Snowflake cloud data-warehouse

“Simplicity is the ultimate sophistication”

Leonardo da Vinci

Snowflake Cloud Data warehouse

Day-5

Snowflake Database Objects

VIEWS

- Standard View
- Secured View
- Materialized View

Snowflake Database Objects- Materialized View

When to Create MVs

- ✓ Underlying data is stable
- ✓ MVs used frequently
- ✓ Underlying query is expensive

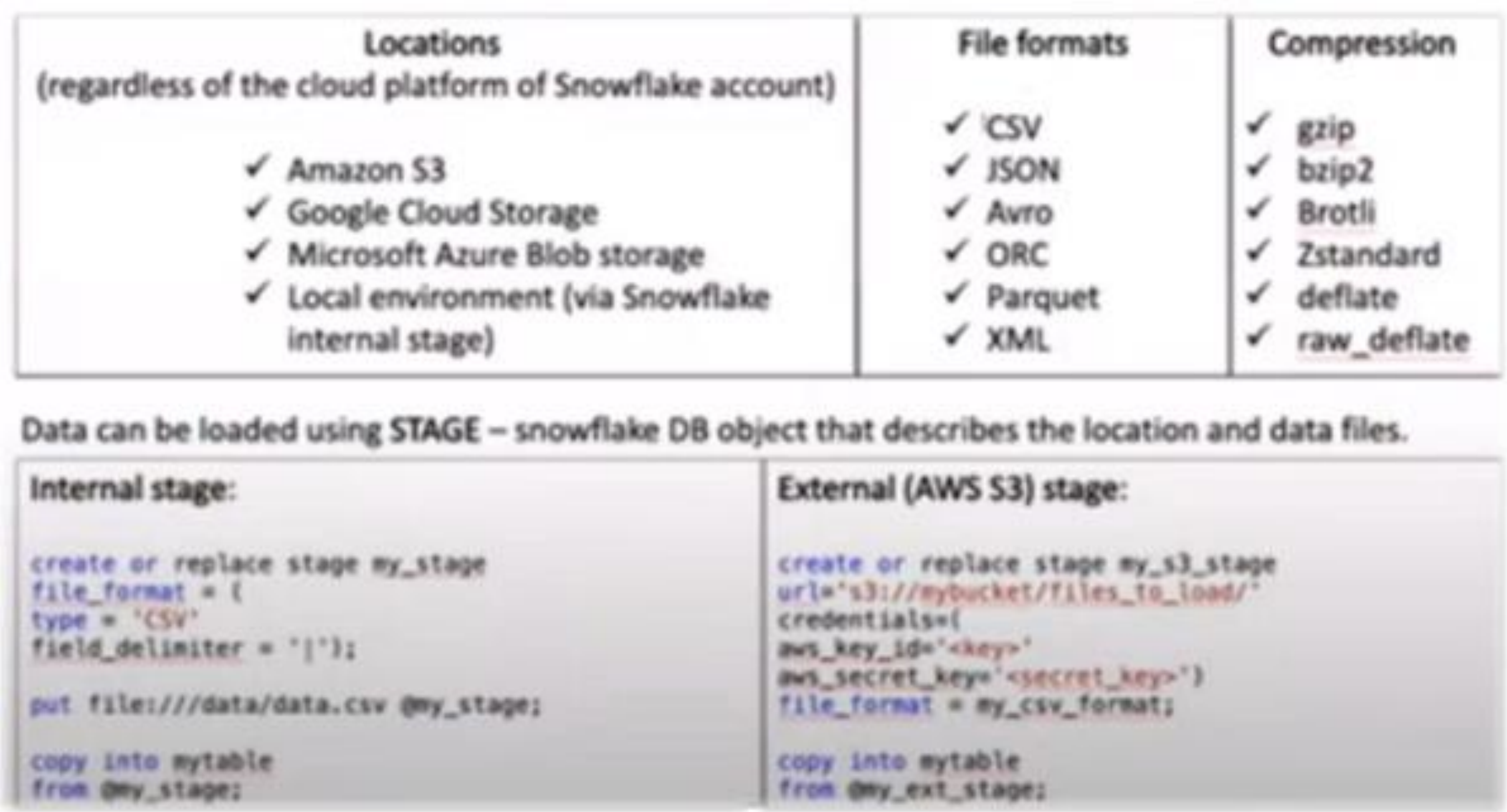
Limitations

- ❑ Only available on the Snowflake Enterprise Edition & above
- ❑ Can query only a single table, Self Join also not supported
- ❑ MV Query can not query another MV, View
- ❑ Limited Query Support, Can not include ORDER BY, HAVING, LIMIT etc.
- ❑ Can not control Refresh rate
- ❑ Not monitored by Resource Monitor
- ❑ DML Operations NOT Allowed
- ❑ Time Travel NOT Allowed

