LOADING DATA INTO SNOWFLAKE

- Key concepts related to data loading : Data File, File Formats, Data Stage
- Loading Using the Web Interface (Limited)
- Bulk Loading using COPY INTO command
- Loading Continuously Using Snowpipe

Key concepts related to data loading : Data File, File Formats, Data Stage

Data Loading in SF

Bulk vs Continuous Loading

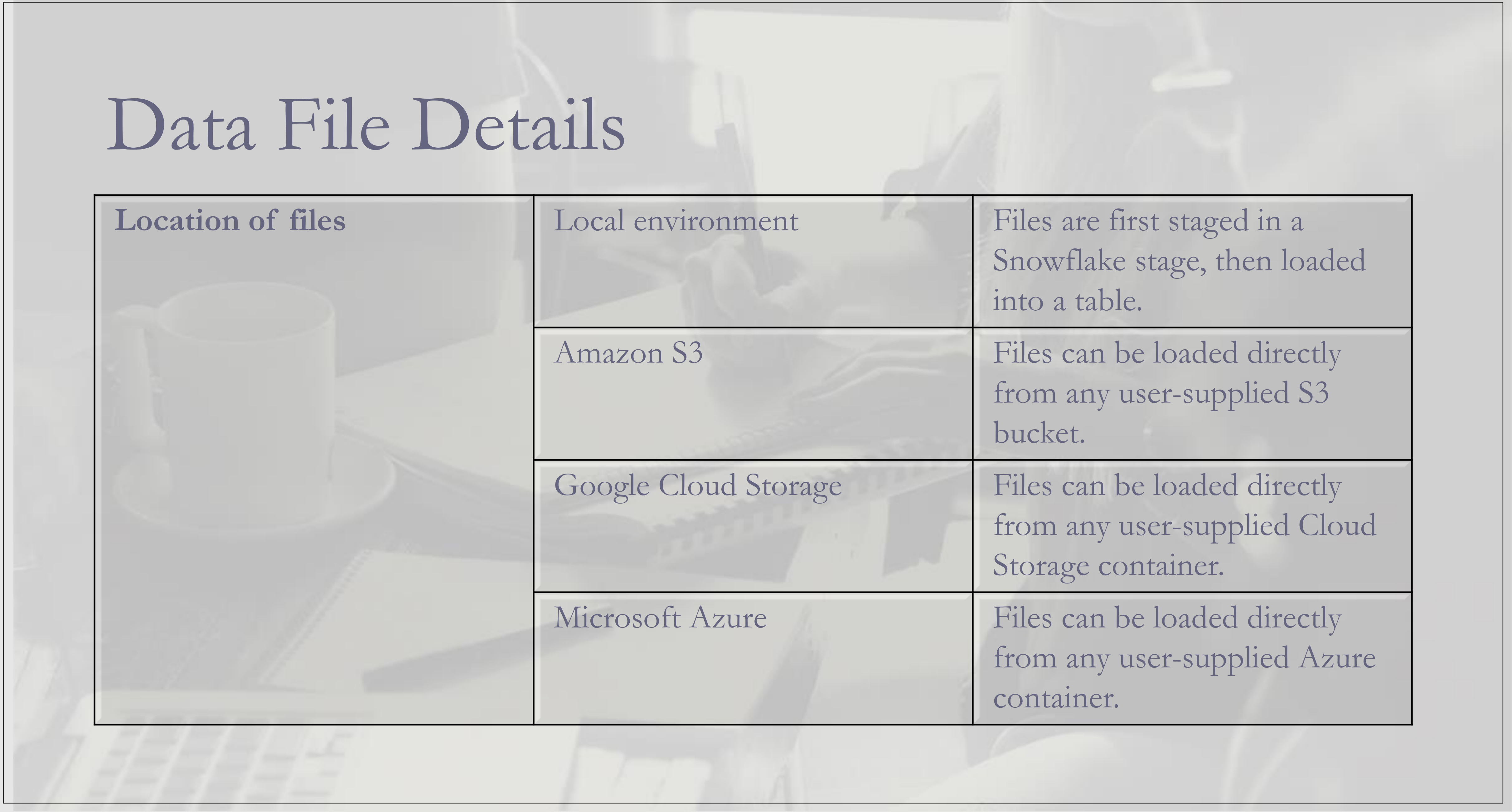
Bulk Loading Using the COPY Command

This option enables loading batches of data from files already available in cloud storage, or copying (i.e. staging) data files from a local machine to an internal (i.e. Snowflake) cloud storage location before loading the data into tables using the COPY command.

Continuous Loading Using Snowpipe

This option enables loading small volumes of data (i.e. micro-batches)

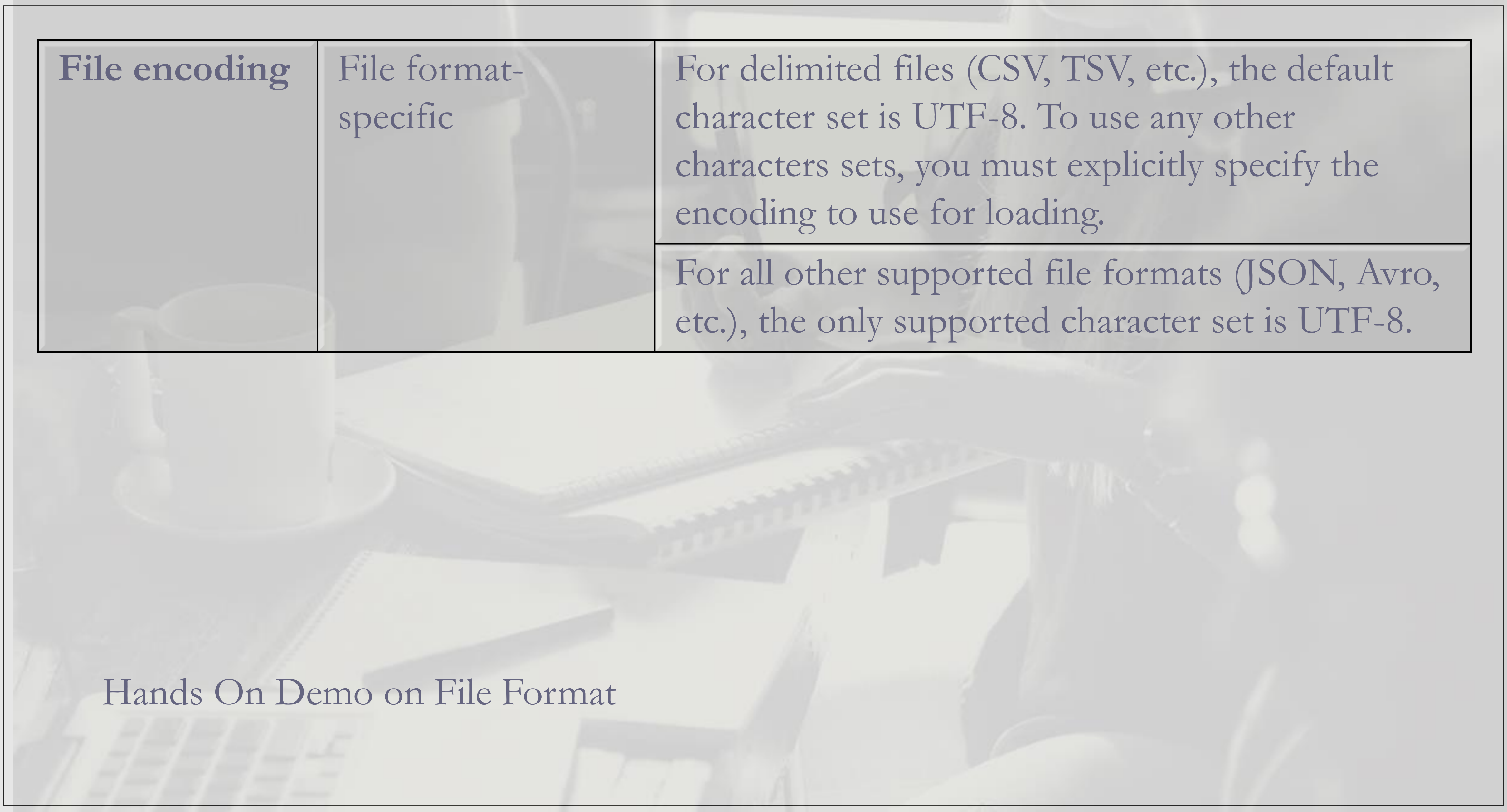
Snowpipe loads data within minutes after files are added to a stage and submitted for ingestion.



