Recap :

Architecture -3 layers

Data storage

Compute

Metadata -cloud service layer

DB -sch-tables

Stages ; user stage ; single user and multiple tables

Table stage : multiple user and single table

Named stage ; internal and external

Multiple user and multiple table

Aws s3 : bucket → container

Object store → 5Tb / unlimited objects

Version →

External stage ; access id and secret key

1---Superstore.csv --copy ---snowtable

new-- Superstore.csv

Eventually consistency :

Copy :

Storage class ; type data/ usage data

Same data ---> Emp.csv → location

2021/jul/21/

====================================

Types ;

Temp trans perm

session/user durable durable

Retention time

0/1 0/1 0/1/90

Failsafe-7 days

=================================

cdp===> continuous data protection

Security :

Undrop

Time travel /failsafe:

1. Timestamp --at /before
2. Offset -- 5 mins / 1 day --300 seconds
3. Statement -query id →

==================================

Cloning ⇒ zero copy cloning

CTAS →

1. Like → structure
2. Clone -duplicate
3. As -duplicate

Dev test prod

Emp 5gb -- > empdup 5gb ⇒ Emp 5gb

--only one store →

emp1 --reference to the micropartition base table

Emp2 -

Dev Testing production

Till cloning

base\_Sales sales\_c1 sales\_c2

1tb --2tb Metadata -500mb metadata

Sales ===>metadata => Micropartition

New mp

================================

base clone1 clone2

Cloning time :

100 100 100

Mp1 mp1 mp1

Base:

120 Mp1 Mp1

MP2

MP1

Clone 1

120 125 130

MP2 MP3 /mp1 Mp1

Cloud services layer :

Base -- **mp2**

Clone -- mp3

Clone - mp4 --

=================================

Base table restore

1--------1000 -snapshot 1---1000

2 4

============================

Table 1 --retention 10 days

Increase : 10 days -- + 10

Decrease : 5 days ---> 5 -- failsafe=> snowflake

==================================

Fast queries ⇒

Remote ===Table data == micropartition

Table ⇒ partition --10 mp

Warehouse ===> compute ==SSD cache /local

Nodes

MPP

**Cloud service layer → statistics**

Min

Max

Distinct

Count

Avg

current \_date()

Show databases;

Show tables;

Cache ====>

==============================

Warehouse : EC2 -HDD -memory -OS

Compute :

Query processing :

Multiple vir warehouse :

Data loading

Query --basics /complex

Reporting

======================================

Caching :

4 7 2 3 1

1 2 3 4 7 storage --cache

Warehouse local ssd cache ;

Remote ----stored --in local

Valid : warehouse resize/modified/suspended

Changes in underlying table

-- local cache get cleared

14 partition (remote) ---- 14 local cache ⇒

linenumber=1 (local)

linenumber=2 (local)

2.Second query -executed after warehouse auto suspend

Linenumber =3 ===no data local cache===hit remote

14 partition (remote) ---- 5 local cache ⇒

3. linenumber==4 == local cache ==5 partition

== 9 p remote

Get both remote and local

=====================================

Query result ⇒ cache ==24 hrs (no changes in table)

28/7 == first query execution

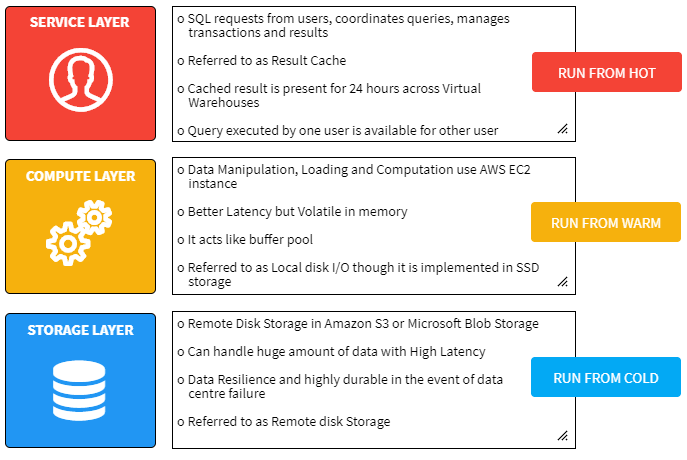
6 am ========

Second time 9pm==== ======

First expiry time : 29/7 6 am

updated expiry time : 29/7 9 pm

===till 30 days ==after ---> query result fully flushed



<https://sonra.io/snowflake/deep-dive-on-caching-in-snowflake/>