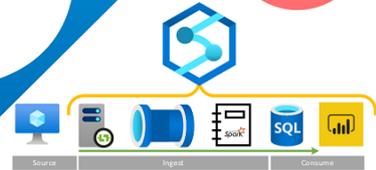


SAP connection tools for process automation: Microsoft, Pentaho, Talend

SAP

talend



pentaho



CONTENT

1.	INTRODUCTION	3
2.	PENTAHO	3
3.	TALEND.....	8
4.	POWER AUTOMATE.....	20
5.	AZURE DATA FACTORY / SYNAPSE ANALYTICS	26
6.	CONCLUSIONS.....	64

1. INTRODUCTION

The document gathers a series of tools used for process automation. A brief introduction of each tool will be made, as well as the most common advantages and disadvantages when working with this tool will be explained and finally for each tool it will be explained how it can be connected to SAP Hana to automate a process.

2. PENTAHO

Introduction to Pentaho

Pentaho is a solution-oriented, process-centric Business Intelligence (BI) platform that includes the components required to implement process-based solutions such as data mining, ETL, reporting etc.

It helps to improve analysis and decision-making capabilities, since it offers solutions that are mainly composed of an infrastructure of analysis and reporting tools integrated with a business process workflow engine.

ADVANTAGES AND DISADVANTAGES

When talking about the **advantages** of this tool we can see that:

- It's open source, so its not necessary to purchase a license for its use.
- It's multi-platform
- It uses standard technologies such as Java, XML or JavaScript
- It has basic development environments in PDI as well as in BI server and other tools.
- Easy to install and configure
- It is a flexible solution that allows the creation of new functionalities or modules that adapt to the needs of the organization for example by the installation of plugins to extend its functionality.
- It has a powerful OLAP query engine, such as Mondrian, which is also Java-based and allows us to perform powerful queries in a tiny amount of time.
- Focused on Big Data and non-relational databases.
- It is compatible with a multitude of data sources and Big Data systems, such as Hadoop and Apache Spark, which makes it very easy to connect it with different sources without having to adapt or go through third-party tools.

When talking about the **disadvantages** of this tool we can see that:

- It is a suite of tools that are designed by different developers, so we are going to have different menus, of different shapes and different visualizations, which can complicate a bit the learning in new Pentaho tools.
- There is not much reliable documentation, so many times we must rely on other users who have had the same error or the same need.
- There is only a basic user manual that does not detail many concepts

CONNECT ETL PROCESSES TO SAP HANA

By default, Pentaho has no connector to SAP databases. Therefore, to connect to these databases it is necessary to use the "Generic Database" connector, which connects via JDBC.

This connection requires a jdbc driver called "ngdbc" that must be inserted in the "lb" folder of the PDI .

It can be obtained from the SAP HANA STUDIO (Eclipse customized to work with SAP) and can also be downloaded from the following URL: https://jar-download.com/?search_box=ngdbc

The fields to be filled in this type of connections are:

- **Custom URL** : In this field you must specify the IP address of the SAP server and the port through which to connect to the corresponding tenant (see documentation or ask the SAP provider for the values). The formula to specify these aspects is as follows: jdbc:sap://<IP address>:<Port>.
- **Driver class**: In this field the following string must be inserted: com.sap.db.jdbc.Driver
- **Username**: The username authorized to access the database.
- **Password**: The password associated with the user.

Settings	
Dialect:	Generic database
Custom connection URL:	jdbc:sap://52.137.7.147:30241
Custom driver class name:	com.sap.db.jdbc.Driver
Authentication	
Username:	SAPHANADB
Password:	●●●●●●●●

IT NOVUM CONECTORS FOR PDI

On the other hand, to connect to SAP from PDI, just know that IT Novum has developed a series of plugins to connect to internal data sources of some SAP solutions, only for SAP ERP, SAP BW and SAP R3/ERP.

All connectors require the same parameters to establish the connection:

- **Host Name:** IP address or URL of the SAP server.
- **System Number:** Two-digit number defining the instance number (usually 80 in systems). To know it you can use the "GET_SYSTEM_NUMBER" function.
- **SAP Client:** The client ID is a three-digit number. It can be obtained with the transaction "SC4" in the SAP GUI
- **Username:** User with permission to access the SAP system.
- **Password:** The password associated with the user.

Settings	
Host Name:	10.0.0.148
System Number:	00
SAP Client:	001
Language:	EN
Username:	DEVELOPER
Password:	●●●●●●●●

In this case for making the installation refer to the IT-Novum documentation.

To make the connection work, PDI must be installed on the virtual machine that had the SAP client and is linked to the server on which the system is running.

DATA EXTRACTION FROM AN SAP ERP OR SAP R3 SYSTEM – IT NOVUM

These connections are used to connect to ERP systems or to R3 systems. The connectors are used by the "SAP ERP Table Input" component, which is responsible for extracting data from the SAP ERP tables.

To use this tool, a connection to one of the two systems must be specified. The connection is made through the different connectors provided by IT Novum in its plugins.

Once the connection is established, the table from which the data is to be extracted must be identified. This identification can be done by entering the complete name of the table or by leaving the field blank and selecting the magnifying glass to the right of the field; in this case a pop-up window will appear with all the tables that you want to extract the data from.

Step name

Connection

Table

When the table is selected, in the lower part of the window will appear all the fields that the table has, the table with its name, the type of data, both for SAP and its transformation into PDI, its length and its field description. The "Convert" column specifies whether the data has been passed to the Java field type (specified in the "PDI-Data Type" field). This conversion can be selected with the different buttons below the fields.

These fields can be removed from the window in order not to include them in the ETL process. To do this, click on the field and press the delete button.

Table /BOBF/D_PR_NAME

#	Name	SAP-DataType	PDI-DataType	Length	Convert	Description
1	MANDT	C	String	3	YES	Client
2	DB_KEY	X	String	16	YES	NodeID
3	PARENT_KEY	X	String	16	YES	NodeID
4	LANGUAGE_CODE	C	String	1	YES	Language Key
5	NAME	C	String	255	YES	EPM: Text field for names and descriptions
6	DESCRIPTION	C	String	255	YES	EPM: Text field for names and descriptions

At the bottom of the window there are three buttons. The "Convert Yes" and "Convert No" buttons are the ones that define the value of the "Convert" column. Finally, the get fields button is used to retrieve the deleted fields.

POSITIVE AND NEGATIVE ASPECTS REGARDING THE SAP CONECTION

Pentaho is a powerful tool when offering ETL solutions that allow to integrate information in Data Lakes, Data Warehouses as well as to create applications and analytical projects. In addition, it also allows to extract all the necessary information of the SAP systems to be able to establish these connections.

However, since it does not have a specific connector for the SAP environment, it can be difficult to find documentation on errors and more specific information on how to use this component in the environment itself.

Although there is the possibility of installing a plugin in PDI to connect specifically to SAP solutions the process is somewhat complex so it is not worth this solution considering that there are other tools that can solve the automation of the process in a simpler way.

3. TALEND

Introduction to Talend

Talend Open Studio (TOS) is a suite that provides a very complex, varied and complete set of tools to perform data integration offered in an open-source version. It is also a Data Integration (ETL) platform that manages and implements new digital processes in companies, which will give you a competitive advantage.

It allows processes such as advanced analytics and decision making, it is a tool that gives Artificial Intelligence to existing or new processes, as well as implementation and improvement of ecommerce processes, digital marketing processes focused on target audiences, etc.

ADVANTAGES AND DISADVANTAGES

When talking about the **advantages** of this tool we can see that:

- It is an all-in-one, what allows us to reduce the number of tools and therefore additional configurations
- It has a relatively long-lived and well-structured documentation and community
- The use of generic components allows Talend to connect to virtually any platform
- Talend provides its own framework for developing custom components.

When talking about the **disadvantages** of this tool we can see that:

- The installation of the tool can be complex
- When first using the tool you may notice a steep learning curve due to its wide range of functionalities
- When using the tool, the consumption of machine resources can be quite high.
- The tool does not support the .ods format

CONNECT ETL PROCESSES TO SAP HANA

Talend does offer a specific connector to the SAP Hana database. This connector is also based on the JDBC connection. This connection requires the same jdbc driver that is used in the PDI connection. This driver must be inserted into the folder "configuration\m2\repository\org\talend\libraries\ngdbc" folder inside the Talend directory. The Driver can be obtained from the SAP Hana Studio, it can also be downloaded at https://jar-download.com/?search_box=ngdbc

In the Community version of Talend, this connection can be made with the component "tDBInput"

The fields to be filled in this type of connections are:

- **DB Type:** "SAPHana".
- **Login:** Authorized username to access the database.
- **Password:** The password associated with the user
- **Server:** IP address or URL of the SAP server.
- **Port:** Port through which to connect to the corresponding tenant.
- **Schema:** Database to connect to.

The screenshot shows a configuration window for a SAP HANA connection. The 'DB Type' is set to 'SAPHana'. The 'String of Connection' field contains 'jdbc:sap://52.137.7.147:30241?'. The 'Login' field contains 'SAPHANADB', the 'Contraseña' field is masked with dots, the 'Servidor' field contains '52.137.7.147', the 'Puerto' field contains '30241', and the 'Esquema' field contains 'SAP_HANA_DEMO'. There is an 'Additional parameters' field which is empty. At the bottom right, there is a 'Test connection' button. At the bottom center, there are two buttons: 'Exportar como contexto' and 'Revertir Contexto'.

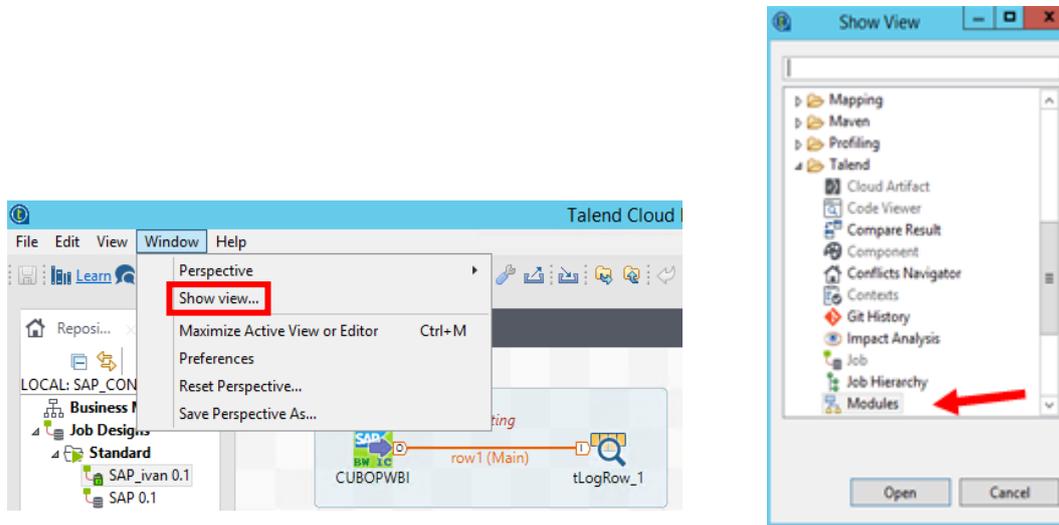
In the paid version of Talend there is a component that performs the same function without having to fill out the DB Type parameter. This component is called "tSAPHanaInput"

This component offers not only the ability to connect to SAP Hana databases but also the functionality to connect to different data structures used by SAP in its SAP Hana databases. With this component it will be possible to connect to any type of SAP system and extract data from its tables, its InfoProviders or its Bapis (SAP standard interfaces)

To be able to establish the talend connections with SAP systems, it is necessary to have in the directory "Talend-Studio v7.2.1\ configuration\ m2\ repository \org\talend\libraries\sapjco3\6.0.0" the driver JAR "sapjco3.jar" and in the "C:\Windows\System32" directory the "sapjco3.dll" file.

Afterwards, open the Talend tool, and in the upper part of the tool, drop down the "Window" option, then select the "Show view" option.

Then a new window will appear with several folders from which you should select the "Talend" folder and click on the "Modules" option.



With the modules window open, select the icon located in the upper right part of the modules window. After clicking on the icon, a new window will open where the path to the "sapjco.jar" file should be displayed.

Status	Context	Module	Description	Required
Insta...	plugin:org.talend.libraries.apache.axis2	activation-1.1.jar		<input checked="" type="checkbox"/>
Insta...	tSAPBWInput	activation-1.1.jar	Required for using this component.	<input checked="" type="checkbox"/>
Insta...	tMicrosoftCrmInput	activation-1.1.jar	Required for using this component.	<input checked="" type="checkbox"/>
Insta...	tMicrosoftCrmOutput	activation-1.1.jar	Required for using this component.	<input checked="" type="checkbox"/>
Insta...	plugin:org.talend.libraries.apache.axis2	activation.jar		<input checked="" type="checkbox"/>
Insta...	tSugarCRMInput	activation.jar	Required for using this component.	<input checked="" type="checkbox"/>
Insta...	tWebService	activation.jar	Required for using this component.	<input checked="" type="checkbox"/>
Insta...	tMDMConnection	activation.jar	Required for using this component.	<input checked="" type="checkbox"/>

On the other hand, with Talend version 7.2 it is not installed by default the SAP connector that is used to extract data from InfoProviders generated by SAP BW. So, without installing this connector you will be able to see that there are InfoProviders created but you will not be able to extract the information stored in those InfoProviders.

The installation process can be quite tedious, as you must work with both the SAP client and the server where it is installed. To install this connector, you must go to the path “\TalendStudio7.2.1\plugins\ org.talend.libraries.sap_7.2.1.20190419_0324 \resources” in it there are several compressed files with different components that can be installed to extend Talend's functionality, and, in addition, there is a text file called "readme"

where the process to be followed to install these features is explained. The file to be used is the "TIDK900022" file which contains two other files: the "K9000220.TID" and "R900022.TID". These two files are the ones you must put in the server where the system is installed.

To access the system, you must connect via SSH, and once inside you must navigate until you reach the directory "trans" to the path "/usr/sap/trans". Inside this directory there are several folders created, but the only ones that are going to be modified are the "data" and "cofiles" folders, since the "cofiles" folder is where the "K900022.TID" file must be inserted and in the "data" folder the "R900022.TID" file must be inserted. Once this operation has been carried out, you will not work with the server again if there are no problems.

When working with the server, you can check with the transaction "AL 11" that the two files have been correctly entered into the system.



In the transaction we will see all the SAP directories and in between of them we will see the "DIR_TRANS" one. If you click in it you can see what it contains

SAP Directories (19.06.2020 10:10:25 NPL vhcainplci)

Name of Directory Parameter	Directory
DIR_ATRA	/usr/sap/NPL/D00/data
DIR_BINARY	/usr/sap/NPL/D00/exe
DIR_CCMS	/usr/sap/ccms
DIR_CT_LOGGING	/usr/sap/NPL/SYS/global
DIR_CT_RUN	/usr/sap/NPL/SYS/exe/uc/linux86_64
DIR_DATA	/usr/sap/NPL/D00/data
DIR_DBMS	/usr/sap/NPL/SYS/SAPDB
DIR_EXECUTABLE	/usr/sap/NPL/D00/exe
DIR_EXE_ROOT	/usr/sap/NPL/SYS/exe
DIR_GEN	/usr/sap/NPL/SYS/exe/dbg
DIR_GEN_ROOT	/usr/sap/NPL/SYS/gen
DIR_GLOBAL	/usr/sap/NPL/SYS/global
DIR_GRAPH_EXE	/usr/sap/NPL/D00/exe
DIR_GRAPH_LIB	/usr/sap/NPL/D00/exe
DIR_HOME	/usr/sap/NPL/D00/work
DIR_INSTALL	/usr/sap/NPL/SYS
DIR_INSTANCE	/usr/sap/NPL/D00
DIR_LIBRARY	/usr/sap/NPL/D00/exe
DIR_LOGGING	/usr/sap/NPL/D00/log
DIR_MEMORY_INSPECTOR	/usr/sap/NPL/D00/data
DIR_PAGING	/usr/sap/NPL/D00/data
DIR_PERF	/usr/sap/tmp
DIR_PROFILE	/usr/sap/NPL/SYS/profile
DIR_PUT	/usr/sap/put
DIR_REORG	/usr/sap/NPL/D00/data
DIR_RSYN	/usr/sap/NPL/D00/exe
DIR_SAPHOSTAGENT	/usr/sap/hostctrl
DIR_SAPUSERS	.
DIR_SETUPS	/usr/sap/NPL/SYS/profile
DIR_SORTTMP	/usr/sap/NPL/D00/data
DIR_SOURCE	/usr/sap/NPL/SYS/src
DIR_TEMP	/tmp
DIR_TRANS	/usr/sap/trans

Directory: /usr/sap/trans/cofiles

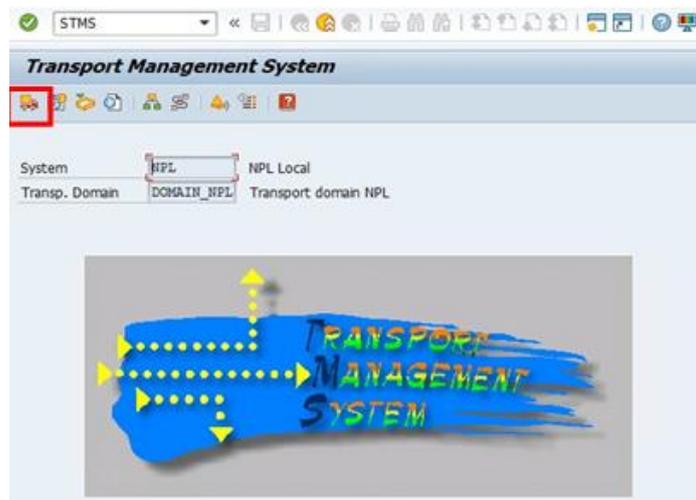
Usable	Viewed	Changed	Length	Owner	Lastchange	Lastchange	File Name
			4096	npladm	17.06.2020	11:17:30	.
			4096	npladm	03.06.2020	17:13:29	.
X			580	npladm	19.06.2020	11:14:05	K900022.TID

