IDE for Data Modelling (python)

Data modelling of the data base tables of Common Core.

|  |  |
| --- | --- |
| Diagrama  Descripción generada automáticamente | **Data Modelling:** is a process of creating a visual or conceptual representation of data structures and their relationships. Data modeling helps to communicate connections between data points and structures, and to organize, measure, and manage data. |

# Install VSCode from Microsoft Store

1. Open the Microsoft Store on Windows 11
2. Download [Visual Studio](https://code.visualstudio.com/Download) or download it from the Microsoft Store on Windows 11 as follows:

|  |  |
| --- | --- |
| Interfaz de usuario gráfica, Aplicación  Descripción generada automáticamente | 1. **Install:** Click the Free button, then the Install button will appear. Click on it and follow the VS Code instructions. |

1. Install python 3.10 - Microsoft Store
2. Open Microsoft Store of Windows 11
3. Download [python3.10](https://www.python.org/ftp/python/3.10.0/python-3.10.0-amd64.exe) or download it from the Microsoft Store on Windows 11 as follows:

|  |  |
| --- | --- |
|  | 1. **Install:** Click the Free button, then the Install button will appear. Click on it and follow the python 3.10 instructions. |

# Python virtual environment

* Create a folder on the desktop called “wsdmv,” which stands for “WorkSpace Data Model Visualization"
* Inside the folder “wsdmv,” create a new folder called “.venvs.”
* Open a Windows terminal and navigate to the folder ./wsdmv/

cd “C:\Users\$REPLACEFORYOURWINDOWSUSER\OneDrive - Otto Group\Desktop\wsdmv\”

* Create the virtual environment by executing the following command

C:\Users\$REPLACEFORYOURWINDOWSUSER\OneDrive - Otto Group\Desktop\wsdmv>**python3.10 -m venv .venvs**

* Go to the .venvs folder

cd .venvs

* Check if the .venvs folder has the following structure:

\Include

\Lib

\pyvenv.cfg

\Scripts

* Activate the Python virtual environment

C:\Users\$REPLACEFORYOURWINDOWSUSER\OneDrive - Otto Group\Desktop\wsdmv\.venvs\Scripts>.\activate

* Add the requirements.txt file to the path C:\Users\$REPLACEFORYOURWINDOWSUSER\OneDrive - Otto Group\Desktop\wsdmv



* Execute the pip install command for the requirements.txt file.

(.venvs) C:\Users\$REPLACEFORYOURWINDOWSUSER\OneDrive - Otto Group\Desktop\wsdmv>pip install -r requirements.txt

|  |  |
| --- | --- |
| Imagen que contiene texto, tabla  Descripción generada automáticamente | **No errors:** Once all Python libraries are installed in the virtual environment, leave the Windows terminal open.  Errors: Review the first one that appears. I highly recommend reviewing the error on [Stack Overflow](https://stackoverflow.com/). |

1. VScode and Python venv
2. Open the VSCode on Windows 11
3. Click the “File” option on the top menu of VSCode and select “Open Folder” option

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

1. Select the path where the project “wsdmv” is located.

C:\Users\$REPLACEFORYOURWINDOWSUSER\OneDrive - Otto Group\Desktop\wsdmv

1. Accept the workspace trust

Texto

Descripción generada automáticamente

1. Check the project “wsdmv” in the left menu

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

# Jupyter notebook through VSCode

Working with Jupyter Notebook on VSCode,

Note: It is highly recommended to work with Jupyter Notebook in Windows Terminal and the URI endpoint in a web browser. It is in the next section F) on this document.

**VSCode:**

1. Open the VSCode on Windows 11
2. Check the project “wsdmv” in the left menu
3. In the left menu of the "wsdmv" project, select the "Extensions" option.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

1. Search for Python, Jupyter Notebook, and Graphviz. Then, click the “Install” button. Finally, restart VSCode.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Python Extension:

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Jupyter Extension:

Interfaz de usuario gráfica, Sitio web

Descripción generada automáticamente

Graphviz Extension:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

1. On the Welcome page, in the "Walkthroughs" section, select the "Jupyter: Get Started" option.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

1. Click the “Create New Jupyter Notebook” button. Also,

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

1. Another way to create a Jupyter Notebook in VSCode is Goto File -> New File, select “Jupyter Notebook .ipynb”

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

1. In the right menu of the untitled Jupyter Notebook, choose the "Select Kernel" option.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

1. Select Python Enviroment

Interfaz de usuario gráfica

Descripción generada automáticamente

1. Select the previously created Python virtual environmentInterfaz de usuario gráfica, Aplicación

   Descripción generada automáticamente
2. Inside the rectangle called "cell," import the Graphviz library with **"from graphviz import Digraph**".On the right, click on "Execute Cell and Below."