Snowflake Database Objects

VIEW Creation Demo

Load into Snowflake



Locations (regardless of the cloud platform of Snowflake account)	File formats	Compression
✓ Amazon S3 ✓ Google Cloud Storage ✓ Microsoft Azure Blob storage ✓ Local environment (via Snowflake internal stage)	✓ CSV ✓ JSON ✓ Avro ✓ ORC ✓ Parquet ✓ XML	✓ gzip ✓ bzip2 ✓ Brotli ✓ Zstandard ✓ deflate ✓ raw_deflate

Data can be loaded using **STAGE** – snowflake DB object that describes the location and data files.

Internal stage:

```
create or replace stage my_stage
file_format = (
type = 'CSV'
field_delimiter = '|');

put file:///data/data.csv @my_stage;

copy into mytable
from @my_stage;
```

External (AWS S3) stage:

```
create or replace stage my_s3_stage
url='s3://mybucket/files_to_load/'
credentials=(
aws_key_id='<key>'
aws_secret_key='<secret_key>')
file_format = my_csv_format;

copy into mytable
from @my_ext_stage;
```

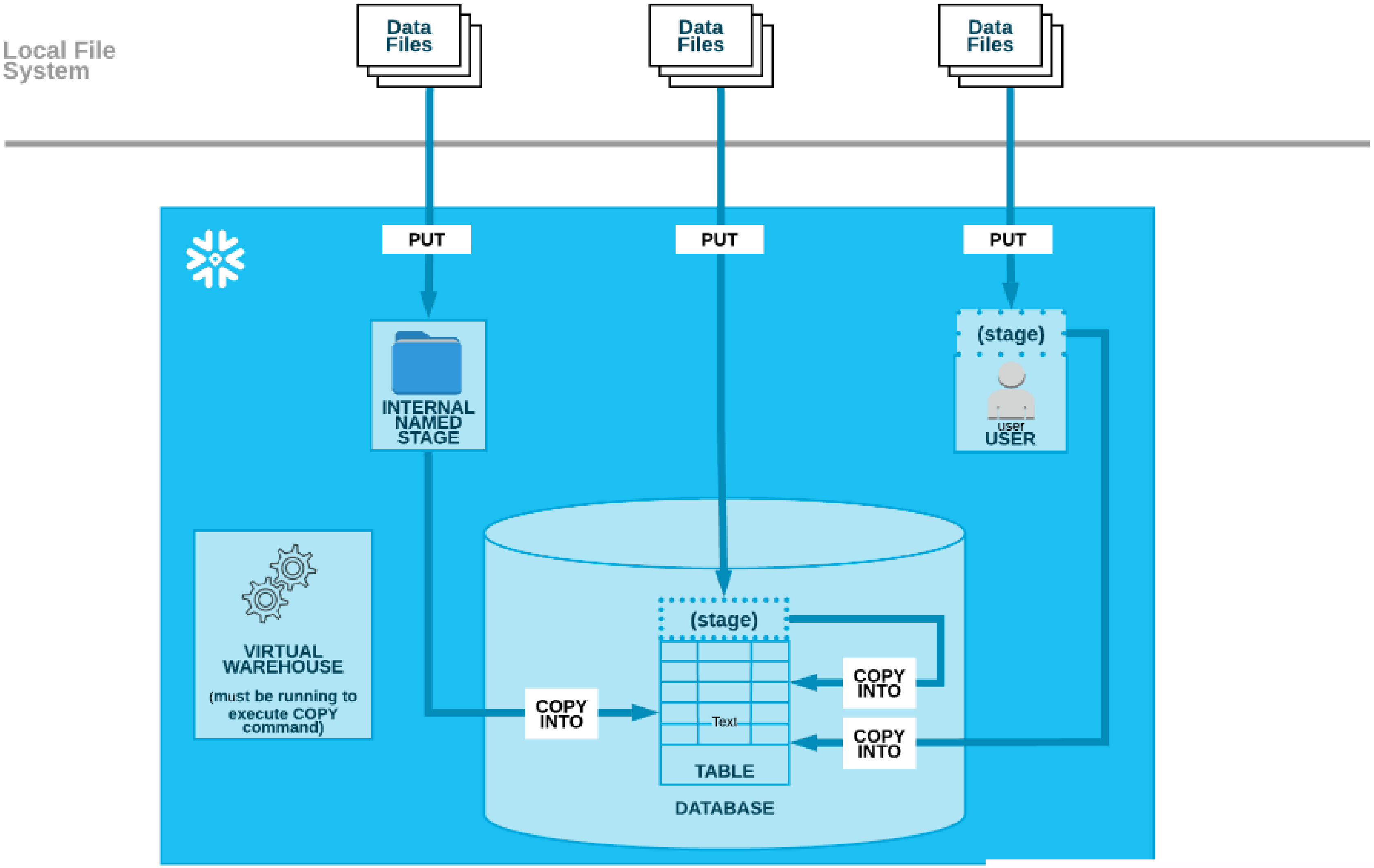
Snowflake Internal Stage

- ❖Internal Named Stage @
- ❖Table Stage @%
- ❖User Stage @~

LOAD into Snowflake

Supported File Formats

- ❖ CSV
- ❖ JSON
- ❖ XML
- ❖ ORC
- ❖ Parquet
- ❖ AVRO



Load Into Snowflake

PUT Command

- Uploads Data into Snowflake Internal Stage
  - Named Internal Stage
  - Internal Table Stage
  - Internal Current User Stage
- does not support uploading files to external stages
- GO,.NET, Node.js Driver not supported
- 16-64 MB limit for parallel file (individual) upload
- More than 64 MB will be sliced into smaller chunks
- Metadata of loaded files stored for 64 days
- Can not be executed from WEBUI-Worksheets
- Ignores Duplicate files automatically for same stage
- Encrypts files as per Configuration

Load Into Snowflake