Usable	Viewed	Changed	Length	Owner	Lastchange	Lastchange	File Name
			4096	npladm	03.06.2020	17:13:29	.
			4096	root	03.06.2020	17:13:28	..
			4096	npladm	19.06.2020	17:13:29	EPS
			4096	npladm	03.06.2020	17:13:28	actlog
			4096	npladm	19.06.2020	10:13:42	bin
			4096	npladm	19.06.2020	10:46:49	buffer
			4096	npladm	17.06.2020	11:17:30	cofiles
			4096	npladm	17.06.2020	11:18:41	data
			4096	npladm	03.06.2020	17:13:28	etc
			4096	npladm	19.06.2020	10:56:51	log
			4096	npladm	16.06.2020	14:06:17	sapnames
			4096	npladm	03.06.2020	17:13:28	storage
			4096	npladm	19.06.2020	12:02:27	tmp

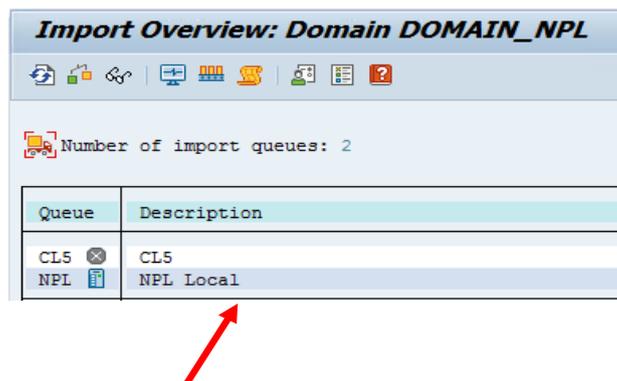
Directory: /usr/sap/trans/data

Usable	Viewed	Changed	Length	Owner	Lastchange	Lastchange	File Name
			4096	npladm	17.06.2020	11:18:41	.
			4096	npladm	03.06.2020	17:13:29	.
X			119306	npladm	17.06.2020	11:16:58	R900022.TID

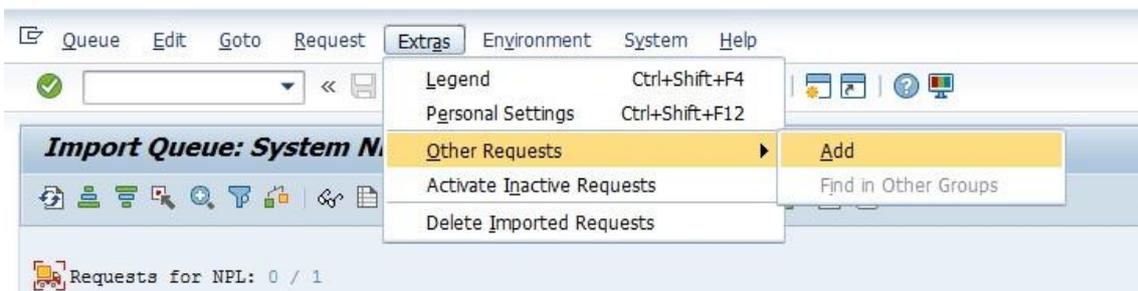
Once you have verified that the two files are in their corresponding directories, you must switch to the "STMS" transaction, where the transaction will be activated with the two files that have been inserted into the system.



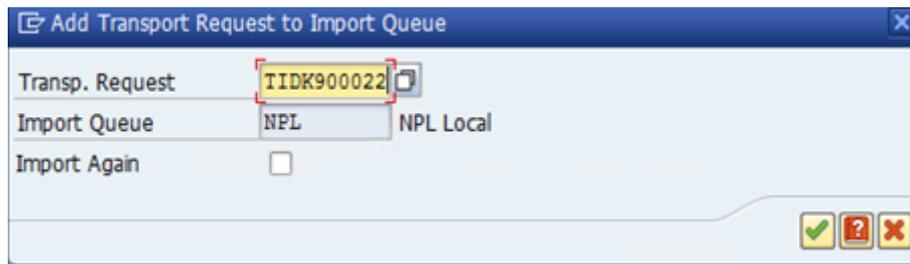
In the new window, select the physical system you are working with by double clicking on the system. You can see the system you are working with from the initial window.



Within the selected system the queue with all transport activation requests is displayed. To create a new transport request, go to the upper part of the window, where you will find the extras option, select the "Other requests" drop-down menu and in the drop-down menu select the "Add" section.



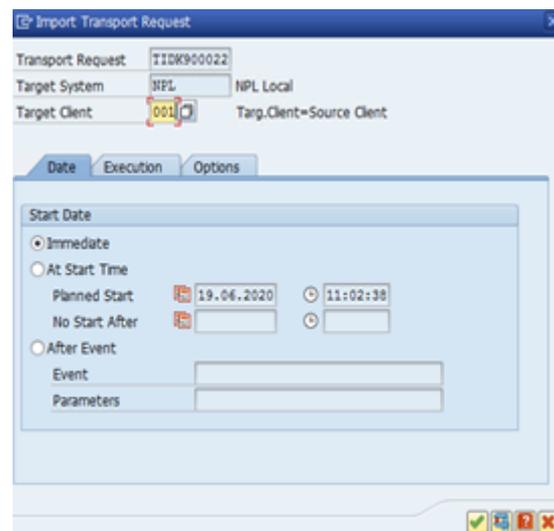
After selecting this option, a pop-up window will appear where the name of the transaction to be added to the queue must be entered. The name of the transaction will be the same as the name of the compressed file you worked with at the beginning of the process: "TIDK900022"



To activate it you must press the f9 key on your keyboard and then press the "ctrl" and "F11" keys at the same time. This request triggers a new pop-up window where you must specify some aspects of the request. In this window it is important to correctly select the user number that is being used in the system.

Requests for NPL: 0 / 1

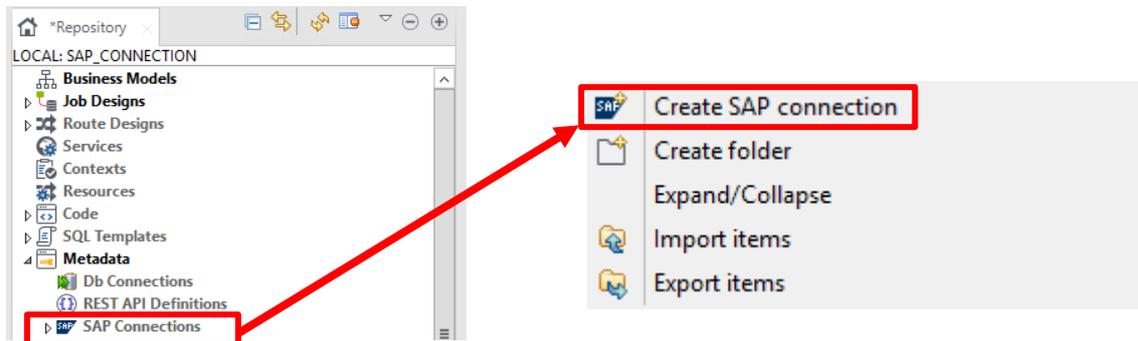
Number	Request	T	QM	RC	I	UMC	Owner	Project	Short Text
1	TIDK900022	R	100	◆	k	1	DKASSEBAUM		



Once the Talend connectors are integrated in SAP you can extract data from the system

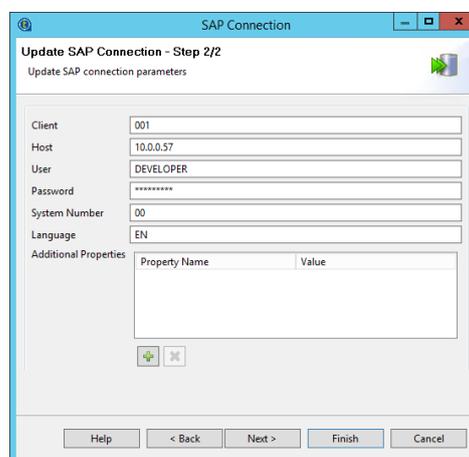
In Talend you can create a connection to an SAP solution in several ways, but the easiest is from the Metadata tool in the "Repository" window (on the left side of the screen). In the Metadata tool there is an option called "SAP Connections", which is the tool that should be used to connect to SAP systems.

Right-click on this tool and you will see a small drop-down menu in which one of the options is "Create SAP connection". Selecting this option will pop up a new window with the parameters to be filled in to make the connection.



The window where the connection parameters must be filled in has three steps:

- The first step is where you name the connection and where you can put a small description of the connection.
- The second step is where you fill in the parameters needed to establish a connection to the system:
 - **Client:** This is the ID of the client and is a three-digit number. It can be obtained with the transaction "SCC4" in the SAP GUI.
 - **Host Name:** IP address or URL of the SAP server.
 - **User:** User with permission to access the SAP system.
 - **Password:** The password associated with the user.
 - **System Number:** Two-digit number that defines the instance number. To know it you can use the GET_SYSTEM_NUMBER function
 - **Language:** The language in which the system is usually English (EN) or German (DE).
 - **Additional Properties:** It is recommended to add the property "api.use_z_talend_read_table" with its value "true" so that there are no problems with the maximum data length supported by Talend when reading tables.



- The third step is optional and is to connect to the SAP Hana database linked to the system.

SAP Connection

Update SAP Connection - Step 3/3

⚠ Db Host must be specified

SAP Hana Database connection parameters

Db Host

Db Port

Db Schema

Db Username

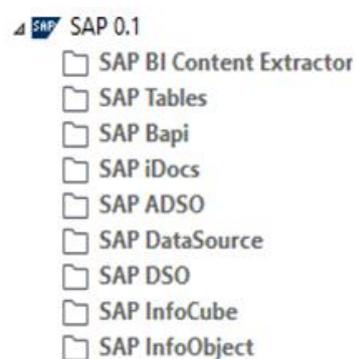
Db Password

Check

Export as context Revert Context

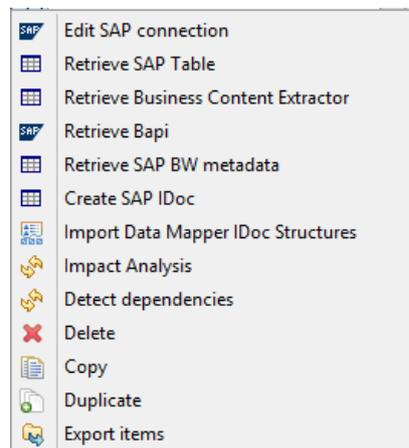
Help < Back Next > Finish Cancel

After creating the connection, if we display the option in the "SAP Connections" tool located at Metadata, you will be able to visualize the connection created. This connection has a drop-down appearance and if it is displayed, all the types of data structures that the SAP system has from which data can be extracted are shown.



To extract the data from these structures, right-click on the connection (in the case of the example on "SAP 0.1") and select the structure from which you want to extract the data.

The option we are going to work with in this document is the "Retrieve SAP Table", whose functionality is to extract data from a table in the SAP system.



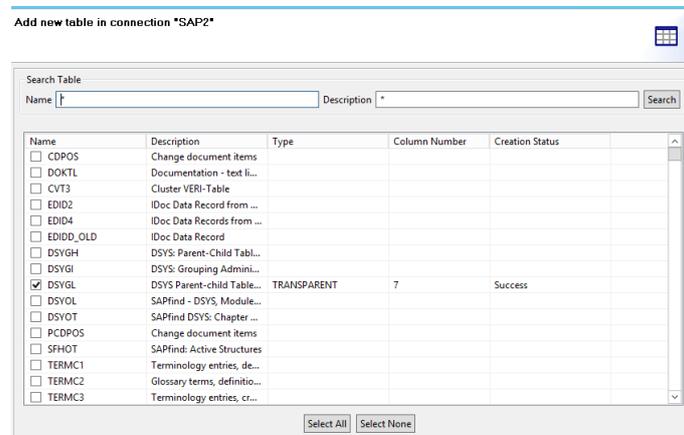
RETRIEVE SAP TABLE

With this tool you can extract data from the internal tables that the SAP solution has on the server.

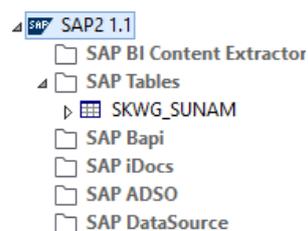
When you select this tool a new window pops up with a Wizard with several steps in which you will specify several parameters to find the data you want to extract.

- In the first step you identify the tables from which you want to extract data. The window where you select these tables has in the upper part two spaces where you can insert the table name or its description. If neither the name nor the description is known, you can fill it with an asterisk and all the tables in the system will be selected. To perform the search, you must click on the search button to the right of these fields. Below the two fields there is a box in which all the tables corresponding to the search will appear. The name and description of these tables are displayed and if selected, the number of columns will also be displayed.

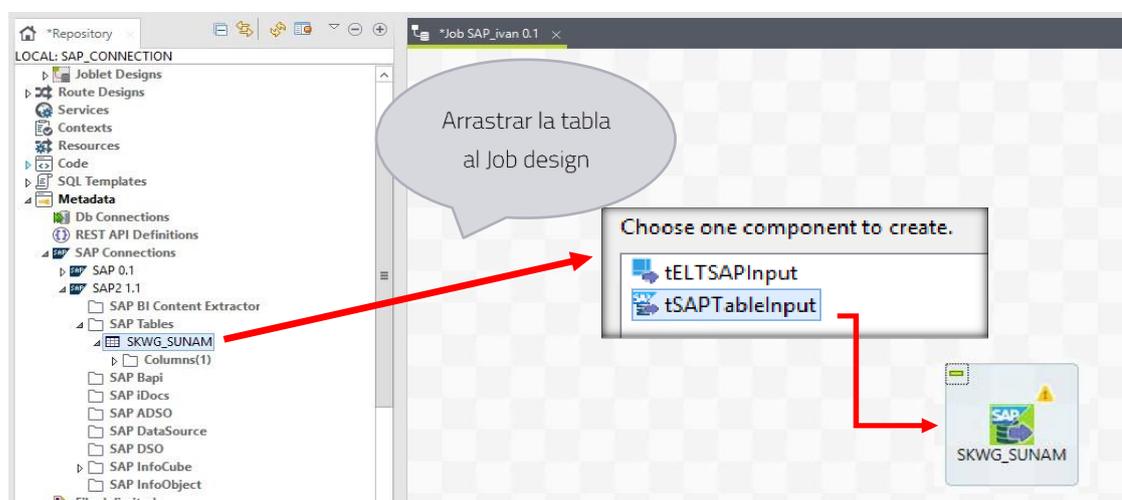
- The second step shows the fields contained in the selected tables and a small "Preview" of their data.



Once the selection of tables is finished, deploying the connection to the system, the section of SAP Tables will also be displayed, and you will be able to see all the tables in the system that have been selected.



To work with them in the ETL process, all you must do is drag the table to the job design you are working on. When the table has been dropped on the job design, a new window will appear showing the tools that can be used to work with the table. Select the option "tSAPTableInput".



POSITIVE AND NEGATIVE ASPECTS REGARDING THE SAP CONNECTION

Talend, therefore, is a complete low-code platform that unifies and manages the organization's data lifecycle, backed by professional services and ecosystem partners and built for cloud, multi-cloud and hybrid environments.

In this case, unlike Pentaho, Talend does not have a specific connector for the SAP environment, both in its community version and in its paid version, which could be a more successful solution when looking for a tool that allows us to connect to SAP to automate a process.

In most cases the community version is enough, it all depends on the SAP solution to which you want to connect and the complexity of the process you want to perform. On the other hand, from the paid version of Talend, the steps to establish the connection itself are somewhat more complex to accomplish so it can be tedious to use this tool if the process to be performed is something simple that could be done with another tool.

In summary, considering that the installation of the tool is complex, that it is even more complicated to use it due to its great variety of functionalities and that the consumption of machine resources can be quite high, it is not worth using this tool for simple processes that could be executed with another tool.

4. POWER AUTOMATE

Introduction to Power Automate

Power Automate is an enterprise system from Microsoft, it is part of Microsoft Power Platform and allows us to integrate and synchronize all our data analysis and applications in an automated way, with the objective of increasing productivity and business efficiency.

In this way, we can simplify repetitive tasks and business processes so that we can focus our attention on what is most needed.

ADVANTAGES AND DISADVANTAGES

When talking about the **advantages** of this tool we can see that:

- There are many templates available
- Is extremely easy to use, even for non-technical users
- Easily integrates with other applications thanks to the plenty of built-in connectors that it gathers and the possibility of creating custom connectors
- Power Automate helps users to automatize and prioritize their tasks
- It allows users to easily share and access data by connecting your favorite services together and creating a flow between them.

