

Submitter: Eric Wolfson, for the Sr. Data Engineer MDM position at HubSpot

**Here are the steps needed to install and run this project on Windows 11.**

- 1) Install Python 3.11 (Dagster works with Python 3.8 – 3.11) For this project I used Python 3.11.7 at <https://www.python.org/downloads/>. Run the python-3.11-amd64 file that you just downloaded.
- 2) To add python to the environment variables, run “Edit System Environment Variables” from the control panel. Click Environment Variables. Add or Edit the path variable adding C:\<user path>\AppData\Local\Programs\Python\Python311 or whichever drive you are using and C:\<user path>\AppData\Local\Programs\Python\Python311\Scripts.
- 3) Install MS SQL Server 2022 and optionally SQ Server Management Studio 19 I used SQL Server Express at <https://www.microsoft.com/en-us/sql-server/sql-server-downloads>. After installation, there will be a button at the final window that says “Install SSMS”. Optionally, click this and it will take you to the download page for SSMS 19. Download and install it.
- 4) In command line create a new directory for the project
- 5) Create the environment: `python -m venv <ENV NAME>`
- 6) Run the environment in the directory with `<ENV NAME>\Scripts\activate.bat`
- 7) Type `pip install dagster dagster-webserver`
- 8) If you get the error “Failed building wheel for pendulum” in the command prompt, type: `“pip install pendulum==2.0.3”`
- 9) Type `pip install pandas`
- 10) Type `pip install pypyodbc`
- 11) To set the server name in the configuration file
  - a. Right Click on “This PC” on the desktop or in windows explorer and click properties.
  - b. Under system – About – Device Specifications check “Device name”
  - c. The name of the default server is <DEVICE NAME>\SQLEXPRESS or <DEVICE NAME>\SQLEXPRESS01. Alternatively, you can open SQL Server Management Studio and in the server connection pop up window check the server name.
  - d. The string SERVER\_NAME in the configuration file for this python project should be changed to the server name that was just discovered
- 12) Create the entityresolution database in SQL Server by:
  - a. Loading a second command prompt
  - b. Typing `sqlcmd -S WOLFSON179\SQLEXPRESS01 -E`
  - c. Typing `select @@Version` and hitting enter
  - d. Typing go and hitting enter
  - e. Type `CREATE DATABASE entityresolution`
- 13) In the command prompt with the environment activated, type: `dagster dev -f entityresolution.py`
- 14) Open a browser (preferably Chrome) and type in the address bar: localhost:3000
- 15) The data pipeline in this project should be displayed.