Interfaz de usuario gráfica

Descripción generada automáticamente

1. Goto File -> Save As, in "Save as Type" select No Extension and save the file with file name "name".ipynb
2. Jupyter in Windows Terminal

Working with Jupyter Notebook in Windows Terminal,

**Windows Terminal:**

1. Activate the Python virtual environment

C:\Users\$REPLACEFORYOURWINDOWSUSER\OneDrive - Otto Group\Desktop\wsdmv\.venvs\Scripts>.\activate

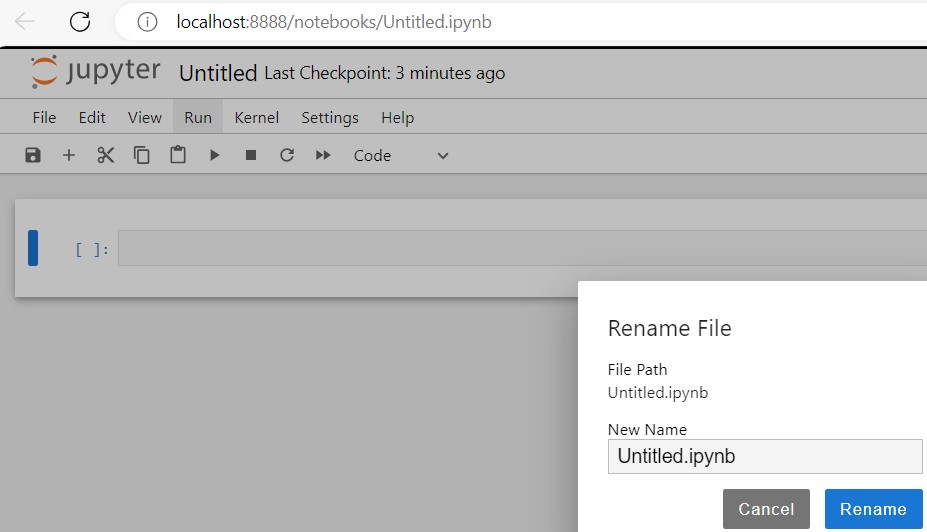
1. Go to the .wsdmv folder

cd (.venvs) C:\Users\$REPLACEFORYOURWINDOWSUSER\OneDrive - Otto Group\Desktop\wsdmv

1. Execute the jupyter notebook

(.venvs) C:\Users\$REPLACEFORYOURWINDOWSUSER\OneDrive - Otto Group\Desktop\wsdmv> jupyter notebook

1. Accept the preferred web browser to work with Jupyter Notebook, for instance, Microsoft Edge. A folder tree appears in the localhost. Create a Jupyter NotebookInterfaz de usuario gráfica, Aplicación

   Descripción generada automáticamente
2. Rename File from Untitled.ipynb to AquaDM
3. Inside the rectangle called "cell," import the Graphviz library with "from graphviz import Digraph".On top menu, click on "Run this cell and advance."

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

1. In case you want to learn more about [Working with Jupyter code cells in the Python Interactive window (visualstudio.com)](https://code.visualstudio.com/docs/python/jupyter-support-py)
2. Shutdown the localhost Jupyter Notebook, Goto File -> Shutdown

# cc\_aqua.. Tables on Oracle T-SQL

We will draw our first data model using Aqua.. Tables. [AQUA](https://confluence.witt-gruppe.eu/display/BAI/AQUA) = a quality management treatment: *“Application for information on the receipt and inspection of articles. The tool helps QM record the stage the item is in in the goods receipt process”.*

1. Execute the next query to count the number of cc\_aqua tables in the database through

SELECT table\_name, owner

FROM all\_tables

WHERE upper(table\_name) LIKE '%CC\_AQUA%';

1. Match between the obtained list from SQL Developer and the list of [IT-BAI Data Catalogue – Home (sharepoint.com)](https://og2gether.sharepoint.com/sites/IT-BAIDataCatalogue/?xsdata=%3D&sdata=ckJiQUEwNUpkSThKT3hDYUNhUjdGU3VqRzFTVlpPR2NOU3RIZE85VWhIUT0%3D&ovuser=8794e153-c3bd-4479-8bea-61aeaf167d5a%2CPatricia.GomezBello%40ottogroup.com)
2. Reviweing all tables of CC\_AQUA.., there are more in the database than the list of IT-BAI Data Catalogue
3. Execute the next query to count the number of cc\_aqua tables in the database per owner. Observation: as admin user of the database, I am getting the same number of tables as the 1st step.

