Model Agnostic Framework

Online Training Session Provincial, Territorial, and University Research Partners

Kasra Keshavarz
Research Scientist
Schulich School of Engineering
University of Calgary

Alain Pietroniro, PhD, PEng
Professor and Canada Research Chair
Schulich School of Engineering
University of Calgary

UNIVERSITY OF CALGARY

January 27th, 2026

Acknowledgements

UNIVERSITY OF CALGARY

UNU INWEH

GLOBAL WATER FUTURES

Environment and Climate Change Canada

Environnement et Changement climatique Canada

Digital Research Alliance of Canada | **Alliance de recherche numérique** du Canada

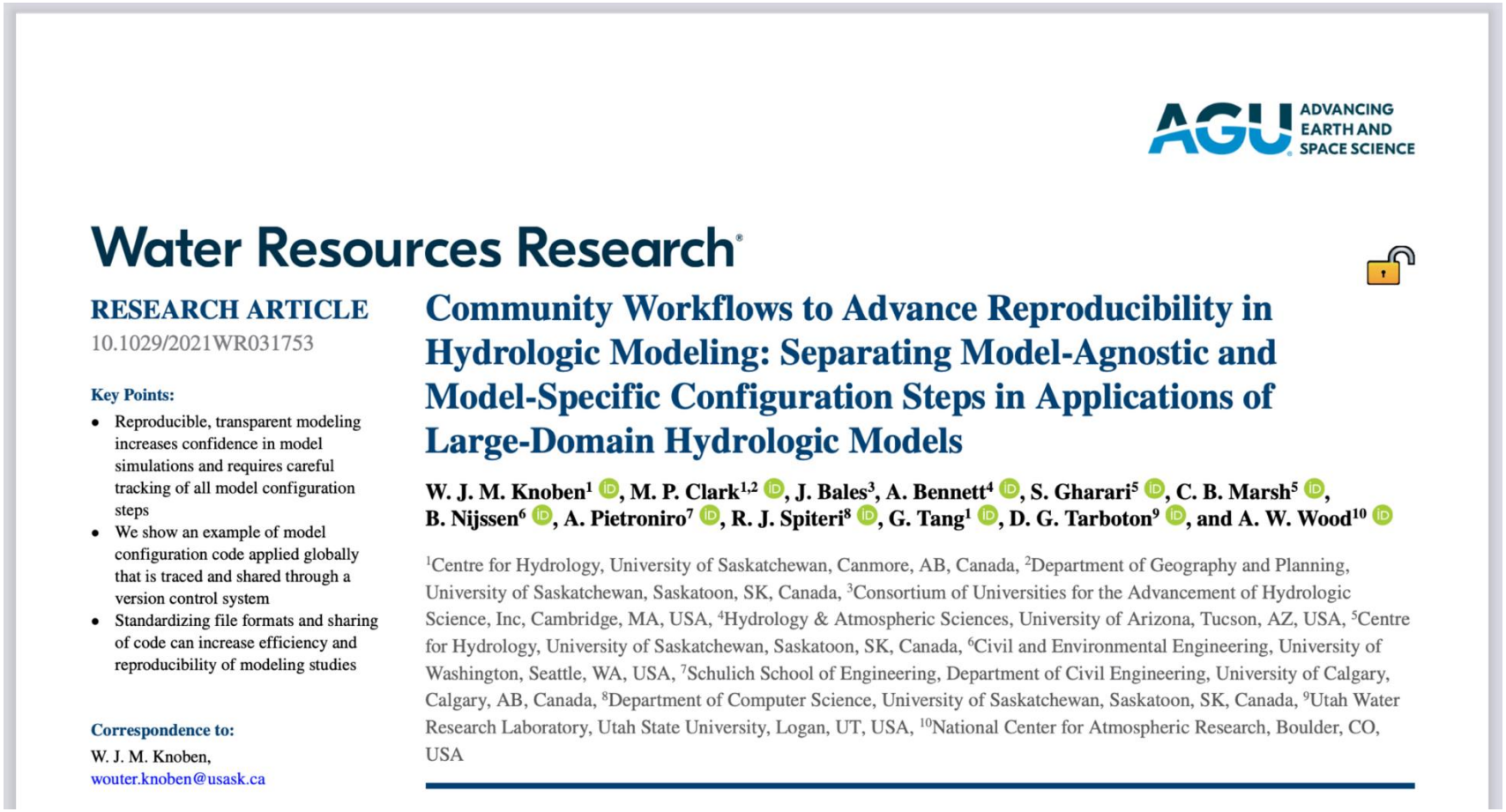
Outline

- Community Hydrological Modelling Recipes
- Computational Resources for the Training Session
- Setting up ECCC's National Water Model—MESH

Outline

- Community Hydrological Modelling Recipes
- Computational Resources for the Training Session
- Setting up ECCC's National Water Model—MESH

Recipe for Community Hydrological Modelling



AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2021WR031753

Key Points:

- Reproducible, transparent modeling increases confidence in model simulations and requires careful tracking of all model configuration steps
- We show an example of model configuration code applied globally that is traced and shared through a version control system
- Standardizing file formats and sharing of code can increase efficiency and reproducibility of modeling studies

Correspondence to:
W. J. M. Knoben,
wouter.knoben@usask.ca

Community Workflows to Advance Reproducibility in Hydrologic Modeling: Separating Model-Agnostic and Model-Specific Configuration Steps in Applications of Large-Domain Hydrologic Models

W. J. M. Knoben[1], M. P. Clark[1,2], J. Bales[3], A. Bennett[4], S. Gharari[5], C. B. Marsh[5], B. Nijssen[6], A. Pietroniro[7], R. J. Spiteri[8], G. Tang[1], D. G. Tarboton[9], and A. W. Wood[10]

[1]Centre for Hydrology, University of Saskatchewan, Canmore, AB, Canada, [2]Department of Geography and Planning, University of Saskatchewan, Saskatoon, SK, Canada, [3]Consortium of Universities for the Advancement of Hydrologic Science, Inc, Cambridge, MA, USA, [4]Hydrology & Atmospheric Sciences, University of Arizona, Tucson, AZ, USA, [5]Centre for Hydrology, University of Saskatchewan, Saskatoon, SK, Canada, [6]Civil and Environmental Engineering, University of Washington, Seattle, WA, USA, [7]Schulich School of Engineering, Department of Civil Engineering, University of Calgary, Calgary, AB, Canada, [8]Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada, [9]Utah Water Research Laboratory, Utah State University, Logan, UT, USA, [10]National Center for Atmospheric Research, Boulder, CO, USA

Knoben, W. J. M., Clark, M. P., Bales, J., Bennett, A., Gharari, S., Marsh, C. B., Nijssen, B., Pietroniro, A., Spiteri, R. J., Tang, G., Tarboton D. G. & Wood, A. W. (2022). Community Workflows to Advance Reproducibility in Hydrologic Modeling: Separating Model-Agnostic and Model-Specific Configuration Steps in Applications of Large-Domain Hydrologic Models. *Water Resources Research*, *58*(11), e2021WR031753.

- Community Workflows to Advance Reproducibility in Hydrologic Modeling (CWARHM)
- Developing first generation of community hydrological modelling workflows
- Keeping workflows open source and reproducible
- Keeping workflows efficient for real-world applications
- Separating model-agnostic and model-specific configuration parts

UNIVERSITY OF CALGARY

Recipe for Community Hydrological Modelling

manuscript submitted to *Environmental Modelling and Software*

Streamlining hydrological model applications by enhancing model-agnostic and model-specific configuration engines

Kasra Keshavarz[1]*, Alain Pietroniro[1], Mohamed Moghairib[1], Wouter J. M. Knoben[1], Paul Coderre[1], Darri Eythorsson[1], Shervan Gharari[1], Martyn P. Clark[1]

[1]Department of Civil Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada

Key Points:

- A generalized model-agnostic framework is proposed to streamline configurations of process-based hydrological models
- The framework advances reproducibility in hydrological modelling by addressing complexities involved in handling dynamic and static datasets
- Prototype applications in a high-performance computing environment demonstrate the effectiveness of the framework in reducing setup time

To be submitted to Environmental Modelling & Software

- Enhancing and proposing model-agnostic and model-specific engines for HPC environments
- Modularized methodology for the model-agnostic step for various data types and preprocessing needs
- Expanding model-specific workflows for other community modelling systems including ECCC MESH and SMHI HYPE
- Proposing strategies to objectively orchestrate the configuration stage of hydrological models

UNIVERSITY OF CALGARY

Outline

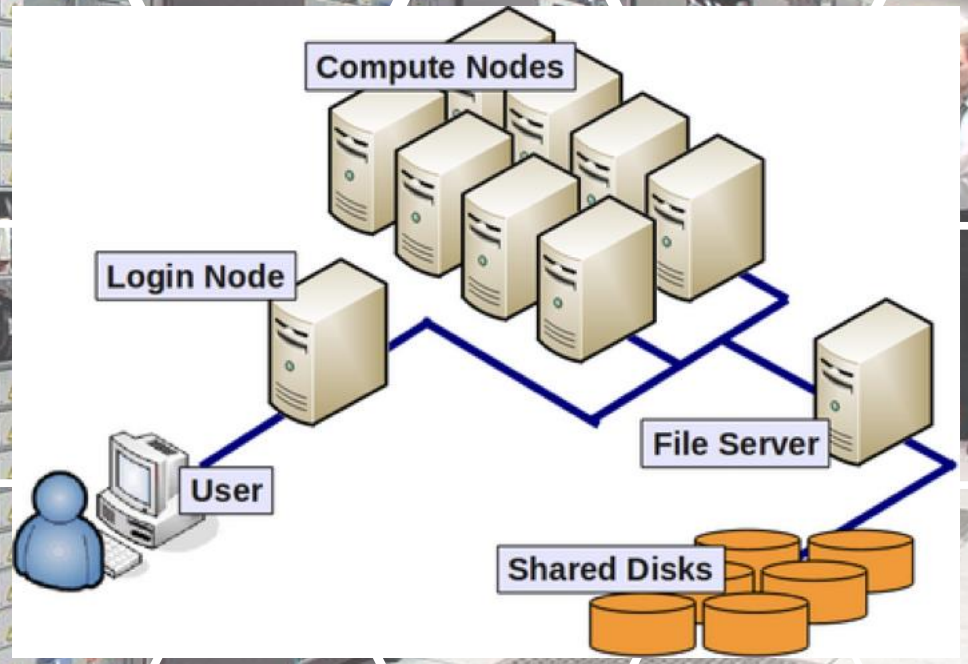
- Community Hydrological Modelling Recipes
- Computational Resources for the Training Session
  - Launch a Jupyter session on Fir
  - Create a “collection of HPC modules”
  - Create a Python virtual environment
- Setting up ECCC’s National Water Model—MESH

Computational Resources

- Wikipedia (accessed January 27th, 2025)
  "High-performance computing (HPC) is the use of supercomputers and computer clusters to solve advanced problems."

