Recap /Quiz :

Mp → 1----------100

Mp1 mp2 mp3

Zonemap:

Header:

Range : min -max

Distinct

Null

----------------clustering key ===

Sorted data → column pruning / micropartition scan

Reclustering ---->old micro partition --new micropartition

Snowflake -->reclustering ⇒ manual reclustering --deprecated

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

70million

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

Depth more =less performance

Depth is near to 1 ===good to go

Change the clusterkeys after checking the depth before executing the frequent used

Cluster key ⇒ business →

Table creation ⇒ natural key =>

Overwrite natural column by user defined cluster key

Alter cluster key(line\_item) === >not immediate

Enabling =>Automatic clustering ⇒

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

Data loading ;

Column number ⇒

Staging ==$1,$2

Emp pattern

Fileformat

Query for staged files:

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

Data loading :

Copy

1-----------------------90

10------------------- 100

Temp db . temp schema temp table

Trans db trans schema trans table

Perm db perm schema perm table

Retention time ⇒ parent ⇒ child objects

Individual child object level retention is not modified if you apply at parent

Sales db --10

Sch1 -20

Tabl1 -20

Tabl2 -5

21/7--superstore== 26/7 =RT => 5

TT -- 31/7=10 -1/8 --5/8

Time travel ; query table --historical

Sch2 -5

Data storage :

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

Cache ==query /compute

2tb ------2tb

-zero copy cloning ⇒ no storage , only once

Modified data ===new micro partition

Unmodified data ==only once store/

After clone ---> individual

Base table dev ----test clone ---prod clone → prodclone1

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

First query 6am 30/7 -----------8pm ------------------- 6am expiry ---8 pm31/st

Max : 31 days--purged

--reusing --> query result cache stored max 30 days

Reuse ⇒ Identical query without any change in the underlying table

Remote : full partition ==> local ⇒ query

Remote : partially partition ⇒ local partially

Query : local + remote

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

Semi structured ; json xml orc parquet avro

Data types : variant , object , array

Max : 16 mb : src variant

Src:salesperson.name

Src [salesperson][name]

Array;

[0]

Column : case insensitive

Element :

------------file format

Encoding file format :

Compression codec : gzip , deflate , snappy

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

File ---zip /gzip /7zip

Compression column ;

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

Managing snowflake account :

Enterprise . business critical and vps

Url : organization ==licensed version

Aws : acountname-region.

account-azure-us-east-1 .snowflake

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

Dw ---> ELT ⇒ extract --snowflake ---> transform

Data pipeline : ETL : informatica /talend

Oracle =======transform =========snowflake

Dbt ---> transform ==less transform

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

Streams and task

Data security

Power bi /integration

Talend Integration

Best practices ;

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

Continuous Data Pipelines :

Batch →collected and store --consumed →

extract → time --batch job

Streaming → process of collecting continuous data

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

Snowflake Streams :

CDC : change data capture :

1------1000

Ins-2

Delete -5

Update -3

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

Employee ===>raw employee ⇒

dept--------------->raw dept =====> curated ===> consumption

Scheduler ⇒ cron tab ⇒ unix shell script

Built -scheduler ⇒ on demand / often run

Control M

Autosys

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

Srcfile → Raw table -streams

1-10 ---> 1-10 1-10 --insert

2 2 2 ---update

1

Streams ---> Curated →

1 --->load 1-10

--updated 2

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

Streams

operation metadata$action metadata$isupdate

Insert insert false

Update

Insert true

Delete