When talking about the **disadvantages** of this tool we can see that:

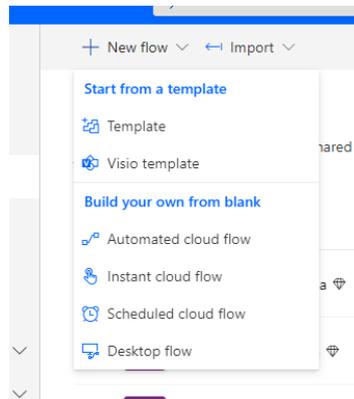
- When creating complex workflows, it affects the performance of the program considerably
- Maximum flow frequency for a free plan is 15 minutes as compared to app-based plan which is 1 minute. Flow frequency is the time it takes for the flow to trigger.
- Power Automate supports only sequential workflows.
- You can only save workflow instances for 30 days which might not be adequate for business-critical processes.
- You need a premium license to access to the SAP connector and use Power Automate Desktop

CONNECT ETL PROCESSES TO SAP HANA

- **Power Automate Cloud service:**
 - To make the connection to SAP HANA it is necessary to install an on-premises gateway as well as create a user, with certain permissions and

a password through which we will connect to the SAP system from Power Automate.

- Then we will select the type of flow that we want for our process to be. In this case we are going to select an instant flow so when a user for examples clicks a button the flow will run.



Build an instant cloud flow

Triggered manually from any device, easy-to-share instant flows automate tasks so you don't have to repeat yourself.

Examples:

- Get an automatic mobile alert whenever a VIP client emails you
- Save all your email attachments to a folder automatically

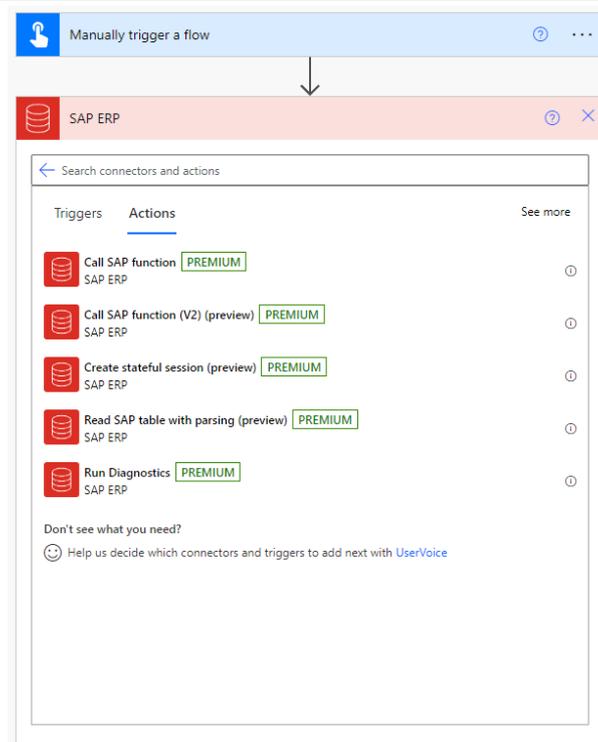
Flow name: SAP Flow

Choose how to trigger this flow *

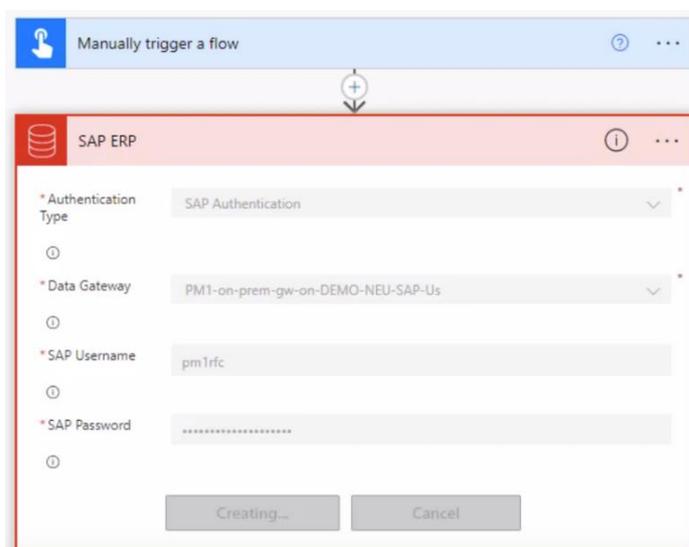
- Manually trigger a flow (Flow button for mobile)
- PowerApps
- When Power Virtual Agents calls a fl... (Power Virtual Agents)
- When a flow step is run from a busin... (Microsoft Dataverse)
- For a selected message (V2) (Microsoft Teams)
- From the compose box (V2) (Microsoft Teams)
- When someone responds to an ada...

Buttons: Skip, Create, Cancel

- Once we create the flow we select the next step option, and we will look for the SAP ERP connector. The SAP ERP connector for Power Automate allows us to connect our workflows to an SAP ERP system (either SAP ECC or SAP S/4HANA). Once we chose the connector we can see that there are plenty of actions available. We need to take into account that we will need a Power Automate license to use this type of connector.

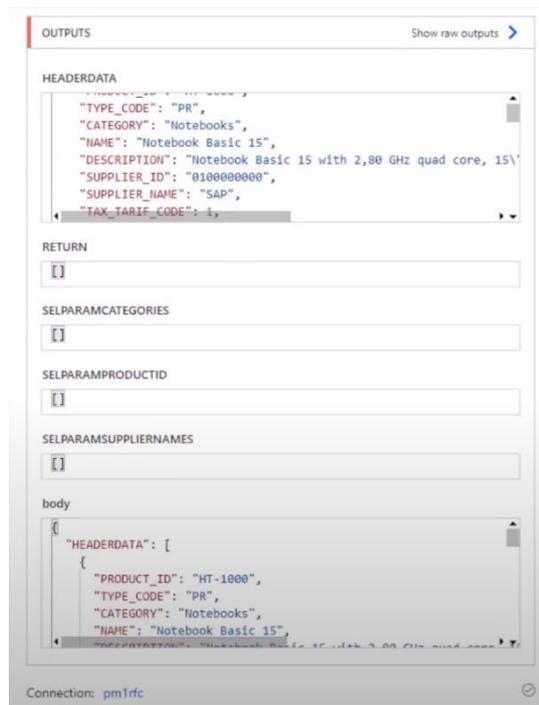


- For testing the connection, we are going to select the one that's called "Call SAP function."
- Before filling the fields for the proper action, we are going to add the SAP connection to the Power Automate flow. For this connection there are different fields:
 - Authentication Type: SAP Authentication
 - Data Gateway: The one-premise gateway that we have created previously
 - SAP Username
 - SAP Password



- With this step we are checking our connectivity to the on-premises data gateway
- Once we establish our connection we can fill in the fields of the action that we have selected previously. The fields that we need to fill in are:
 - **As Host:** This is the Internal Ip address or internal host name
 - **Client:** This is the SAP Client ID
 - **As System:** This is the SAP System Number, that it is a number ranging from 00 to 99
 - **SAP function name:** You can select different functions for your flow depending on what you want to do with your data from SAP. When you select a specific function Power Automate automatically fetches all the relevant properties or fields that are in the system for that specific function.

- Once we finish configuring our action we can test our flow with the test button and see not only if it works properly, but also the different data that it returns when calling the SAP function.



- Once we have set this action and verified that the flow returns data from the SAP environment we can continue to configure our data flow, depending on our business requirements.

In terms of licensing for using this solution we will need a Power Platform license (which in a lot of cases is already part of our Microsoft 365 license) and a license for using premium connectors because the SAP ERP connector is a premium connector. And on the SAP side we will need to have also the appropriate licenses for accessing the environment

POSITIVE AND NEGATIVE ASPECTS REGARDING THE SAP CONNECTION

Power Automate is a tool to create simple workflows both in the cloud and in local environments with the Power Automate Desktop application. For our connection with SAP, we need only the cloud Power Automate tool.

As for the connection with SAP, if what you want is to create a simple workflow that connects to SAP and returns a file you could use this tool because it works for simple flows, and it is extremely easy to use even for non-technical users. However, it should be noted that a paid license is required to use it, since the connector that connects to SAP to take the data is a premium one.

On the other hand, it should also be noted that it requires an on-premises gateway (required version: December 2019 or higher) and an SAP .NET Connector SDK 3.0 from SAP. The access to the download requires a valid S-user. If you do not have an S-User account in the SAP environment it will be impossible to make the connection as it is

required as well as the on-premises gateway and therefore you will not be able to use the tool to create the workflow.

Also, there are some known issues and limitations for the SAP ERP connector:

- The connector supports only RFC's and BAPIs
- The connector does not support receiving messages from SAP Server
- Transactional RFCs (tRFCs) are not supported

If there is a lot of data to load into the file or the flow has a lot of steps, it can take much longer than expected and the flow can become an inefficient process.

5. AZURE DATA FACTORY / SYNAPSE ANALYTICS

Introduction to Azure Data Factory / Synapse Analytics

Azure Data Factory is Azure's cloud ETL service for horizontally scalable serverless data integration and transformation. It offers a code-free user interface that supports intuitive creation, monitoring and management from a single pane of glass.

Azure Synapse Analytics is a limitless analytics service that brings together data integration, enterprise data warehousing, and big data analytics. It gives you the freedom to query data on your terms, using either serverless or dedicated options—at scale.

ADVANTAGES AND DISADVANTAGES

- When talking about the **advantages** of this tool we can see that:
 - **No-code data workflows:** Configure Azure Data Factory to collect and integrate data from most data sources without having to write a single line of code.
 - **Large collection of data collectors:** Azure Data Factory currently offers nearly 100 prebuilt data connectors to import data from external sources. A large amount of Azure Data Factory's online data collectors can be set up instantly.
 - **Built-in monitoring and alerting:** Azure Data Factory offers built-in monitoring visualization. These native visibility features mean you can easily keep track of the status of data integration operations. On top of this, it helps the user be proactive about identifying and reacting to problems, such as a failed data transformation, that could disrupt workflows. You can also set up alerts to warn about such failed operations.
 - **Consumption-based pricing:** Unlike on-premises data integration tools, which typically require a large upfront investment, Azure Data Factory offers pay-as-you-go pricing.
 - **Fully integrated with Azure Data services**
 - **Drag and drop UI**
- When talking about the **disadvantages** of this tool we can see that:
 - **Custom data collectors:** While you can create data pipelines based on a variety of common sources without writing code in Azure Data Factory, you'll need to write custom code to configure nonstandard data sources.
 - **Focus on Azure:** Azure Data Factory supports some data sources hosted outside of Azure, but it's designed first and foremost for building

integration pipelines that connect to Azure or other Microsoft resource types.

- **Long-term expense:** While consumption-based pricing is attractive in some ways, its long-term total cost of ownership may be higher than that of on-premises options. If you plan to run data integration services for years, you may save money by hosting it on your own infrastructure.

CONNECT ETL PROCESSES TO SAP HANA

CONNECTOR SAP HANA

The SAP HANA connector is suitable for ingestion of data from SAP Hana databases. It works with SAP on premise and SAP on cloud, and it can connect also to information models like analytics and calculated views and has the capability of connecting to row column tables.

It runs on the Azure Data Factory self-hosted integration runtime, so it needs to be run on a virtual machine, specifically in a Windows Virtual Machine.

For connecting to SAP HANA, we need to install two things:

- SAP HANA ODBC driver installed in the virtual machine
- Azure Data Factory self-hosted integration runtime installed

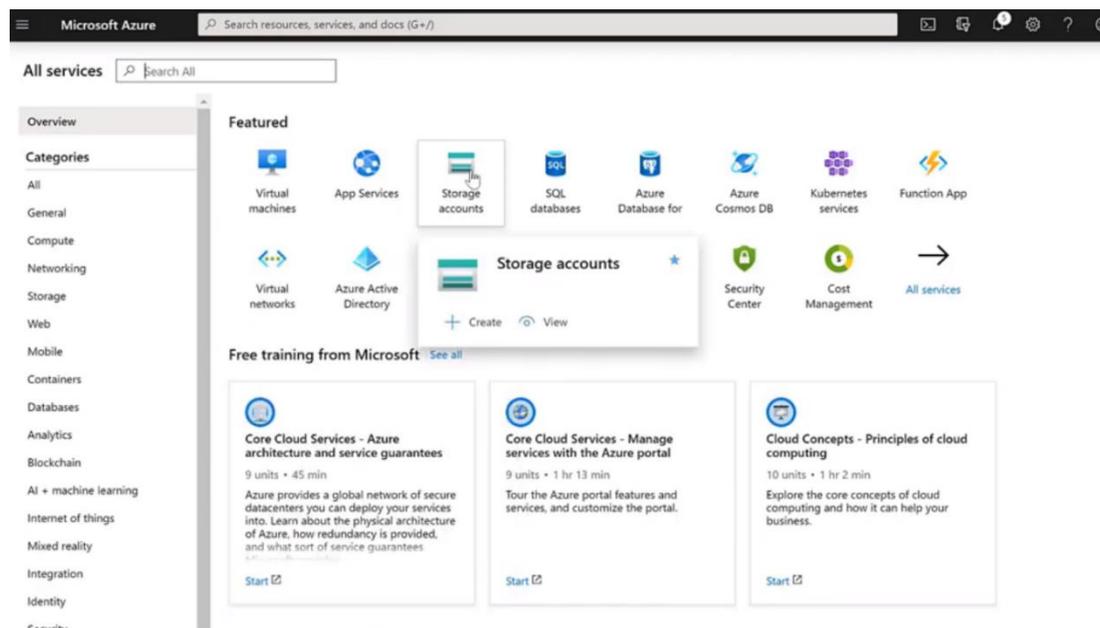
The virtual machine needs to be in the same VirtualNet with SAP, and it must be able to connect to SAP using the HANA ODBC driver, so it must be in the same network within the SAP environment.

In summary, to be able to connect to S/4HANA we must follow these requirements:

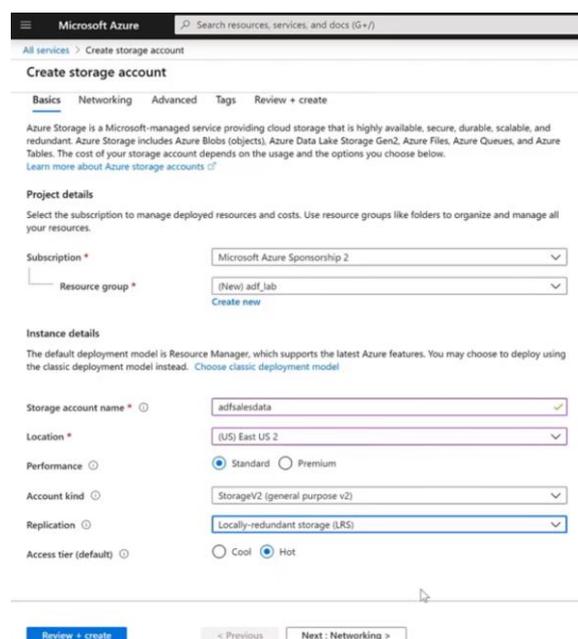
- We need an Azure subscription
- We need access to our SAP HANA environment
- We need to have installed the SAP HANA client 2.0 which is the ODBC driver to connect to HANA (<https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-sap-hana>) and you will need an Azure user ID on the SAP marketplace in order to download this

CONNECT AND COPY SAP HANA DATA WITH AZURE DATA FACTORY

- First thing that we need to do is go to Azure subscription and create a storage account and fill in all the fields

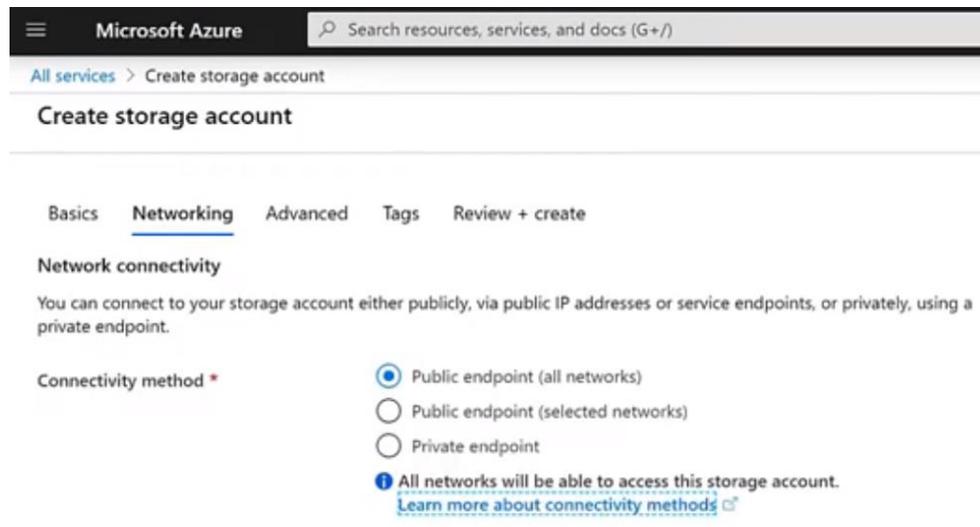


- For the Basics page we will need to configure:
 - **Resource Group:** A container that holds related resources for an Azure solution. In our case we are going to create a new one.
 - **Storage Account Name:** A name for the storage account
 - **Location:** Where the SAP HANA instance is at, so that would be within the same data center
 - **Replication:** How you can replicate the data in your storage

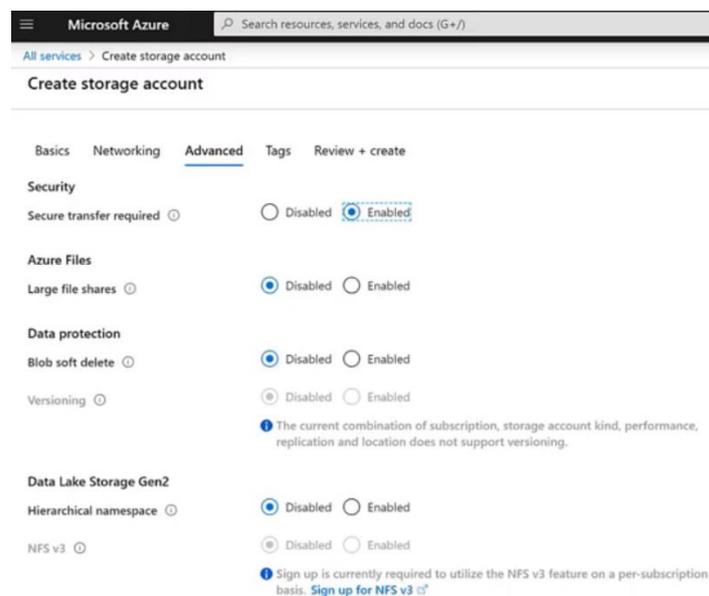


- For the Networking page we will need to configure:

- **Connectivity method:** Where you can connect your data



- For the Advanced page we will need to configure:
 - **Security**
 - **Large files shares**
 - **Blob soft delete**
 - **Hierarchical namespace**



- For the Tag page we will not need to configure anything

Microsoft Azure Search resources, services, and docs (G+)

All services > Create storage account

Create storage account

Basics Networking Advanced **Tags** Review + create

Tags are name/value pairs that enable you to categorize resources and view consolidated billing by applying the same tag to multiple resources and resource groups. [Learn more about tags](#)

Note that if you create tags and then change resource settings on other tabs, your tags will be automatically updated.