- The idea is to expand computational performance horizontally (multiple CPUs), instead of vertically (one massive CPU)

- High Performance Computing (HPC) systems provide multiple login and compute nodes for users



TOP 500
The List.

https://top500.org/

Computational Resources

Our recommendation is to connect to HPCs using either of the following methods:

jupyterlab

- Multi-user environment
- Resource efficient
- Collaborative environment
- Reproducibility mechanisms
- Strong community support
- Ease of use

Terminal

```
$ ssh USERNAME@HPC_URL
```

- Maximal flexibility
- Remote Access
- Steep learning curve
- Full control of command sessions
- Strong community support

MobaXTerm PuTTY MacOS Terminal

UNIVERSITY OF CALGARY

Computational Resources

How can we launch a Jupyter session on HPCs? A few examples…

ACCESS

Anvil at Perdue University

OPEN OnDemand

Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

Fir at Simon Fraser University

jupyterhub

UNIVERSITY OF CALGARY

ARC at the University of Calgary

OPEN OnDemand

UNIVERSITY OF CALGARY

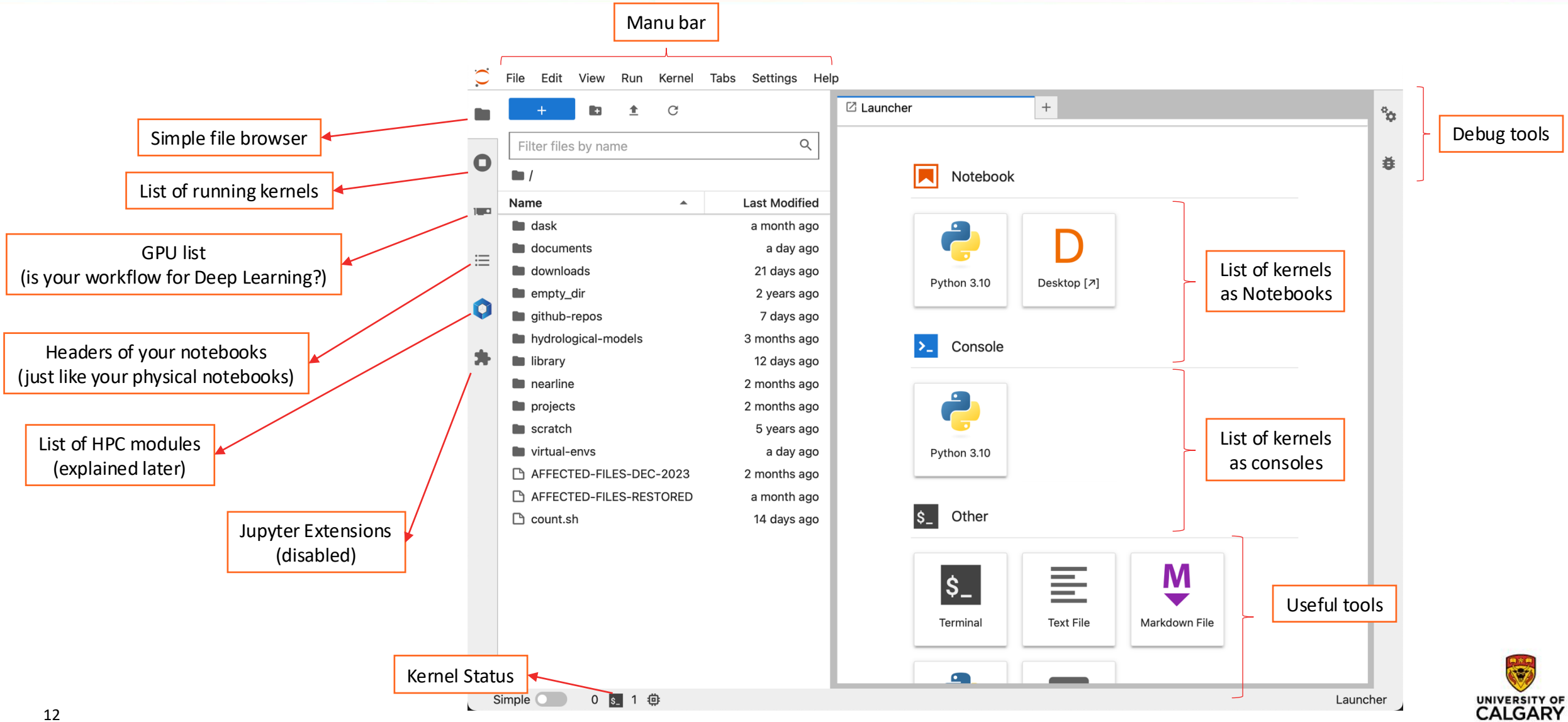
Any questions so far?

Let’s try launching a Jupyter session!

Use the following link in your browser:

https://jupyterhub.fir.alliancecan.ca

Computational Resources

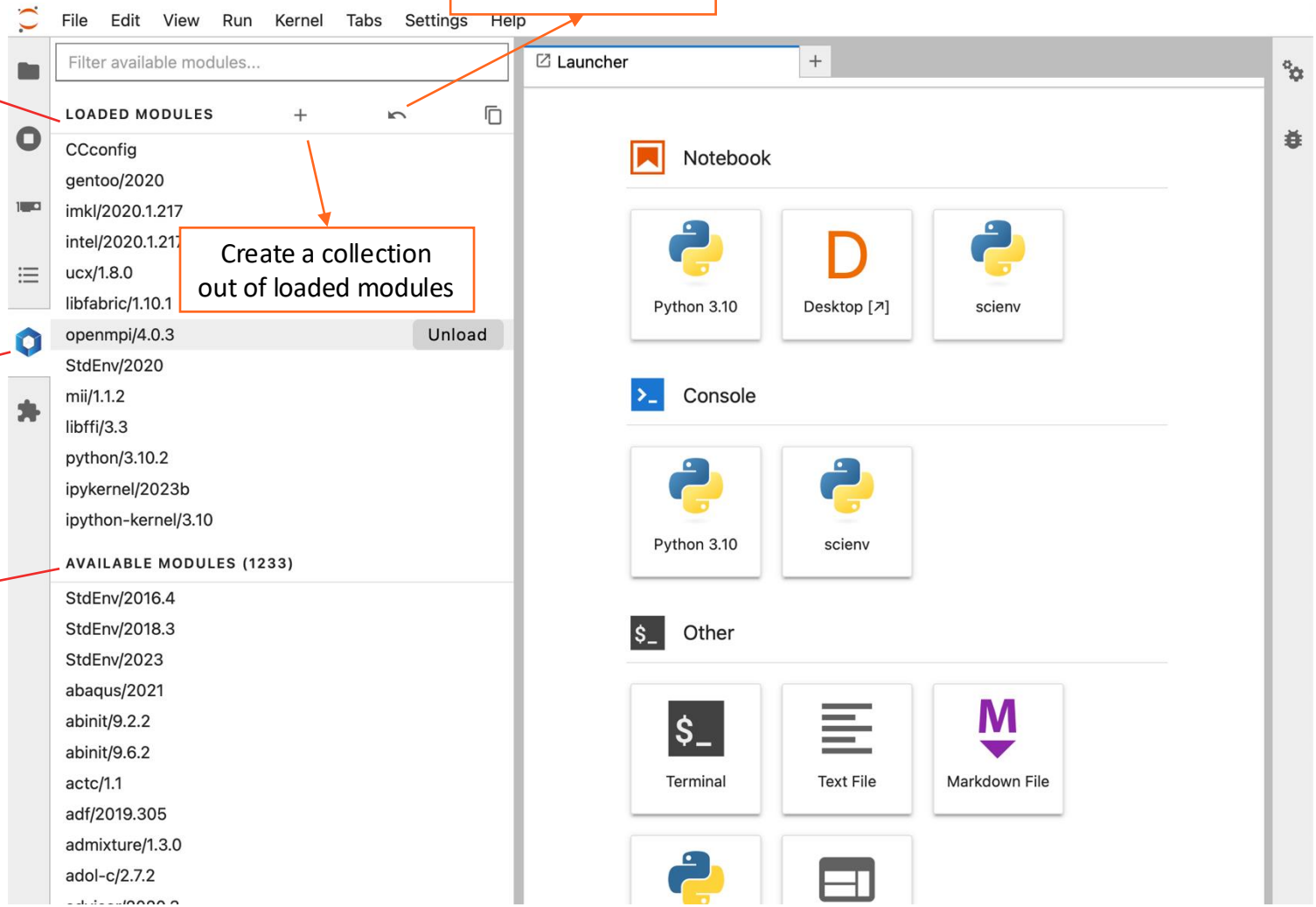


UNIVERSITY OF CALGARY

Computational Resources

Restore a previously saved collection

Loaded modules



Create a collection out of loaded modules

List of HPC modules (LMOD Extention)

Unloaded modules

A saved list of modules is called a “collection”

