Different Date Formats

DD-MM-YYYY

24-12-2021

MM-DD-YYYY

DD\MM\YYYY

24\12\2021

DD\MM\YY

24\12\21

Through Wizard we can load only load files that are less than 50 MB in size

FILE FORMATS

A named file format object describes and stores the format information required to load data into Snowflake tables. You can specify different parameters, for example, the delimiter of the file, whether you want to skip the header or not…



Some supported File Formats in Snowflake .

Snowflake supports both Structured and Semi-Structured Data, so just as an example, you can store JSON files into tables. Let’s see the differences between them:

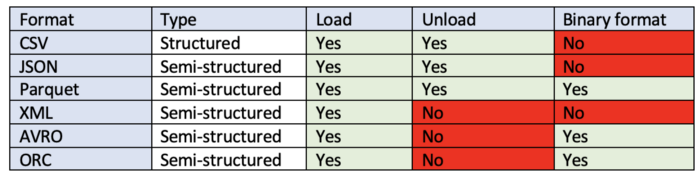
**Structured Data**

* **CSV** → You can Load and Unload files in CSV format. It’s the fastest file format to load data.

**Semi-structured Data**

**Semi-structured data is saved as Variant type in Snowflake tables, and it can be queried using JSON notation**. You can store arrays, objects… Non-native values, such as dates and timestamps, are stored as strings when loaded into a VARIANT column.

* **JSON** → It’s used for both loading & unloading data.
* **Parquet** → Binary format used for both loading & unloading data.
* **XML** → You can only load data in Snowflake using the XML format.
* **Avro** → Binary format used to load data.
* **ORC** → Binary format used to load data.



Different file formats in Snowflake

Supported File Formats and Encoding

Snowflake supports most of the common file formats used for loading data. These file formats include:

• Delimited files (any valid delimiter is supported; the default is a comma)

• JSON and XML

• Avro, including the automatic detection and processing of staged Avro files that were compressed using Snappy (Snappy (previously known as Zippy) is a fast data compression and decompression library)

• ORC, including the automatic detection and processing of staged ORC files that were compressed using Snappy or zlib

• Parquet, including the automatic detection and processing of staged Parquet files that were compressed using Snappy

For delimited files, the default character set is UTF-8. To use any other characters set, you must explicitly specify the encoding to use for loading.

For other supported file formats (JSON, Avro, etc.), the only supported character set is UTF-8.

Note Many character encoding sets are supported for the loading of delimited files.

Snowflake also allows you to configure a file format object for reuse.

It’s useful for formats that are frequently used by many load jobs. Here is an example of a file format named "DEMO\_DB"."PUBLIC".sample\_file\_format.

CREATE FILE FORMAT "DEMO\_DB"."PUBLIC".sample\_file\_format

TYPE = 'CSV' COMPRESSION = 'AUTO' FIELD\_DELIMITER = ',' RECORD\_

DELIMITER = '\n' SKIP\_HEADER = 0 FIELD\_OPTIONALLY\_ENCLOSED\_BY =

'NONE' TRIM\_SPACE = FALSE ERROR\_ON\_COLUMN\_COUNT\_MISMATCH = TRUE

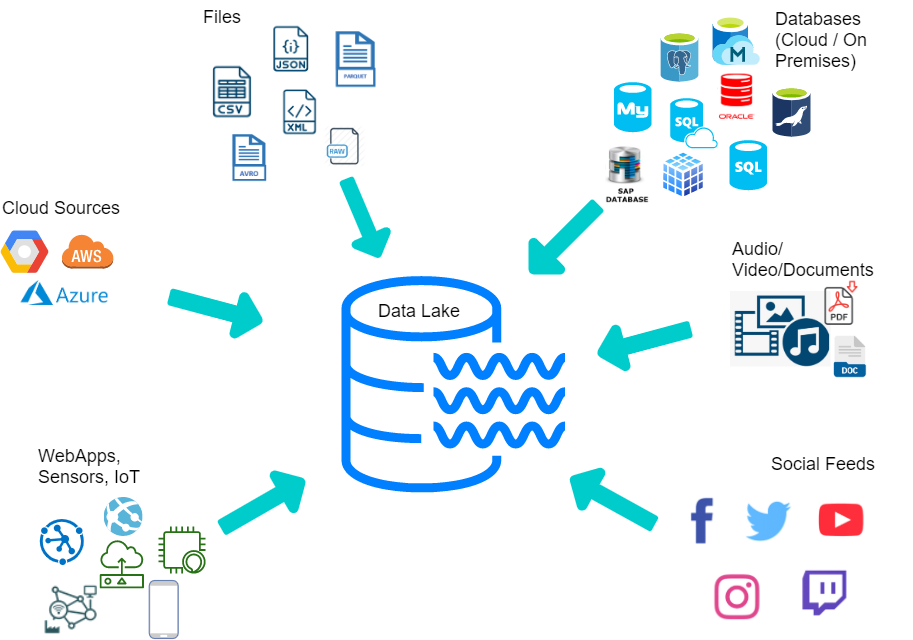
ESCAPE = 'NONE' ESCAPE\_UNENCLOSED\_FIELD = NONE DATE\_FORMAT =

'AUTO' TIMESTAMP\_FORMAT = 'AUTO' NULL\_IF = ('\\N');

The file format object name can then be referenced during loads

**Regular Data flow in Snowflake projects**

**Oracle/SQL Server/MySQL/MongoDB/Applications 🡪 AWS S3 (Datalake) – raw (Taken care by Different Team) --- generate data in CSV files**



Data Lake is a system or **repository** of data stored in its **natural/raw format** either in **structured, semi structured or unstructured** format

Cloud Data Lake is a **cloud-hosted centralized repository** that allows you to store all your structured, semi structured and unstructured data at any scale, typically using an **object store** such as Amazon S3 or Microsoft Azure Data Lake Storage (ADLS), BLOB STORAGE or Google cloud storage

Structured files: Delimited files, CSV (comma separated values), TSV (Tab separated value)..

Semi structured files: JSON, XML, PARQUET, ORC, AVRO..

Unstructured files: Audio files, video files, PDF, Excel, word documents, Images…

**S3->External stage -> copy commands -> Snowflake (Snowflake Developer/Lead’s responsibility)**

**Local drives/servers -> load files to Snowflake Internal stage/User or table stage using PUT command -> copy commands to load data into Snowflake tables**

**Location of data files where data can be loaded from is called stage**

**AWS=> AWS S3 (storage)**

**Azure => Azure Blob storage (storage) or Azure Data Lake storage(ADLS )**

**Google cloud=> Google cloud storage (storage)**