[SAP ODP] Manage DIA work process consumption on SAP Server

Last updated by | Zhangyi Yu | Jan 19, 2023 at 12:04 AM PST

Contents

Issue

When customers use SAP in ADF dataflow to read/extract initial or delta data, he may find too many dialog work process are built on its SAP server side, even meet throttling error "too many work process required".

Root Cause

When customers use ADF dataflow to read/extract SAP data, it will generate connection when ADF call SAP server to read data through RFC every time, then one dialog work process (per connection) is created on customer's SAP server side.

Today we always use 4 connections as default value to connect SAP server to read data in parallel, so 4 work process is required.

Here we use reading 8 SAP tables' data in parallel as example:

- If customer doesn't enable SAP source partition feature, then 8 tables * 1 internal copy (single partition per table, 1 partition 1 internal copy) * 4 connections (default connection number in each internal copy) = 32 work process is required.
- If customer enabled SAP source partition feature, then 8 tables * 4 internal copy (if 4 partitions per table, 1 partition 1 internal copy) * 4 connections (default connection number in each internal copy) = 128 work process is required.

For most customers, if their SAP server is poor/weak, it cannot handle so many work process at the same time, then up issue is happened.

Resolution

Today We have different solutions to solve this issue. Except asking customer to reduce the parallelism count (table number) or scale up the server to be powerful. Now ADF dataflow supports hidden feature to help customer manage DIA work process consumption.

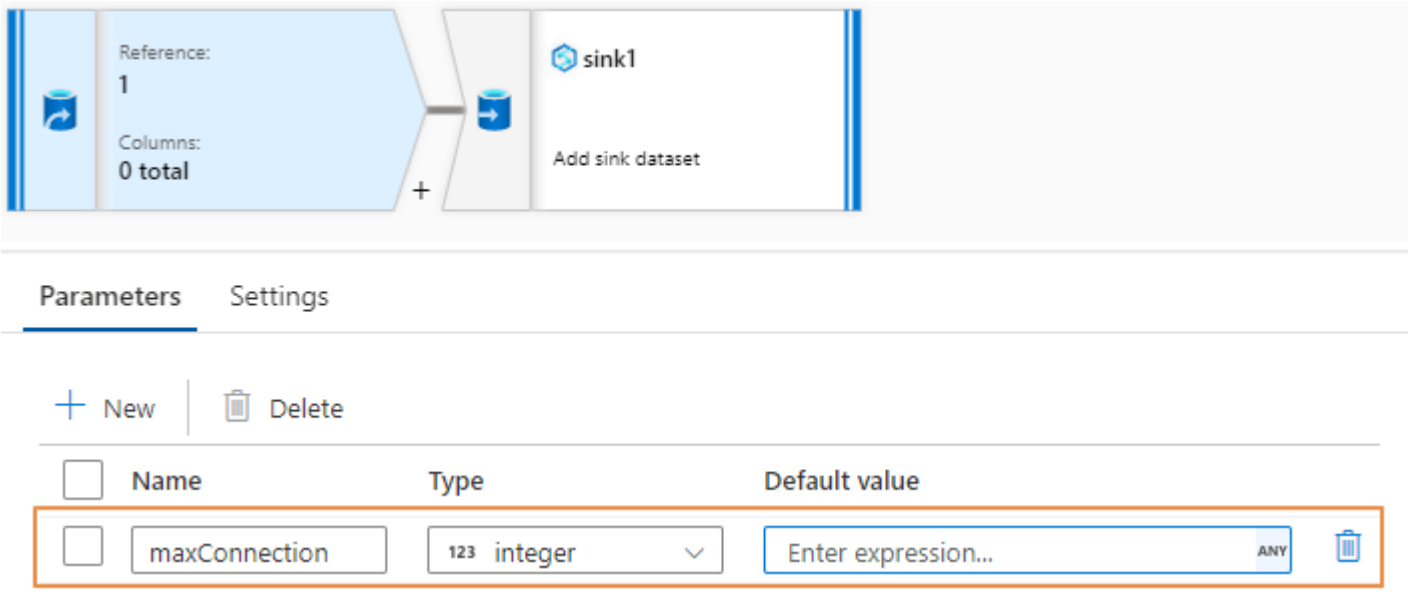
- For SAP full-data load or one-time initial load, mostly it needs to read whole table's data and data size is big. The customer can use more connections (e.g. 4/8/16/...) to speed up this read/write data.
- For SAP delta-data load, mostly it is touching and handling small data, so few connection (e.g. 1) is enough to reduce stress on SAP server.

The customer can manually modify ADF dataflow script by adding following key-value ("parallelCopies") to control the connection (work process) number. The value of this key means max connections (work process) used to read SAP data per table per partition.

```
source(allowSchemaDrift: true,
    validateSchema: false,
    store: 'SapOdp',
    format: 'sapObject',
    staged: true,
    context: 'SAPI',
    objectName: 'ZODPTEST2',
    readMode: 'fullLoad',
    enableCdc: false,
    parallelCopies: 1 ) ~> source1
```

If customer wants to provide dynamic value on it, he can use parameter.

- Step-1: Create dataflow reference with parameter definition.



- Step-2: Manually edit dataflow script and adding parameter, then save it.

```
parameters{
    maxConnection as integer
}
source(allowSchemaDrift: true,
    validateSchema: false,
    store: 'SapOdp',
    format: 'sapObject',
    staged: true,
    context: 'SAPI',
    objectName: 'ZODPTEST2',
    readMode: 'fullLoad',
    enableCdc: false,
    parallelCopies: ($maxConnection)) ~> source1
source1 sink(allowSchemaDrift: true,
    validateSchema: false,
    format: 'table',
    store: 'synapseanalytics',
    schemaName: 'dbo',
    tableName: 'testWorkProcess',
    insertable: true,
    updateable: false,
    deletable: false,
    upsertable: false,
    allowCopyCommand: true,
    staged: true,
    skipDuplicateMapInputs: true,
    skipDuplicateMapOutputs: true,
    errorHandlingOption: 'stopOnFirstError') ~> sink1
```

1

2 OK | Copy as single line | Cancel

Additional Information:

- Icm Reference: N/A
- Author: Zhangyi Yu
- Reviewer: Zhuoyang Zhang; Yanchen Shi