snowflake Platform

By:

Bhavuk Chawla

Founder, DataCouch

Snowflake Certified Professional

Google Certified ML Engineer

Google Certified Data Engineer

Agenda

1 Challenges with Traditional/Big Data DWs

2 What is Snowflake?

3 Why Snowflake?

4 Snowflake Architecture

5 Working with Snowflake

6 Data Caching in Snowflake

About Bhavuk Chawla

Bhavuk has over 17 years of experience in IT, more than 8 years of experience Implementing Cloud/ML/AI/Big Data Science related projects. He is an official instructor for Google, Confluent, Snowflake and Cloudera. He has delivered and continues to deliver his knowledge sharing sessions in various companies like Google Singapore, Microsoft Bangalore, Starbucks Coffee Seattle, Adobe India and EMEA Region, etc.

He was recognized by Cloudera as the Instructor of the Year 2016 (APAC) and Pluralsight Elite instructor 2020 & 2021 for his exceptionally high ratings received in various training sessions.



About Datacouch

1

100K+ Participants Transformed

2

100+ Courses

3

Focused in Big Data, Data Science, AI, DevOps and Cloud

4

Industry Experts / Globally Renowned / Certified by Google



About Datacouch



AR / VR



IoT & Devices



5G Technology



Blockchain



Digital Twins



P I O N E E R S I N B L E E D I N G E D G E T E C H N O L O G I E S



AI / ML



Data Fabric



Decentralized
Computing



Semantic Web



Cybersecurity
Mesh

Sharing Knowledge Globally



Sapient - Gurgaon



Axiata - Malaysia



Deutsche Bank - Pune



Google - Singapore



Intel - Malaysia



BCBS - Chicago



NCell (Axiata) - Nepal



IBM - Malaysia



TCS - Chennai

What is a Data Warehouse?

- ❑ A central repository of primarily structured data accumulated from a wide range of data sources
- ❑ Used for reporting and data analysis and is considered a core component of business intelligence



Source: SAP

Challenges with Traditional DWs

- *Handle only Structured Data*
- *Performance for large Datasets is usually slow*
- *Storage*
 - *Not Infinite*
 - *Not Elastic (Scalable)*
 - *Cost per TB is usually very high*
- *Operational Burden on DBA, need to take care of a lot of activities -*
 - *Backup & Restore*
 - *Multi-Tenancy*
 - *Analyze*
 - *Upgrading*
 - *Indexing*
 - *Memory Management*
 - *Patching*
 - *Partitioning*
 - *Workload*
 - *Capacity Planning*
 - *Ordering*
 - *Security Management*
 - *Initial Setup*
 - *Vacuuming*

Challenges with Big Data DWs *(such as Hive)*

- *Batch Data Analysis Framework*
- *HiveQL lacks a lot of features such as Indexing support limit*
- *No Materialized Views*
- *No Stored Procedure support*
- *Updating data in Hive is very painful*
- *RDBMS is required to store Metadata*

What is Snowflake?

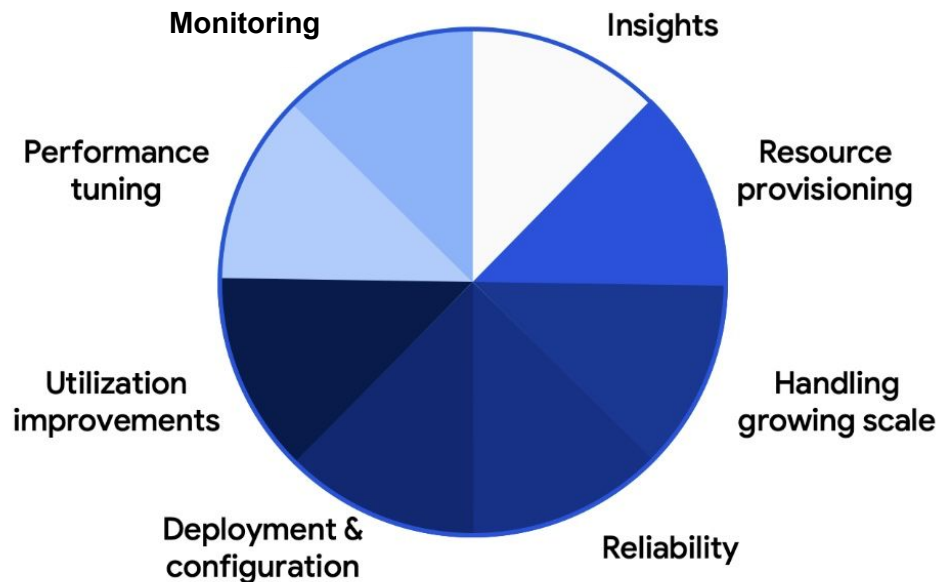
- ❑ Fully Managed Cloud Native Data Warehouse
- ❑ Founded in 2012 in San Mateo, California by three data warehousing experts: Benoit Dageville, Thierry Cruanes and Marcin Zukowski
- ❑ Having AWS support since Inception
- ❑ Now it can run on top of any of the three major cloud providers i.e. AWS, GCP or Azure

Why Snowflake?