Data File Details

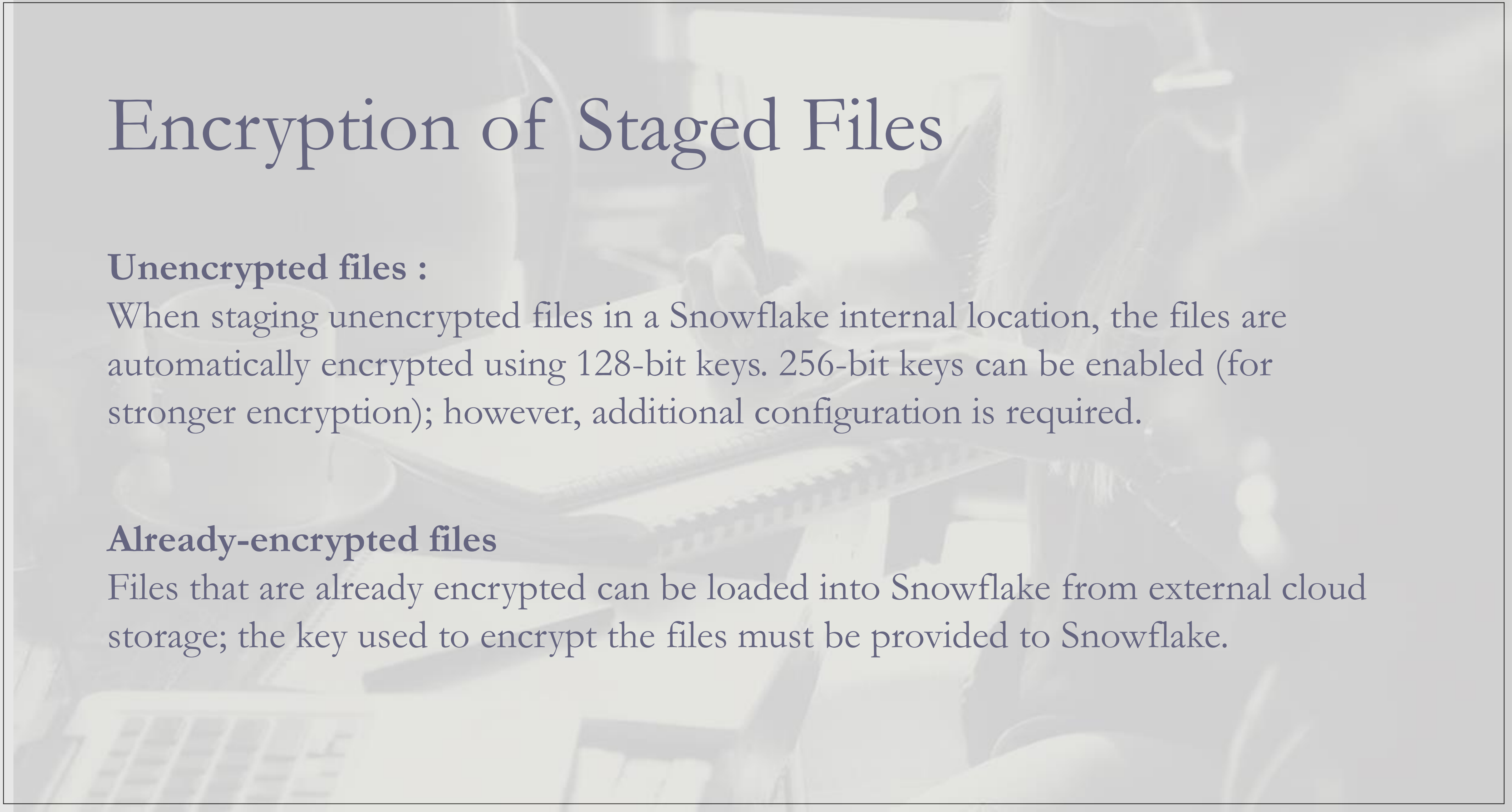
Location of files	Local environment	Files are first staged in a Snowflake stage, then loaded into a table.
	Amazon S3	Files can be loaded directly from any user-supplied S3 bucket.
	Google Cloud Storage	Files can be loaded directly from any user-supplied Cloud Storage container.
	Microsoft Azure	Files can be loaded directly from any user-supplied Azure container.

File formats	Delimited files (CSV, TSV, etc.)	Any valid delimiter is supported; default is comma (i.e. CSV).
	JSON	
	Avro	Includes automatic detection and processing of staged Avro files that were compressed using Snappy.
	ORC	Includes automatic detection and processing of staged ORC files that were compressed using Snappy or zlib.
	Parquet	Includes automatic detection and processing of staged Parquet files that were compressed using Snappy.
	XML	Supported as a preview feature



File encoding	File format-specific	For delimited files (CSV, TSV, etc.), the default character set is UTF-8. To use any other characters sets, you must explicitly specify the encoding to use for loading.
		For all other supported file formats (JSON, Avro, etc.), the only supported character set is UTF-8.

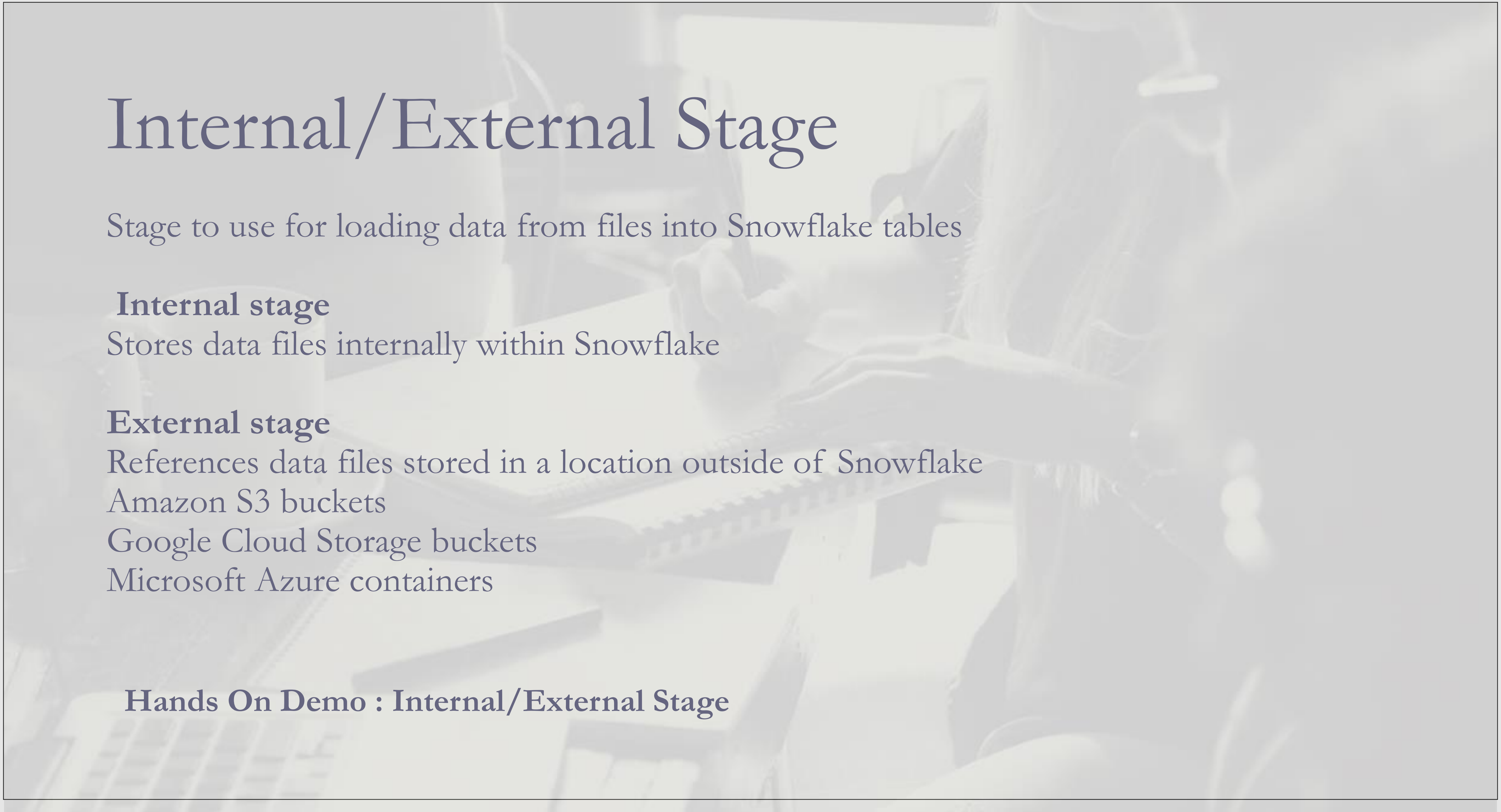
Hands On Demo on File Format



Encryption of Staged Files

Unencrypted files :
When staging unencrypted files in a Snowflake internal location, the files are automatically encrypted using 128-bit keys. 256-bit keys can be enabled (for stronger encryption); however, additional configuration is required.

Already-encrypted files
Files that are already encrypted can be loaded into Snowflake from external cloud storage; the key used to encrypt the files must be provided to Snowflake.



