[CosmosDB & JSON] Support map type in reading data

Last updated by | Jackie Huang | Jan 4, 2022 at 12:24 AM PST

Contents

- Issue
- Root Cause
- Resolution
- Map type support
- Additional Information

Issue

When dataflow customers try to use map data type in Cosmos DB or JSON source, it cannot be directly supported while customers can not get map data type under "import projection".
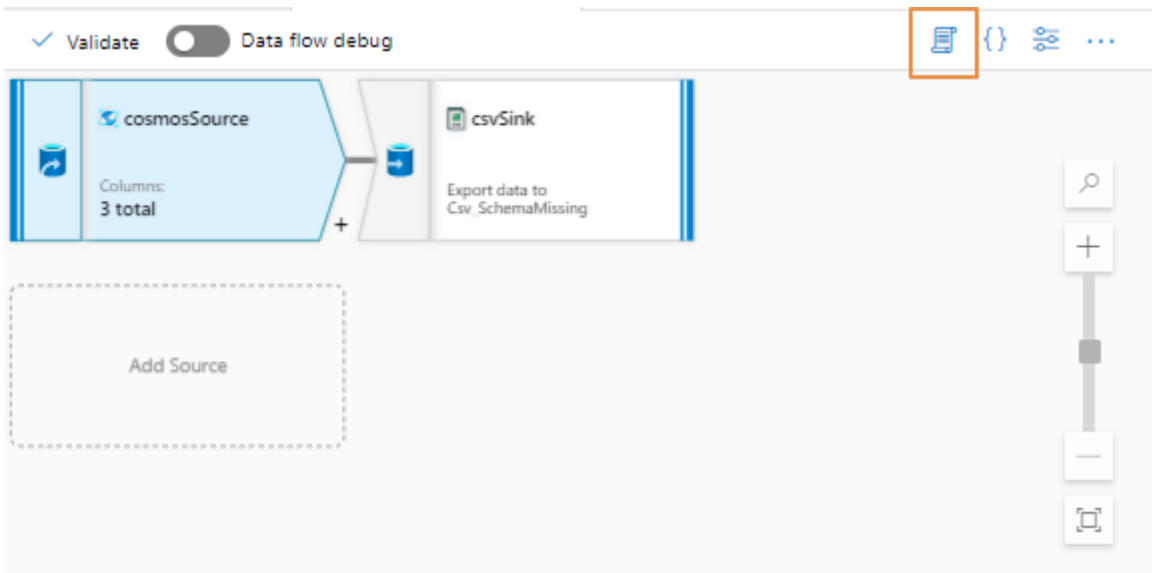
Root Cause

