**Data Integration and Data Flow Modeling with SAP Data Warehouse Cloud**

ANA365

Exercises  
Jascha Kanngiesser / SAP SE

Table of ContentS

[Headline 1 3](#_Toc478041256)

[Headline 2 3](#_Toc478041257)

[Headline 3 3](#_Toc478041258)

Thank you for participating in this hands-on session about Data Integration and Data Flow Modeling with SAP Data Warehouse Cloud! The session is divided into three main parts:

1. Space creation and connection setup
2. Data loading and modeling
3. Data visualization

# Systems

For working your way through the hands-on materials you need access to a SAP Data Warehouse Cloud tenant. Depending on whether you join the live session or watch the session on demand or replay the materials at a later point in time, different systems and credentials can be used.

Participating live

In case you are reading this manual while participating in the live session during the SAP TechEd 2020 program, you can use any of the following three tenants for participating in this hands-on session:

* EMEA region: <https://dwc-teched2020.us10.hcs.cloud.sap>
* APJ region: <https://dwc-teched2020.eu10.hcs.cloud.sap>
* NA region: <https://dwc-teched2020.ap10.hcs.cloud.sap>

To receive login credentials please send an e-mail to [jascha.kanngiesser@sap.com](mailto:jascha.kanngiesser@sap.com?subject=Please%20share%20login%20credentials%20for%20session%20ANA365%20-%20Data%20Flow) mentioning the session ID and the tenant you would like to log in to.

On Demand

In case you are working on this hands-on materials after the live session you cannot use any of the three tenants listed above. Instead, you need to get yourself a SAP Data Warehouse Cloud trial tenant here: <https://saphanajourney.com/data-warehouse-cloud/trial/>

You need to set up all the required connectivity yourself, including the Data Provisioning Agent, uploading certificates, creating the data sets in the remote sources (or uploading the data sets to SAP Data Warehouse Cloud).

Please note that the exercise below is written for attending the live session. Therefore, for example the user credentials and tenant information might not match in case you are following the materials later on demand. In this case please substitute any live session-specific information with the individual information available to your in your trial tenant.

# Things you will learn in this session

* Create connections to sap & non-sap sources
* Data load external tools
* Csv file upload
* Federated data access
* Data extraction using data flow
* Combine everything in a single model
* Visualize in sac