Internal/External Stage

Stage to use for loading data from files into Snowflake tables

Internal stage

Stores data files internally within Snowflake

External stage

References data files stored in a location outside of Snowflake

Amazon S3 buckets

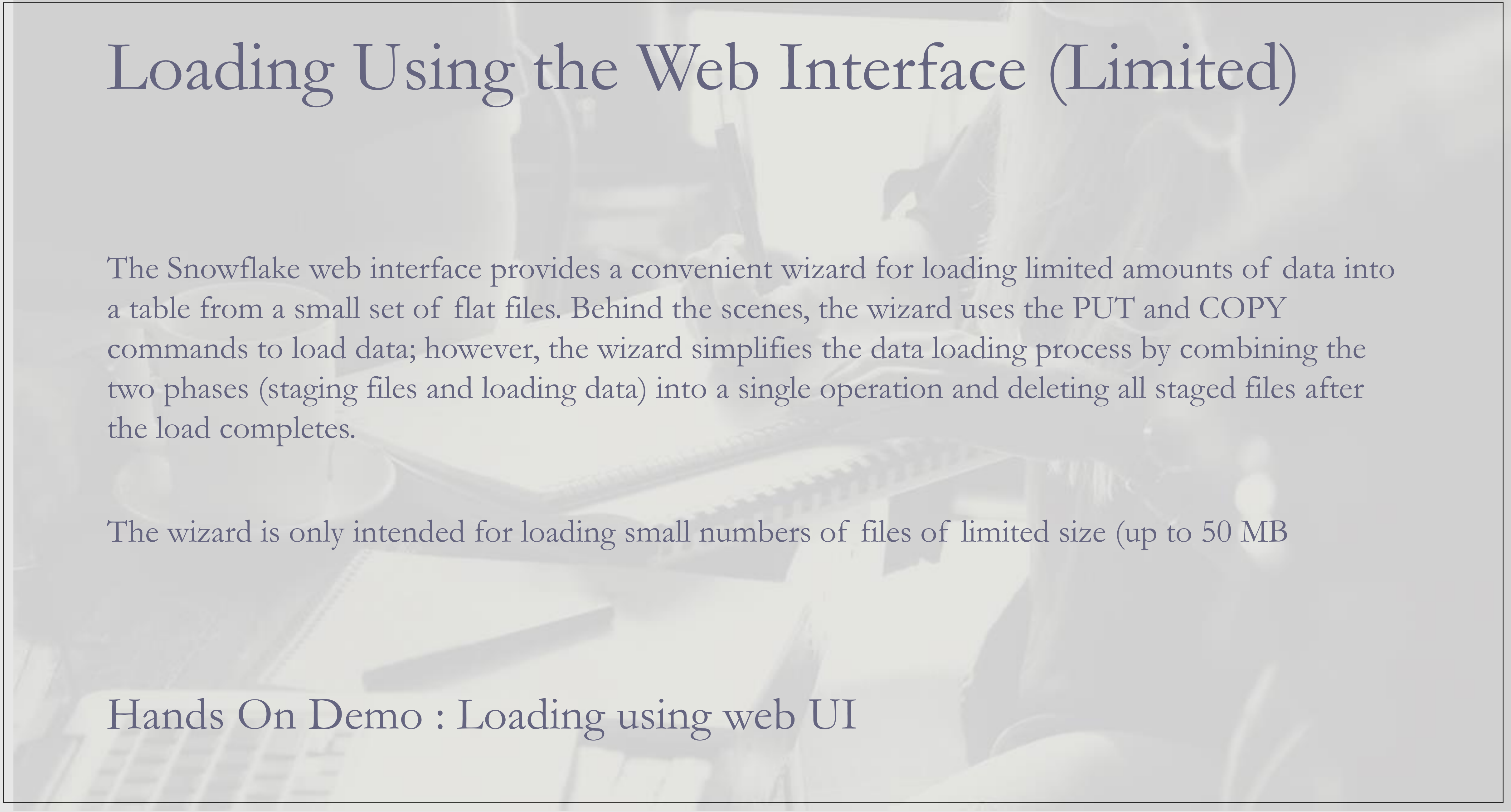
Google Cloud Storage buckets

Microsoft Azure containers

Hands On Demo : Internal/External Stage

Loading Using the Web Interface (Limited)

Loading Using the Web Interface (Limited)

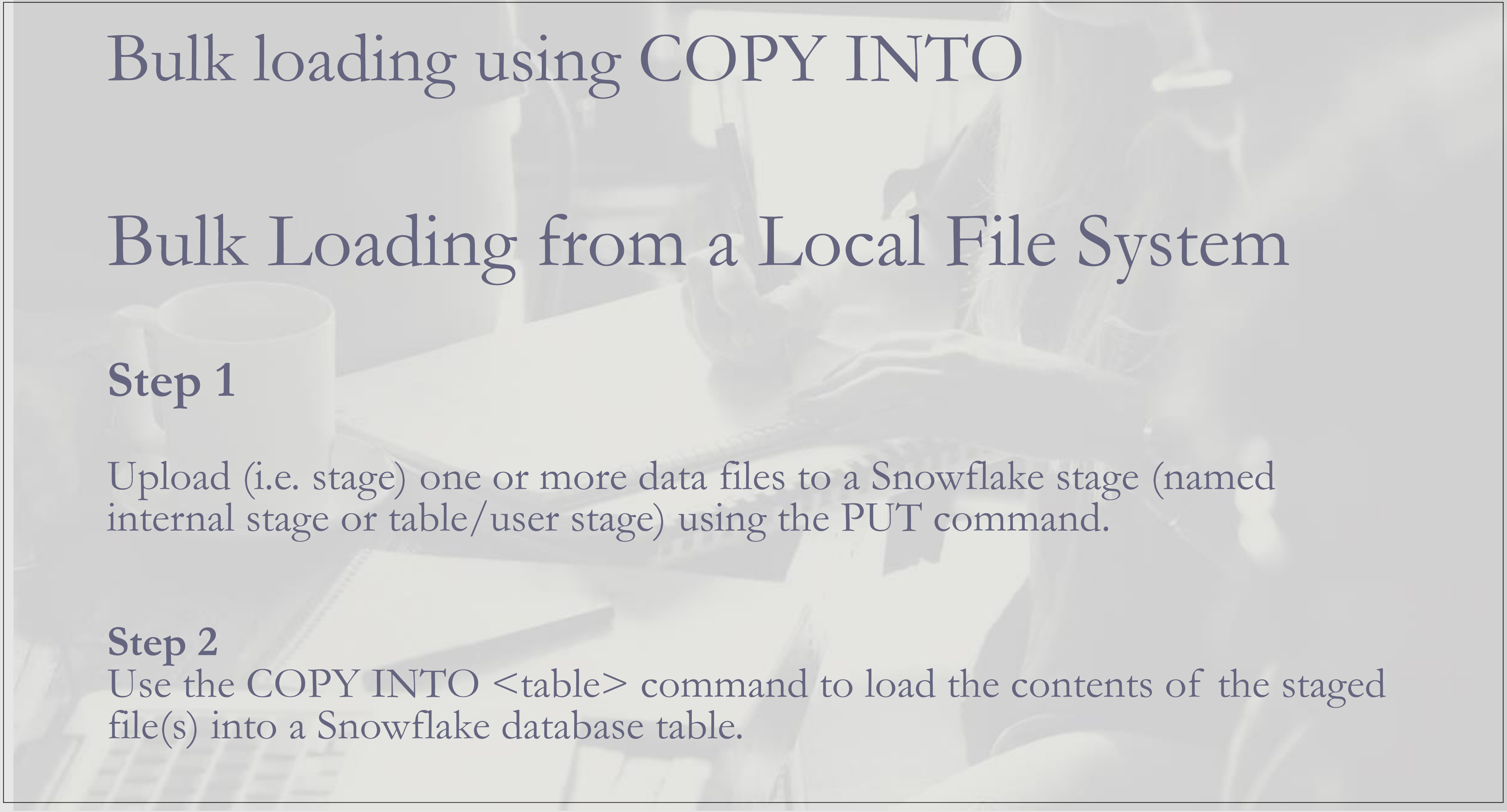


The Snowflake web interface provides a convenient wizard for loading limited amounts of data into a table from a small set of flat files. Behind the scenes, the wizard uses the PUT and COPY commands to load data; however, the wizard simplifies the data loading process by combining the two phases (staging files and loading data) into a single operation and deleting all staged files after the load completes.

The wizard is only intended for loading small numbers of files of limited size (up to 50 MB