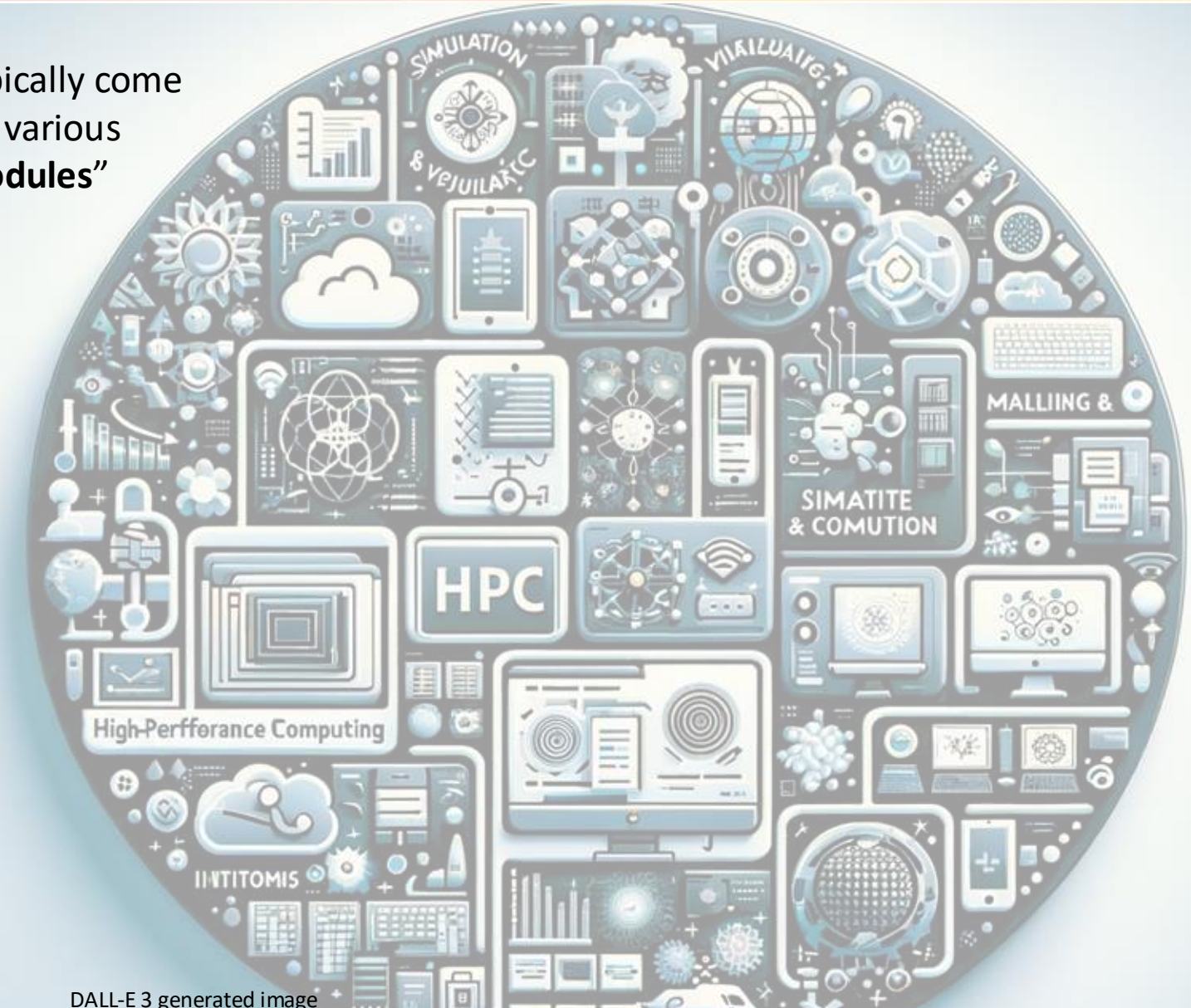
UNIVERSITY OF CALGARY

Any questions so far?

Let’s learn about HPC modules!

Computational Resources

HPCs typically come with various "**modules**"



DALL-E 3 generated image

Flexibility

Different users have different requirements

Customization

Users may need specific versions of software or libraries

Optimization

software packages are often optimized for the underlying hardware architecture

Productivity

Having commonly used software readily available, saves users time and effort

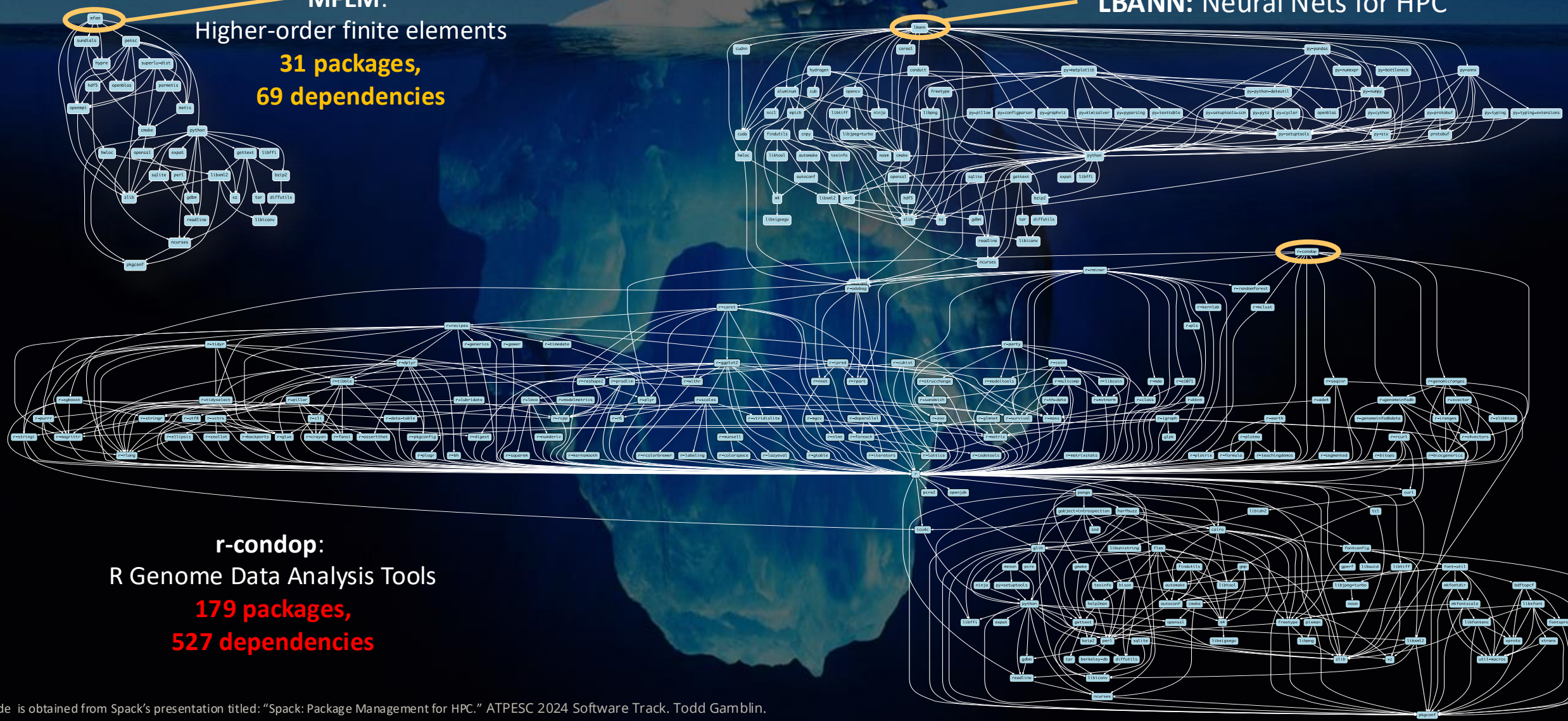
UNIVERSITY OF CALGARY

Modern scientific codes rely on icebergs of dependency libraries

MFEM:
Higher-order finite elements
31 packages,
69 dependencies

LBANN: Neural Nets for HPC
71 packages
188 dependencies

r-condop:
R Genome Data Analysis Tools
179 packages,
527 dependencies



Slide is obtained from Spack's presentation titled: "Spack: Package Management for HPC." ATPESC 2024 Software Track. Todd Gamblin.

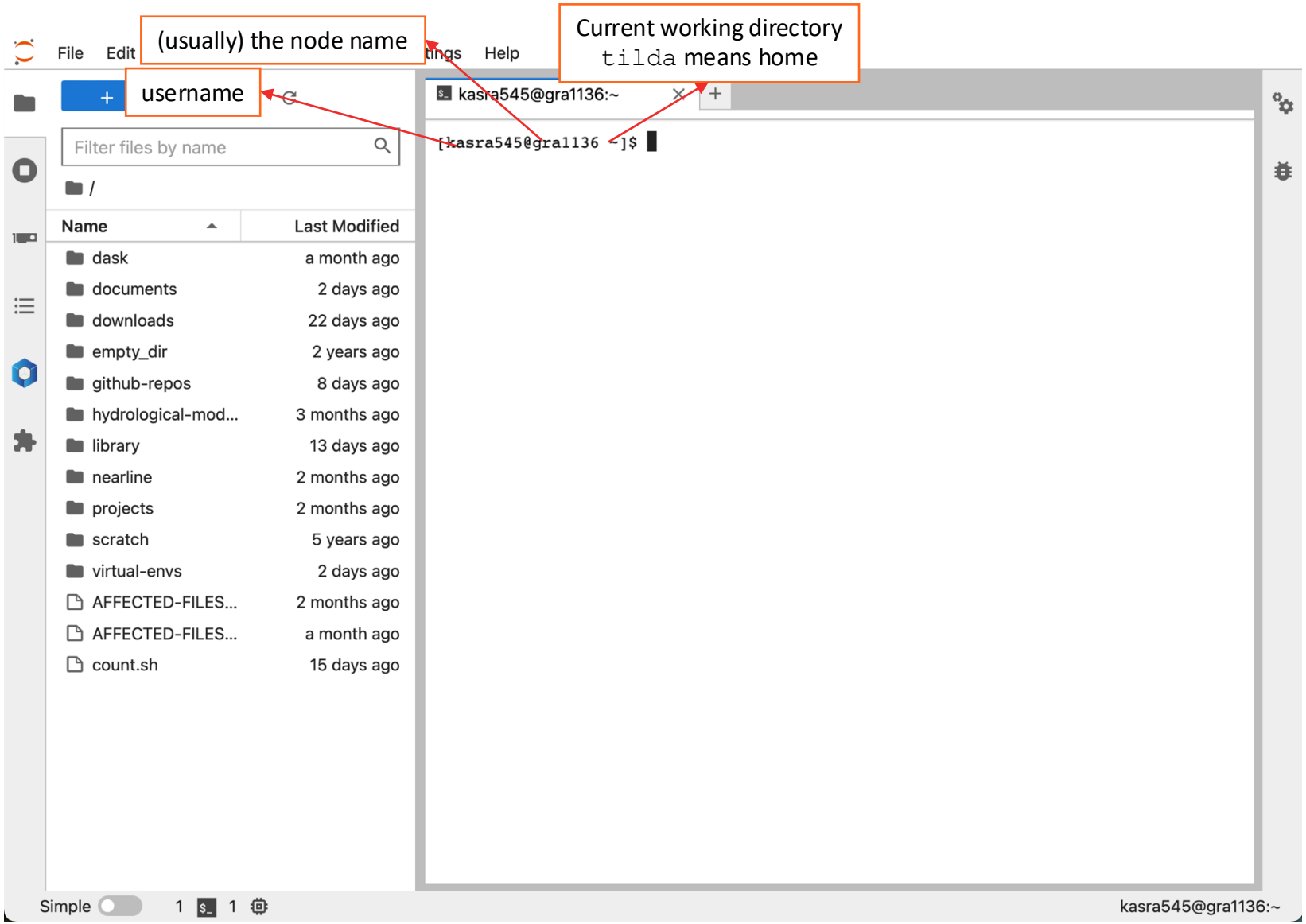
Any questions so far?

Let’s go back to our Jupyter session and get some hands-on experience with HPC modules!