Name	Value	Resource
<input type="text"/>	<input type="text"/>	Storage account

- For finishing configuring the storage account we will go to the Review and create page where we can see a resume of the storage account that we have created and once we click the button create it will be automatically created

Create storage account

✓ Validation passed

Basics Networking Advanced Tags **Review + create**

Basics

Subscription	Microsoft Azure Sponsorship 2
Resource group	(New) adf_lab
Location	East US 2
Storage account name	adfsalesdata
Deployment model	Resource manager
Account kind	StorageV2 (general purpose v2)
Replication	Locally-redundant storage (LRS)
Performance	Standard
Access tier (default)	Hot

Networking

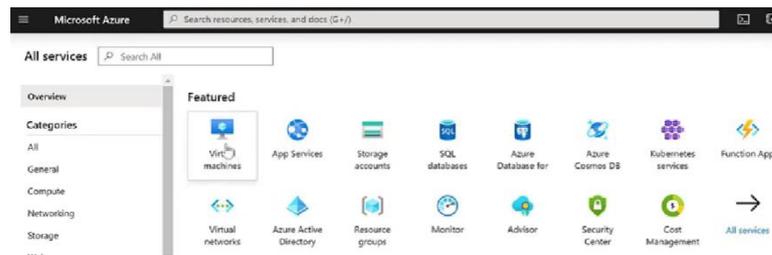
Connectivity method	Public endpoint (all networks)
---------------------	--------------------------------

Advanced

Secure transfer required	Enabled
Large file shares	Disabled
Blob soft delete	Disabled
Blob change feed	Disabled
Versioning	Disabled
Hierarchical namespace	Disabled
NFS v3	Disabled

Create < Previous Next > **Download**

- Now we need to create the virtual machine. We will go back to the “All services” page in Azure and select the virtual machine option.



- For the Basics page we will need to configure:
 - **Resource Group:** A container that holds related resources for an Azure solution.
 - **Virtual Machine Name:** A name for the storage account
 - **Region:** Where the SAP HANA instance is at, so that would be within the same data center
 - **Image:** The type of virtual machine that we want.
 - **Size:** The size of our virtual machine
 - **Username:** The username for log in the virtual machine
 - **Password:** The password for log in the virtual machine
 - **Selected inbound ports:** Configure the ports that we want to select for our virtual machine

Create a virtual machine

Instance details

Virtual machine name * ✓

Region * ✓

Availability options ✓

Image * ✓
[Browse all public and private images](#)

Azure Spot instance Yes No

Size * ✓
4 vcpus, 16 GiB memory (\$274.48/month)
[Change size](#)

Administrator account

Username * ✓

Password * ✓

Confirm password * ✓

Inbound port rules

Select which virtual machine network ports are accessible from the public internet. You can specify more limited or granular network access on the Networking tab.

Public inbound ports * None Allow selected ports

Select inbound ports * ✓

- For the Disks page we will need to configure:
 - **OS disk type**
 - **Encryption type**

The screenshot shows the 'Create a virtual machine' page in the Microsoft Azure portal, specifically the 'Disks' tab. The breadcrumb trail is 'All services > Virtual machines > Create a virtual machine'. The page title is 'Create a virtual machine'. Below the title are tabs for 'Basics', 'Disks', 'Networking', 'Management', 'Advanced', 'Tags', and 'Review + create'. A paragraph explains that Azure VMs have one operating system disk and a temporary disk for short-term storage, and that the size of the OS disk determines the type of storage and the number of data disks allowed. The 'Disk options' section includes:

- OS disk type ***: A dropdown menu set to 'Standard SSD'.
- Encryption type ***: A dropdown menu set to '(Default) Encryption at-rest with a platform-managed key'.
- Enable Ultra Disk compatibility**: Radio buttons for 'Yes' and 'No', with 'No' selected. A note below states: 'Ultra Disk compatibility is not available for this VM size and location.'

 The 'Data disks' section explains that additional data disks can be added or existing ones attached. Below this is a table with columns: 'LUN', 'Name', 'Size (GiB)', 'Disk type', and 'Host caching'. There are two links: 'Create and attach a new disk' and 'Attach an existing disk'. At the bottom, there is a 'Advanced' section that is currently collapsed.

- For the Networking page we will need to configure:
 - **Virtual Network**
 - **Subnet**
 - **Public IP**
 - **NIC network security**
 - **Select inbound ports**

The screenshot shows the 'Create a virtual machine' page in the Microsoft Azure portal, specifically the 'Networking' tab. The breadcrumb trail is 'All services > Virtual machines > Create a virtual machine'. The page title is 'Create a virtual machine'. Below the title are tabs for 'Basics', 'Disks', 'Networking', 'Management', 'Advanced', 'Tags', and 'Review + create'. A paragraph explains that network connectivity is defined by configuring network interface card (NIC) settings, including control ports, inbound and outbound connectivity with security group rules, or placement behind an existing load balancing solution. The 'Network interface' section includes:

- Virtual network ***: A dropdown menu set to '(new) adf_lab-vnet' with a 'Create new' link below it.
- Subnet ***: A dropdown menu set to '(new) default (10.19.0/24)'.
- Public IP**: A dropdown menu set to '(new) adfvm-ip' with a 'Create new' link below it.
- NIC network security group**: Radio buttons for 'None', 'Basic', and 'Advanced', with 'Basic' selected.
- Public inbound ports ***: Radio buttons for 'None' and 'Allow selected ports', with 'Allow selected ports' selected.
- Select inbound ports ***: A dropdown menu set to 'HTTP (80), HTTPS (443), RDP (3389)'. A warning message below states: 'This will allow all IP addresses to access your virtual machine. This is only recommended for testing. Use the Advanced controls in the Networking tab to create rules to limit inbound traffic to known IP addresses.'
- Accelerated networking**: Radio buttons for 'On' and 'Off', with 'On' selected.

 At the bottom, there is a 'Load balancing' section and navigation buttons: 'Review + create', '< Previous', and 'Next: Management >'.

- For the Management page we will need to configure:
 - **Boot diagnosis**
 - **OS guest diagnosis**

- Diagnosis storage account
- System assigned managed identity

Create a virtual machine

Basics Disks Networking **Management** Advanced Tags Review + create

Configure monitoring and management options for your VM.

Azure Security Center
Azure Security Center provides unified security management and advanced threat protection across hybrid cloud workloads.
[Learn more](#)

✔ Your subscription is protected by Azure Security Center standard plan.

Monitoring

Boot diagnostics On Off

OS guest diagnostics On Off

Diagnostics storage account * [Create new](#)

Identity

System assigned managed identity On Off

Azure Active Directory

Login with AAD credentials (Preview) On Off

- For the Advanced and Tag page we are going to leave the default settings and then we are going to go to the Review + Create page for finally creating the virtual machine:

Microsoft Azure Search resources, services, and docs (G+)

All services > Virtual machines > Create a virtual machine

Create a virtual machine

✔ Validation passed

Basics Disks Networking Management Advanced Tags **Review + create**

PRODUCT DETAILS

Standard D4 v3
by Microsoft
[Terms of use](#) | [Privacy policy](#)

Subscription credits apply
0.3760 USD/hr
[Pricing for other VM sizes](#)

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

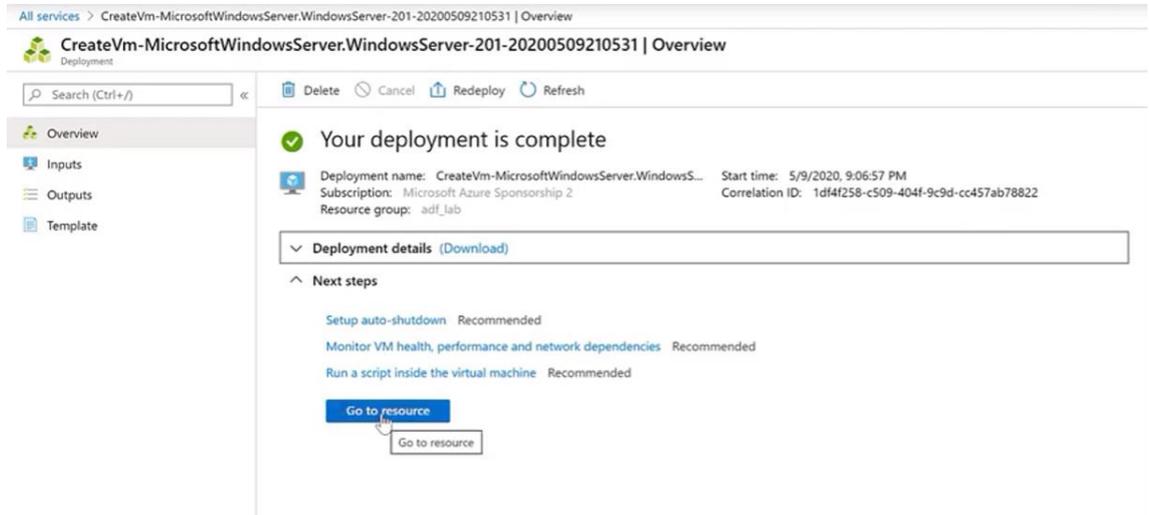
⚠ You have set RDP port(s) open to the internet. This is only recommended for testing. If you want to change this setting, go back to Basics tab.

Basics

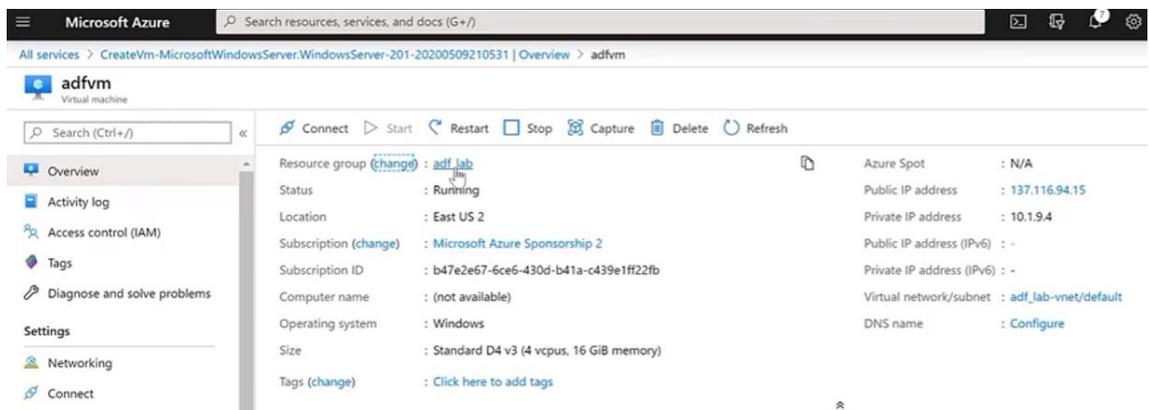
Subscription	Microsoft Azure Sponsorship 2
Resource group	adf_lab
Virtual machine name	advvm
Region	East US 2
Availability options	No infrastructure redundancy required
Username	adadmin

[Create](#) [< Previous](#) [Next >](#) [Download a template for automation](#)

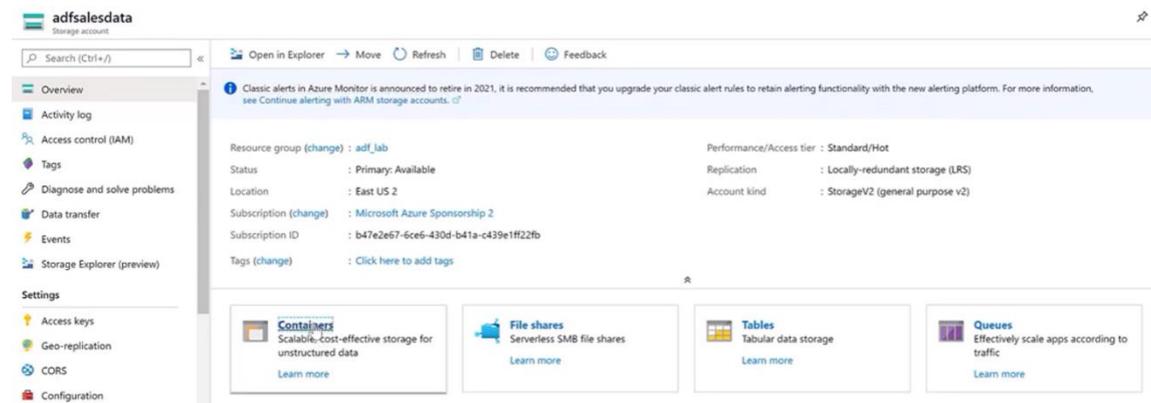
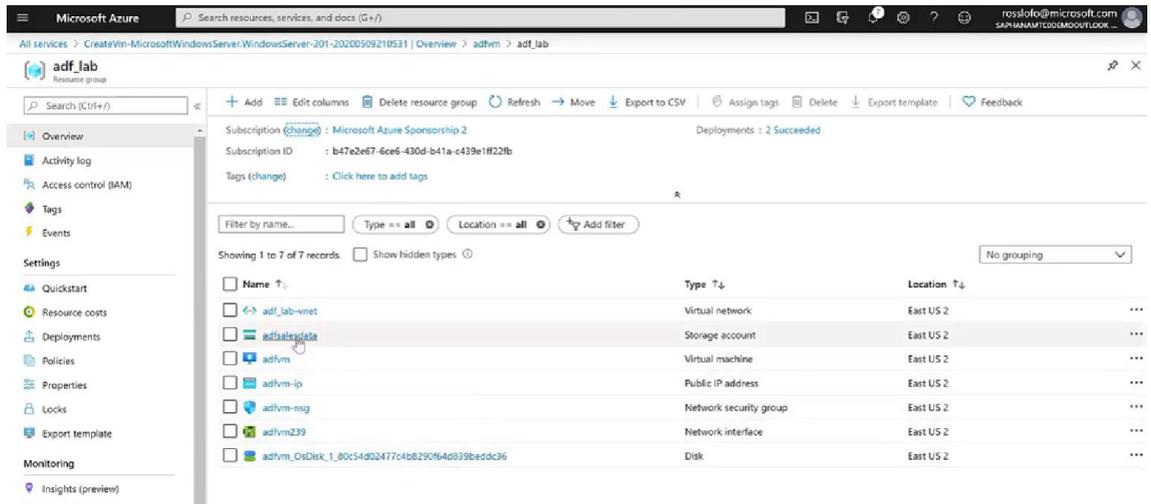
- Now that we have created the storage account and the virtual machine we want to create a container for the storage account. For doing that we are going to click the button “Go to resource”



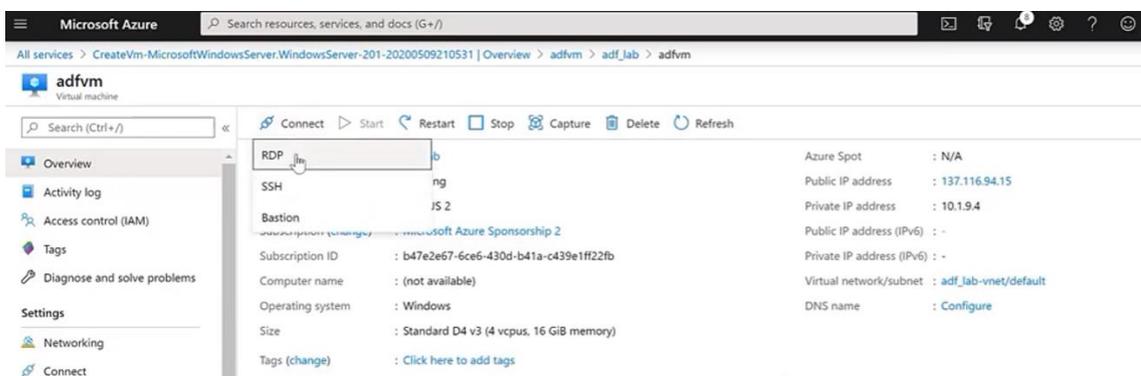
- Then we are going to select the name of the resource group that we have created