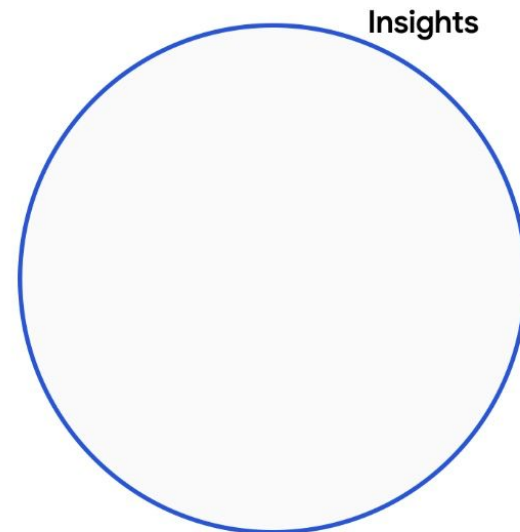
Snowflake Warehouse is a single integrated system with fully independent scaling for compute and storage

- *Snowflake is Cloud Agnostic Data Warehouse*
- *Supports modern features like auto-scaling, auto suspend, big data workloads, and secure data sharing*
- *Pricing is based on the amount of data you store, and the compute hours you use*
- *No need to worry about managing, scaling multi-cluster systems, or tuning clusters for fast performance*
- *Native support for Semi-Structured Data*
- *Time Travel and Zero Copy Clone enables us to implement automated Backup and Restore seamlessly*
- *Highly Secure - Have full control over who has access to the data*

Summary: Why Snowflake?



Other Data Warehouses

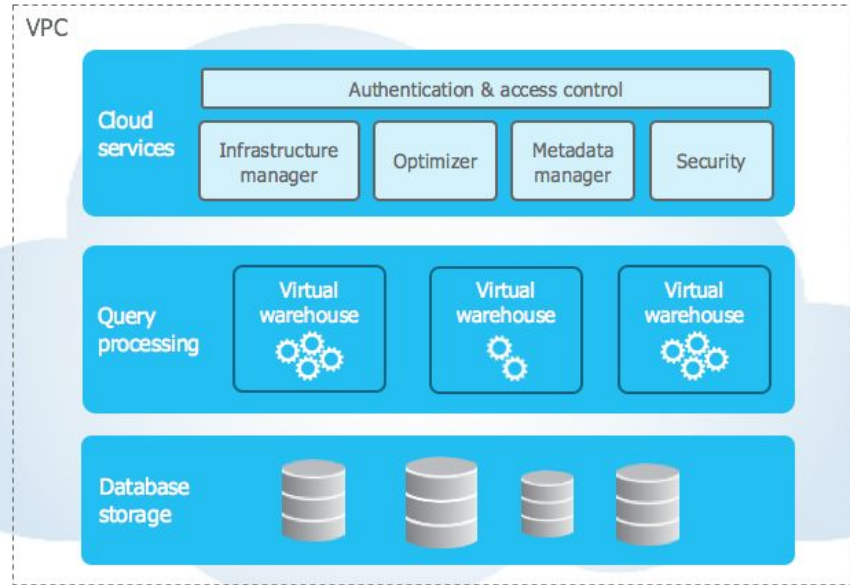


Snowflake

Snowflake Architecture

Snowflake's unique architecture consists of three key layers:

