**Weather Data Pipeline Setup Guide**

This guide covers:

1. Installing Miniconda (Python 3.11) via the command line
2. Creating & activating the weather\_project conda environment using environment.yml
3. Installing any additional pip packages
4. Setting up your NOAA CDS API key
5. Tailoring get\_data2.py for custom date ranges

0. **Create project directory (I created a project in WSL ubuntu)**place the get\_data2.py and the environment.yml file inside the directory

**1. Install Miniconda (Python 3.11)**

# Download the Miniconda installer for Linux (adjust URL for macOS/Windows)

curl -LO https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86\_64.sh

# Run the installer

bash Miniconda3-latest-Linux-x86\_64.sh

# Follow prompts: accept license, confirm install location (e.g. $HOME/miniconda3)

# Initialize conda for your shell

source "$HOME/miniconda3/etc/profile.d/conda.sh"

Verify:

conda --version # should report conda 4.x.x

**2. Create & Activate the Conda Environment**

Use the provided **environment.yml** (located at your project root):

conda env create -f environment.yml

conda activate weather\_project

Confirm key versions:

python -c "import sys, xarray, dask; print(sys.version.split()[0], xarray.\_\_version\_\_, dask.\_\_version\_\_)"

**3. (Optional) Install Additional Pip Packages**

If you need any extras not in the YAML (e.g. a new library), run:

pip install <package-name>

**4. Set Up NOAA CDS API Key**

1. Register at <https://cds.climate.copernicus.eu/>
2. In **My Account → API key**, copy the lines:
3. url: https://cds.climate.copernicus.eu/api/v2
4. key: <username>:<your-api-key>
5. Create ~/.cdsapirc with those contents:
6. cat <<EOF > ~/.cdsapirc
7. url: https://cds.climate.copernicus.eu/api/v2
8. key: <username>:<your-api-key>
9. EOF
10. Add ~/.cdsapirc to your ~/.gitignore to keep it private.

**5. Tailor get\_data2.py for Custom Date Ranges**

Open **get\_data2.py** and locate the date‑range block near the top:

# Define range: start (inclusive) to end (exclusive)

years = range(1960, 1965)

# List of two‑digit month strings:

months = [f"{m:02d}" for m in range(1, 13)]

* **Change years**: e.g. for 1950–2020 use range(1950, 2021).
* **Restrict months**: e.g. only summer: months = ["06","07","08"].

Save and run:

python get\_data2.py

The script will download GRIB files into data/ for your specified period.