Hands On Demo : Loading using web UI

Bulk Loading using COPY INTO command



Bulk loading using COPY INTO

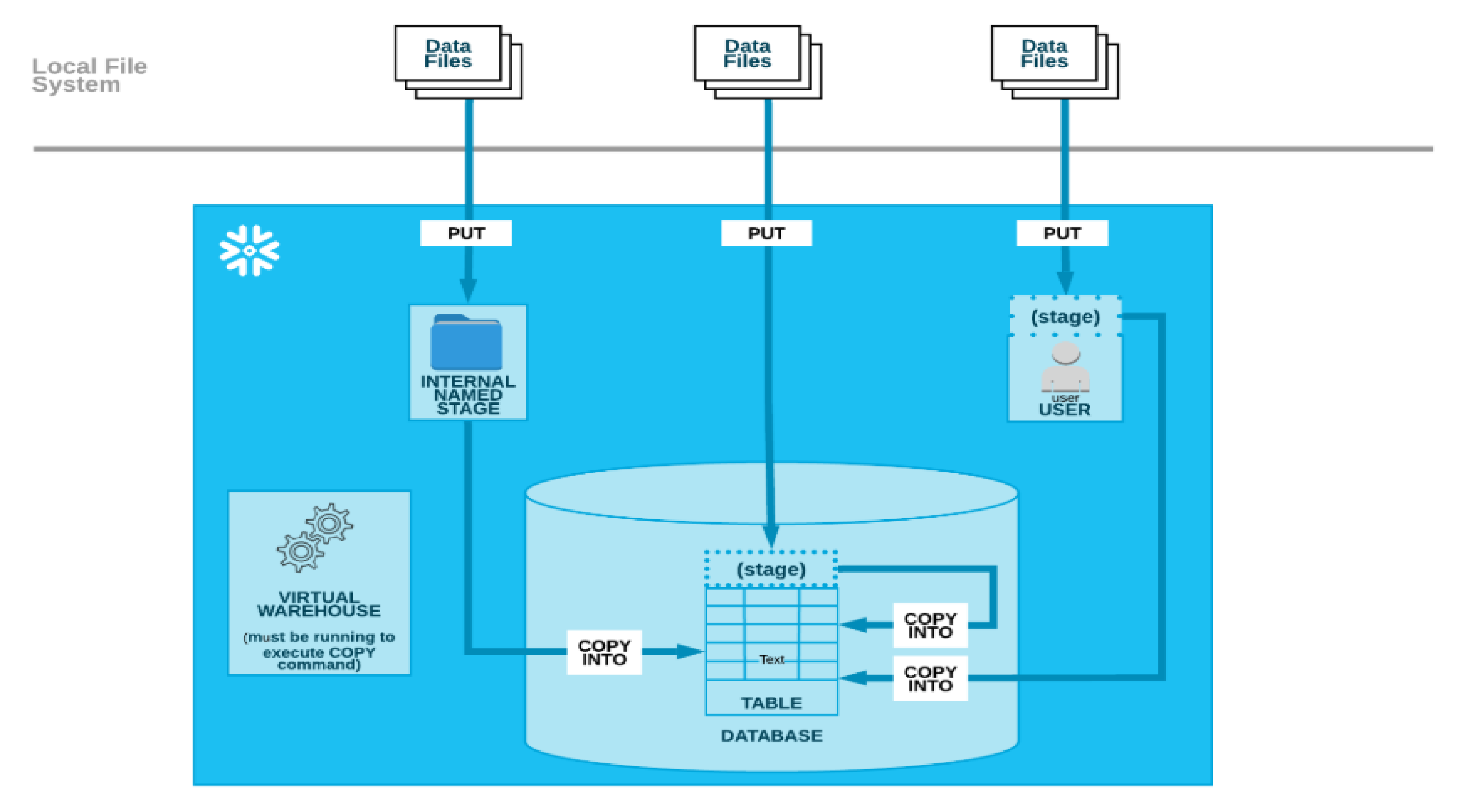
Bulk Loading from a Local File System

Step 1

Upload (i.e. stage) one or more data files to a Snowflake stage (named internal stage or table/user stage) using the PUT command.

Step 2

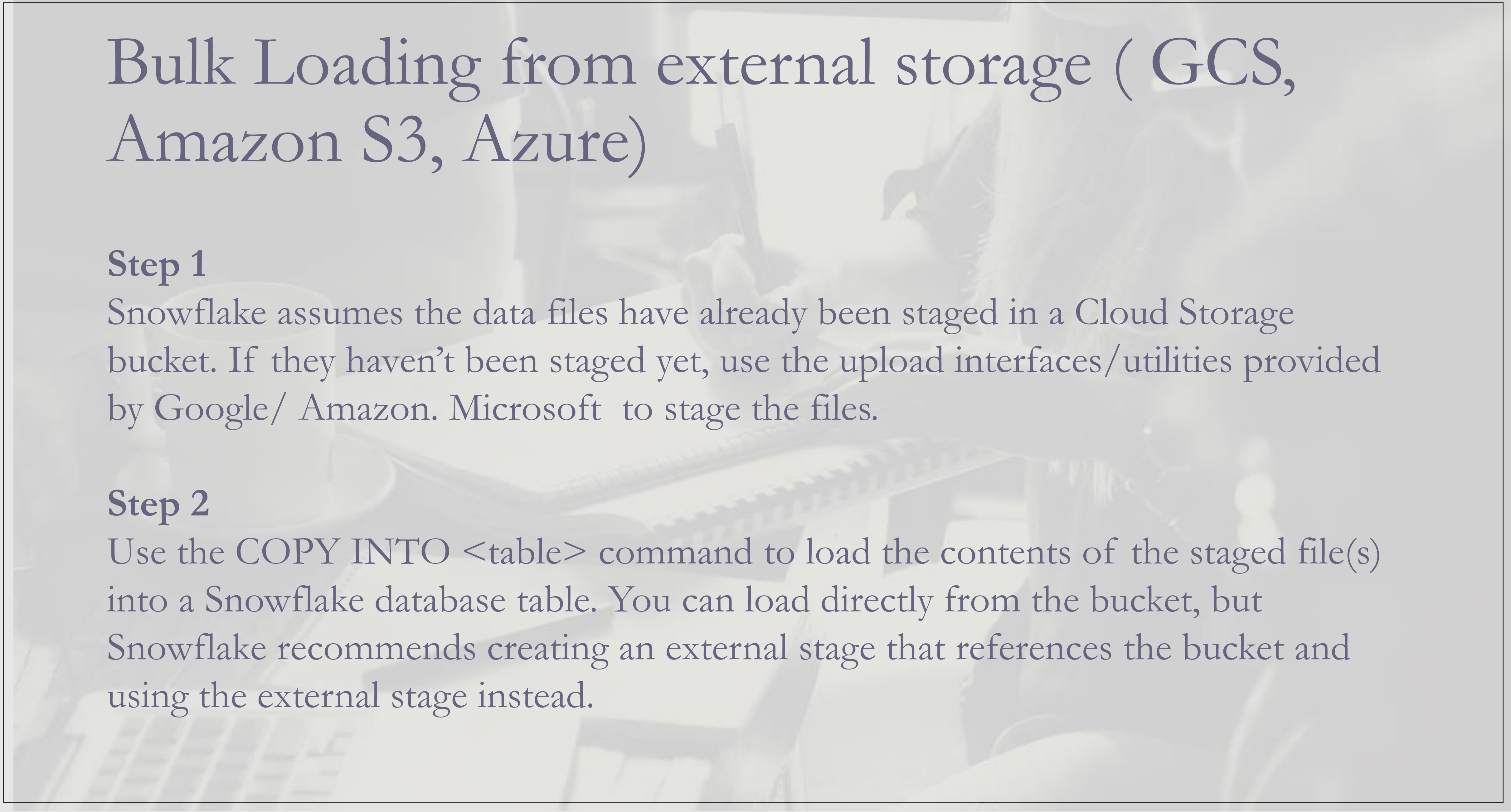
Use the COPY INTO <table> command to load the contents of the staged file(s) into a Snowflake database table.


Local File System
Data Files
Data Files
Data Files
PUT
PUT
PUT
INTERNAL NAMED STAGE
(stage)
user
USER
VIRTUAL WAREHOUSE
(must be running to execute COPY command)
(stage)
COPY INTO
COPY INTO
COPY INTO
Text
TABLE
DATABASE

Hands On Demo : Snowsql(CLI Client)

SnowSQL is the next-generation command line client for connecting to Snowflake to execute SQL queries and perform all DDL and DML operations, including loading data into and unloading data out of database tables.

Download and install snowsql:
https://sfc-repo.snowflakecomputing.com/snowsql/bootstrap/index.html



Bulk Loading from external storage (GCS, Amazon S3, Azure)

Step 1

Snowflake assumes the data files have already been staged in a Cloud Storage bucket. If they haven’t been staged yet, use the upload interfaces/utilities provided by Google/ Amazon. Microsoft to stage the files.

Step 2

Use the COPY INTO <table> command to load the contents of the staged file(s) into a Snowflake database table. You can load directly from the bucket, but Snowflake recommends creating an external stage that references the bucket and using the external stage instead.

Hands On Demo : Bulk Loading from Amazon S3)