PUT Command

```
PUT file://<path_to_file>/<filename> @<Internal Stage Path>
    [ PARALLEL = <4> ]
    [ AUTO_COMPRESS = TRUE | FALSE ]
    [ SOURCE_COMPRESSION = AUTO_DETECT | GZIP | BZ2 | BROTLI | ZSTD | DEFLATE | RAW_DEFLATE | NONE ]
    [ OVERWRITE = TRUE | FALSE ]
```

Load Into Snowflake

FILE FORMAT

```
CREATE [ OR REPLACE ] FILE FORMAT [ IF NOT EXISTS ] <name>
                    TYPE = { CSV | JSON | AVRO | ORC | PARQUET | XML } [ formatTypeOptions ]
                                        [ COMMENT = '<string_literal>' ]
```

```
[formatTypeOptions ] for CSV
COMPRESSION = AUTO | GZIP | BZ2 | BROTLI | ZSTD | DEFLATE | RAW_DEFLATE | NONE
RECORD_DELIMITER = '<character>' | NONE
FIELD_DELIMITER = '<character>' | NONE
SKIP_HEADER = <integer>
SKIP_BLANK_LINES = TRUE | FALSE
DATE_FORMAT = '<string>' | AUTO
TRIM_SPACE = TRUE | FALSE
ERROR_ON_COLUMN_COUNT_MISMATCH = TRUE | FALSE
FIELD_OPTIONALLY_ENCLOSED_BY = '<character>' | NONE
```

Load Into Snowflake

COPY Command

```
 COPY INTO [<namespace>.]<table_name> [ ( <col_name> [ , <col_name> ... ] ) ]
    FROM ( SELECT [<alias>.]$<file_col_num>[.<element>] [ , [<alias>.]$<file_col_num>[.<element>] ... ]
           FROM { internalStage | externalStage } )
[ PATTERN = '<regex_pattern>' ]
[ FILE_FORMAT = ( { FORMAT_NAME = '[<namespace>.]<file_format_name>' |
                   TYPE = { CSV | JSON | AVRO | ORC | PARQUET | XML } [ formatTypeOptions ] } ) ]
[ copyOptions ]
[ VALIDATION_MODE = RETURN_<n>_ROWS | RETURN_ERRORS | RETURN_ALL_ERRORS ]
```

```
[ copyOptions ]
ON_ERROR = { CONTINUE | SKIP_FILE | SKIP_FILE_<num> | SKIP_FILE_<num>% | ABORT_STATEMENT }
PURGE = TRUE | FALSE
TRUNCATECOLUMNS = TRUE | FALSE
FORCE = TRUE | FALSE
```