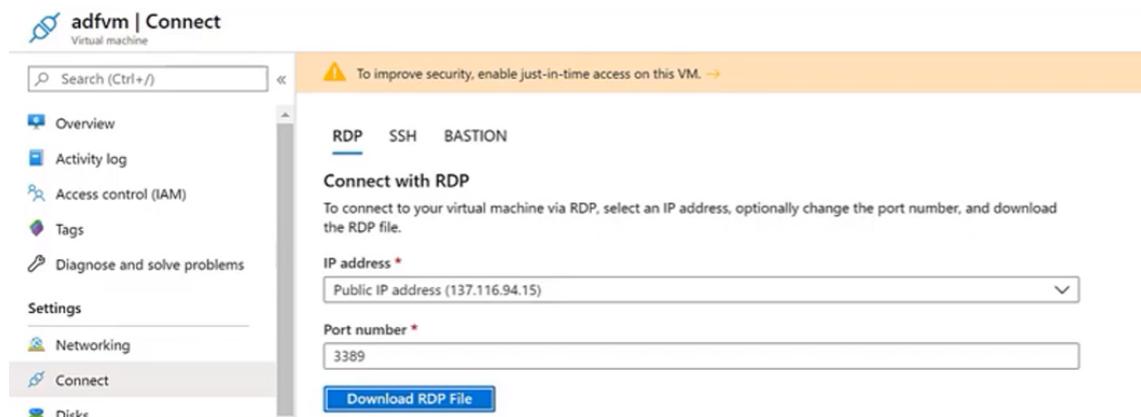
- The next step is to look for the Storage account resource in our resource group



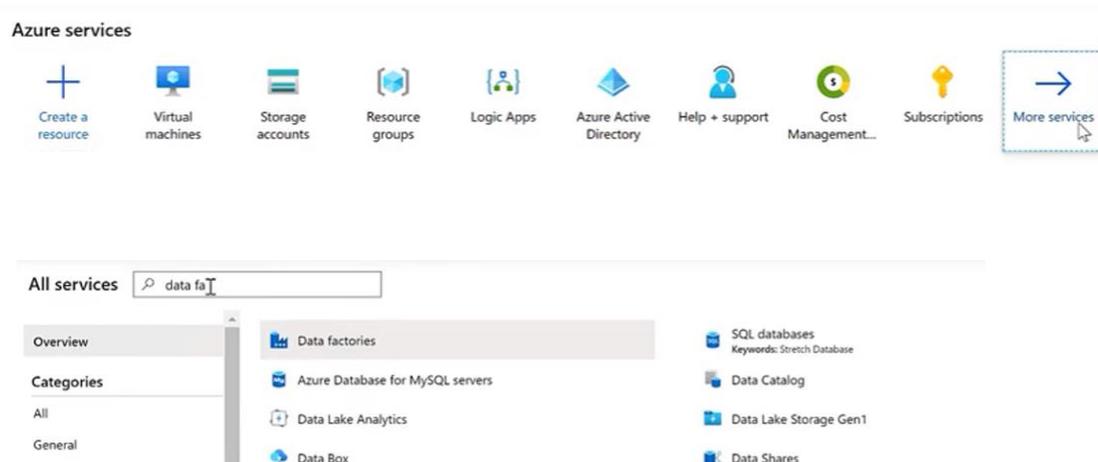
- Now we are going to go back to the virtual machine to connect it to the rest of the work from within the virtual machine. For doing that we are going to click the connect button.



- Then we are going to select the RDP option and download the RDP file. Once it is downloaded we will open the file and we will log in the virtual machine with the user and password that we created previously.



- Now that we are log in our virtual machine we are going to access to the azure portal to do the rest of the work from there. Now we are going to create and configure the azure data factory environment from creating our ETL process.
- For doing that we are going to select the “more services” option and then search for the data factories resource.



- Now we are going to create and configure the new data factory resource in order to create our ETL process.
- We will need to configure:
 - **Name:** Name of the data factory
 - **Version:** Version of the data factory
 - **Subscription:** Type of subscription

- **Resource group:** The resource group which it belongs to
- **Location:** Where the SAP HANA instance is at, so that would be within the same data center

New data factory

Name *
adflabforsap ✓

Version ○
V2 ✓

Subscription *
Microsoft Azure Sponsorship 2 ✓

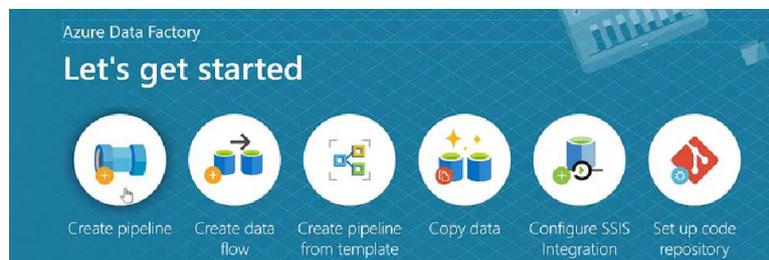
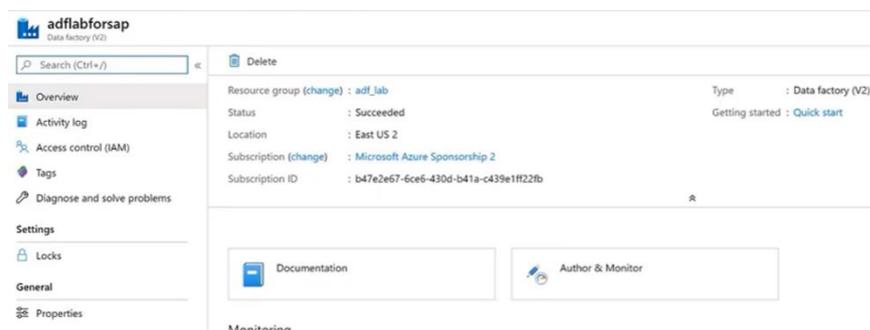
Resource Group *
adflab ✓
[Create new](#)

Location * ○
(US) East US 2 ✓

Enable GIT ○

[Create](#)
Create

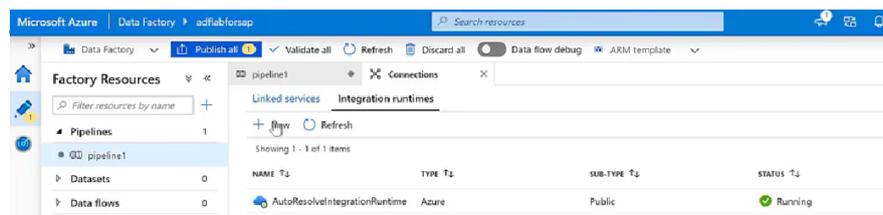
- If we go to the resources page we can see that the data factory was created successfully. Now we are going to create the pipeline for our ETL process. For doing that we need to click the “author and monitor option” and then select the create pipeline button.



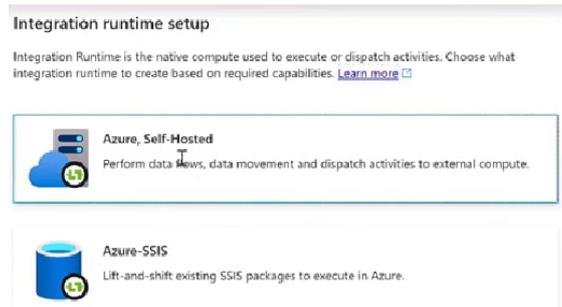
- Once we create the pipeline we need to configure the connection to SAP HANA. For doing that, we need to create a gateway to the SAP environment.



- Once we click the connections button we are going to go to the integration runtimes tab to add another integration runtimes.



- Now we need to select which type on integration runtimes we want. In our case we are going to select Azure, Self-Hosted



- Next thing we must do is to give a name to our integration runtime.

Integration runtime setup

Private network support is realized by installing integration runtime to machines in the same on-premises network/VNET as the resource the integration runtime is connecting to. Follow below steps to register and install integration runtime on your self-hosted machines.

Name * ⓘ

integrationRuntimesap

Description

Enter description here...

Type

Self-Hosted

- The next window that appears is for making the installation of the integration runtime in our case we are going to make the installation with the option on express that will automatically make the installation. The reason that we are doing it within the virtual machine is so that it just stores automatically into it.

Integration runtime setup

Settings Nodes Auto update Sharing

Install integration runtime on Windows machine or add further nodes using the Authentication Key.

Name ⓘ

integrationRuntimesap

Option 1: Express setup

[Click here to launch the express setup for this computer](#)

Option 2: Manual setup

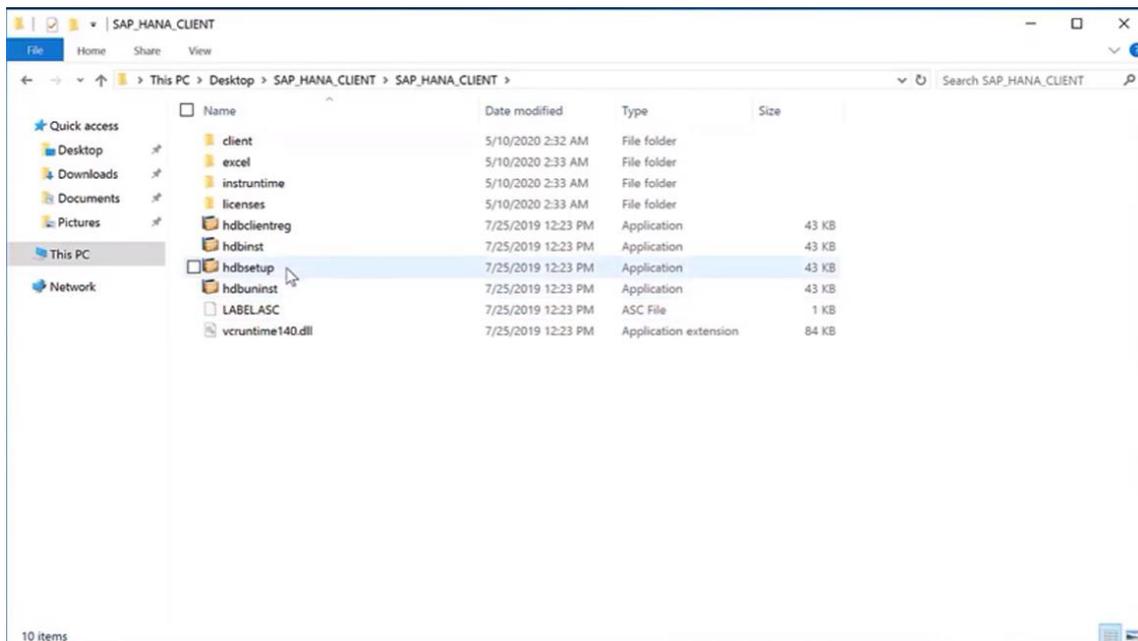
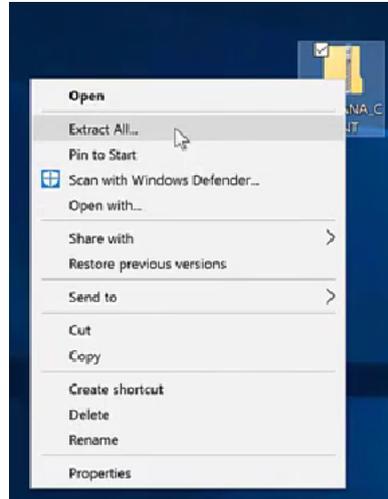
Step 1: [Download and install integration runtime](#)

Step 2: Use this key to register your integration runtime

NAME	AUTHENTICATION KEY
Key1	IR@890481f4-dbdd-4ba3-91b5-256ae17bece8@adflabforsap@eu2@FylJ3  
Key2	IR@890481f4-dbdd-4ba3-91b5-256ae17bece8@adflabforsap@eu2@Zo4l  

- Once the integration runtime is downloaded we click the install button to install it. It is important to know that an integrated runtime could only be associated to one azure data factory, this means that the integration runtime itself could be connected to many data sources but only to one Azure Data Factory.

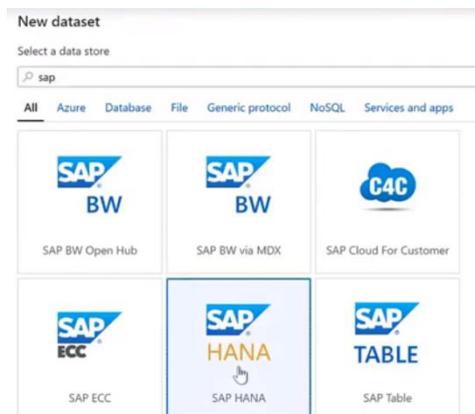
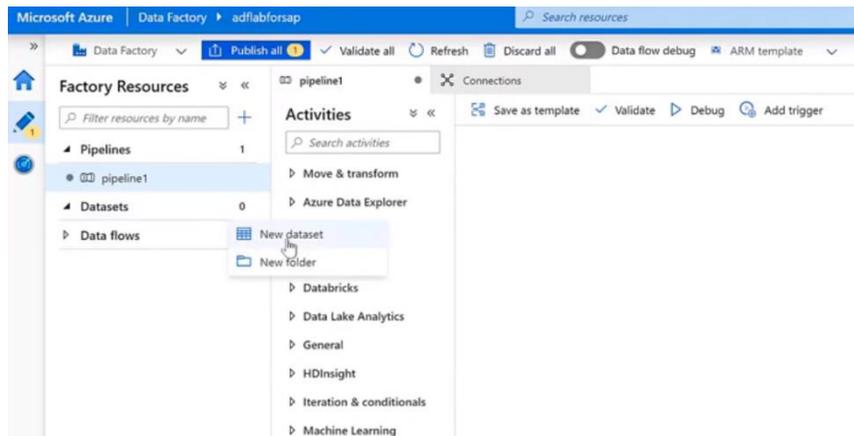
- Now we'd copy the driver in



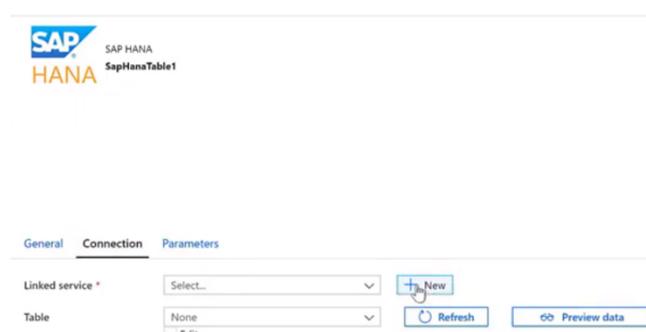
AUTOMATION ETL PROCESS

Once we have set up the necessary configuration we can begin to automate our process with Azure Data Factory. What we are going to do is copy some data from SAP to Azure Blob storage.

- The next step is to create a new dataset and then select the SAP HANA connector.



- Then we are going to create an SAP connection



- For the configuration of the new connection, we need to fill in the different fields:
 - **Name:** The name of the connection
 - **Connection via integration runtime:** The integration runtime that we have created previously
 - **Server name:** The name of the server that we want to connect to.
 - **Authentication type:** The type of authentication that you want to establish
 - **Username**
 - **Password**

New linked service (SAP HANA)

Name *
SapHana1

Description

Connect via integration runtime *
integrationRuntimeSap

Edit integration runtime

Server name *
52.251.31015

Authentication type *
Basic authentication

User name *
DEMOUSER

Password *

Additional connection properties
+ New

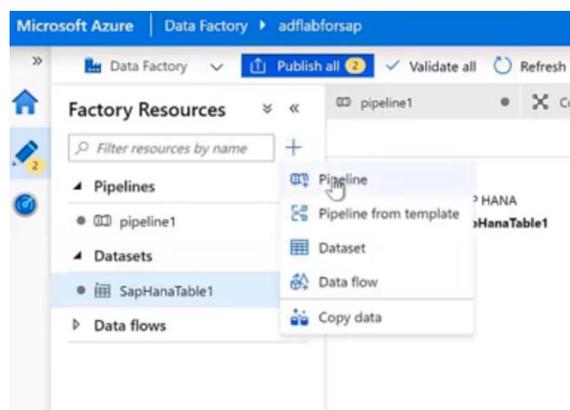
Annotations
+ New

Advanced

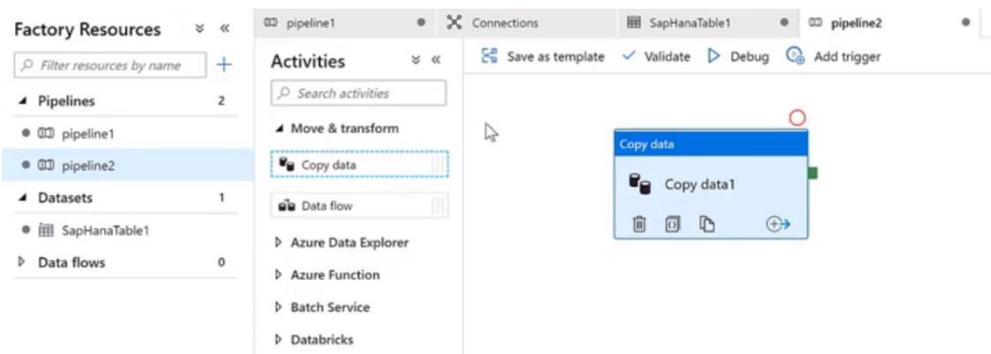
Connection successful

Create Test connection Cancel

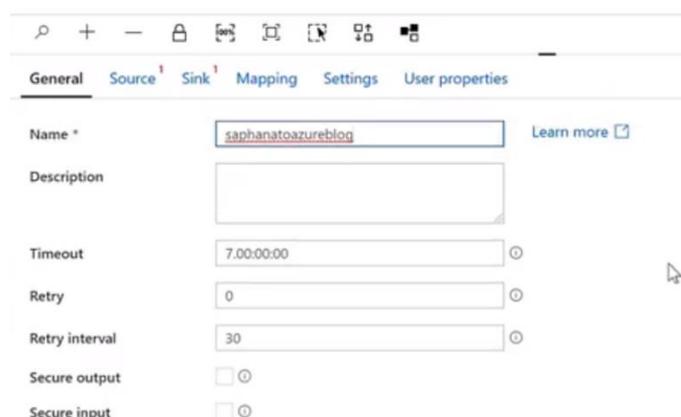
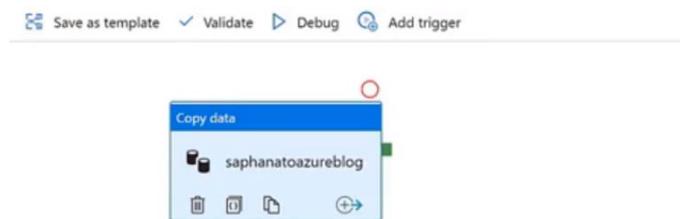
- The next step is to create the pipeline for our process.