Computational Resources

All instructions are available on this presentation's repository:

https://github.com/kasra-keshavarz/maf-training



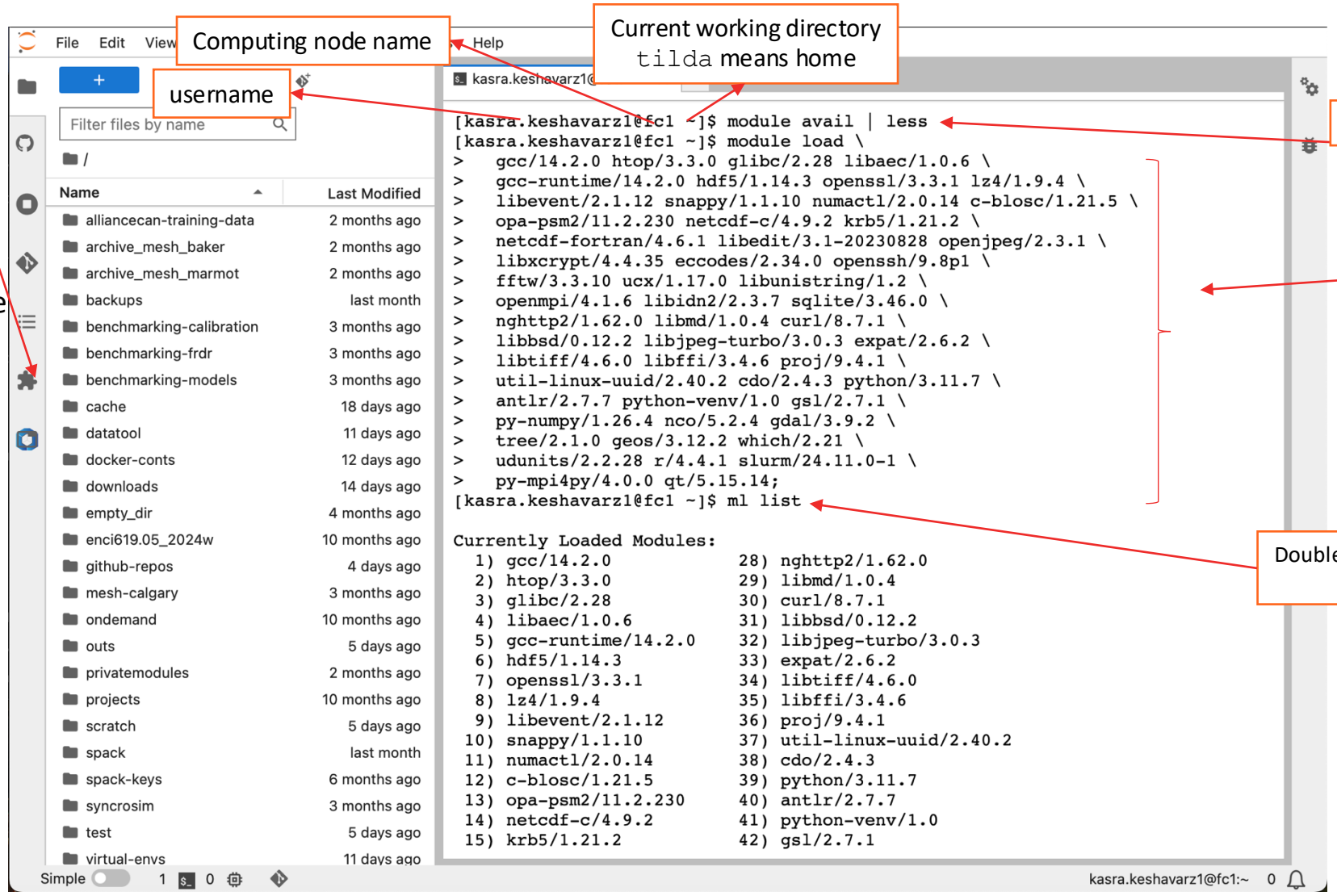
UNIVERSITY OF CALGARY

Computational Resources

Exactly similar to what we can do in this tab

All instructions are available on this presentation's repository:

https://github.com/kasra-keshavarz/maf-training

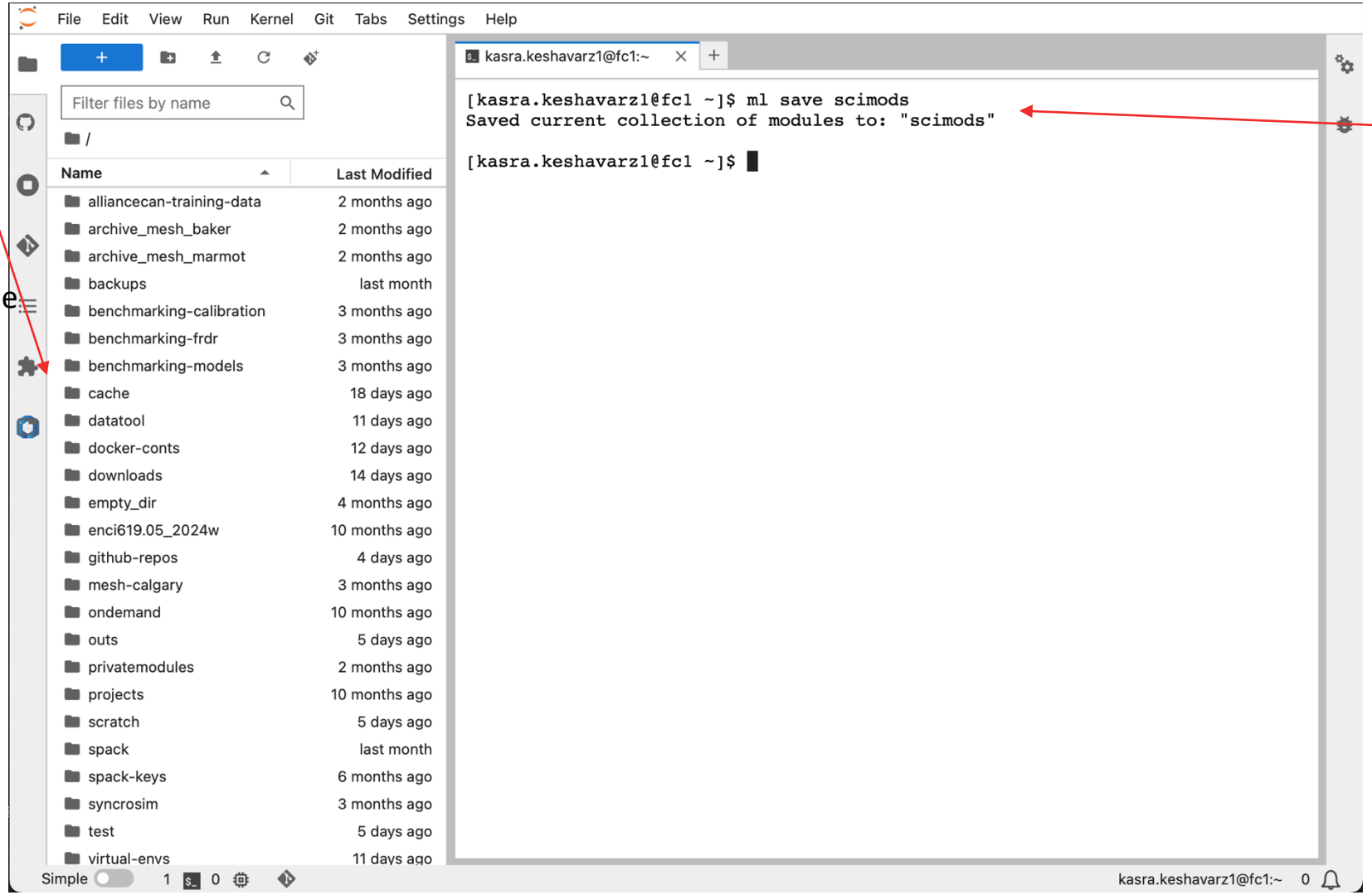


Computational Resources

Exactly similar to what we can do in this tab

All instructions are available on this presentation's repository:

https://github.com/kasra-keshavarz/maf-training



Saving the modules as a “collection”

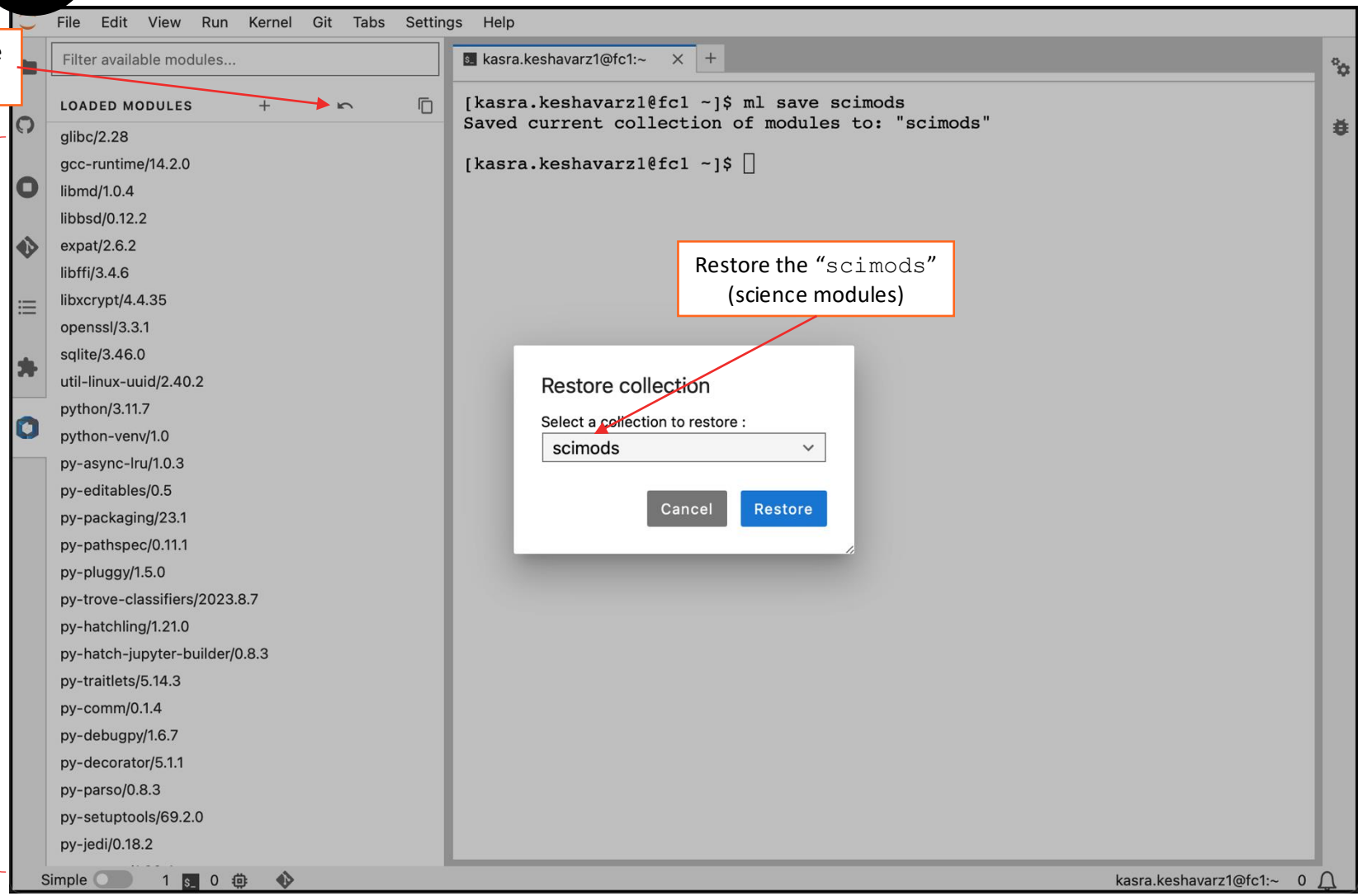
UNIVERSITY OF CALGARY

Computational Resources

Refresh the JupyterLab session webpage, and see if you can find the collection you just saved:

Check out what module collections you have

Wait for a few seconds so Jupyter loads all modules saved in "`scimods`" collection



This is only a one-time process.
You can always access all the collections you saved from Jupyter's module extension tab

UNIVERSITY OF CALGARY

Any questions so far?