Load Into Snowflake- Load Best Practices

Points to consider while loading Data

- ❑ **Produce Data files between ~100 MB to 250 MB Compressed size**
- ❑ **Maintain UTF-8 Character set wherever possible**
- ❑ **Maintain consistent column count for each row for delimited file**
- ❑ **Avoid NULL values in semi-structured data, USE STRIP_NULL_VALUES if NULL value remains**
- ❑ **Avoid having multiple data type for single column e.g. Value: “10” != Value:10**
- ❑ **Organize Data files by Path**
- ❑ **Metadata of loaded files stored for 64 days**

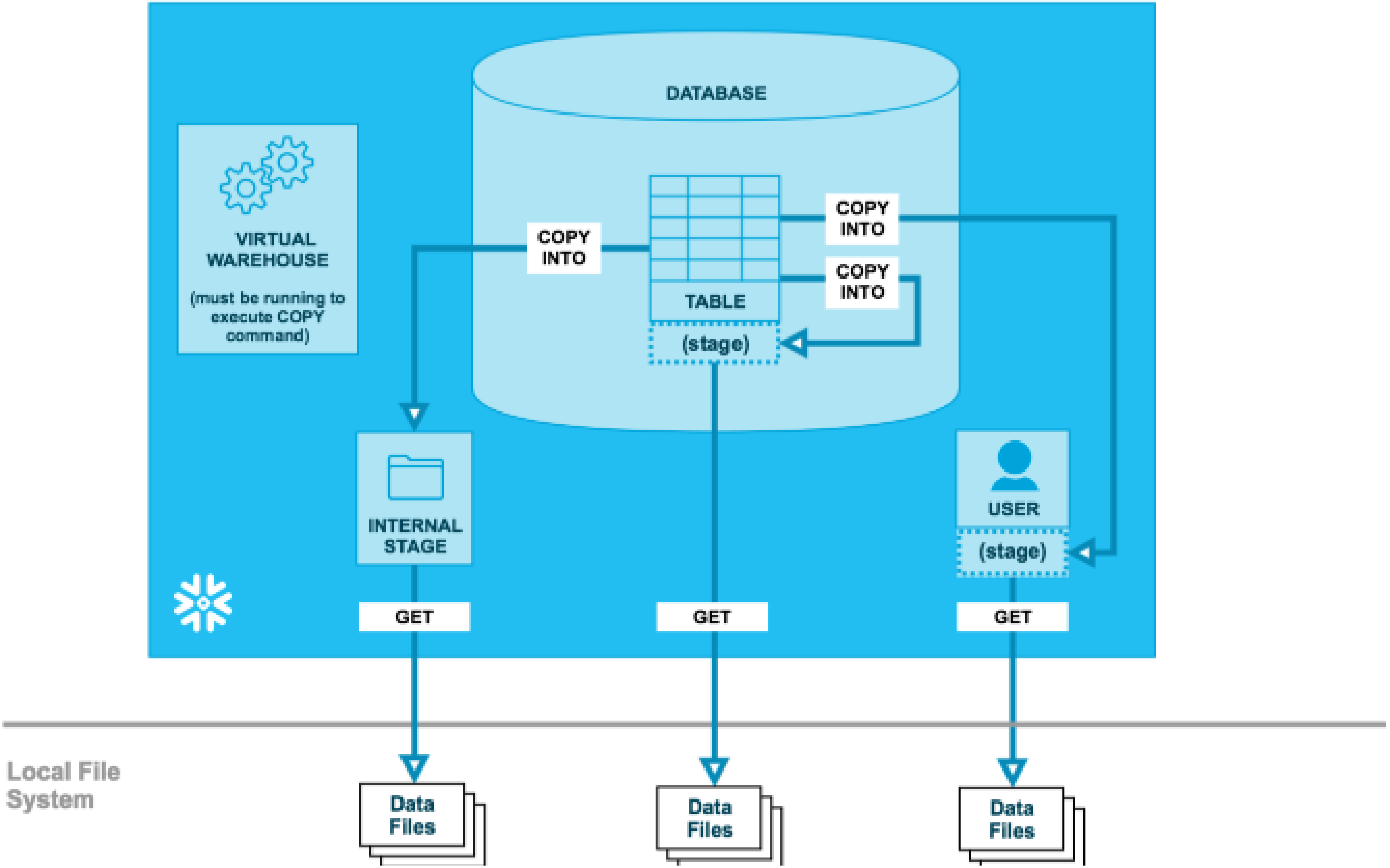
Load Into Snowflake - LAB

Load CSV Files

Load Into Snowflake - LAB

Load JSON Files

Unload from Snowflake



Unload from Snowflake

COPY Command

```
COPY INTO { internalStage | externalStage | externalLocation }
     FROM { [<namespace>.]<table_name> | ( <query> ) }
[ FILE_FORMAT = ( { FORMAT_NAME = '[<namespace>.]<file_format_name>' |
                    TYPE = { CSV | JSON | PARQUET } [ formatTypeOptions ] } ) ]
[ copyOptions ]
[ HEADER ]
```

[copyOptions]

OVERWRITE = TRUE | FALSE

SINGLE = TRUE | FALSE

MAX_FILE_SIZE = <num>

INCLUDE_QUERY_ID = TRUE | FALSE