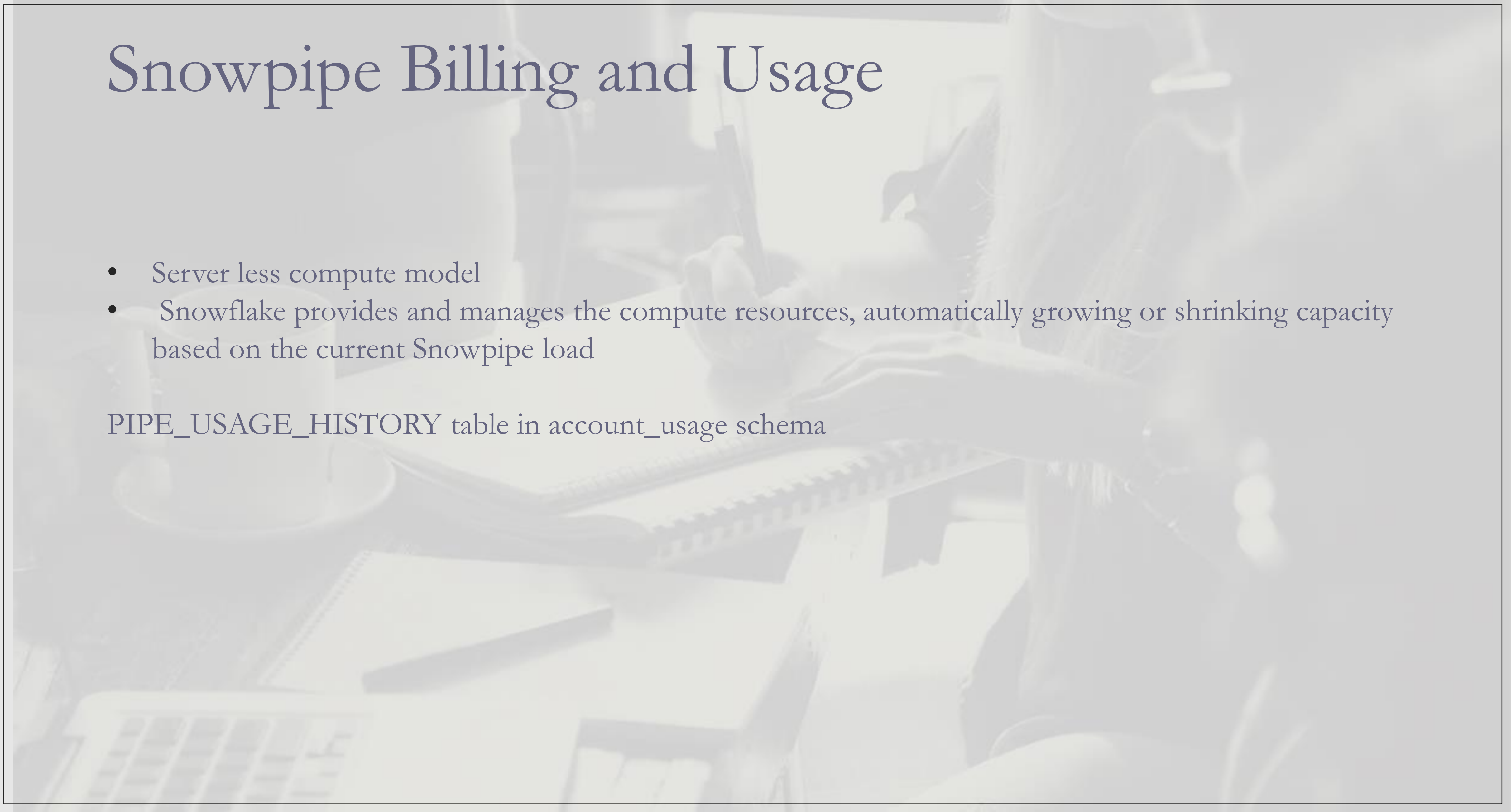
Loading Continuously Using Snowpipe

Loading Continuously Using Snowpipe

Snowpipe is Snowflake’s continuous data ingestion service. Snowpipe loads data within minutes after files are added to a stage and submitted for ingestion.

Mechanisms for detecting the staged files are available:
Automating Snowpipe using cloud messaging
Calling Snowpipe REST endpoints

Snowpipe Billing and Usage

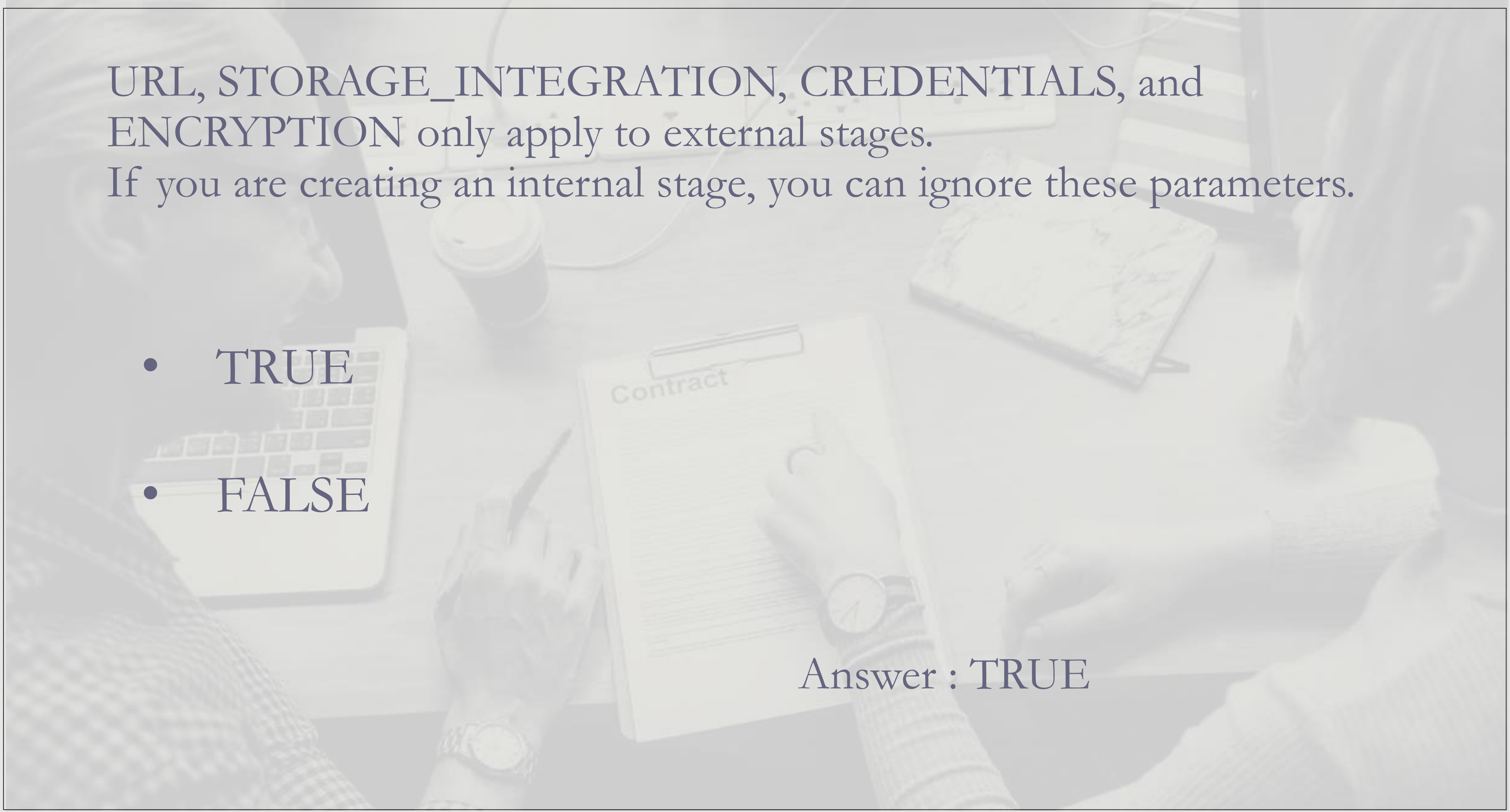


- Server less compute model
- Snowflake provides and manages the compute resources, automatically growing or shrinking capacity based on the current Snowpipe load

PIPE_USAGE_HISTORY table in account_usage schema

Hands on Demo : Automating Snowpipe for Amazon S3

QUIZ



URL, STORAGE_INTEGRATION, CREDENTIALS, and ENCRYPTION only apply to external stages.
If you are creating an internal stage, you can ignore these parameters.

