Extracting Data from SAP ECC

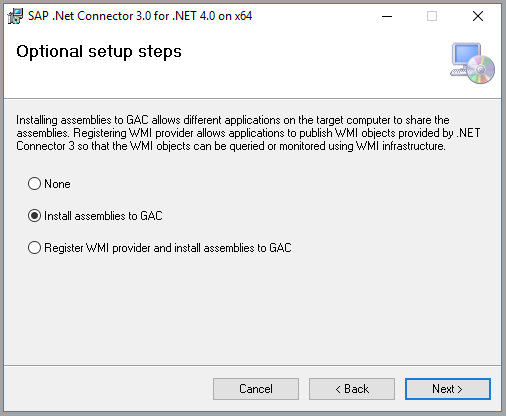
using Azure Data Factory Sap Table Connector

# Prerequisites:

1. Download the Private self-Hosted IR from [Here](https://wenbofdemo.blob.core.windows.net/sapselfir/0425/IntegrationRuntime_3.16.7046.47%20(64-bit).msi?st=2019-04-25T09%3A06%3A29Z&se=2019-05-26T09%3A06%3A00Z&sp=rl&sv=2017-07-29&sr=b&sig=M8r26BQkkRjRxQkTnorHmBv3jZgpAXdI7LxzVp3z%2Bh4%3D)[[1]](#footnote-1)
2. Supported Version: SAP products on Netweaver version 7.01 or higher (in a recent SAP Support Package Stack released after the year 2015).
3. SAP user needs to have following permissions:
   1. Authorization for RFC and SAP System(such as ECC if you are going to extract data from the SAP ECC system).
   2. Permissions to the “Execute” Activity of Authorization Object “S\_SDSAUTH”.
4. [download the SAP .NET Connector 3.0](https://support.sap.com/en/product/connectors/msnet.html)  from SAP’s website using the following link. Please use the 64-bit one since our self-hosted IR only has 64 bit.

[SAP .NET Connector 3.0](https://support.sap.com/en/product/connectors/msnet.html)

1. Install the Sap .NET Connector 3.0 on the same machine as where the self-hosted IR is installed. When installing, in the optional setup steps window, please make sure you select the Install Assemblies to GAC option as shown in the following image.



# Private Entry URL of ADF Portal for SapTable Preview

Go to Author & Monitor of the Data Factory, append the flag as below into the url of the ADF authoring page.

&feature.saptable=true

//the URl will look like below:

https://ms-adf.azure.com/authoring/pipeline/pipeline1?factory=%2Fsubscriptions%<xxxxxxxx>%2FresourceGroups%2F<xxxxxxx>%2Fproviders%2FMicrosoft.DataFactory%2Ffactories%2F<xxxxxx>&feature.saptable=true

# Payload Definition

**SapTable Linked service Properties**

|  |  |  |
| --- | --- | --- |
| **Property** | **Description** | **Required** |
| type | The type property must be set to SapTable. | Yes |
| server | Name of the server on which the SAP instance resides. | Yes |
| systemNumber | System number of the SAP system. Allowed value: two-digit decimal number represented as a string. | Yes |
| clientId | Client number of the system. Allowed value: three-digit decimal number represented as a string. | Yes |
| language | Language that the system uses. By default, it is EN for English. | No |
| userName | The user name used in Basic authentication type. | Yes |
| password | The password used in Basic authentication type. | Yes |
| connectVia | The [Integration Runtime](https://docs.microsoft.com/en-us/azure/data-factory/concepts-integration-runtime) to be used to connect to the data store.  A self-hosted Integration Runtime is required for this connector. | Yes |

**SapTableResource Dataset Properties**

|  |  |  |
| --- | --- | --- |
| **Property** | **Description** | **Required** |
| type | The type property must be set to SapTableResource. | Yes |
| tableName | The name of the SAP Table. | Yes |

**SapTableSource Properties**

|  |  |  |
| --- | --- | --- |
| **Property** | **Description** | **Required** |
| type | The type property must be set to SapTableSource. | Yes |
| rowCount | The number of rows to be retrieved. | No |
| rowSkips | The number of rows that will be skipped. | No |
| rfcTableFields | The fields of the SAP table that will be retrieved. For example, “column0, column1”. | No |
| rfcTableOptions | The options for the filtering of the SAP Table. For example, “COLUMN0 EQ ‘SOME VALUE’” | No |

# Partition Options

since there  is performance latency brought by reading batch after batch when splitting the whole table by hard-coded row count, we provide the partition options such as below for you to split the table so that the data can be read by partitions. It can also support read multiple partitions in parallel.

**Note: currently the parallel will copy into multiple files which also means you should not set filename of the binary(e.g. file, ADLS, blob etc) sink if you would like to leverage parallel. Otherwise, only 1 parallel will be initiated to process your run.**

There are four partition options you can choose:

PartitionOnInt

PartitionOnCalendarYear

PartitionOnCalendarMonth

PartitionOnCalendarDate

|  |  |
| --- | --- |
| Properties | Description |
| partitionColumnName | The name of the column that will be used for proceeding range partitioning. |
| partitionUpperBound | The maximum value of column specificed in partitionColumnName that will be used for proceeding range partitioning. |
| partitionLowerBound | The minimum value of column specificed in partitionColumnName that will be used for proceeding range partitioning. |
| maxPartitionsNumber | The maximum value of partitions the table will be splitted into.  Better to keep the average row count for each partition to be equal with the batch size you would like to set. |

Here is a sample usage payload

|  |
| --- |
| "type": "Copy",  "typeProperties": {    "source": {      "type": "SapTableSource",      "rfcTableOptions": "COLNAME EQ '122460' ",  "partitionOption": "SapPartitionOnCalendarMonth",  "partitionSettings": {      "partitionColumnName": "CALMONTH",      "partitionUpperBound": "201212",      "partitionLowerBound": "201201",  "maxPartitionsNumber": 6  }    },    "sink": {      "type": "BlobSink"    },    "parallelCopies": 3  } |

# Best Practice on Setting the partition settings:

The goal of the partition settings is to get the row count of each partition to be some splits that have similar size of rows which is allowed to be retrieved by your server during one RFC call.