| Explanation | Screenshot |
| --- | --- |
| 1. Make sure you have the login credentials available to you. You should have received a username like saptechedana365+**xyz**@sapcom and a password. Replace the value **xyz** with the number assigned to you. |  |
| 1. Let’s get going – first part! | As mentioned above, this session consists of three main areas: Creating your space and required connections, load the data and build your data models and visualizing the results. Let’s get started with creating your space and setting up the required connectivity. |
| 1. Click on the menu bar to expand the navigation menu. |  |
|
| 1. Select Space Management in the bottom left corner. |  |
| 1. Hit the Create Space + - button in the top right corner to create a new space. |  |
| 1. Enter a Space Name and Space ID. Make sure that at least the Space ID follows the syntax ANA365\_<your three-digit number>. Then hit Create to create your space. |  |
| 1. Reduce the space size in the Overview section at the top. Make sure to specify both the Disk (GB) and In-Memory (GB) storage assignment as 0,1 (GB). |  |
| 1. Hit Save in the top right corner to sabe your changes. |  |
| 1. Head on to the Users section and hit the Add button on the right to add yourself to your space. |  |
| 1. Search for your user ANA365\_<your three digit number> in the dialog, select your user and hit Add to close the dialog. |  |
| 1. Head on to the Connections (Local Connections) section and hit the + - button to créate a new connection. |  |
| 1. Select the Google Cloud Storage tile. |  |
| 1. Or use the filter and search for Google Cloud storage in case you cannot find it in the list of connections. |  |
| 1. Enter a meaningful business name and technical name as well as a useful description. |  |
| 1. Hit Next Step to navigate to the next screen. |  |
| 1. Enter the Project as flash-ocean-262507 |  |
| 1. Enter the root path as /product-reviews-01/ |  |
| 1. Download the key gcs-key.json from the assets folder in the Github repository here: <https://github.com/SAP-samples/teched2020-ANA365/assets/step-17> | **tbd** |
| 1. Hit the Browse button next to the Key entry field and select the downloaded file. |  |
| 1. Hit Create Connection to finish the dialog. |  |
| 1. Select the created connection from the list of connections and hit the Validate Connection button on the right. |  |
| 1. Make sure that the connection is valid for data flow building in the toast message show non the bottom of the screen. |  |
| 1. Hit the + button again to créate another connection. |  |
| 1. This time select the SAP BW connection. |  |
| 1. In case you cannot find it, search for SAP BW in the filter bar. |  |
| 1. Enter a meaningful business name and technical name as well as a descriptive description. |  |
| 1. Make sure that the right Data Provisioning Agent is selected. The ending (in this case us10) should match the tenant you are logged in to. |  |
| 1. Hit Next Step. |  |
| 1. Fill out the Connection Details. Use ld2529 as the Application Server. |  |
| 1. Enter Client 100. |  |
| 1. Enter Systen Number 20. |  |
| 1. Enter user CLOUDCON with password Teched2020 |  |
| 1. Hit Create Connection to finish the dialog. |  |
| 1. Select the newly created SAP BW connection and again hit the Validate Connection button on the right. |  |
| 1. Make sure that the toast message on the bottom of the screen mentions that the connection can be used in the view builder. |  |
| 1. Head on to the Database Access (Database Users) section. Hit the Create button to create a new database user. |  |
| 1. Enter a meaningful Database User Name suffix and make sure to select the Enable Data Ingestion checkbox in the Privileges section. Then hit Create to close the dialog. |  |
| 1. Make sure to copy the Host Name, Port, Database User Name and Password to a safe place, for example your local notes on Mac or Notepad on Windows. Then hit Close to finish the user creation. |  |
| 1. The created database user should be shown with status **Active**. |  |
| 1. Congratulations! | You successfully created your space and created connections to a non-SAP source, Google Big Query! Also you integrated a SAP BW system and enabled your space to get connected to external SQL clients! Before we continue with the second parts, let’s quicky make sure that the space is still empty. |
| 1. Before we start loading data into our space, hit the Monitor Space button in the top right corner. |  |
| 1. Make sure that the space does not yet contain any data. |  |
| 1. Hit Edit Space in the top right corner to navigate back to the Management of your Space. |  |
| 1. Now it’s time to prep your space with some data! | As part of the second section of this hands-on we will equip your space with all the required data and setup the data models. We will ingest some data via an external SQL client (SAP HANA Database Explorer, but can be any client actually), replicate data from some of the data sets available in the SAP BW source, upload a flat file, use the Data Flow to ETL-like move data from Google Cloud Storage into your space and combine the locally available data from these different steps with some data we access virtually from the SAP BW system. |
| 1. Navigate back to the Database Access section and select the Database User you created. |  |
| 1. Hit the Open Database Explorer button on the right. This takes you to the SAP HANA Database Explorer website. |  |
| 1. Within the SAP HANA Database Explorer you can execute DML and DDL statements to create tables or views or insert table into tables. |  |
| 1. Hit the + button in the top left corner. |  |
| 1. From the Database Type select SAP HANA Database. |  |
| 1. Enter the information copied earlier to a safe place when creating the database user like the Host Name, Port, Database User and Password. |  |
| 1. Make sure to check the Save password (stored in the SAP HANA secure store) and Connect to the database securely using TLS/SSL (prevents data eavesdropping) options. Uncheck the Verify the server’s certificate using the trusted certificate below option. |  |
| 1. Hit the OK button to add the database to your list of databases. |  |
| 1. Right-click the newly created database and select Open SQL Console. |  |
| 1. Get the SQL statements from the folder in the Github repository to create and fill the table that holds the Sales Orders: <https://github.com/SAP-samples/teched2020-ANA365/assets/step-51> | **Tbd** |