For Cosmos DB and JSON, they are schema free connectivity and related spark connector uses sample data to infer schema, then use that schema as Cosmos DB/JSON source schema. During infer schema, Cosmos DB/JSON spark connector can only infer object data as struct not map data type, that's why map type cannot be directly supported.
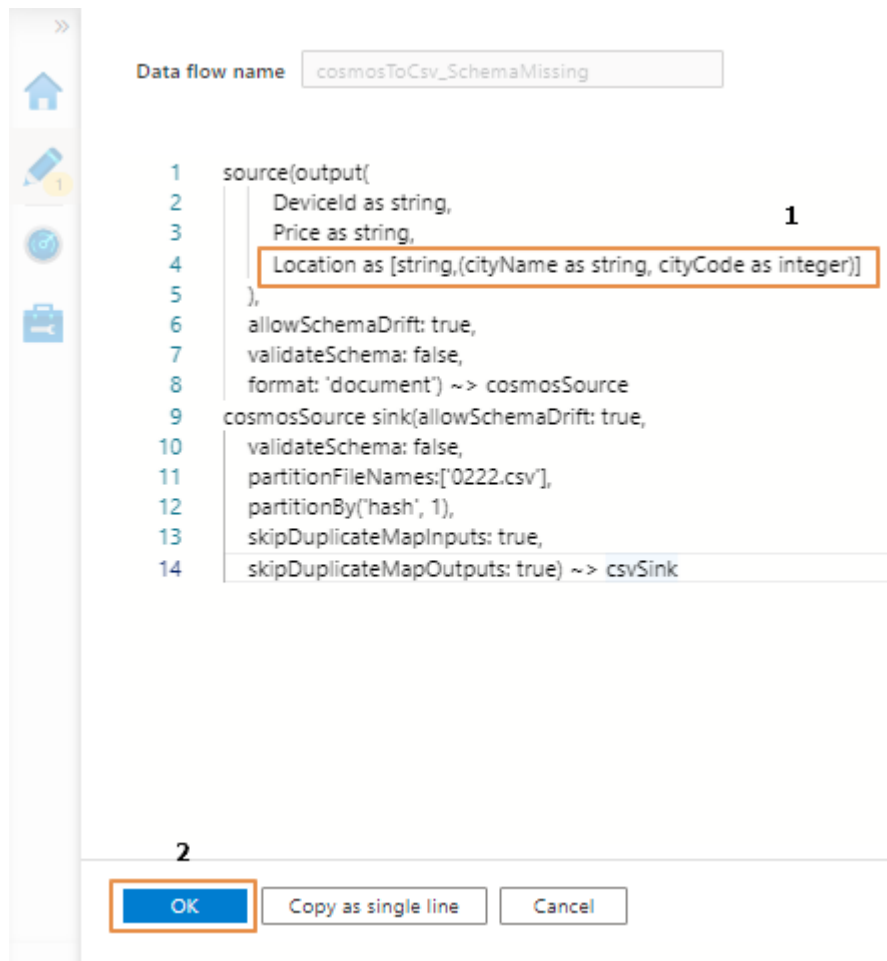
Resolution

In order to solve up known issue/limitation, here we give some examples to guide customers to manually update the script (DSL) of Cosmos DB/JSON source to get map data type support.

Example	Expression	DSL
```{ "location": { "Code": "123", "Date": "2021-05-21" } }```	Map[String, String]	location as [String, String]
```{ "location": { "Code": { "China": "20100", "USA": "20101" } "Date": { "ChinaTime": "2021-05-21", "USATime": "2021-05-20" } } }```	Map[String, Map[String, String]] (map inner map)	location as [String, [String, String]]
```{ "location": { "Code": { "ChinaCode": "20100", "USACode": "20101" } "Date": { "ChinaTime": "2021-05-21" } } }```	Map[String, Struct] (struct inner map)	location as [String, (ChinaCode as string, USACode as string, ChinaTime as Date)]
```{ "location": { "code": [{ "cityName": "Beijing", "cityCode" "2001" }, { "cityName": "Shanghai", "cityCode": "2002" }] } }```	Map[String, Array] (array inner map)	location as [String, (cityName as string, cityCode as string)[]]
```{ "location": { "Code": { "ChinaCode": { "ShanghaiCode": "2001", "BeijingCode": "2002" } } } }```	Map[String, Map[String, Struct]] (struct inner map and map)	location as [String, [String, (ShanghaiCode as string, BeijingCode as string)]]

Step-1: Open the script of dataflow activity.

Step-2: Update DSL to have map type support by referencing up examples.

Note: In order to provide better user experience, ADF dataflow is planning to have UX improvement to support customers editing the schema under "import projection" to get map type support.

Map type support

Type	Is map type supported	Comments
Excel, CSV	No	Both are tabular data source with primitive type, no need to support map type
Orc, Avro	Yes	
JSON	Yes	Map type cannot be directly supported, please follow up TSG to update script (DSL) under source projection
Cosmos DB	Yes	
Parquet	Yes	Today complex data type is not supported on parquet dataset, so customer needs to have import projection under dataflow parquet source to get map type.
XML	No	TBD based on customers' feedback

Additional Information

- Icm Reference: N/A
- Author: Zhangyi Yu
- Reviewer: Zhangyi Yu; Shawn Xiao
- Keywords: map data type