We can now proceed with creating a Python virtual environment

Computational Resources

Python Programming Language

- MAF: much of our workflows are offered in Python package formats
- Python is Powerful: Python always can offer you a useful workflow
- Reproducible Environments: Python's Virtual Environments enables this requirement
- Supported on HPCs and Commercial Cloud regardless of your Operating System: it can be used everywhere
- Fir is no exception: Python is fully supported on many HPCs including Fir, by default.

Computational Resources

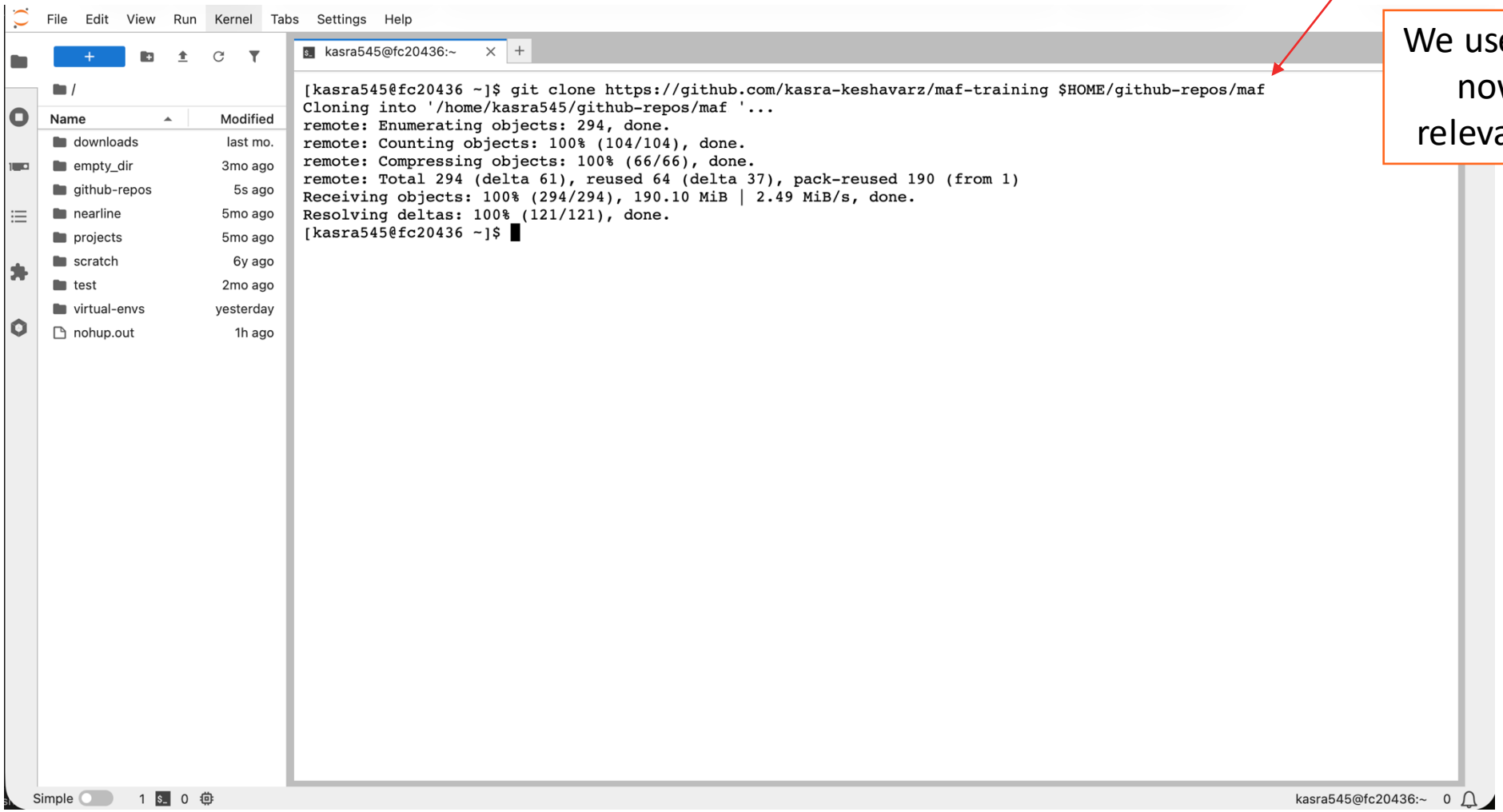
Before setting up the Python Environment, we

need to clone (download) the training session's GitHub repository

Computational Resources

"clone"-ing the repository

We use this terminal session now to download the relevant GitHub repository



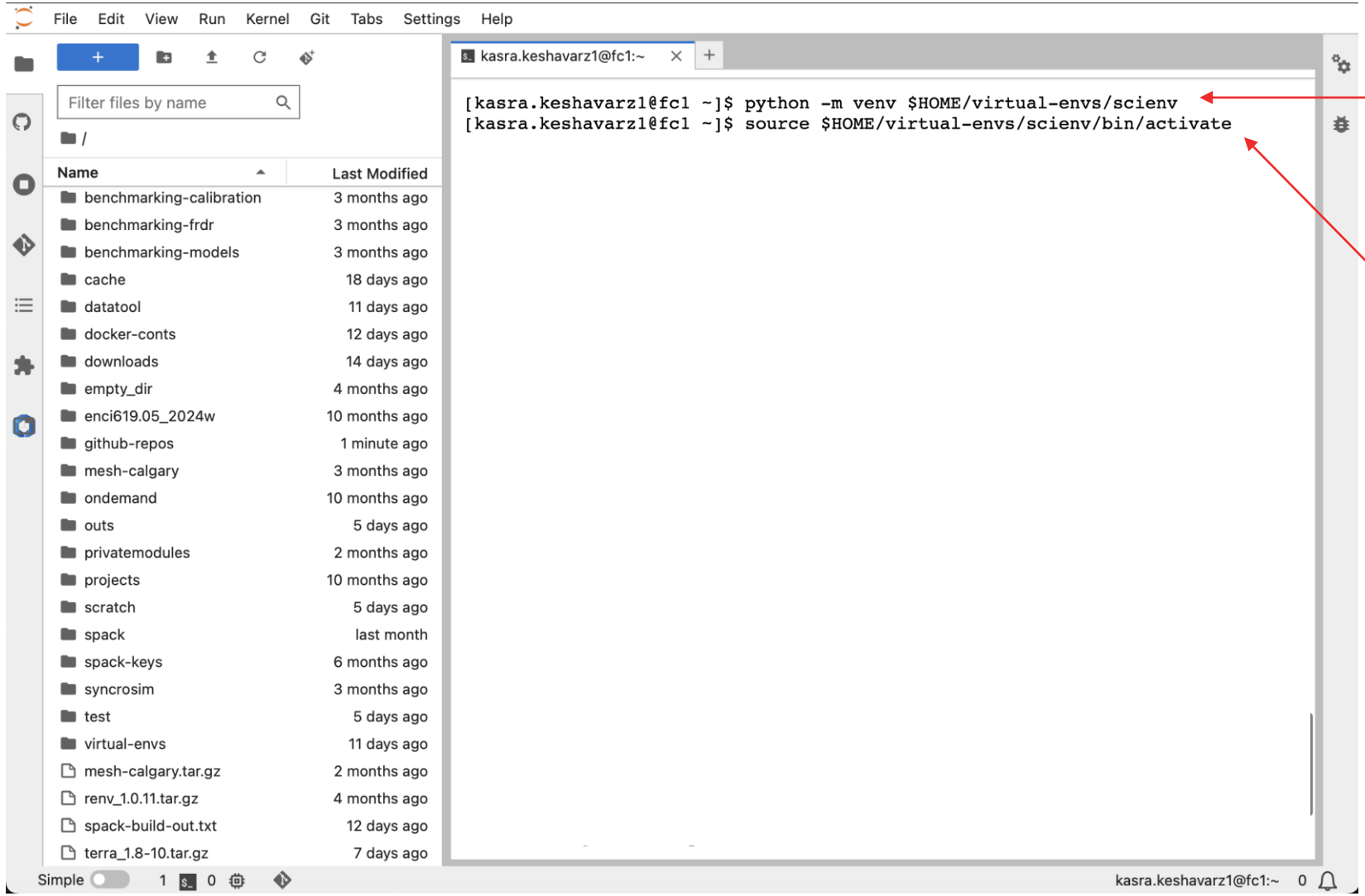
UNIVERSITY OF CALGARY

Computational Resources

Now we can set up a virtual environment using the list of packages described in the training session's repository

Let's go back to our Jupyter session!