~~DETAILED_OUTPUT = TRUE | FALSE~~

Unload from Snowflake

Points to consider while Unloading

❑ **File format supported are Delimited, JSON, Parquet**

❑ **Compression method supported Gzip, bzip2, Brotli & Zstandard**

❑ **File unloaded in Internal Stage always encrypted with 128- or 256-bit keys**

❑ **File unloaded in external stage is encrypted with User-Supplied Keys**

❑ **File unloaded are always encoded with UTF8 character set**

❑ **Default File Size is 16 MB**

❑ **Follow Warehouse size recommendation i.e. small VW = small parallel thread count**

GET Command

```
GET internalStage file://<path_to_file>/<filename>
    [ PARALLEL = <10> ]
    [ PATTERN = '<regex_pattern>'' ]
```

Unload from Snowflake

GET Command

- ❑ Downloads Data into Snowflake Internal Stage
    - ❖ Named Internal Stage
    - ❖ Internal Table Stage
    - ❖ Internal Current User Stage
- ❑ GO,.NET, Node.js Driver not supported
- ❑ Downloaded files are automatically decrypted
- ❑ Does not support downloading files to external stages
- ❑ Can not be executed from WEBUI-Worksheets
- ❑ Can not rename files
- ❑ Maximum of 99 parallel thread supported