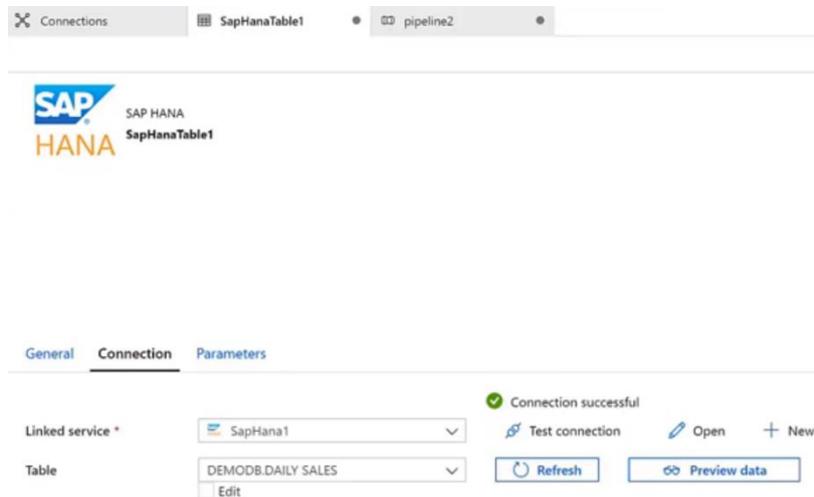
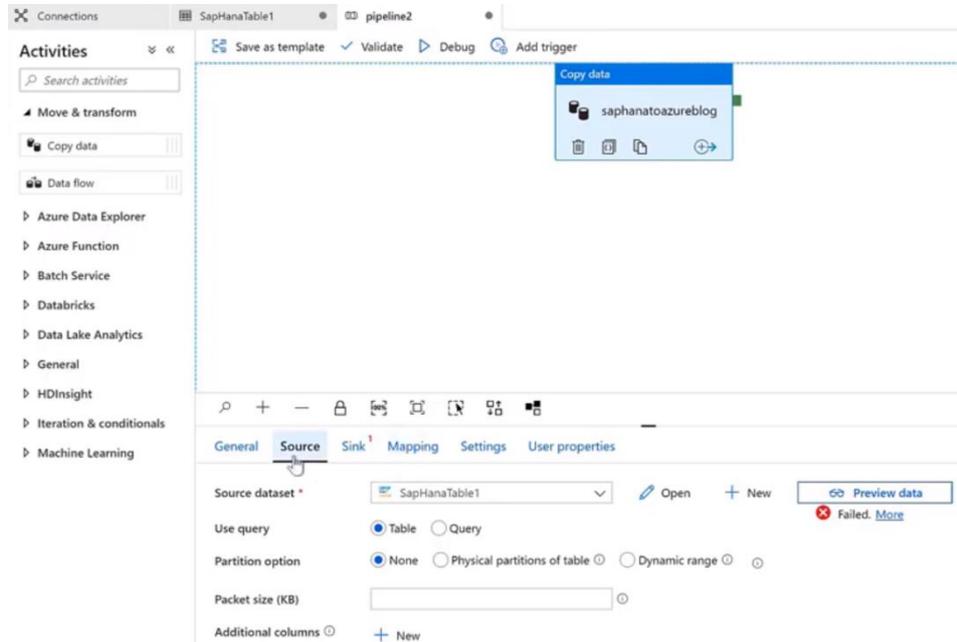
- Once the pipeline is created, since we want to copy the data from SAP to Azure Blob Storage, we will use the activity called Copy data, which is in the Move and Transformation section.



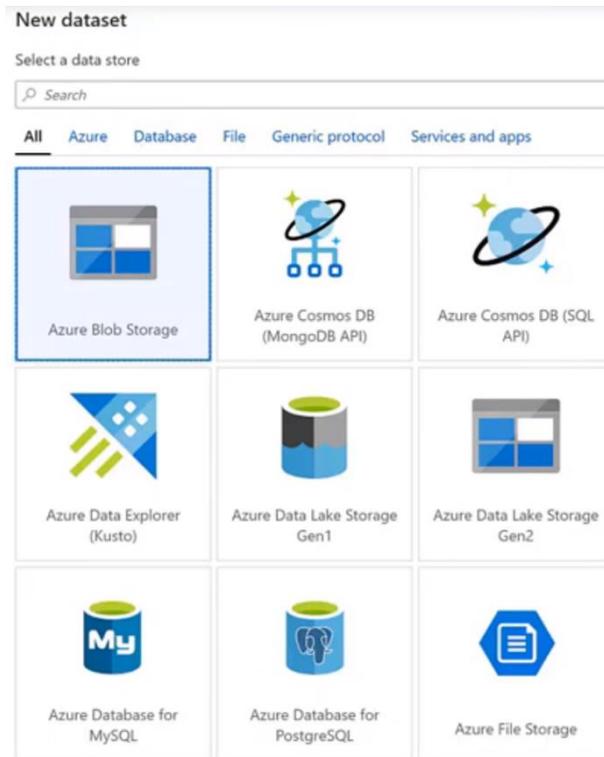
- Now we are going to configure our component. For the General tab we are going to fill in the name of our component.



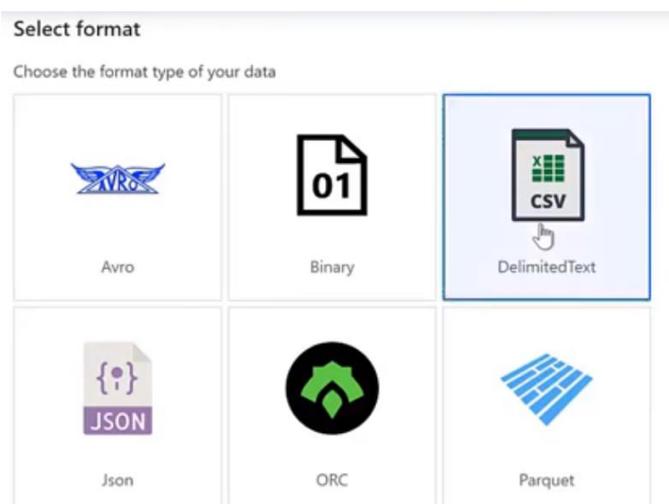
- For the **source** tab we are going to select the SapHanaTable1 that we have created previously when we created our dataset, and we are going to click the open option to select a specific table of our dataset.



- For the **sink** tab we are going to create a new sink that it's where we are going to copy our data from SAP. First we need to create the dataset where we are going to copy our data. In this case its going to be de Azure blob storage.



- And them we need to select the format type of our data



- For the CSV format type, we must set its properties:
 - **Name**
 - **Linked Service:** We are going to create a new linked service that its going to be the Azure Blob Storage where we want to copy our data.

The screenshot shows a 'Set properties' dialog box. The 'Name' field is filled with 'sapdailysales'. Below it, the 'Linked service' dropdown menu is open, displaying a search bar with the placeholder 'Filter...' and a '+ New' button at the bottom left. The dropdown also shows 'Select...' options.

- For the configuration of the new linked service, we need to fill in:
 - **Name**
 - **Connect via integration runtime**
 - **Storage account name**

The screenshot shows the 'New linked service (Azure Blob Storage)' configuration form. It includes a warning message: 'hosted integration runtime is higher than version 4.0 if connecting via self-hosted integration runtime.' The form fields are:

- Name ***: AzureBlobStorage1
- Description**: (empty text area)
- Connect via integration runtime ***: AutoResolveIntegrationRuntime
- Authentication method**: Account key
- Account selection method**: From Azure subscription (selected)
- Azure subscription**: Microsoft Azure Sponsorship 2 (b47e2e67-6ce6-430d-b41a-c439e1ff22fb)
- Storage account name ***: adfsalesdata

 At the bottom, there is a 'Test connection' section with 'To linked service' selected, and a 'Create' button highlighted with a mouse cursor. A 'Connection successful' message is visible at the bottom right.

- To finish configuring our component we need to set the las properties for the linked service that we have just created.

Set properties

Name
sapdailysales

Linked service *
AzureBlobStorage1

Edit connection

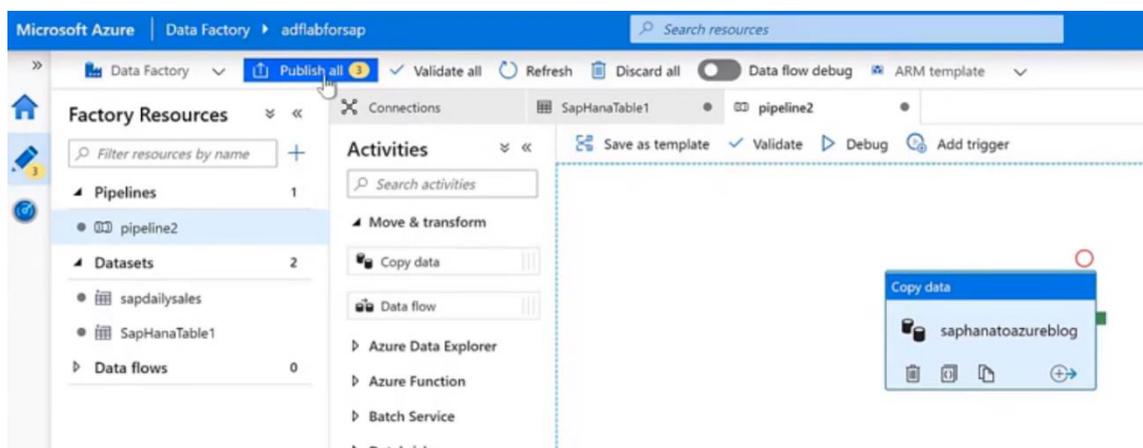
File path
salesdata / Directory / File Browse

First row as header

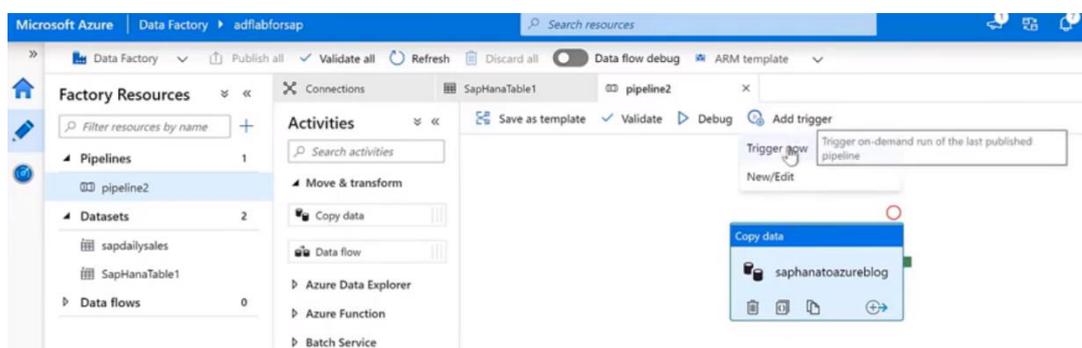
Import schema
 From connection/store From sample file None

Advanced

- The last step is to publish our process so we can test it properly.



- Once we have published it, we can execute it directly from Add trigger.



- And if we go to the monitor tab we can see how our process has been executed correctly and therefore the data has been copied from SAP to Azure blob storage.

PIPELINE NAME	RUN START	DURATION	TRIGGERED BY	STATUS	PARAMETERS	ANNOTATION
pipeline2	5/10/20, 2:43:29 AM	00:00:07	Manual trigger	In progress		

CONNECTOR ODP

The new SAP ODP connector leverages SAP Operational Data Provisioning (ODP) framework, which is an established best practice for data integration within SAP landscapes.

ODP provides access to a wide range of sources across all major SAP applications and comes with built-in CDC (change data capture) capabilities.

Most of the connectors can only extract data in batches, where each batch treats old and new data equally without identifying data changes. This extraction mode isn't optimal when dealing with large data sets, such as tables with millions or even billions of records, that change often. To keep your copied SAP data fresh, frequently extracting it in full is expensive and inefficient.

This connector can extract only data changes using CDC capabilities provided by SAP systems. It can connect to all SAP systems that support ODP, such as ECC, S/4HANA, BW, and BW/4HANA, directly at the application layer or indirectly using SAP Landscape Transformation (SLT) replication server as a proxy.

The connector can fully or incrementally extract SAP data that includes not only physical tables, but also logical objects created on top of those tables, such as Extractors or ABAP Core Data Services (CDS) views, without watermarking.

SAP ODP connector can extract various data source types, such as:

- SAP extractors, originally built to extract data from SAP ECC and load it into SAP BW
- ABAP CDS views, the new data extraction standard for SAP S/4HANA
- InfoProviders and Info Objects in SAP BW or BW/4HANA
- SAP application tables, when using SLT replication server as a proxy

CONNECT AND COPY SAP HANA DATA WITH AZURE DATA FACTORY

Prerequisites:

- Configure SAP systems to use [SAP's Operational Data Provisioning \(ODP\) framework](#). (Acts as a central data persistency layer and supports extraction and replication scenarios for various target SAP applications)
- Become familiar with Data Factory concepts such as integration execution environments, linked services, datasets, activities, data flows, pipelines, and triggers.
- Configure a self-hosted integration execution environment to be used for the connector.
- Configure an SAP CDC linked service.
- Debug SAP CDC connector issues by sending logs from the self-hosted integration runtime environment to Microsoft.
- Become familiar with monitoring data extractions in SAP systems.
- Configuring SAP systems to use the SAP ODP framework

Requirements for SAP systems:

- To support ODP, run SAP systems on SAP NetWeaver 7.0 SPS 24 or later. ([Transferring Data from SAP Source Systems via ODP \(Extractors\)](#)).
- To support full SAP Advanced Business Application Programming (ABAP) Core Data Services (CDS) extractions via ODP, run SAP systems on NetWeaver 7.4 SPS 08 or later. To support SAP ABAP CDS differential extractions, run SAP systems on NetWeaver 7.5 SPS 05 or later. ([Transferring Data from SAP Systems via ODP \(ABAP CDS Views\)](#)).
- 1521883: to use ODP API 1.0
- 1931427: to use ODP API 2.0 supporting SAP hierarchies
- 2481315: to use ODP for the extraction of data from SAP source systems to BW or BW/4HANA systems

SAP user configuration

- Data extractions via ODP require a correctly configured user in SAP systems. The user must be authorized to invoke the ODP API via Remote Function Call (RFC) modules. For more information, please refer to these SAP Support Notes:
 - 2855052: to authorize the use of the ODP API.
 - 460089: to authorize the invocation of ODP RFCs

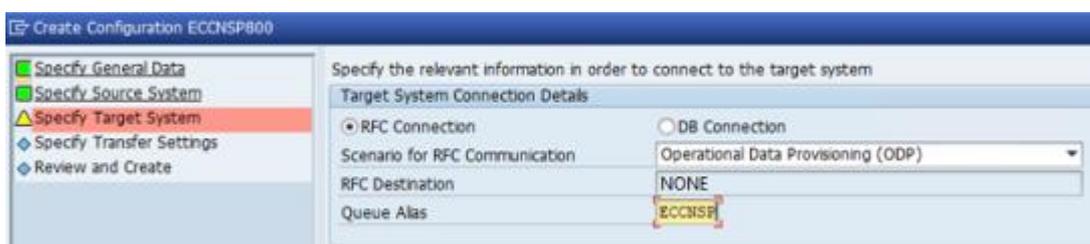
Configuration of SAP Data Origins

To extract data sources, make sure the following requirements are met:

- Ensure that the data sources are activated in the SAP source systems. This requirement only applies to data sources provided by SAP or its partners. Data origins created by customers are activated automatically. If data sources have been or are being extracted by SAP BW or BW/4HANA, the data sources are already activated. ([Installing BW Content Data Sources](#))
- Ensure that the data sources are released for extractions via ODP. This requirement only applies to data sources that are created by customers. Data sources delivered by SAP, or its partners are automatically released. For more information, see these SAP Support Notes:
 - 1560241: to release data sources for the ODP API.