Computational Resources

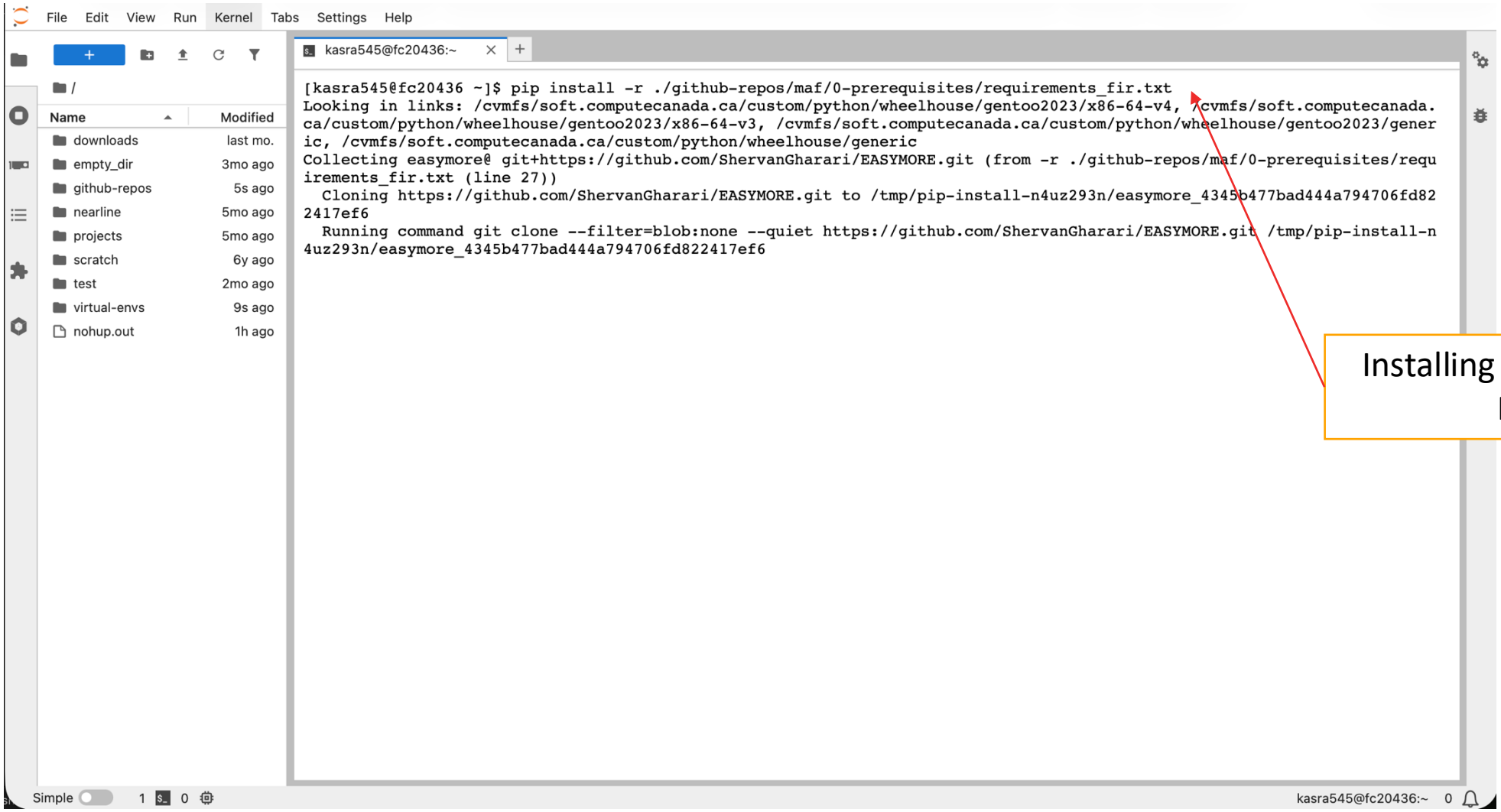


Creating an empty Virtual Environment

Activating the empty Virtual Environment

UNIVERSITY OF CALGARY

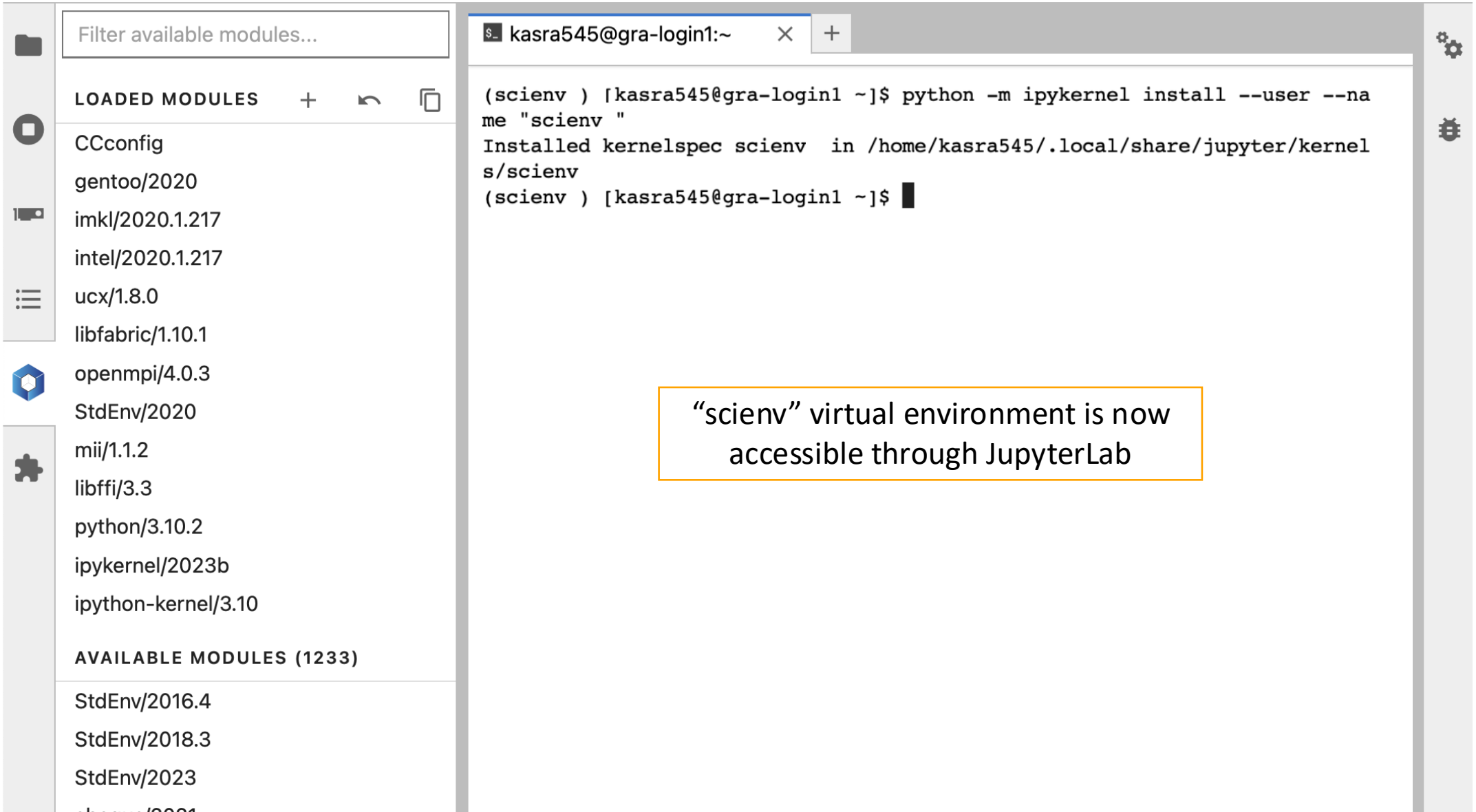
Computational Resources



Installing necessary Python Packages

UNIVERSITY OF CALGARY

Computational Resources

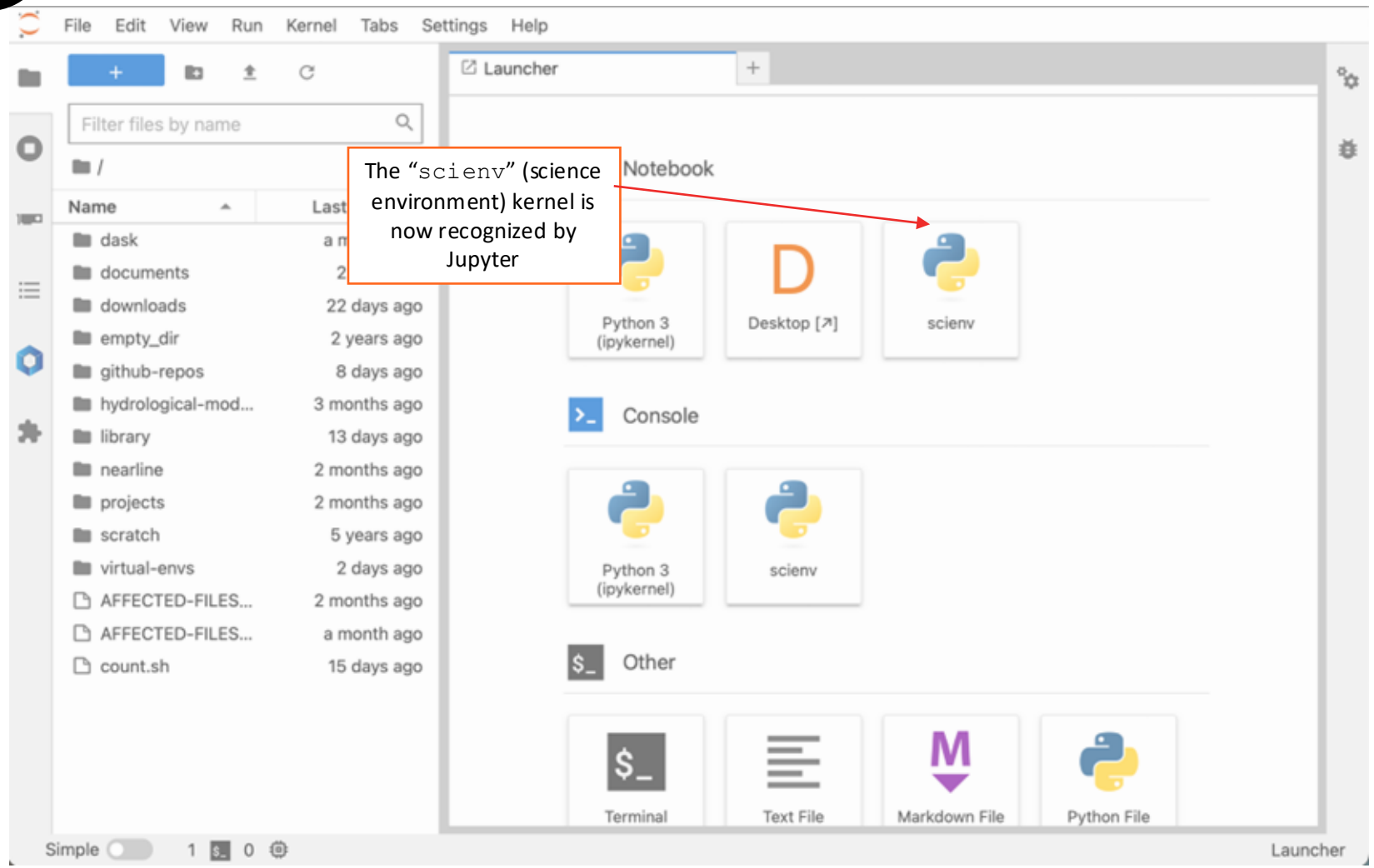


"scienv" virtual environment is now accessible through JupyterLab

UNIVERSITY OF CALGARY

Computational Resources

Refresh the JupyterLab session webpage, and see if you can find the collection you just saved:

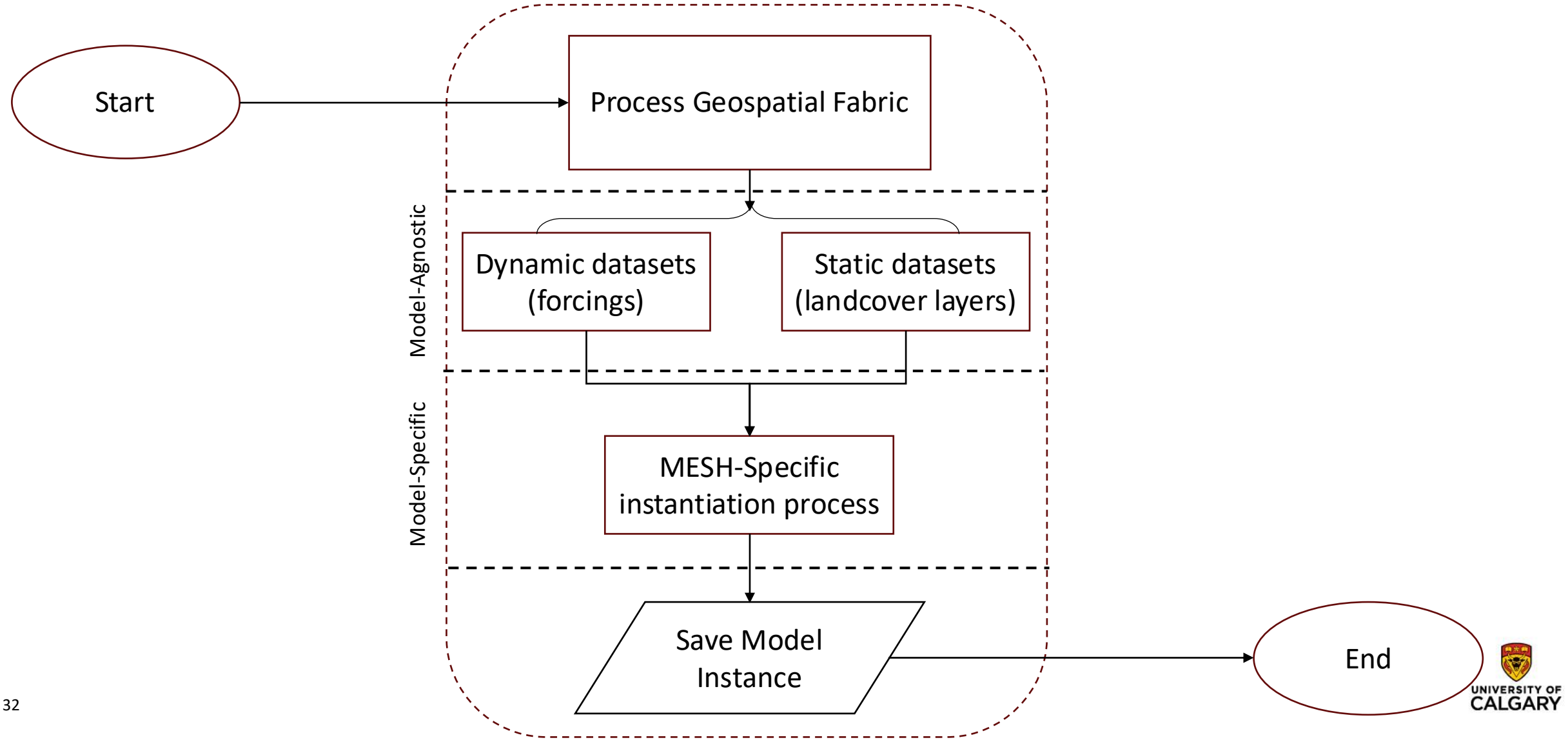


UNIVERSITY OF CALGARY

Outline

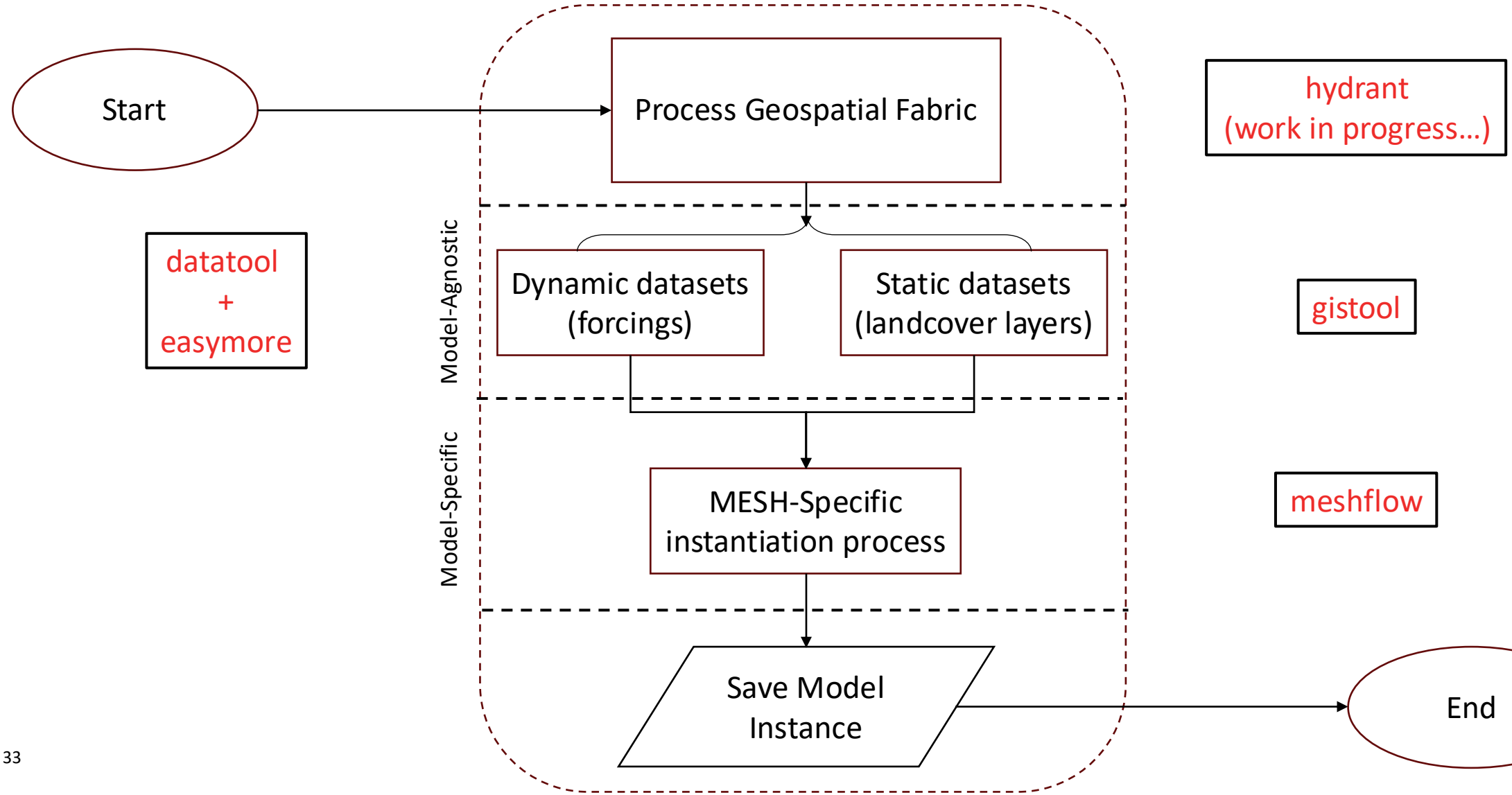
- Community Hydrological Modelling Recipes
- Computational Resources for the Training Session
- Setting up ECCC's National Water Model—MESH
  - General overview of steps
  - Introducing tools for each step
  - Running the steps together for the MESH instance