SELECT table\_name, OWNER

FROM all\_tables

WHERE upper(table\_name) LIKE '%CC\_AQUA%'

ORDER BY owner DESC;

1. Execute the next query to get primary of a CC\_AQUA… table

SELECT cols.table\_name, cols.column\_name, cols.position, cons.status, cons.owner

FROM all\_constraints cons, all\_cons\_columns cols

WHERE cols.table\_name = 'CC\_AQUA TABLE ...'

AND cons.constraint\_type = 'P'

AND cons.constraint\_name = cols.constraint\_name

AND cons.owner = cols.owner

ORDER BY cols.table\_name, cols.position;

1. Execute the next query to get all constraints of a CC\_AQUA… table

SELECT constraint\_name FROM all\_constraints

WHERE UPPER(table\_name) = UPPER('CC\_AQUAQMQZAHL')

1. Save these keys in an Excel file named "AQUA\_BUSINESS\_UNIT.xlsx" because it will be used to create the data model (diagram) using Jupyter Notebook and Graphviz.



Note: *After reviewing the 1st draft of the data model “AQUA Business Unit”, there is a decision to deep the Aqua Business Unit in accordance with provider, invoice, article, shipment.*

# cc\_aqua.. Tables on BAI - DSV

There are 14 tables that belong to the [AQUA](https://confluence.witt-gruppe.eu/display/BAI/AQUA) app. These tables need to be integrated with other tables in the Oracle Data Warehouse. As consequence, the analysis should be based on data lineage to track data as it flows through various data systems, applications, and processes. **Data governance** refers to the rules and processes imposed on maintaining data in a company. Data lineage is the part of data governance that records the movement of data from its original source through any system in between that source and the data’s destination. **Data lineage** (a granular view of data flow) is closely connected to data governance as a vital component in managing an organization’s data ecosystem. Data lineage solutions help data governance teams ensure data complies to these standards, providing visibility into how data changes within the ETL pipeline.

**BAI - DSV:**

1. Open the BAI -DSV on the Citrix Virtual Machine (VDI)

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

1. Goto Verwaltung->Skripte, in "Datenbank-Skripten" select UDWH->UDWH1->Aqua.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

1. For each CC\_AQUA job, verify the primary key, which is commonly the CC\_TABLENAME\_PK, and the surrogate key, **“dwh\_id”**. Extract all these transformed Common Core tables from the Stage Area and Source Core. These are CC\_AQUA… without “…abschluss” or “..restore”. Both types of jobs are part of Oracle ETL pipeline in BAI - DSV.
2. Analyze how these CC\_.. tables are related to CC\_AQUA.. For instance, the job of cc\_ww\_aqualgsendung is using the struture of the next tables:

cc\_ aqualgsendung

cc\_frankatur\_akt\_v

cc\_sendungstatus\_akt\_v

cc\_sendungtarif\_akt\_v

cc\_sendungmarge\_akt\_v

cc\_sendungpalgebuehr\_akt\_v

cc\_sendungtransport\_akt\_v

cc\_sendungtransfer\_akt\_v

cc\_sendungerfassung\_akt\_v

cc\_pruefmusterziehungsart\_akt\_v

* 1. Review all the joins in the job “cc\_ww\_aqualgsendung,” not only the comments but also the JOINs in the SELECT statement.
  2. Describe every table CC\_.. that belongs to the job “cc\_ww\_aqualgsendung,”

DESC cc\_sendungtarif\_akt\_v

* 1. Determine which columns are the same among these tables (metadata). In the job “cc\_ww\_aqualgsendung,” these are the same columns. In general, these columns represent the properties of shipment.

DWH\_CR\_LOAD\_ID

DWH\_UP\_LOAD\_ID

DWH\_STATUS

DWH\_VALID\_FROM

DWH\_VALID\_TO

WERT

BEZ

SG

FR

RU

US

* 1. Limiting the search of the other business domain per CC\_.. tables and UDWH1 owner.

SELECT table\_name, owner

FROM all\_tables

WHERE upper(table\_name) LIKE 'CC\_%RECHNUNG%'

AND OWNER LIKE 'UDWH1'

ORDER BY owner DESC;

* 1. Make the data model of the AQUA Business Unit more detailed and granular. This is a process of iteration and refinement.

