

[Cosmos DB, DataDuplicate] Resource with specified id or name already exists

Last updated by | Zehua Chen | Mar 31, 2023 at 11:57 PM PDT

Background Knowledge

1. In CosmosDB sink, if there is no "id" property in customer's data, CosmosDb Sink will generate id automatically.
2. Different logic partition can have the same id, but there can't be same id in one logic partition.
3. In cosmosNoSql sink, we have logic to avoid duplicate insert.

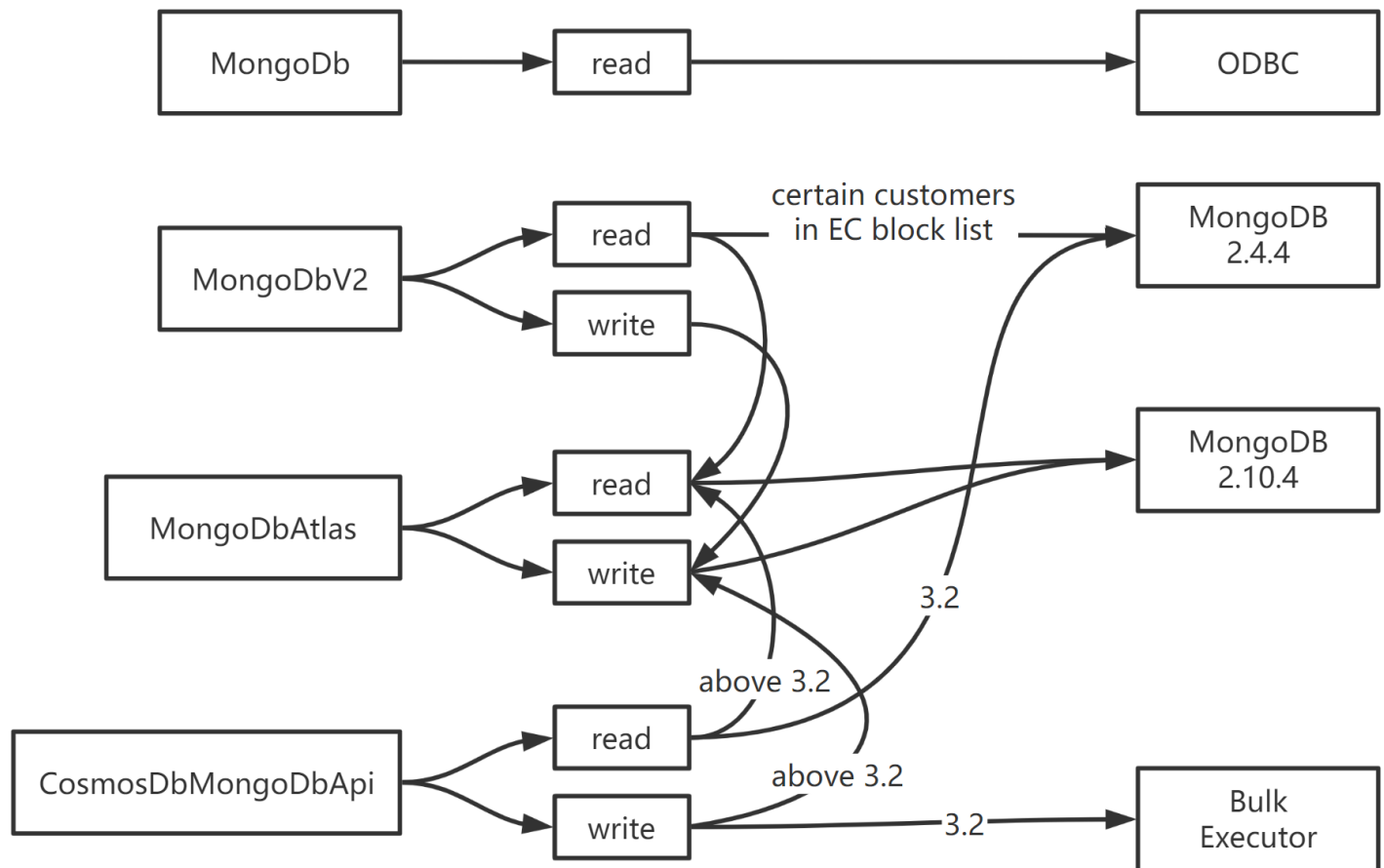
The screenshot shows the Azure Data Factory interface. The left sidebar has a 'DATA' section with 'SampleDB' expanded, showing 'Scale', 'Items', 'Settings', 'Stored Procedures', 'User Defined Functions', 'Triggers', 'leases', 'SQLSample01HTAP', 'SQLSampleLogicApp', and 'SQLSubPartitioning'. The 'Items' folder is selected. The main area shows a table with columns 'id' and '/category'. The table contains three rows: 'bf36ada3-c562-...', '5fa34efd-92f6-4...', and 'sampleid-001'. The 'sampleid-001' row is highlighted in yellow, and a red 'ok' label is next to it. The 'Load more' button is visible below the table.

id	/category
bf36ada3-c562-...	demo
5fa34efd-92f6-4...	demo
sampleid-001	demo
sampleid-001	show

Check whether it's a known issue

Issue1: [CosmosMongoApi sink] Resource with specified id or name already exists.

Root cause



If the version of CosmosDbMongoDbApi is 3.2, it will use Bulk Executor in backend. If the customer's data volume is too large, this SDK may generate duplicate id.

Resolution

Upgrade CosmosDbMongoDbApi above 3.2 or switch to MongoDB/MongoDbAtlas

Issue 2 [CosmosNoSql sink] Resource with specified id or name already exists.

Root cause

ObjectId(_id) is generated by a component called BulkExecutor within the MongoDB SDK. BulkExecutor generates _id automatically and it is quite old version, causing the duplicate _id issue.

Workaround

1. Switch from CosmosSqlApiV2 to CosmosSqlApiV3 by following TSG
https://dev.azure.com/supportability/AzureDataFactory/_wiki/wikis/AzureDataFactory/621186/-CosmosDB-How-to-use-CosmosDbSqlApi-v3 []
2. Try upsert instead of insert as a workaround.

Additional Information

1. Remember to check whether customer data has field "id".

2. Repro this issue, then find PG team for help.