SAP Landscape Transformation Replication Server Configuration

- SAP Landscape Transformation Replication Server (SLT) is a database trigger-enabled CDC solution that can replicate SAP application tables and simple views in near real-time. You can use SLT as a proxy for ODP data extraction. To use SLT as a proxy, complete the following steps:
 - Install NetWeaver 7.4 SPS 04 or later and the DMIS 2011 SP 05 snap-in on the replication server. For more information, see Transferring data from SLT using operational data provisioning.
 - Run the SAP Landscape Transformation Replication Server Cockpit (LTRC) transaction code on the replication server to configure SLT:
 - Under Specify Source System, type the RFC destination that represents the SAP source system.
 - Under Specify Target System, complete these steps:
 - Select RFC Connection.
 - Under Scenario for RFC communication, select Operational Data Provisioning (ODP).
 - Under Queue Alias, type the queue alias to be used to select the context for data extractions via ODP in Data Factory. Use the format SLT-<your queue alias>.



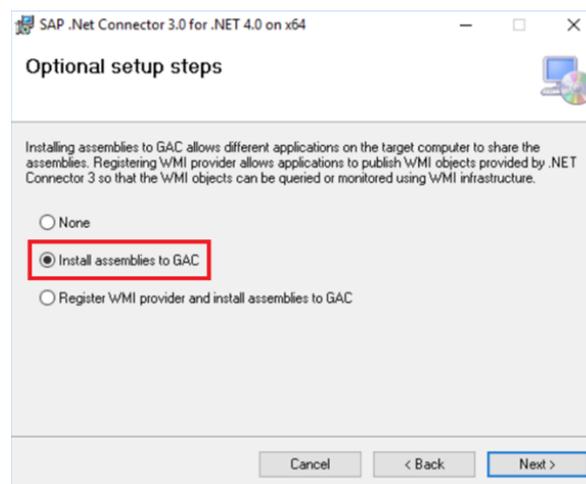
Setting up a self-hosted integration runtime environment for the SAP CDC Connector

Creating and managing a self-hosted integration runtime environment

- In Azure Data Factory Studio, create and configure a self-hosted integration runtime environment. For downloading the self-hosted integration runtime environment, you can go to this [page](#)

Download and install the SAP .NET Connector

- Download the latest version of the [64-bit SAP .NET Connector \(SAP NCo 3.0\)](#) and install it on the machine running the self-hosted integration runtime environment.
- In the Optional Installation Steps dialog box, select Install Assemblies in GAC, and then select Next.



Adding a network security rule

If the SAP system is on an Azure virtual machine, to add the rule you must:

- Source IP addresses/CIDR ranges: Set Source IP Addresses or CIDR ranges to the IP address of the self-hosted integration runtime environment machine.
- Set Destination Port Ranges in 3200,3300.

Checking Connectivity

On the machine running the self-hosted integration runtime environment, run the following PowerShell cmdlet to ensure that you can connect to the SAP systems:

- Test-NetConnection <SAP system IP address> -port 3300

Editing hosts files

Edit the hosts file on the machine that is running the self-hosted integration runtime environment to add the SAP IP addresses to the server names

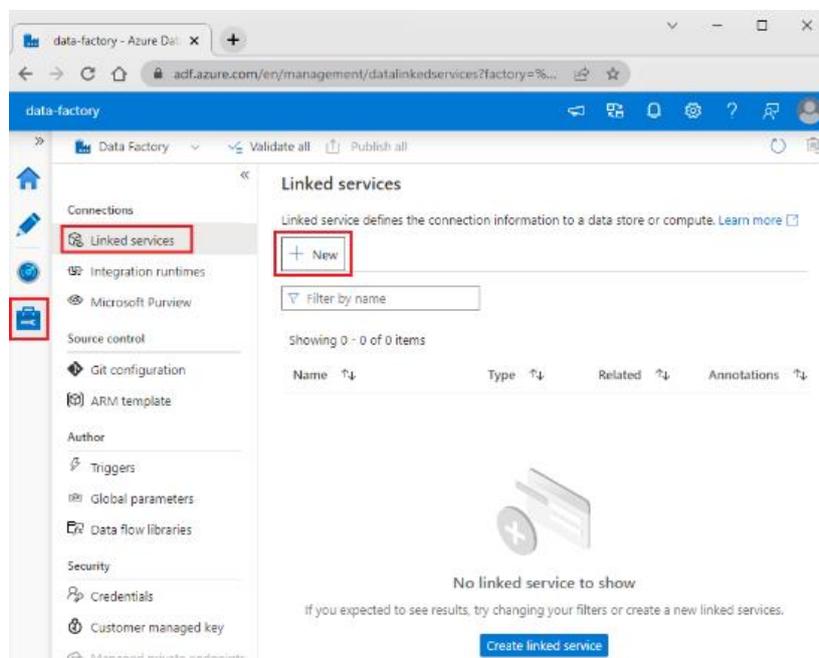
- On the machine running the self-hosted integration runtime environment, edit C:\Windows\System32\drivers\etc\hosts to add mappings of the SAP system IP addresses to the server names. For example:

```
# SAP ECC
xxx.xxx.xxx.xxx sapecc01
# SAP BW
yyy.yyy.yyy.yyy sapbw01
# SAP SLT
zzz.zzz.zzz.zzz sapnw01
```

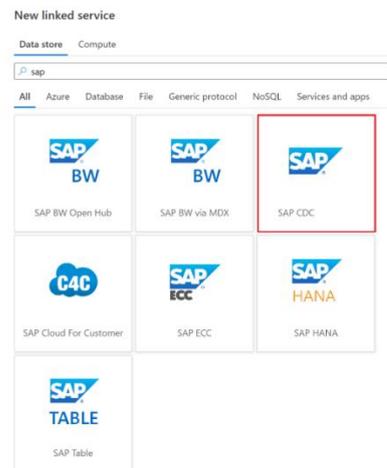
Configuration of a linked service and a data set for the SAP CDC Connector

Configuring linked Services

- In Azure Data Factory Studio, go to the Data Factory Administration center. From the Connections menu, select Linked Services. Select New to create a linked service.



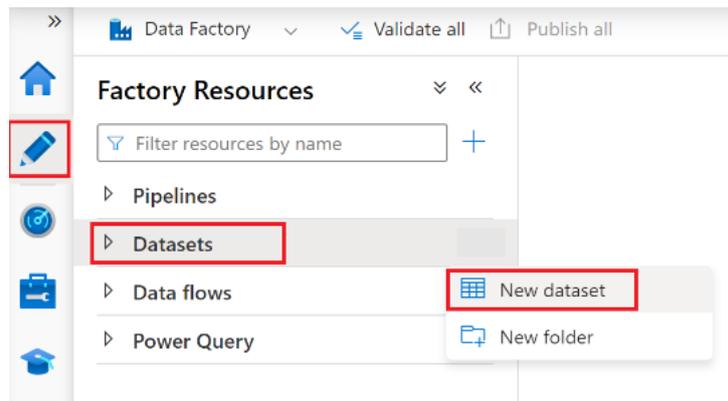
- Under New Linked Service, search for SAP. Select SAP CDC and then choose Continue.



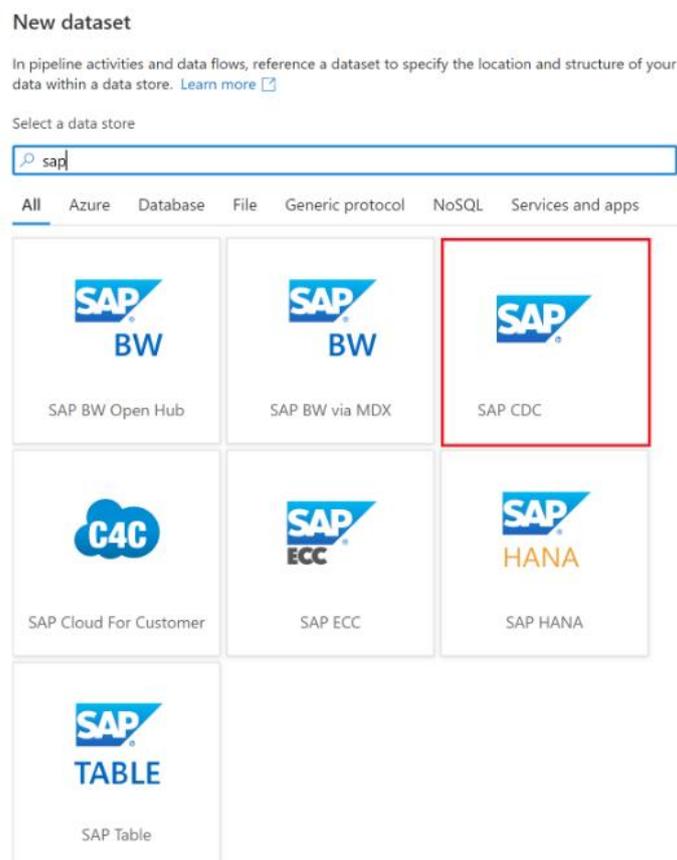
- Set the properties of the linked service:
 - **Name:** Type a unique name for the bound service.
 - **Connect via integration runtime:** select the configured integration runtime environment.
 - **Server name:** name of the assigned server for the SAP system.
 - **Subscriber name:** Unique name to register and identify this Data Factory connection as a subscriber that consumes data packets that are generated in the Operational Delta Queue (ODQ) by the SAP system.

Configuring source data

- In Azure Data Factory Studio, go to the Data Factory Authoring centre.
- In Factory Resources select New Data Set.



- Under New Data Set, search for SAP.
- Select SAP CDC and then choose Continue.



- Set Properties: enter a name for the SAP CDC linked service data source and under Linked Service, select the drop-down list, then select New.
- Select the SAP CDC linked service for the new source data set and set the rest of the properties for the linked service:
 - Under Connect via integration runtime, select the configured integration runtime environment.

- Under ODP context, select the ODP data extraction context. Here are some examples:
 - To extract ABAP CDS views from S/4HANA, select ABAP_CDS.
 - To extract InfoProviders or InfoObjects from SAP BW or BW/4HANA, select BW.
 - To extract SAP extractors from SAP ECC, choose SAPI.
 - To extract SAP application tables from SAP source systems using the SLT replication server as a proxy, select SLT_<queue alias>.

If you want to extract tables from SAP applications, but do not want to use SAP Landscape Transformation Replication Server (SLT) as a proxy, you can create SAP extractors using the RSO2 transaction code or Core Data Services (CDS) views with the tables. Then extract the tables directly from the SAP source systems via a SAPI or an ABAP_CDS context.

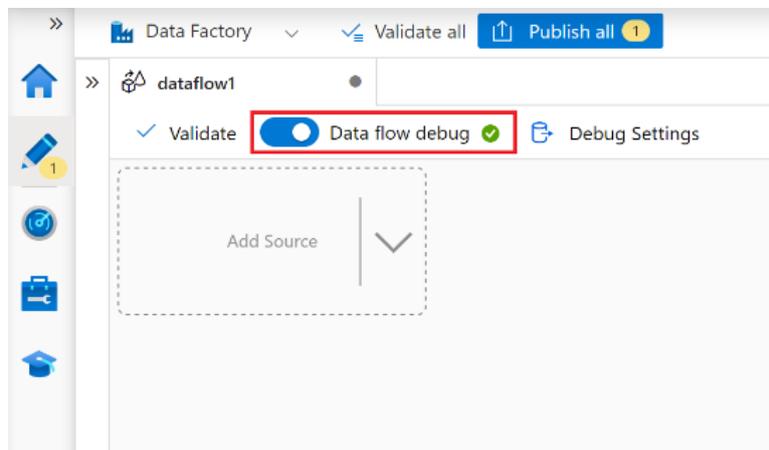
- Under ODP Name, in the selected data extraction context, select the name of the data source object to be extracted. If you connect to the SAP source system via SLT as a proxy, the Data Preview feature is currently not supported.

Data transformation with the SAP CDC connector

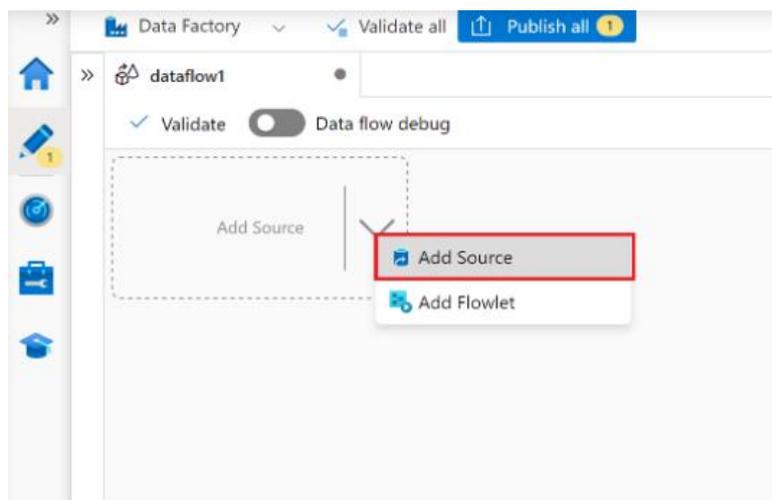
Mapping Data Flow Instance Mapping Properties

- To create a mapping data flow using the SAP CDC connector as the source, complete the following steps:
 - In ADF Studio, go to the Pipeline section of the Creation centre

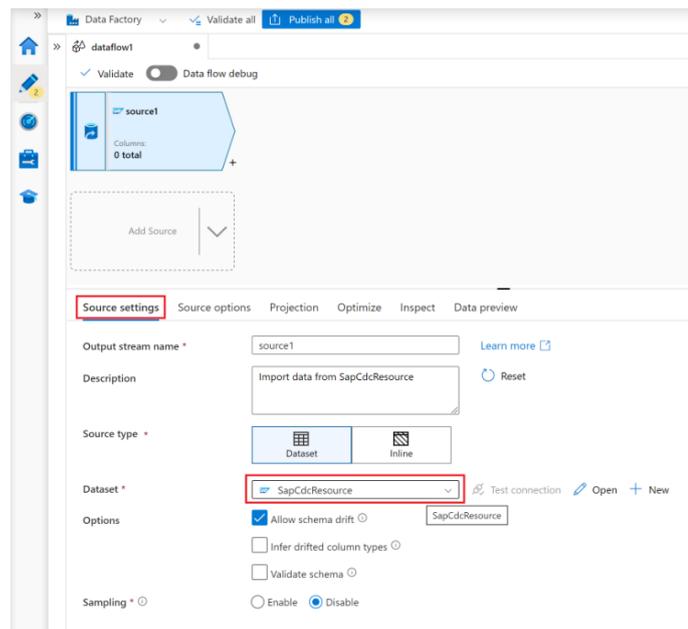
- Select the ... button to open the Pipeline Actions drop-down menu.
- Select the New Pipeline item.
- Activate the debug mode using the Debug Data Flow button in the top bar of the data flow canvas.



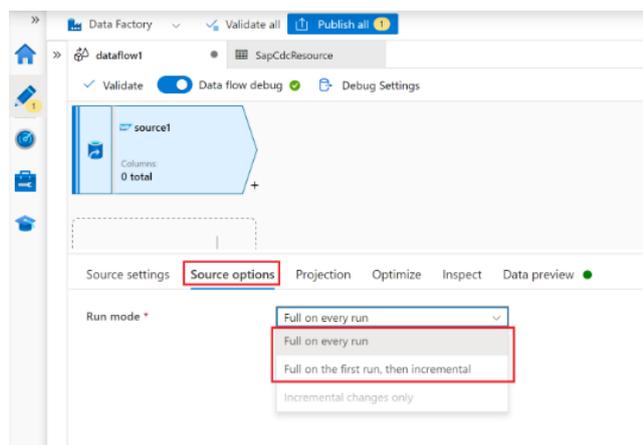
- In the mapping data flow editor, select Add source.



- In the Source Configuration tab, select a prepared SAP CDC dataset or select the new button to create a new one. Alternatively, you can also select Inline in the Source Type property and continue without defining an explicit dataset.



- On the Source Options tab select:
 - Full on every run option if you want to load full snapshots on every run of the mapping data (pipeline) stream
 - Full on first run, incremental, if you want to subscribe to a source of changes from the SAP source system. In case of incremental loads, it is necessary to specify the keys of the ODP source object in the Key Columns property.

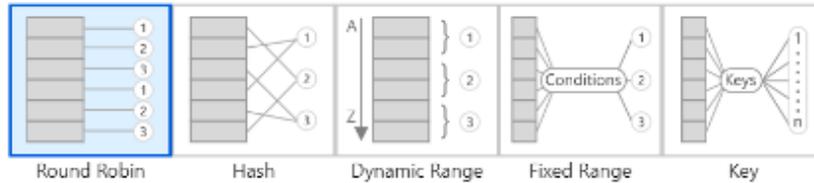


- On the optimize tab we can see the optional values for configuring partitioning schemes.

Aggregate settings **Optimize** Inspect Data preview

Partition option * Use current partitioning Single partition Set Partitioning

Partition type *



Number of partitions *

- On the Inspect tab we can see a view of the metadata of the data stream being transformed. You can see the number of columns, columns that have changed, columns that have been added, data types, column order, and column references. Inspect is a read-only view of the metadata.

