



FLATTEN VARIANT IN MATILLION

The purpose of this component is to separate values contained by a variant column type into individual columns.

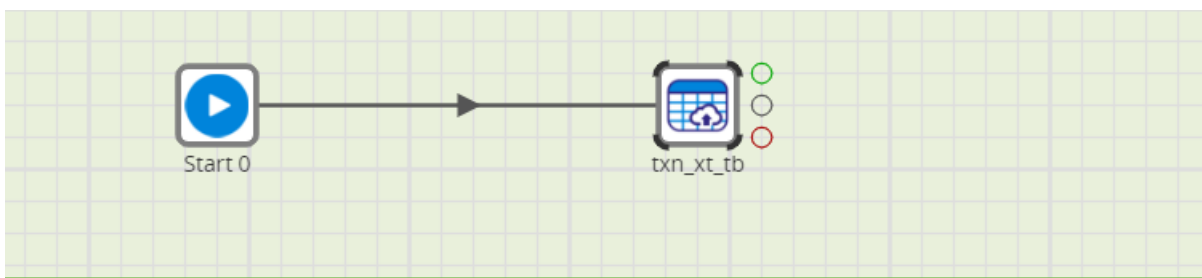
OR

This component is used especially to transform semi-structured data into structured format.

Let's verify this with an example...

Let us first convert the data from structured to semi-structured format, by utilizing the external table component which acts as a pointer to the data (structured) stored on S3 Bucket (Data Lake).

FLATTEN VARIANT IN MATILLION



Start 0 → txn_xt_tb

Properties | SQL | Export | Help

Create External Table OK

Name	Value	Status
Name	txn_xt_tb	...
Create/Replace	Create if not exists	...
Database	DEMO_DB	...
Schema	PUBLIC	...
New Table Name	txn_xt	...
Partition Columns		...
Stage Database	DEMO_DB	...
Stage Schema	PUBLIC	...
Stage	ext-txn-stage	...
Relative Path		...
Pattern		...
Format	[Custom]	...

The external table ("txn_xt") gets created in snowflake (data warehouse) as a **VARIANT** type ("VALUE") via **STAGE** ("ext-txn-stage").

This table is pointing towards the structured data in s3 bucket and looks like:



FLATTEN VARIANT IN MATILLION

	VALUE
1	{ "c1": "2375", "c10": "1", "c11": "0", "c12": "0", "c2": "26984851472", "c3..."
2	{ "c1": "2375", "c10": "1", "c11": "0", "c12": "0", "c2": "26984851472", "c3": "1", "c4":...
3	{ "c1": "2375", "c10": "1", "c11": "0", "c12": "0", "c2": "26984851472", "c3": "1", "c4":...
4	{ "c1": "2375", "c10": "1", "c11": "0", "c12": "0", "c2": "26984851472", "c3": "1", "c4":...
5	{ "c1": "2375", "c10": "1", "c11": "0", "c12": "0", "c2": "26984851472", "c3": "1", "c4":...
6	{ "c1": "2375", "c10": "1", "c11": "0", "c12": "0", "c2": "26984851516", "c3": "1", "c4":...
7	{ "c1": "2375", "c10": "1", "c11": "0", "c12": "0", "c2": "26984851516", "c3": "1", "c4":...
8	{ "c1": "2375", "c10": "1", "c11": "0", "c12": "0", "c2": "26984851516", "c3": "1", "c4":...
9	{ "c1": "2375", "c10": "1", "c11": "0", "c12": "0", "c2": "26984851516", "c3": "1", "c4":...
10	{ "c1": "2375", "c10": "1", "c11": "0", "c12": "0", "c2": "26984851516", "c3": "1", "c4":...
11	{ "c1": "2375", "c10": "1", "c11": "0", "c12": "0", "c2": "26984851516", "c3": "1", "c4":...
12	{ "c1": "1364", "c10": "1", "c11": "0", "c12": "0", "c2": "26984896261", "c3": "1", "c4":...
13	{ "c1": "1364", "c10": "1", "c11": "0", "c12": "0", "c2": "26984896261", "c3": "1", "c4":...
14	{ "c1": "1364", "c10": "1", "c11": "0", "c12": "0", "c2": "26984896261", "c3": "1", "c4":...

Here, c1-c12 are the columns and the rest are the values associated with them, row-wise.

Using the **FLATTEN VARIANT** component under transform directory/folder of a transformation job we will be able to convert the above semi-structured into structured format in Matillion.

The diagram shows a transformation job flow. A source component labeled 'txn_xt' (represented by an orange icon) is connected by an arrow to a target component labeled 'Flatten-Variant-txn_xt' (represented by a blue icon). Below this, a configuration window for the 'Flatten Variant' component is shown. The window has tabs for Properties, Sample, Metadata, SQL, Plan, and Help. The 'Properties' tab is active, showing a table with the following data:

Name	Value	Status
Name	Flatten-Variant-txn_xt	...
Include Input Colu...	No	...
Column Mapping	VALUE, c1, NUMBER...	...
Column Flattens

The important step during configuration of properties is the **Column Mapping**.



FLATTEN VARIANT IN MATILLION

Column Mapping

Column	Property	Type	Alias
VALUE	c1	NUMBER	HOUSEHOLD_KEY
VALUE	c2	NUMBER	BASKET_ID
VALUE	c3	NUMBER	DAY
VALUE	c4	NUMBER	PRODUCT_ID
VALUE	c5	NUMBER	QUANTITY
VALUE	c6	FLOAT	SALES_VALUE
VALUE	c7	NUMBER	STORE_ID
VALUE	c8	FLOAT	RETAIL_DISC
VALUE	c9	NUMBER	TRANS_TIME
VALUE	c10	NUMBER	WEEK_NO
VALUE	c11	NUMBER	COUPON_DISC
VALUE	c12	NUMBER	COUPON_MATCH_DISC



☐ Text Mode

☐ Use Grid Variable

OK

Cancel

- The “COLUMN” in the figure represents the variant column type (VALUE).
- The “PROPERTY” reflects the columns present in the semi-structured VARIANT type.
- The “TYPE” refers to the data type each column holds.
- The “ALIAS” is the new column created after being transformed.

After the job is run, we could refresh the sample data to check for the results we expected i.e., the semi-structured data has been converted to structured format and now the data looks like:



FLATTEN VARIANT IN MATILLION

```
graph LR; txn_xt --> Flatten-Variant-txn_xt
```

Properties | Sample | Metadata | SQL | Plan | Help

Data

1,146,421 rows

Filter Not Set

HOUSEHOLD_KEY	BASKET_ID	DAY	PRODUCT_ID	QUANTITY	SALES_VALUE	STORE_ID	RE
2375	26984851472	1	1004906	1	1.39	364	-0.
2375	26984851472	1	1033142	1	0.82	364	0.0
2375	26984851472	1	1036325	1	0.99	364	-0.
2375	26984851472	1	1082185	1	1.21	364	0.0
2375	26984851472	1	8160430	1	1.5	364	-0.
2375	26984851516	1	826249	2	1.98	364	-0.
2375	26984851516	1	1043142	1	1.57	364	-0.
2375	26984851516	1	1085983	1	2.99	364	-0.
2375	26984851516	1	1102651	1	1.89	364	0.0
2375	26984851516	1	6423775	1	2.0	364	-0.
2375	26984851516	1	9487839	1	2.0	364	-0.

ANALYTICS WITH ANAND