Here is a simple sample:

Suppose you have a table like below

|  |  |  |
| --- | --- | --- |
| IntColumn | OtherColumn1 | OtherColumn2 |
| 1 |  |  |
| … |  |  |
| … |  |  |
| 50,000,000 |  |  |

The best number of rows count allowed by your system in each single RFC call is 500,000 (which is not only allowed by the memory of your system but also can ensure a certain level of performance.)

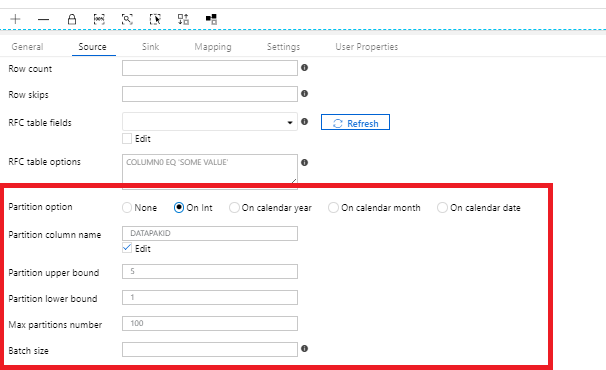
Then you can set the partitionOption as PartitionOnInt with settings as below

partitionColumnName: IntColumn

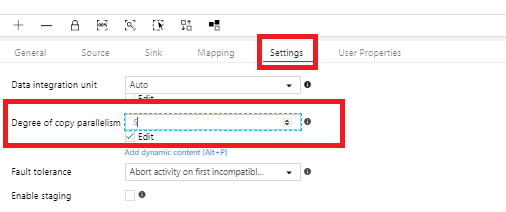
partitionUpperBound:50000000

partitionLowerBound:1

maxPartitionsNumber: 100 //here the 100 is from 50,000,000/500,000 = 100



parallelCopies = 5 //or 10, use some value that can be divided by 100(maxPartitionsNumber)



1. Last versions of IR: [Version0](https://wenbofdemo.blob.core.windows.net/sapselfir/IntegrationRuntime_3.15.6995.1%20(64-bit).msi?st=2019-03-05T06%3A36%3A05Z&se=2019-04-04T06%3A36%3A00Z&sp=rl&sv=2017-07-29&sr=b&sig=ZF4X9bNWq0l5AFEe9Xl1IAs0iUmerojDayDimyWjjB0%3D), [Version1](https://wenbofdemo.blob.core.windows.net/sapselfir/saptablewithbatch0307/IntegrationRuntime_3.15.6997.10%20(64-bit).msi?st=2019-03-07T07%3A46%3A42Z&se=2019-03-08T07%3A46%3A42Z&sp=rl&sv=2017-07-29&sr=b&sig=gW9cR7auqTFChFwYtgwgr4bHCPfDcT6%2FvJyjZ901d9U%3D), [Version2](https://wenbofdemo.blob.core.windows.net/sapselfir/saptablewithbatch0307/IntegrationRuntime_3.15.6998.6%20(64-bit).msi?st=2019-03-07T13%3A52%3A23Z&se=2019-04-01T13%3A52%3A00Z&sp=rl&sv=2017-07-29&sr=b&sig=G490uno8phdjPYrV%2Fj2rg%2FoWvUHfOEqlHTzOacXVHfE%3D)，[Version3](https://wenbofdemo.blob.core.windows.net/sapselfir/saptablepreview/IntegrationRuntime_3.15.7003.2%20(64-bit).msi?st=2019-03-13T02%3A57%3A48Z&se=2019-04-14T02%3A57%3A00Z&sp=rl&sv=2017-07-29&sr=b&sig=0Csh6shEr3%2BtrrQ%2F1MDryv445PMwILwaiA7NBfWWyXA%3D), [Version4](https://wenbofdemo.blob.core.windows.net/sapselfir/saptablepreview/IntegrationRuntime_3.15.7004.17%20(64-bit).msi?st=2019-03-14T08%3A12%3A54Z&se=2019-04-15T08%3A12%3A00Z&sp=rl&sv=2017-07-29&sr=b&sig=MYsoAmBclWUbQVll09UsT7DG%2FvgUVYGO%2BGsskVYbK%2BM%3D)， version5: [Here](https://wenbofdemo.blob.core.windows.net/selfir/0411/IntegrationRuntime_3.16.7032.30%20(64-bit).msi?st=2019-04-11T22%3A30%3A33Z&se=2019-05-12T22%3A30%3A00Z&sp=rl&sv=2017-07-29&sr=b&sig=mcb0a69TyP45iUnoCq4t1IwstamNB9%2BYmYyzT7hi2hM%3D) [↑](#footnote-ref-1)