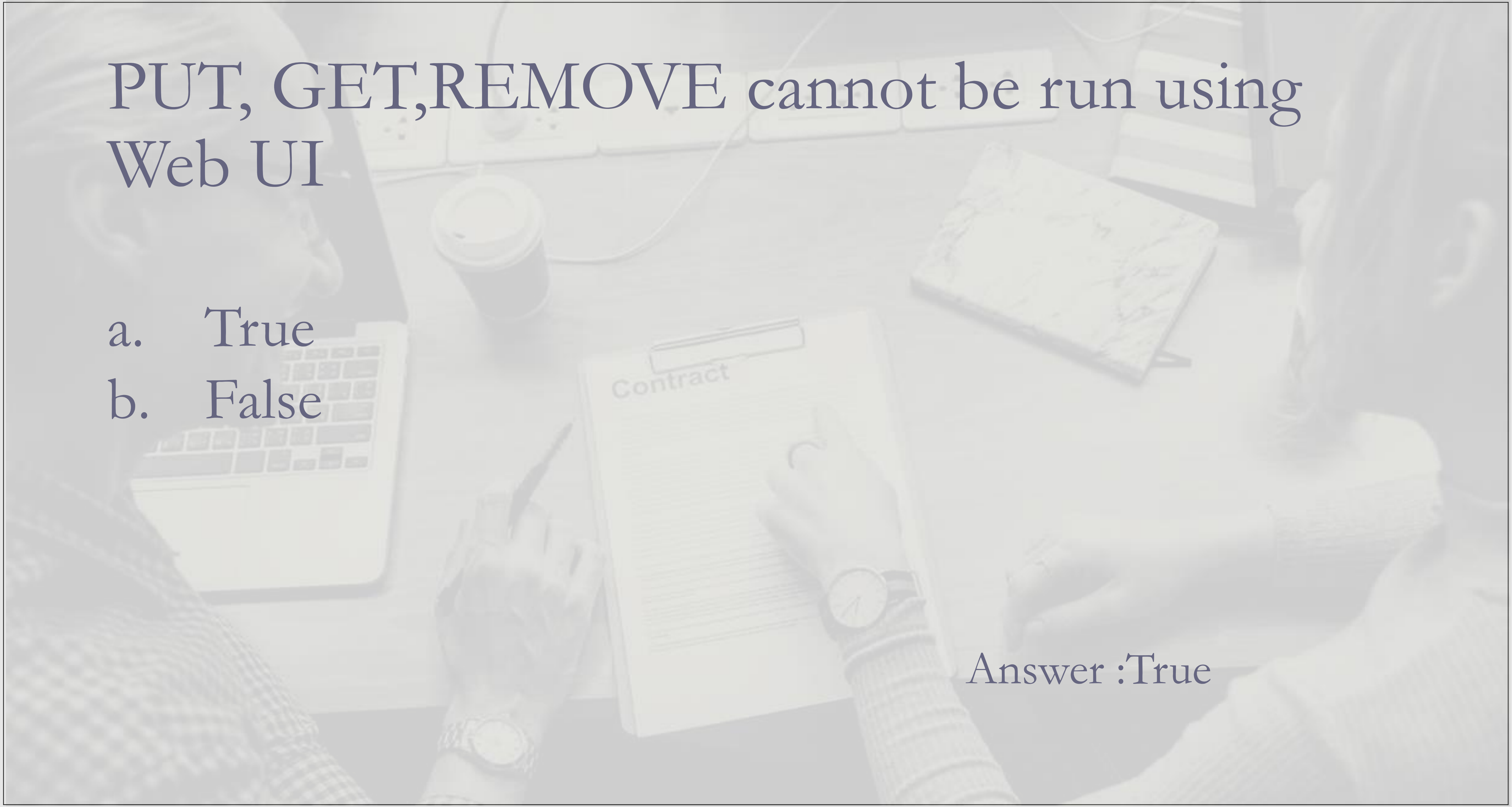
- TRUE
- FALSE

Answer : TRUE

Different data files supported by Snowflake

a. AVRO
b. ORC
c. PARQUET
d. JSON

Answer : All of the above

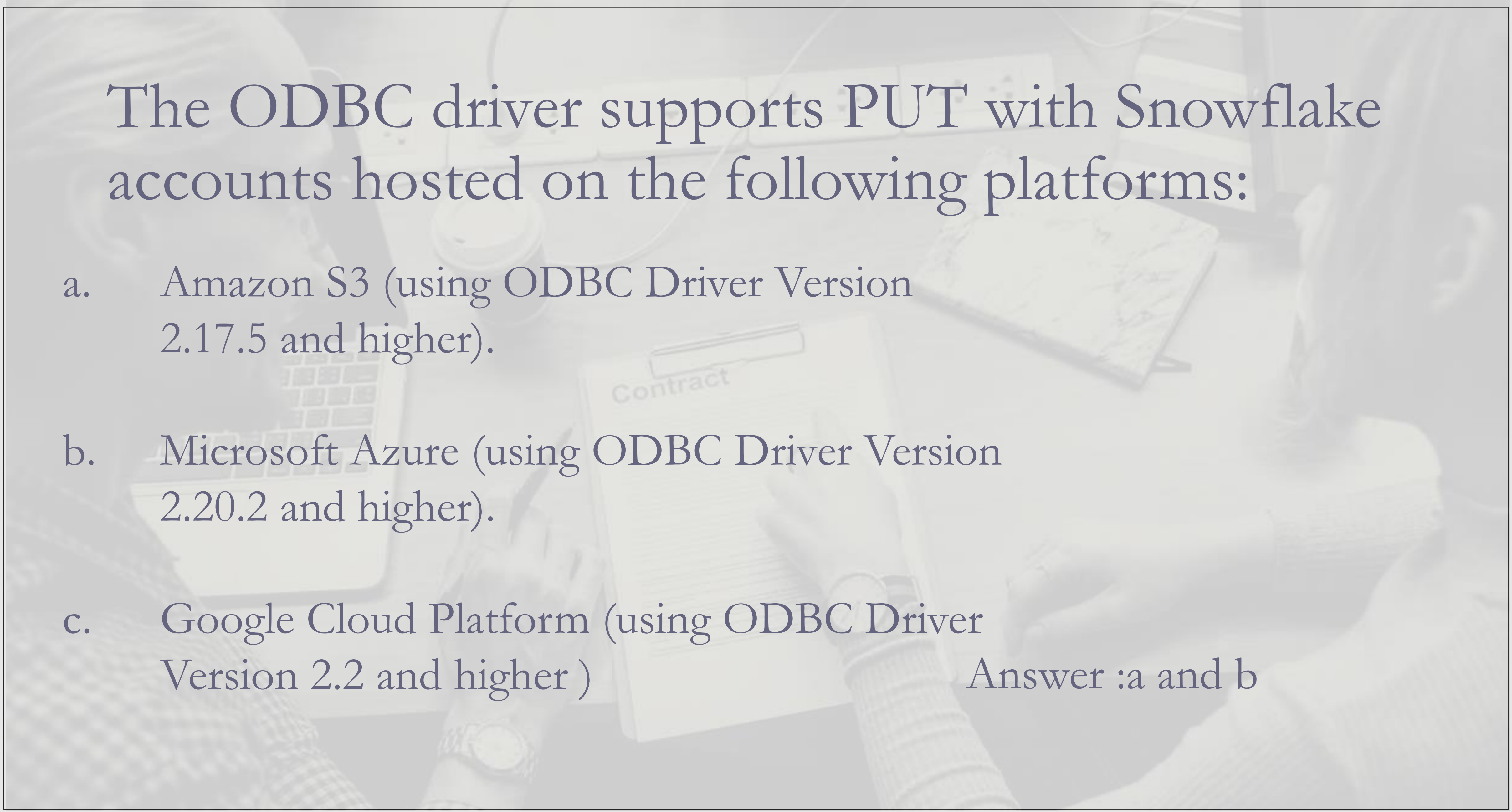


PUT, GET,REMOVE cannot be run using Web UI

a. True

b. False

Answer :True



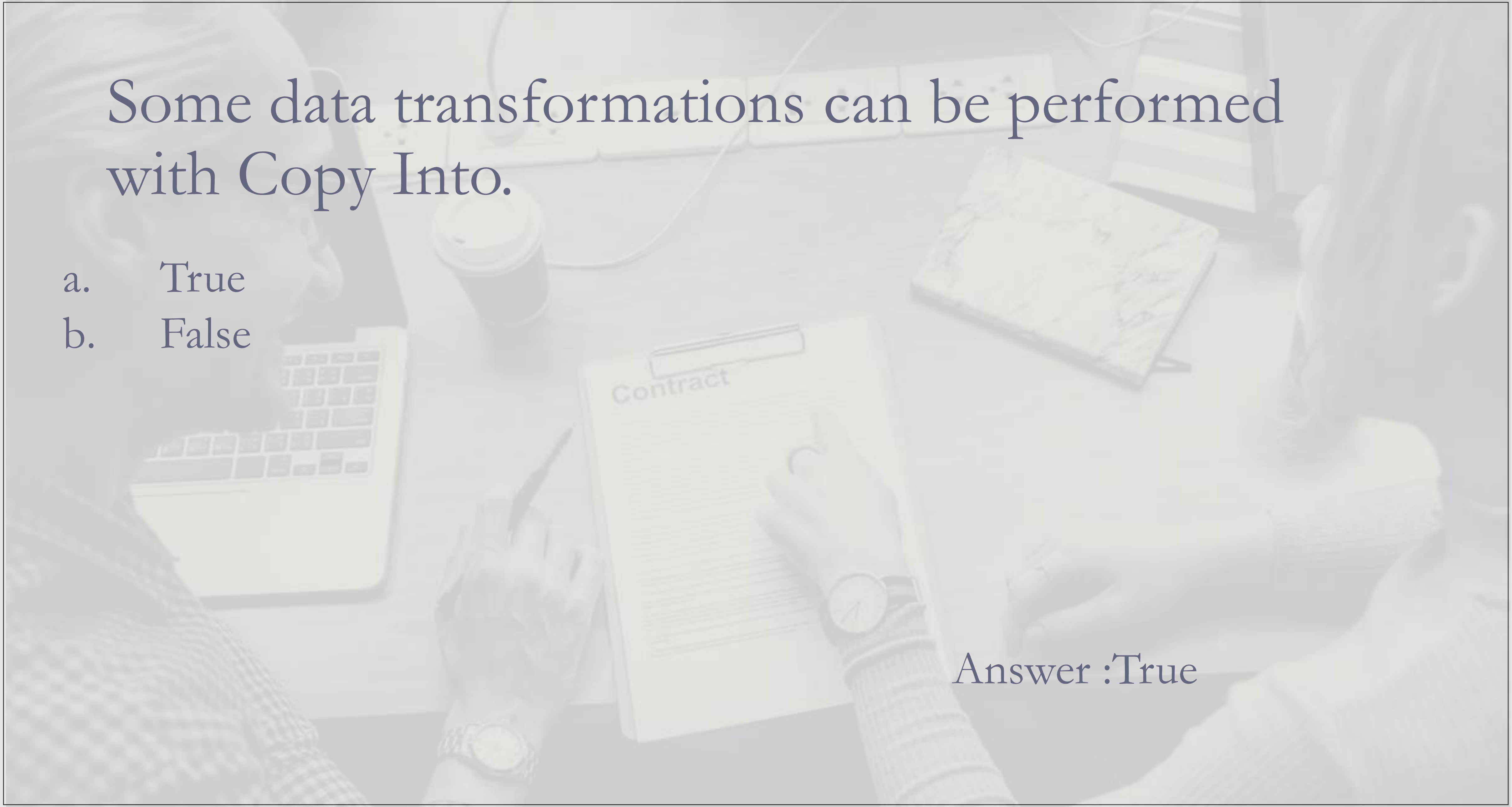
The ODBC driver supports PUT with Snowflake accounts hosted on the following platforms:

a. Amazon S3 (using ODBC Driver Version 2.17.5 and higher).