1. Save the analysis of these relationships among the CC\_AQUA tables and others in an Excel file named "AQUA\_BUSINESS\_UNIT\_DSV.xlsx," as it will be used to create the data model (diagram) using Jupyter Notebook and Graphviz.



1. Data Model - AQUA Business Unit

Go to E) section of this document.

**VSCode:**

1. Open the VSCode on Windows 11
2. Select the Jupyter Notebook -> AquaDM.ipynb
3. Write the following python code to get a pdf file from the data model

from graphviz import Digraph

import os

os.environ["PATH"] += os.pathsep + 'C:/Users/pgomezbe/OneDrive - Otto Group/Escritorio/wsdmv/.venvs/windows\_10\_cmake\_Release\_Graphviz-11.0.0-win64/Graphviz-11.0.0-win64/bin/'

dot = Digraph('AQUABUSINESSUNIT1')

dot.attr(rankdir='LR', splines='ortho')

dot.attr(labelloc='t')

dot.attr(label=r'\nData Model of AQUA Business Unit by Data Team')

dot.attr(fontsize = '12')

dot.attr(fontcolor = 'blue')

metadata = {

'cc\_art': {

'cohere': ('art\_id')

},

'cc\_qmauspruefprot': {

'cohere':('id')

},

'cc\_emailreportticket': {

'cohere': ('ticket\_id')

},

'cc\_aqualgwaeingpos': {

'cohere':('art\_id')

},

'cc\_aqualgwaeing': {

'cohere': ('waeingnr')

},

'cc\_aquaqmwaeing': {

'cohere': ('qmwaeing\_id')

},

'cc\_aquaqmwaeingpos': {

'cohere': ('qmwaeingpos\_id')

},

'cc\_aquaqmklaerfall': {

'cohere': ('qmklaerfall\_id')

},

'cc\_aquaqmreklamation': {

'cohere': ('qmreklamation\_id')

},

'cc\_aqualgsendung': {

'cohere': ('sendungsnr')

},

'cc\_aqualgsendungpos': {

'cohere': ('sendungsnr')

},

'cc\_aquaqmmangel': {

'cohere': ('qmmangel\_id')

},

'cc\_aquaqmmasstab': {

'cohere': ('qmmasstab\_id')

},

'cc\_aquaqmnabfirma': {

'cohere': ('qmnabfirma\_id')

},

'cc\_aquaqmpruefer': {

'cohere': ('qmpruefer\_id')

},

'cc\_aquaqmqzahl': {

'cohere': ('qmqzahl\_id')

},

'cc\_aquaqmtextbaustein': {

'cohere': ('id')

},

'cc\_aquaqmweprio': {

'cohere': ('qmweprio\_id')

}

}

for table, info in metadata.items():

label = f'''<<TABLE BORDER="1" CELLBORDER="1" CELLSPACING="0" CELLPADDING="4">

<TR><TD COLSPAN="2"><FONT POINT-SIZE="10"><B>{table}</B></FONT></TD></TR>

<TR><TD COLSPAN="2"><FONT POINT-SIZE="10">{info['cohere']}</FONT></TD></TR>

</TABLE>>'''

dot.node(table, label=label, shape='plaintext')

relationship = [

('cc\_art', 'cc\_aqualgwaeingpos', ''),

('cc\_art', 'cc\_aqualgwaeing', ''),

('cc\_art', 'cc\_aquaqmwaeing', ''),

('cc\_art', 'cc\_aqualgsendungpos', ''),

('cc\_art', 'cc\_aqualgsendung', ''),

('cc\_aquaqmwaeing', 'cc\_aquaqmklaerfall', ''),

('cc\_emailreportticket', 'cc\_aquaqmklaerfall', 'dwh\_id'),

('cc\_aquaqmwaeing','cc\_aquaqmwaeingpos', ''),

('cc\_emailreportticket', 'cc\_aquaqmreklamation', 'dwh\_id'),

('cc\_emailreportticket', 'cc\_aqualgsendung', 'dwh\_id'),

('cc\_art', 'cc\_aquaqmweprio', 'artnr'),

('cc\_emailreportticket', 'cc\_aqualgsendungpos', 'dwh\_id'),

('cc\_qmauspruefprot', 'cc\_aquaqmmangel', 'dwh\_id'),

('cc\_emailreportticket', 'cc\_aquaqmmasstab', 'dwh\_id'),

('cc\_aquaqmklaerfall', 'cc\_aquaqmmangel', ''),

('cc\_emailreportticket', 'cc\_aquaqmpruefer', 'dwh\_id'),

('cc\_qmauspruefprot', 'cc\_aquaqmnabfirma', 'dwh\_id'),

('cc\_emailreportticket', 'cc\_aquaqmqzahl', 'dwh\_id'),

('cc\_aquaqmwaeingpos', 'cc\_aquaqmmasstab', ''),

('cc\_emailreportticket', 'cc\_aquaqmtextbaustein', 'dwh\_id')