| 1. Open the create-sales-orders.sql file and copy over the content in the SQL console opened in the SAP HANA Database Explorer. |  |
| 1. Hit the green Execute button to create the Sales Order table. |  |
| 1. Make sure that the table was created correctly by checking the log. |  |
| 1. Empty the SQL console by removing the executed statements. |  |
| 1. Open the insert-sales-orders.sql file and copy over the SQL to fill the table you just created. |  |
| 1. Again hit the green Execute button and check the log for the successful execution. |  |
| 1. Congratulations! | You just used an external tool to connect to your SAP Data Warehouse Cloud space via SQL, created a table and inserted data! This is one example how you can use actually any external 3rd party tool, for example also open source applications like DBeaver, to connect to your SAP Data Warehouse Cloud space from the outside and ingest data using SQL. This way you can for example integrate your SAP Data Warehouse Cloud tenant in your already existing ETL processes. |
| 1. Head back to your SAP Data Warehouse Cloud tenant and navigate to the Data Builder. |  |
| 1. Select your space. |  |
| 1. From the list of tiles select the Import CSV File tile. |  |
| 1. Get the required CSV file SalesOrderItems.csv from the Github repository: <https://github.com/SAP-samples/teched2020-ANA365/assets/step-65> Hit the Select Source File button and select the SalesOrderItems.csv file you just downloaded. |  |
| 1. Leave all options as is and hit the Upload button. |  |
| 1. Wait until the file was uploaded to the tenant. |  |
| 1. Make sure that the data is shown and the headers are identified correctly. |  |
| 1. Select the columns SALESORDERID and SALESORDERITEM as key fields. |  |
| 1. Make sure that both columns are enabled as key fields. |  |
| 1. Hit Deploy to open the dialog for deploying the table into your space. |  |
| 1. Enter a Business Name and Technical Name and deploy the table into your space. |  |
| 1. Make sure that the table was correctly created in your space and is visible in the Data Builder overview. |  |
| 1. Congratulations! | You have successfully uploaded data into your SAP Data Warehouse Cloud space using the CSV file upload! As a next step we will continue loading data into your space using the data flow feature. |
| 1. Select the New Data Flow tile. |  |
| 1. Switch to the Sources tab and navigate to the product reviews folder: Connections > Google Cloud Storage > product reviews |  |
| 1. Select the combined reviews.csv file and drag it from the tree on the left onto the canvas. |  |
| 1. Make sure that the CSV properties are set correctly by selecting the added node and expanding the CSV properties section on the left. |  |
| 1. In the Operators button bar above the canvas you can use several different operators for working with the data sets in your data flow. |  |
| 1. First, we want calculate a sentiment from the available product reviews. Add the Script operator as a target node after the combined reviews source node. |  |
| 1. Connect the two nodes by dragging a line from the left to the right node. |  |
| 1. Select the Sript node and navigate to the Script details by selecting the Edit button on the right. |  |
| 1. Select the Columns section. |  |
| 1. Add a new column named SENTIMENT with data type INT8. |  |
| 1. Confirm the details by clicking Save. |  |
| 1. Switch to the Script section and start entering your python script for a sentiment analysis. |  |
| 1. It’s up to you! | Here’s the thing: Are you a Python expert? We’ve got a challenge for you!  You can either try to figure out the right script yourself using the pandas library. Pro-tip: Check out the pandas data frame apply method documentation here: <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.apply.html>  To make your life easier, we have already provided you with a list of predefined key words and assigned sentiments in the file assets/step-data-flow-script/sentiment-key-words.py  Using the list of words and assigned sentiment values (1: positive; -1: negative; word not mentioned: neutral / 0) you want to do two things:   1. Define a function that takes a row from the input data and checks whether the column REVIEW\_TITLE contains any of the words from the words sentiment array. If a match is found, you want to return the sentiment associated with this word, otherwise (no match) you want to return zero using the string.lower() and string.find() methods. 2. You want to use the pandas dataframe apply() function to loop over the data and call the function defined in 1) for each row. Make sure to specify the axis as 1 and the result\_type=reduce! |
| 1. Want to take the quick route? | Ok, fair enough – here’s the deal. You can use the pre-build solution from the assets/step-data-flow-script/sentiment-analysis-script.py file. If you do so, don’t put the blame on us that the solution is not the nicest or you finished the hands-on well before the scheduled 2hrs! ;-) |
| 1. Add a Projection node to filter out unwanted columns or to apply a filter on the data to be extracted. |  |
| 1. Select the added Projection node and hide the following columns from the Details screen on the right:  * REVIEW\_ID * CUSTOMER\_ID * COUNTRY * REVIEW\_DATE * REVIEW\_TITLE * REVIEW\_TEXT |  |
| 1. After removing all columns only the following two columns should be available. |  |
| 1. Next, add an Aggregation node. |  |
| 1. Select the node and from the Details pane on the right set the aggregation for the column SENTIMENT. |  |
| 1. Select the Add Table operator or use the context menu from the added aggregation node to add the target table in which we will persist the data. |  |
| 1. Select the newly added node and define a label, business name and technical name for the target table. |  |
| 1. Select the Create and Deploy Table button to create the table in the respository. |  |
| 1. Confirm the dialog to create the table. |  |
| 1. Save the data flow. |  |
| 1. Provide a name for the data flow and hit Save. |  |
| 1. After saving the data flow execute it. |  |
| 1. Navigate to the data flow monitor by selecting the navigation icon from the left to navigate to the run details. |  |
| 1. Hit the Refresh button until the data flow run completes. |  |
| 1. If you wanted to, you could also define a schedule to run the data flow repeatedly. This is however not needed for this session. ☺ You can define schedules by selecting the Schedule item from the left next to the Refresh icon. |  |
|  |  |