b. Microsoft Azure (using ODBC Driver Version 2.20.2 and higher).

c. Google Cloud Platform (using ODBC Driver Version 2.2 and higher)

Answer :a and b

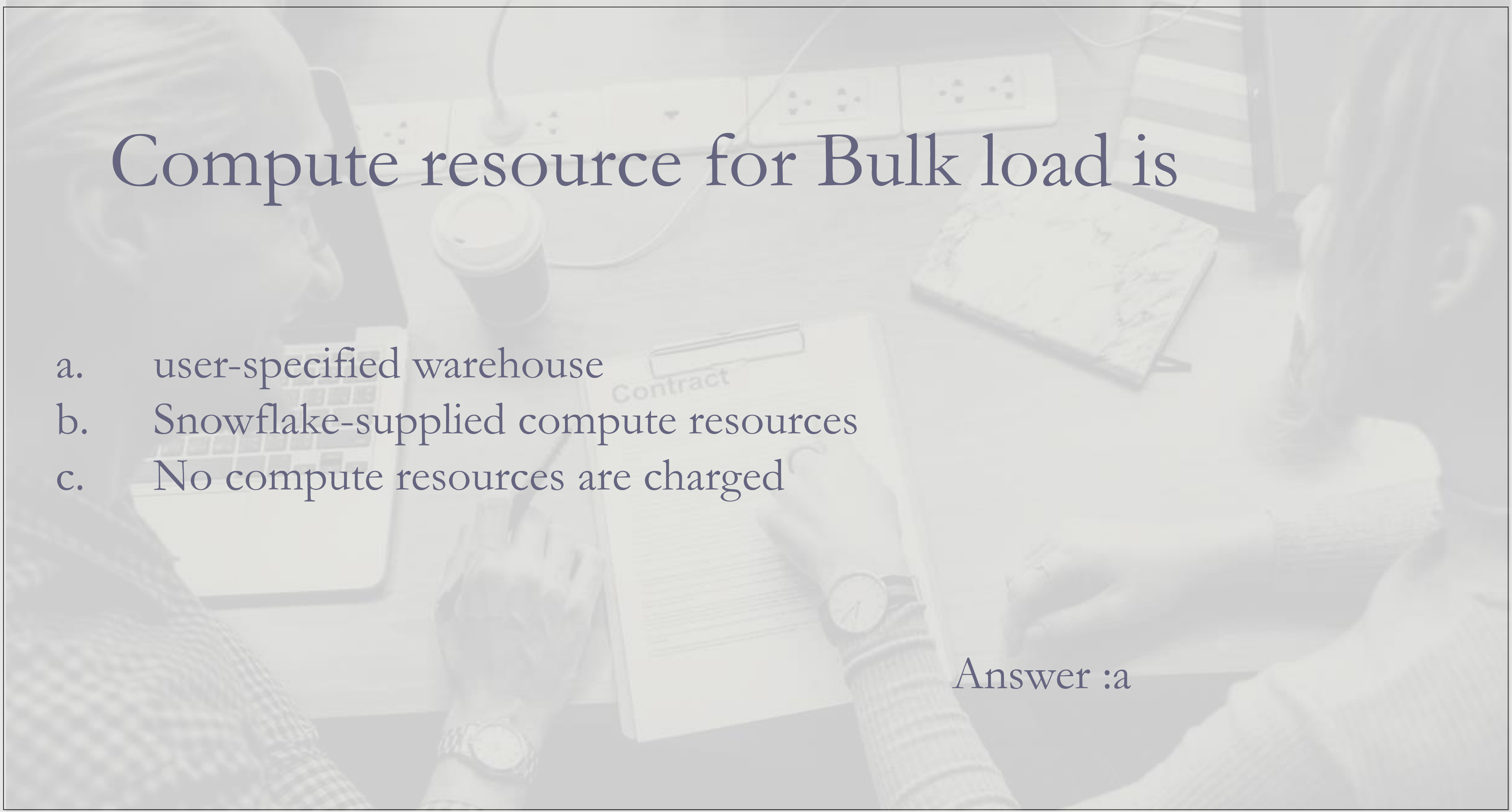


Some data transformations can be performed with Copy Into.

a. True

b. False

Answer :True

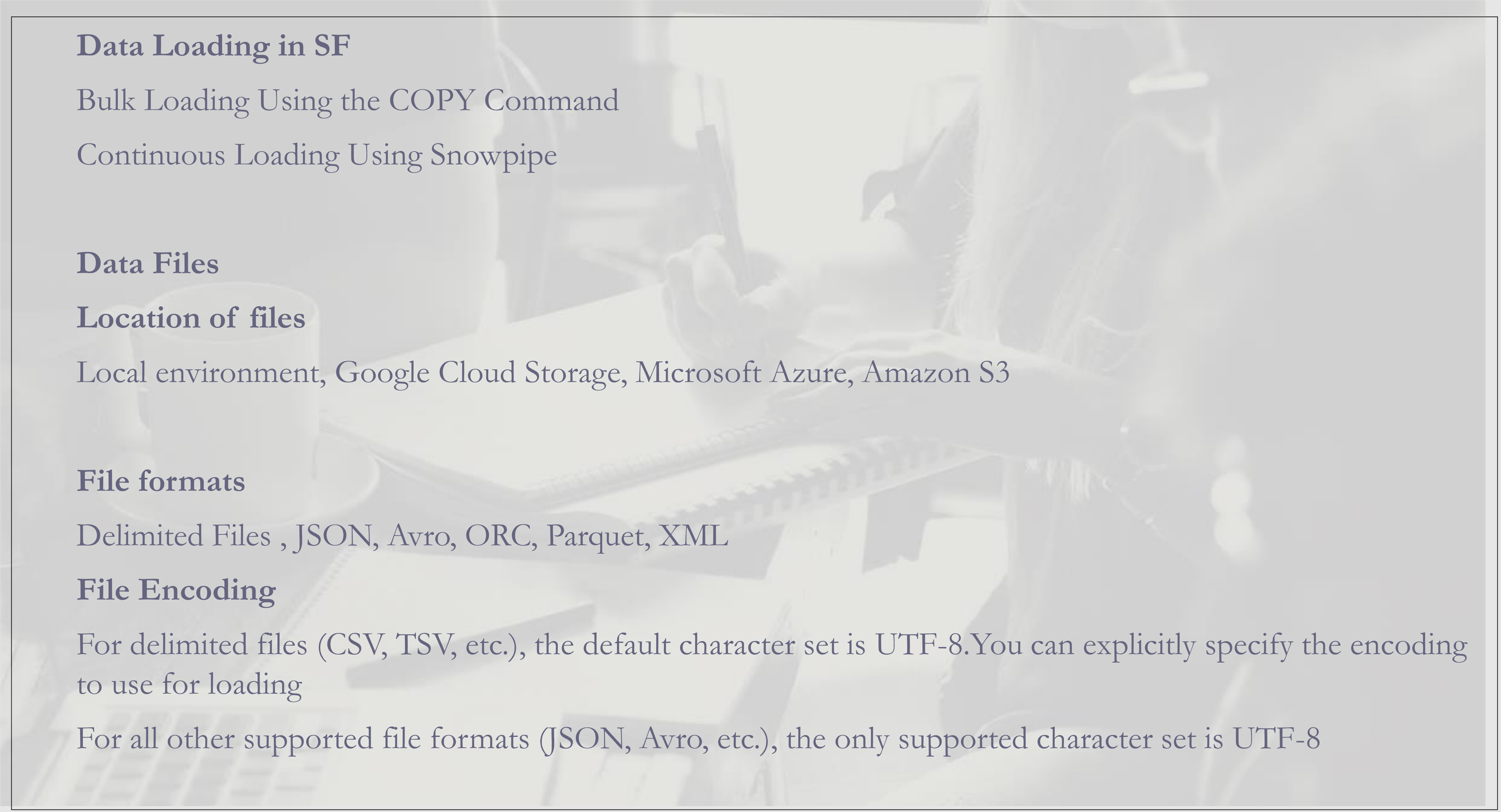


Compute resource for Bulk load is

a. user-specified warehouse
b. Snowflake-supplied compute resources
c. No compute resources are charged