- ❑ Database Storage
- ❑ Query Processing
- ❑ Cloud Services



Source: Snowflake

Snowflake Architecture



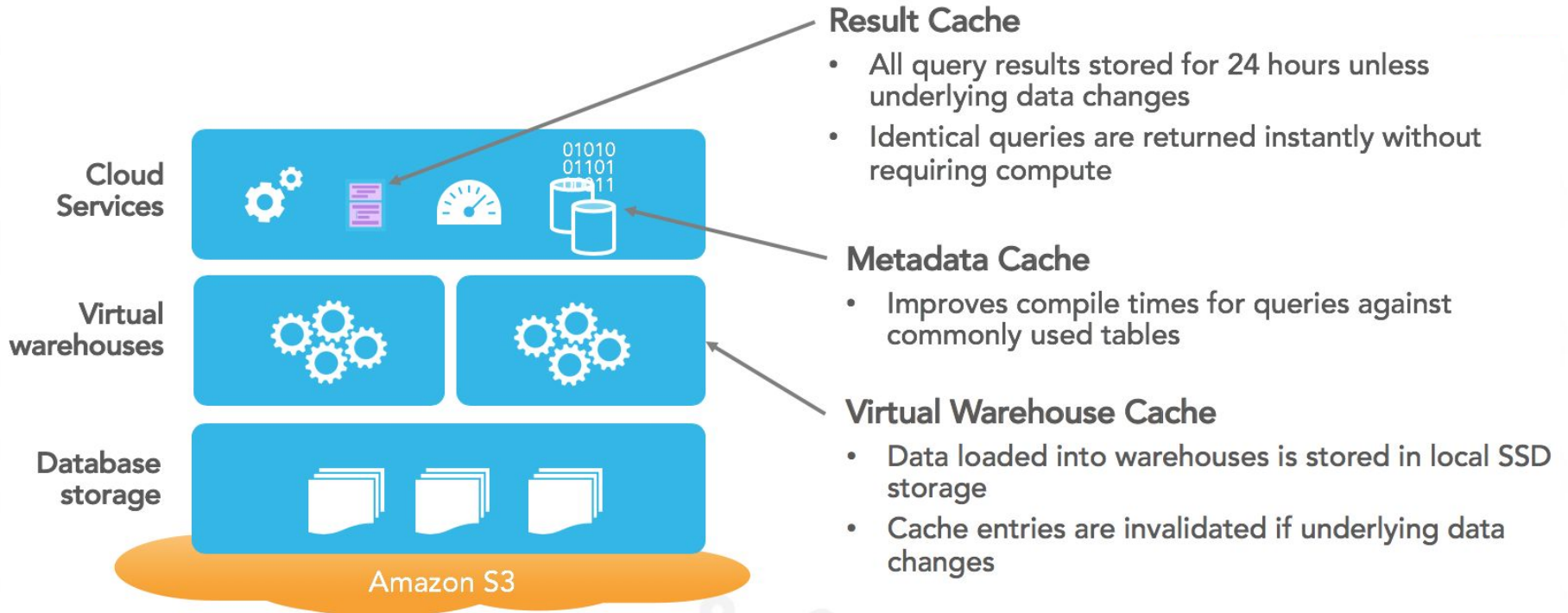
Source: Snowflake

Working with Snowflake

The screenshot shows the Snowflake web interface with several annotations:

- MAIN MENU:** A red box highlights the navigation bar containing icons for Databases, Shares, Data Marketplace, Warehouses, Worksheets, History, and Account.
- Partner Connect:** A red box highlights the 'Partner Connect' button in the top right.
- Snowsight:** A red box highlights the 'Snowsight' button in the top right.
- Annotations:**
 - 'CAN CONNECT THE 3RD PARTY TOOLS FROM HERE' points to the Partner Connect button.
 - 'LATEST VERSION OR UI FOR SNOWFLAKE' points to the Snowsight button.
 - 'DATABASES' points to the 'SNOWFLAKE' database in the left sidebar.
 - 'SCHEMA' points to the 'INFORMATION_SCHEMA' under 'SNOWFLAKE_SAMPLE_DATA'.
 - 'TABLES' points to the 'DAILY_14_TOTAL' table under the 'Tables' section.
 - 'RESULTS OF THE QUERY CAN BE SEEN HERE' points to the 'Results' tab.
 - 'CAN PREVIEW THE DATA WITHIN TABLE FOR BETTER UNDERSTANDING' points to the 'Data Preview' tab.
 - 'CAN CHANGE ROLE, WAREHOUSE, DATABASE & SCHEMA FROM HERE' points to the configuration modal.
 - 'CAN SEE THE QUERIES GOT EXECUTED IN HHISTORY' points to the 'Open History' button.
- Configuration Modal:** A modal window is open showing settings for 'ACCOUNTADMIN' role, 'TEST_WAREHOUSE (XS)' warehouse (Suspended), and 'Schema'.
- Query Results:** The main area shows a query ID '1' and a 'Results' tab. Below the tabs, it says 'Query results will appear here.'

Data Caching in Snowflake



Source: Snowflake

Connect on LinkedIn!!

Here are the details -

- <https://in.linkedin.com/company/datacouch>
- <https://www.linkedin.com/in/bhavuk-chawla/>



Thank You!

Q&A



<https://www.meetup.com/all-things-data/>



<https://www.linkedin.com/company/datacouch>



<https://www.youtube.com/c/datacouch>