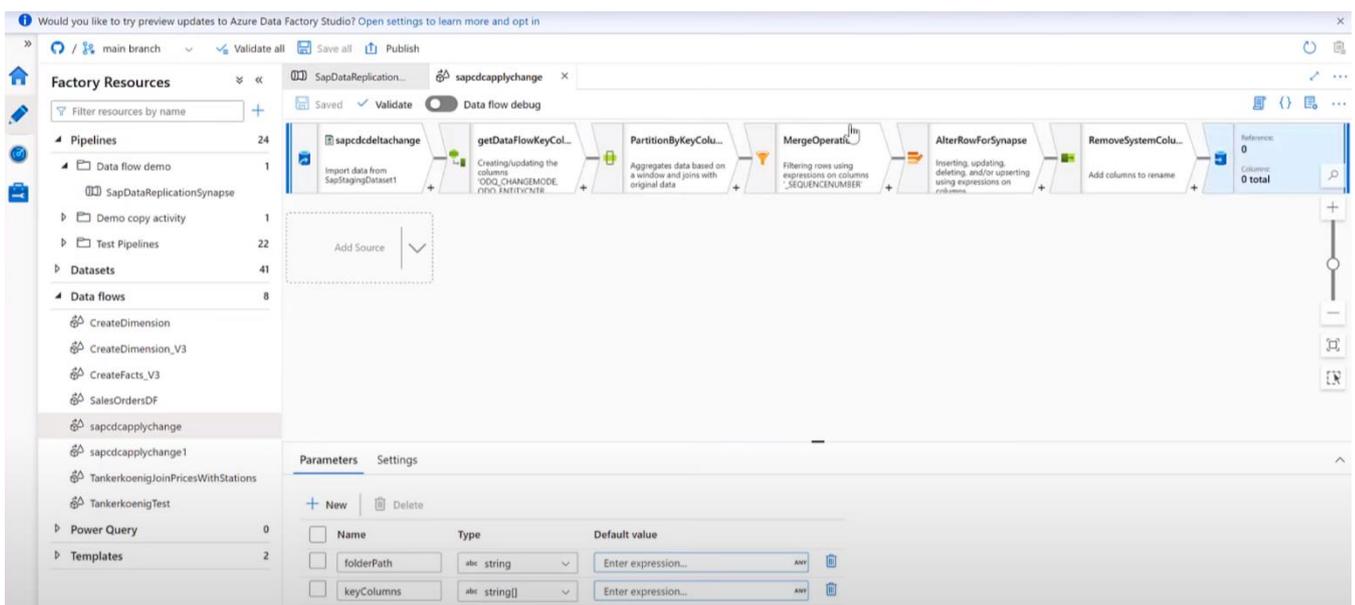
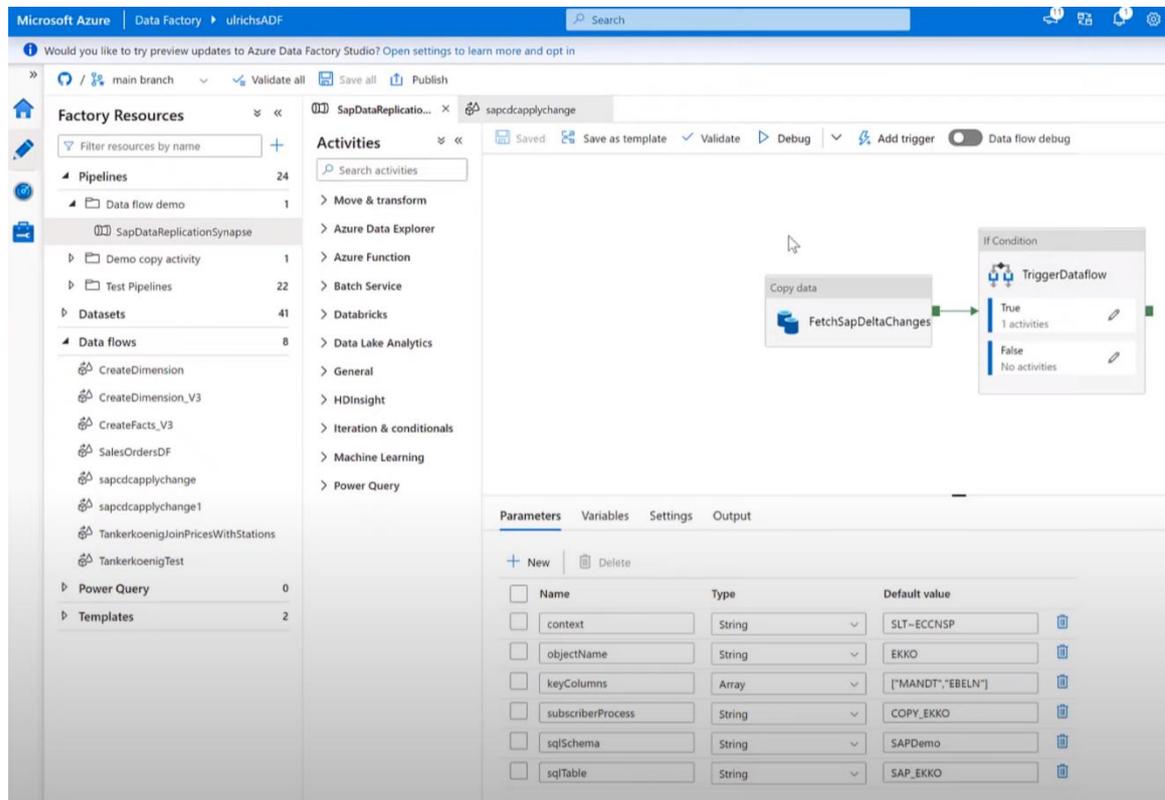
Derived column's settings **Optimize** **Inspect** Data Preview Description

Output schema Input schema

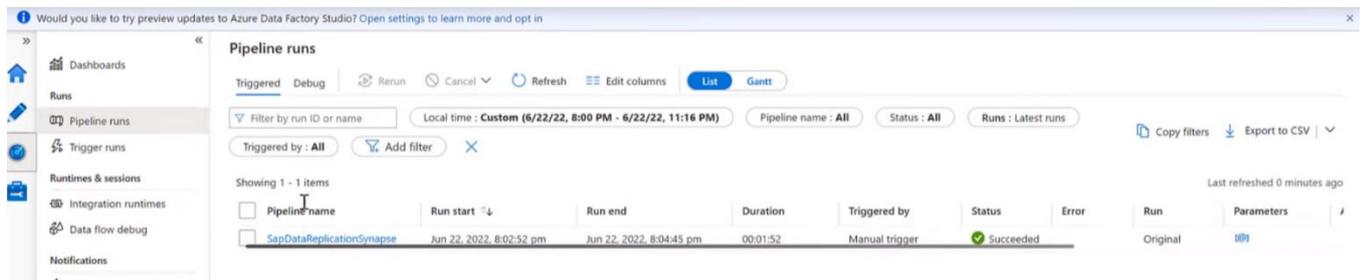
Number of columns		New ⁺ 1	Updated ⁺ 2	Unchanged 4	Total 7
Order ↓	Column ↓	Type ↓	Updated ↓	Based on ↓	
1	movie	string			
2	title	string	*	title	
3	genres	string			
4	year	long	*	year	
5	Rating	string			
6	Rotten Tomato	string			
7	Rotten Tomato	long	*	Rotten Tomato	

AUTOMATION ETL PROCESS

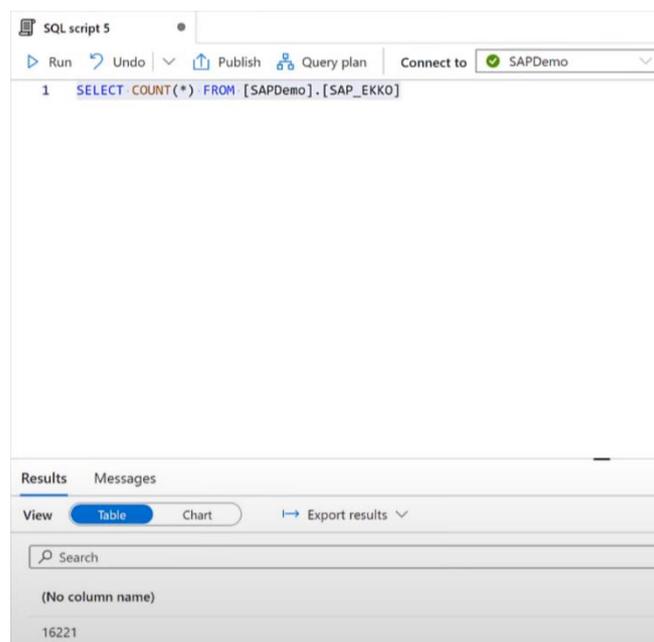
- We will see an example using the connector to copy activity from a table in SAP followed by a template mapping data flow that processes the changes and updates it in an SQL table in Azure Synapse.



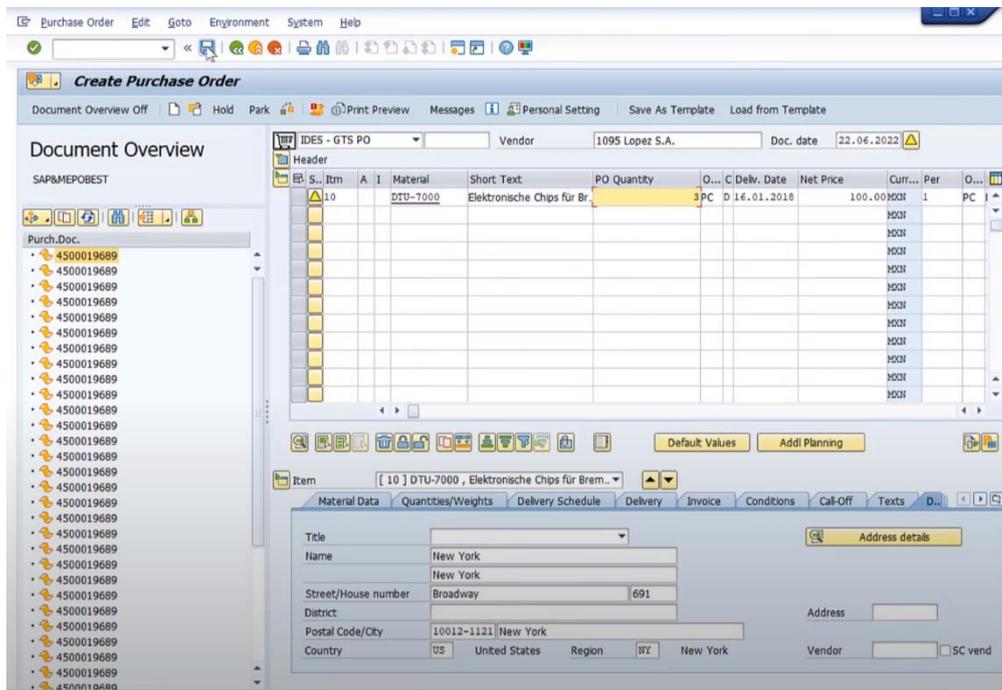
- In the monitoring tool you can see that an initial load has been done



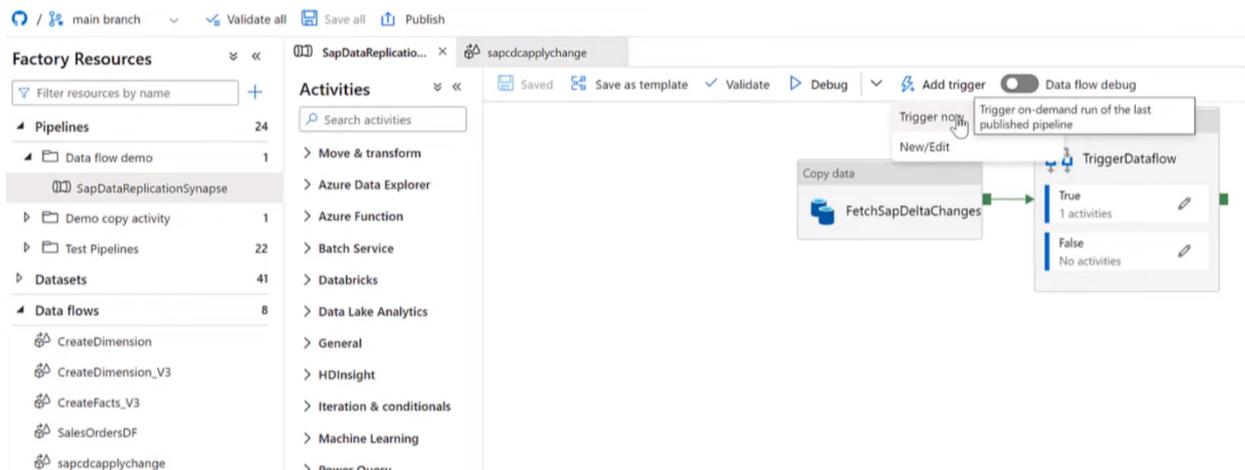
- If we go to Azure synapse we can see that 16221 columns have been loaded into the table



- On the sap side, we can verify that this first pipeline run has initiated the Delta process. On this feed we will now be able to read change as we do in SAP purchasing
- Let's move on to the SAP purchasing application to create a new purchasing order. We will use an existing order made as a template for better simplicity.



- Once we saved our new purchasing order we are going to trigger the process again to see what's happening.



- We are going to see that for the copy activity the process has read only one row from the SAP system.

Details Refresh

[Learn more on copy performance details from here.](#)

Activity run id: 55e9b32a-f552-4782-bf2f-1a01368d17cd

 SAP ODP (Preview) Succeeded  Azure Data Lake Storage Gen2
Region: West Europe

Data read: 652 bytes	Data written: 1,791 KB
Rows read: 1	Files written: 1
Peak connections: 2	Rows written: 1
	Peak connections: 1

Copy duration: 00:00:10
Throughput: 65,536 bytes/s

▼ SAP ODP (Preview) → Azure Data Lake Storage Gen2

Start time: Jun 22, 2022, 8:21:54 pm
Used parallel copies: 1
▼ Duration: 00:00:10

Details	Working duration	Total duration						
Queue		00:00:06						
Transfer	<table border="1"> <tr> <td>Listing source</td> <td>00:00:00</td> </tr> <tr> <td>Reading from source</td> <td>00:00:00</td> </tr> <tr> <td>Writing to sink</td> <td>00:00:00</td> </tr> </table>	Listing source	00:00:00	Reading from source	00:00:00	Writing to sink	00:00:00	00:00:01
Listing source	00:00:00							
Reading from source	00:00:00							
Writing to sink	00:00:00							

Data consistency verification: Not verified

- Finally, if we move to the synapse side to verify the data that we have retrieved we can see that now if we execute the query the result involves one more column, the purchase order that we have created previously.

SQL script 5

Run Undo Publish Query plan Connect to

```
1 SELECT* FROM [SAPDemo].[SAP_EKKO]
```

Results Messages

View Table Chart Export results

Search

(No column name)

16222

POSITIVE AND NEGATIVE ASPECTS REGARDING THE SAP CONECTION

Azure Data Factory is an interesting tool when connecting to SAP for automating a process. On the one hand, the configuration that must be followed to be able to use this type of connectors is somewhat tedious, but on the other hand, the two connectors presented above connect and interact directly with the SAP database.

In terms of data loading efficiency, it is worth mentioning that the ODP is much more efficient when working with large amounts of data, since for each trigger call it will only use the new data entered in the SAP tables to which it is connected instead of the entire table directly.

On the other hand, although it is more efficient, it must be taken into account that if the process to be created does not require much data load that could slow down the execution or if you want to make a simple data copy process, it would be better to use the usual SAP HANA connector since that de ODP, was released in July 2022, so it does not have as much documentation as the SAP HANA connector.

In general, in comparison to the rest of the tools, using Azure Data Factory to connect to SAP is more efficient in terms of data partitioning because it can load terabytes of data, very fast

Finally, keep in mind that to access to certain documentation related to SAP it is necessary to have an S-user account so sometimes it is difficult to access certain documentation.

6. CONCLUSIONS

Once we have seen the automation tools, we can see how there are big differences when choosing one tool or another to automate our process. In general terms, the choice of one tool or another will be dictated by the type of process we want to develop, as well as the type of licenses we have in our organization.

In general terms we can see how Azure Data Factory and Power Automate could be the most appropriate when automating processes. Since both tools have enough documentation to investigate and consult in case of any problem. However, Azure Data Factory would be more desirable in case the process to be automated had some complexity or followed a type of workflow that could not be created with Power Automate, since this tool allows to create much simpler flows.

On the other hand, Pentaho, although it is not a bad option, it is worth mentioning that it does not have extensive documentation and that it does not have a specific connector to connect to SAP HANA so it would be necessary to install the IT NOVUM plug-in, which is a tedious and complex process. In resume, it would be more advisable and, considered that Pentaho is not being used from the beginning, to use another tool.

Talend could be another option, and even if the specifications of the company require it when creating the process, it would not be necessary to buy the paid version of the tool. On the other hand, if the process to be automated is a simple process, it is not worth it because the installation of the tool is somewhat complicated and although there is a lot of documentation about it Talend is a complicated tool to use if you are new to it.

As for Power Automate, it should be noted that this tool is very intuitive and easy to use, but with it you cannot perform flows as complex as those you can perform for example in Talend. In this case to connect to SAP and extract information from the tables it would require a license because the connector is premium.

Finally, regarding Azure Data Factory, as we have seen there are several types of connectors with which we can connect to SAP, in this case we have explained the SAP HANA connector and the ODP connector for SAP. Using one or another connector would depend in part on the amount of data that would be handled in the tables that are in SAP, as well as the type of process that you want to carry out. Choosing this tool to automate processes or another tool will depend not only on the amount of data to be handled in the SAP tables, as ADF is capable of loading terabytes of data very fast but also in the frequency with which the process will be used because we must consider that with Azure Data Factory we pay for what we use.