]

for from\_table, to\_table, key in relationship:

dot.edge(from\_table, to\_table, headlabel=key, fontsize='10', fontcolor='green')

dot.view()

dot.render('Aqua\_Data\_Model\_v1', format='pdf', view=True)

1. Graphviz.exe on Windows 11-64bit

To save the data model diagram as a PDF file, it is necessary to make Graphviz executable.

1. Download [graphviz](https://graphviz.org/download/) as follows:

|  |  |
| --- | --- |
| Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico  Descripción generada automáticamente | 1. **Install:** Click the “graphviz-11.0.0 (64-bit) ZIP archive [sha256]” link, then the download button will appear. Click on it and follow the graphiz’s instructions. |

1. Unzip the compressed file “windows\_10\_cmake\_Release\_Graphviz-11.0.0-win64” from the “dowonloads” folder.
2. Move the uncompressed folder from .\downloads to .\.venvs

C:\Users\$REPLACEFORYOURWINDOWSUSER\Downloads\windows\_10\_cmake\_Release\_Graphviz-11.0.0-win64\Graphviz-11.0.0-win64 to C:\Users\$REPLACEFORYOURWINDOWSUSER\OneDrive - Otto Group\Desktop\wsdmv\.venvs

1. Configuring the environment variable for Graphviz in your Jupyter Notebook on Windows 11 as follows:

import os

os.environ["PATH"] += os.pathsep + 'C:/Users/\$REPLACEFORYOURWINDOWSUSER/OneDrive - Otto Group/Desktop/wsdmv/.venvs/windows\_10\_cmake\_Release\_Graphviz-11.0.0-win64/Graphviz-11.0.0-win64/bin/'

1. Execute Graphviz's dot command, and the PDF should be created, resulting in the following log:

Texto

Descripción generada automáticamente

1. Git on Windows 11-64bit & VSCode

To work with Git repositories, it is necessary to make Git.exe executable in VScode.

1. Open VSCode
2. On the left menu, choose “Source Control” and click “Download Git for Windows”

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

1. Select “64-bit Git for Windows Setup” option

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

1. On download folder, click “Git-2.45.2-64-bit” file



1. Enter your laptop username and password, and the executable will begin installing. Click “Next” in all default values.
2. Restart your laptop
3. To use VSCode and Git Otto Seat Repository, please follow the next instructions of these URLs:
   1. [Introduction to Git in Visual Studio Code](https://code.visualstudio.com/docs/sourcecontrol/intro-to-git)
   2. [Collaborate on GitHub (visualstudio.com)](https://code.visualstudio.com/docs/sourcecontrol/github)
4. Git.exe on Windows Terminal

To work with Git repositories in Windows Terminal:

1. Clone the respository on your Desktop directory

C:\Users\YOURUSERLAPTOP\OneDrive - Otto Group\Desktop>**git clone --single-branch --branch feature/pgb https://github.com/pgomezbe/datamodelling.git**

1. Go to the cloned directory

C:\Users\YOURUSERLAPTOP\OneDrive - Otto Group\Desktop>cd datamodelling

1. Copy your changes from wsdm directory to data modelling
2. Verify if you are in the correct branch

C:\Users\YOURUSERLAPTOP\OneDrive - Otto Group\DESKTOP\datamodelling>**git branch**

\* feature/pgb

1. After you have done the previous step 3, please verify all your changes, they appear in red

C:\Users\YOURUSERLAPTOP\OneDrive - Otto Group\DESKTOP\datamodelling>**git status**

...

1. To commit your changes
   1. C:\Users\YOURUSERLAPTOP\OneDrive - Otto Group\DESKTOP\datamodelling>**git add .**
   2. C:\Users\YOURUSERLAPTOP\OneDrive - Otto Group\DESKTOP\datamodelling>**git commit -m “Aqua DM - Creating two pdfs for revision”**
2. To push your changes from local to remote

C:\Users\YOURUSERLAPTOP\OneDrive - Otto Group\DESKTOP\datamodelling>**git push -u origin feature/pgb**