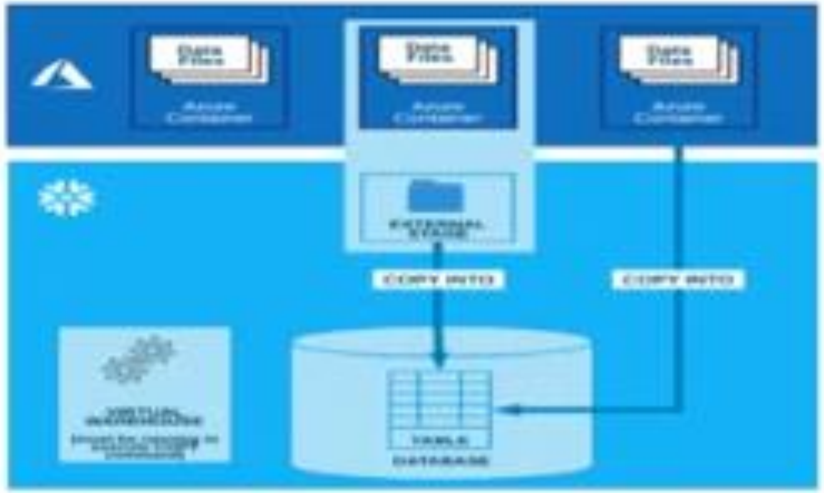
Unload from Snowflake - LAB

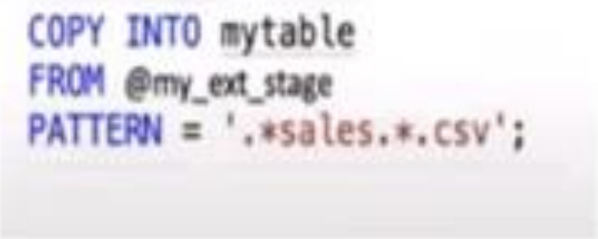
Unload CSV File

Unload from Snowflake - LAB

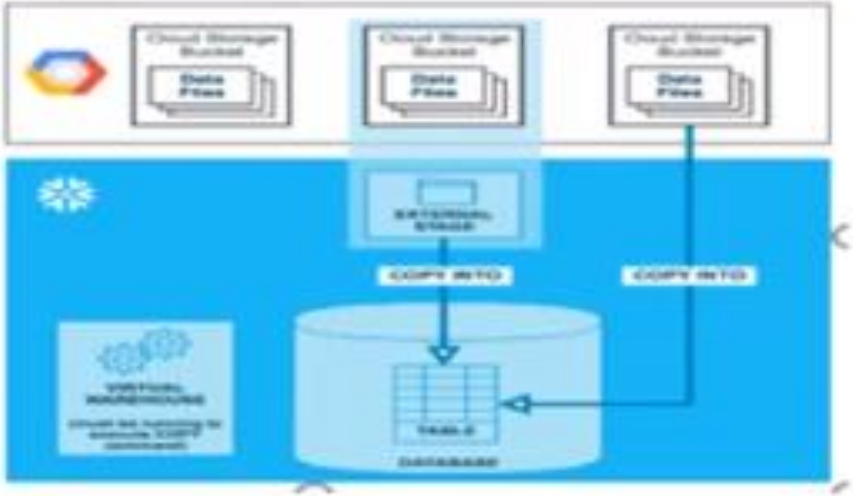
Unload JSON

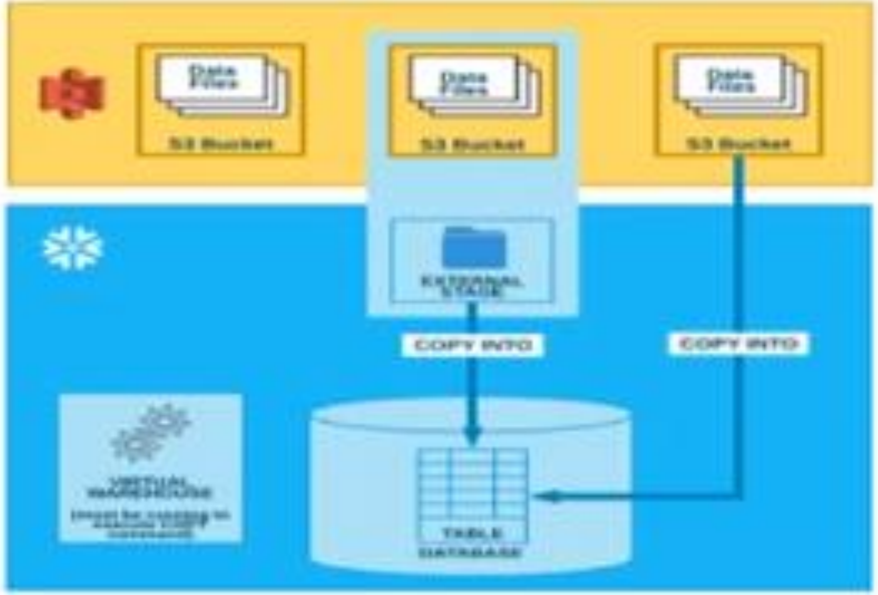
Load Data from External Stage





```
COPY INTO mytable
FROM @my_ext_stage
PATTERN = '.*sales.*.csv';
```


Cloud Storage Bucket
Data Files
Cloud Storage Bucket
Data Files
Cloud Storage Bucket
Data Files
EXTERNAL STAGE
COPY INTO
COPY INTO
VIRTUAL WAREHOUSE
(must be running to execute COPY command)
TABLE
DATABASE


Data Files
S3 Bucket
Data Files
S3 Bucket
Data Files
S3 Bucket
EXTERNAL STAGE
COPY INTO
COPY INTO
VIRTUAL WAREHOUSE
(must be running to execute COPY command)
TABLE
DATABASE