

Unstructured Data

Sujith Nair

Cloud Data Architect

Snowflake Snowpro Certified

Can you unload data from a relational table to a JSON stage with the *COPY* command

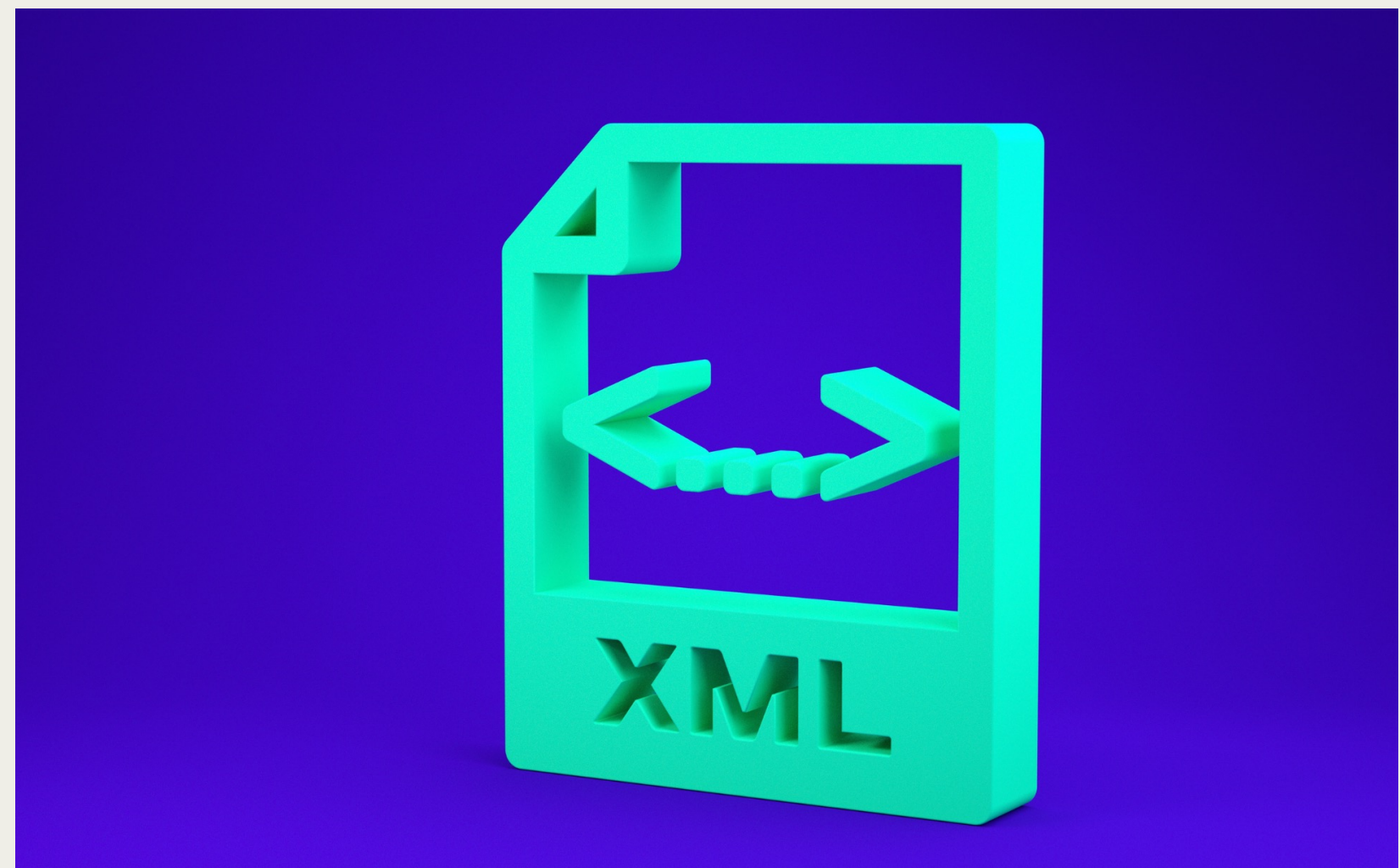
If we create a JSON stage and load relational data into it the load process will fail, because the stage is expecting JSON format.

We will need to use the `object_construct` function convert the relational data to JSON format.



What data type do you recommend to store JSON and XML data ?

Snowflake natively supports storing of semi-structured data like JSON and XML with the `VARIANT` datatype.



Can you explain a scenario in your project where you needed to use variant data type?

We receive IOT data files in JSON format and we copy this data into snowflake tables called the RAW layer which has VARIANT data type. We use snowpipe to copy the data to snowflake tables. Once data is added to the raw tables, we parse the data from the variant columns and load the data to the final tables in snowflake.



#What are the semi-structured data supported by snowflake?

- JSON
- XML
- PARQUET
- AVRO
- ORC



What does the FLATTEN Command do ?

The flatten command is used with Array, Object VARIANT datatypes and is used to convert data present in these data types from rows to columns.

	STUDENT_ID	FIRST_NAME	LAST_NAME	COURSES	...	CONTACT_INFO
1	1	John	Doe	["Math", "English", "History"]		{ "email": "john.rambo@example.com", "phone": "+1234567890" }

	STUDENT_ID	FIRST_NAME	LAST_NAME	THIS	VALUE	...
1	1	John	Doe	["Math", "English", "History"]	"Math"	
2	1	John	Doe	["Math", "English", "History"]	"English"	
3	1	John	Doe	["Math", "English", "History"]	"History"	

```
SELECT
  student_id,
  first_name,
  last_name,
  cert.this,
  cert.value
FROM students stu, LATERAL FLATTEN (input => stu.courses) cert;
```



Thank you!