Answer :a

Recap



Data Loading in SF

Bulk Loading Using the COPY Command

Continuous Loading Using Snowpipe

Data Files

Location of files

Local environment, Google Cloud Storage, Microsoft Azure, Amazon S3

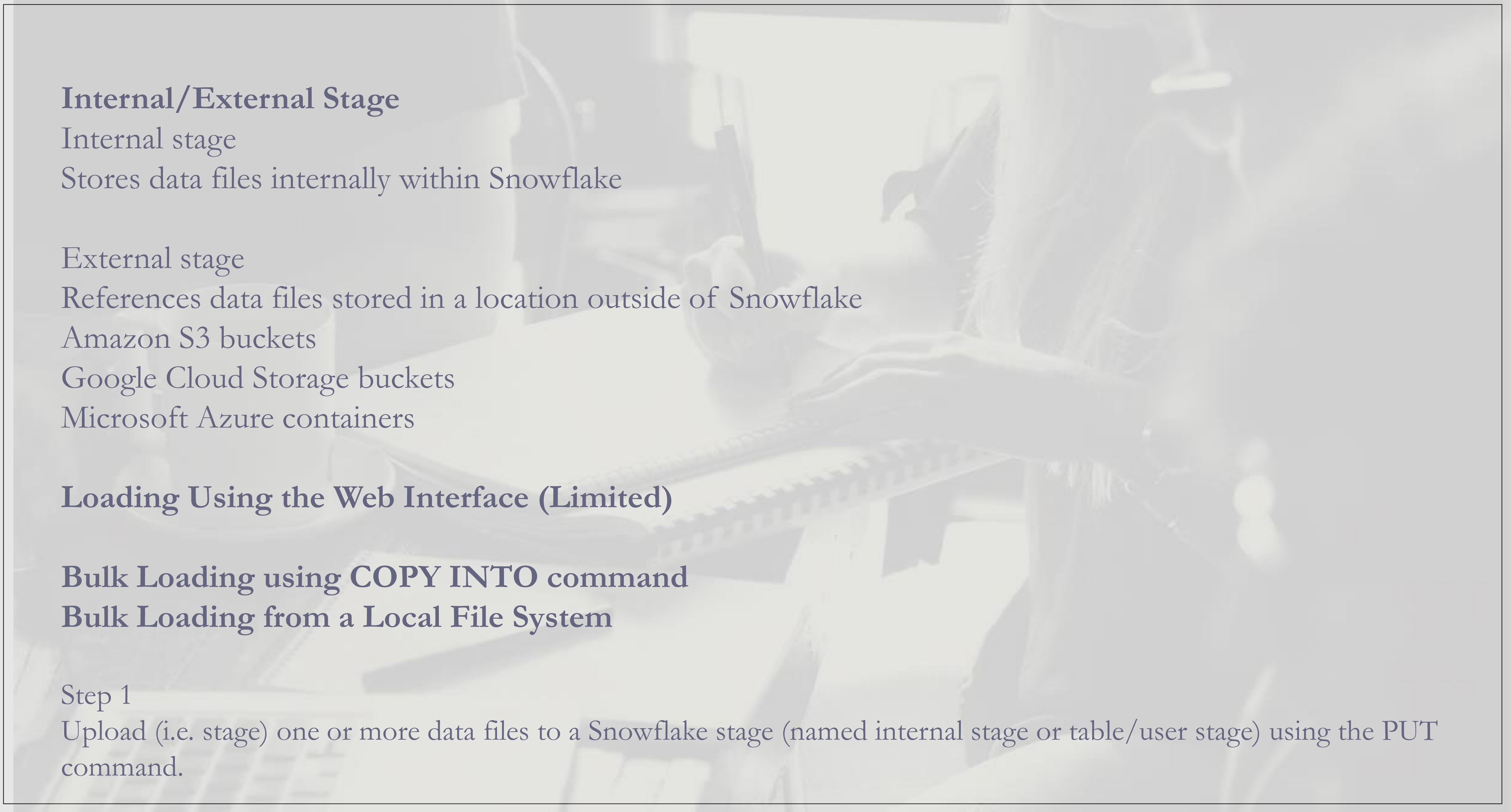
File formats

Delimited Files , JSON, Avro, ORC, Parquet, XML

File Encoding

For delimited files (CSV, TSV, etc.), the default character set is UTF-8.You can explicitly specify the encoding to use for loading

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



Internal/External Stage

Internal stage

Stores data files internally within Snowflake

External stage

References data files stored in a location outside of Snowflake

Amazon S3 buckets

Google Cloud Storage buckets

Microsoft Azure containers

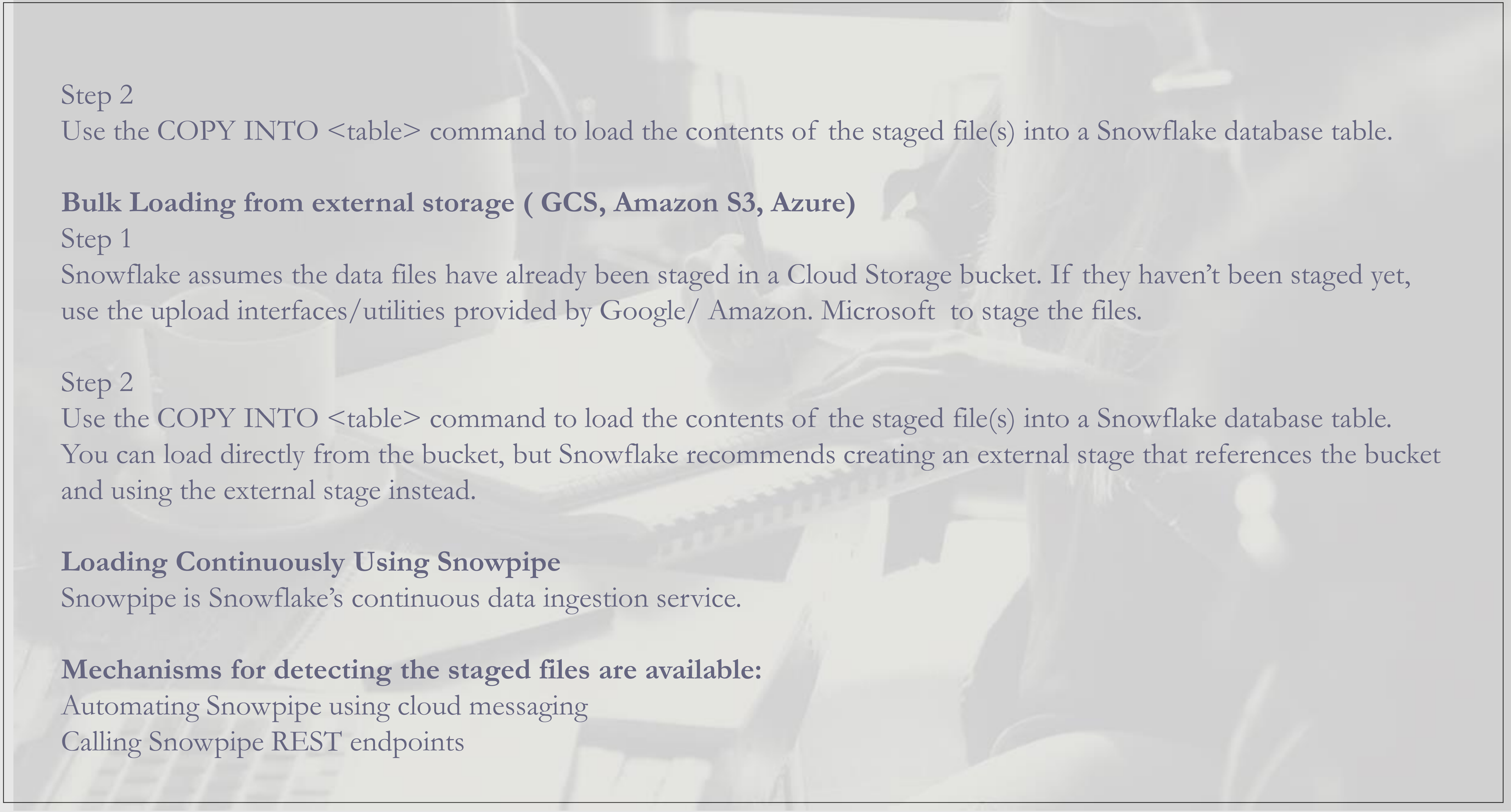
Loading Using the Web Interface (Limited)

Bulk Loading using COPY INTO command

Bulk Loading from a Local File System

Step 1

Upload (i.e. stage) one or more data files to a Snowflake stage (named internal stage or table/user stage) using the PUT command.



Step 2

Use the COPY INTO <table> command to load the contents of the staged file(s) into a Snowflake database table.

Bulk Loading from external storage (GCS, Amazon S3, Azure)

Step 1

Snowflake assumes the data files have already been staged in a Cloud Storage bucket. If they haven't been staged yet, use the upload interfaces/utilities provided by Google/ Amazon. Microsoft to stage the files.

Step 2

Use the COPY INTO <table> command to load the contents of the staged file(s) into a Snowflake database table. You can load directly from the bucket, but Snowflake recommends creating an external stage that references the bucket and using the external stage instead.

Loading Continuously Using Snowpipe

Snowpipe is Snowflake's continuous data ingestion service.

Mechanisms for detecting the staged files are available:

Automating Snowpipe using cloud messaging

Calling Snowpipe REST endpoints