Setting up ECCC's National Water Model—MESH



UNIVERSITY OF CALGARY

Setting up ECCC's National Water Model—MESH



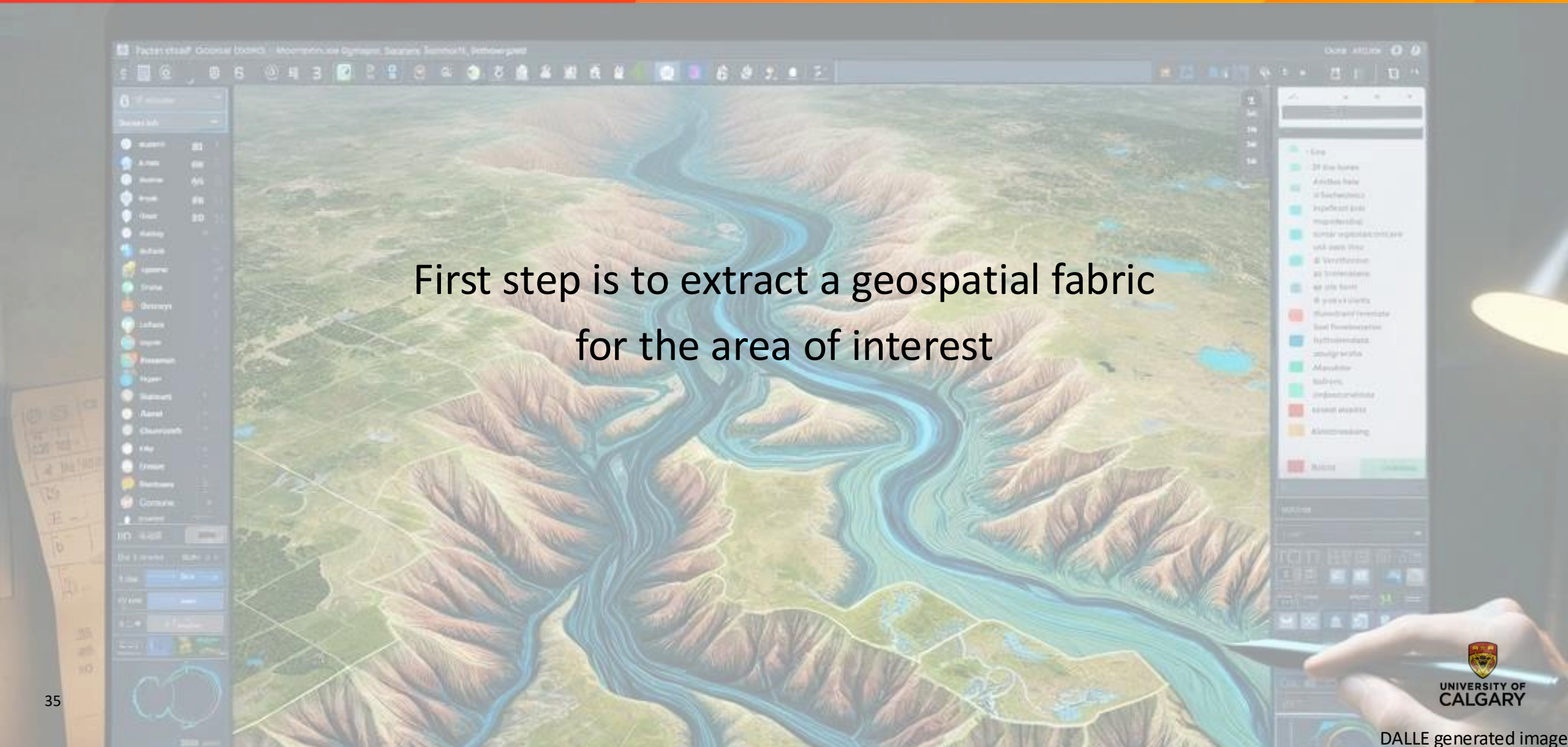
UNIVERSITY OF CALGARY

Setting up ECCC's National Water Model—MESH

Configuring ECCC's MESH with the following datasets for the Bow River at Calgary:

Meteorological Forcing Dataset	Landcover Dataset	Geospatial Fabric Dataset	Study Area
Canadian Surface Reanalysis version 3.2 (CaSRv3.2)	North American Land Change Monitoring System 2020 (NALCMS 2020)	MERIT-Basins	Bow River at Calgary

Setting up ECCC's National Water Model—MESH

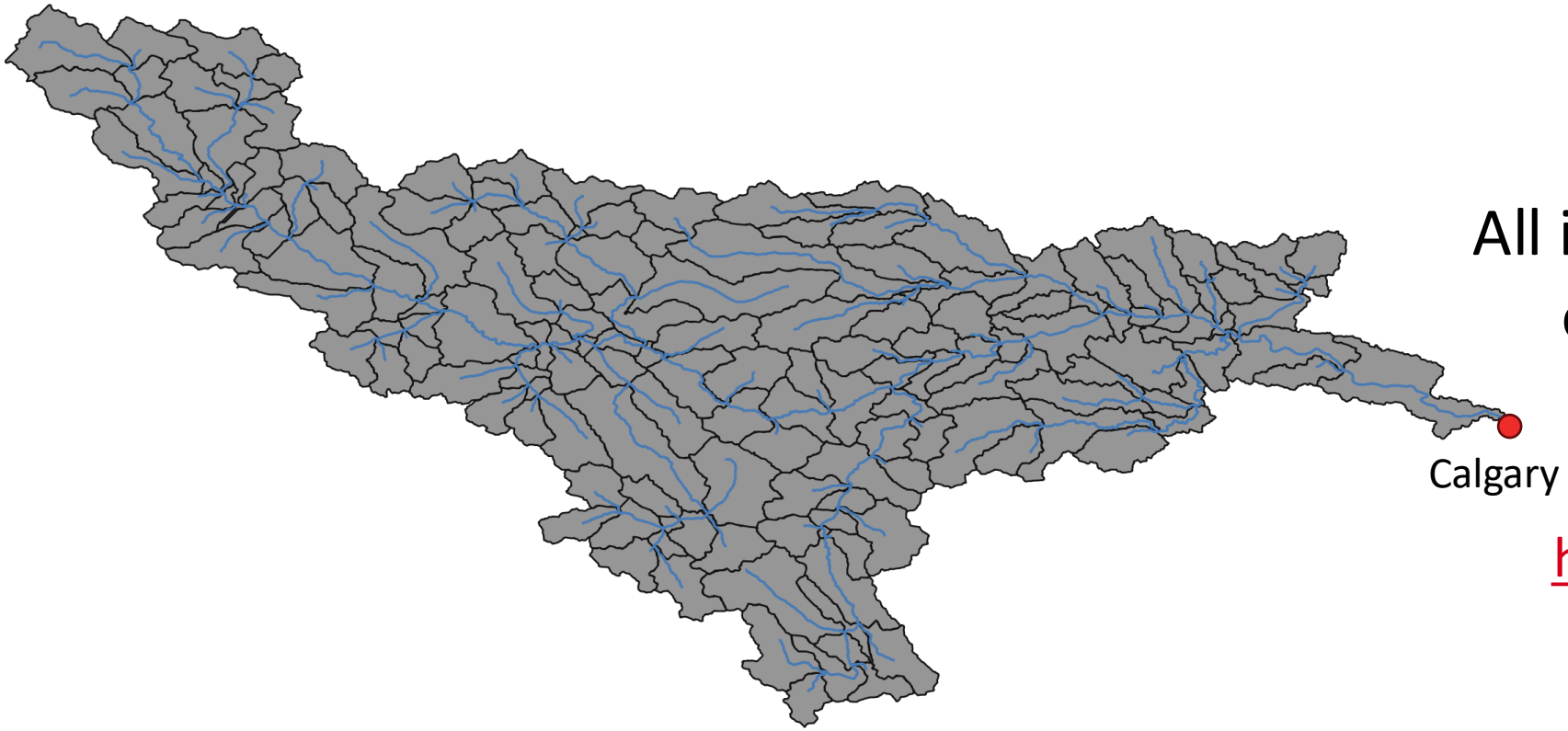


First step is to extract a geospatial fabric for the area of interest

UNIVERSITY OF CALGARY

DALLE generated image

Setting up ECCC's National Water Model—MESH



All instructions are available on this presentation's repository:

https://github.com/kasra-keshavarz/maf-training

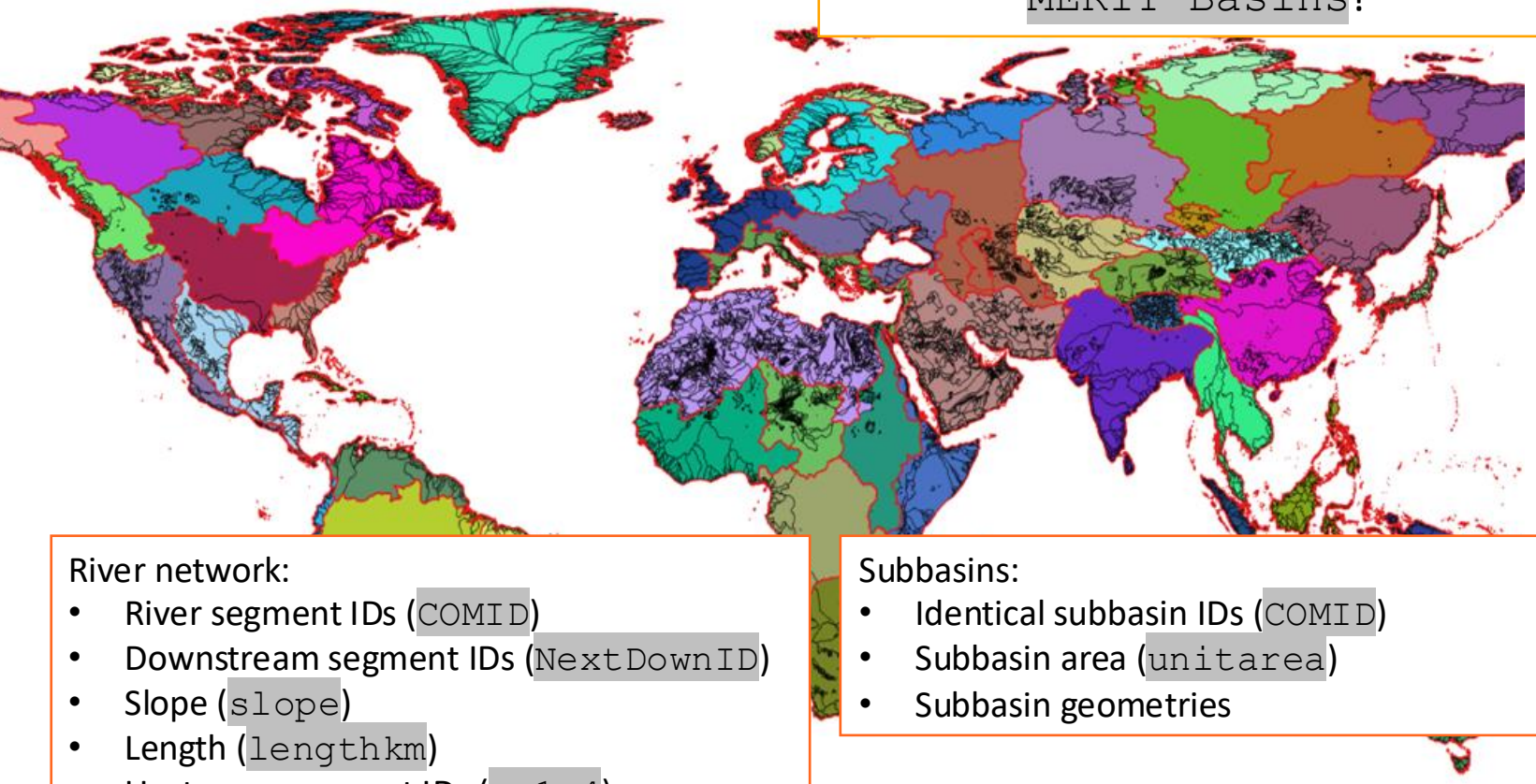
Bow River at Calgary

Sub-basins and river network extracted from `MERIT-Basins` using `Hydrant v0.1.0-dev0`

UNIVERSITY OF CALGARY

Setting up ECCC's National Water Model—MESH

What information is provided through `MERIT-Basins`?



- 25 km^2 drainage area threshold for channels
- Corrected/refined basin/region definitions
- Below 25 km^2 non-channelized areas along the coast or some endorheic areas (incomplete catchments or hillslopes) excluded
- ~2.94 million river reaches (unit catchments)

River network:
- River segment IDs (`COMID`)
- Downstream segment IDs (`NextDownID`)
- Slope (`slope`)
- Length (`lengthkm`)
- Upstream segment IDs (`up1-4`)
- Upstream drainage area (`uparea`)
- River geometries

Subbasins:
- Identical subbasin IDs (`COMID`)
- Subbasin area (`unitarea`)
- Subbasin geometries

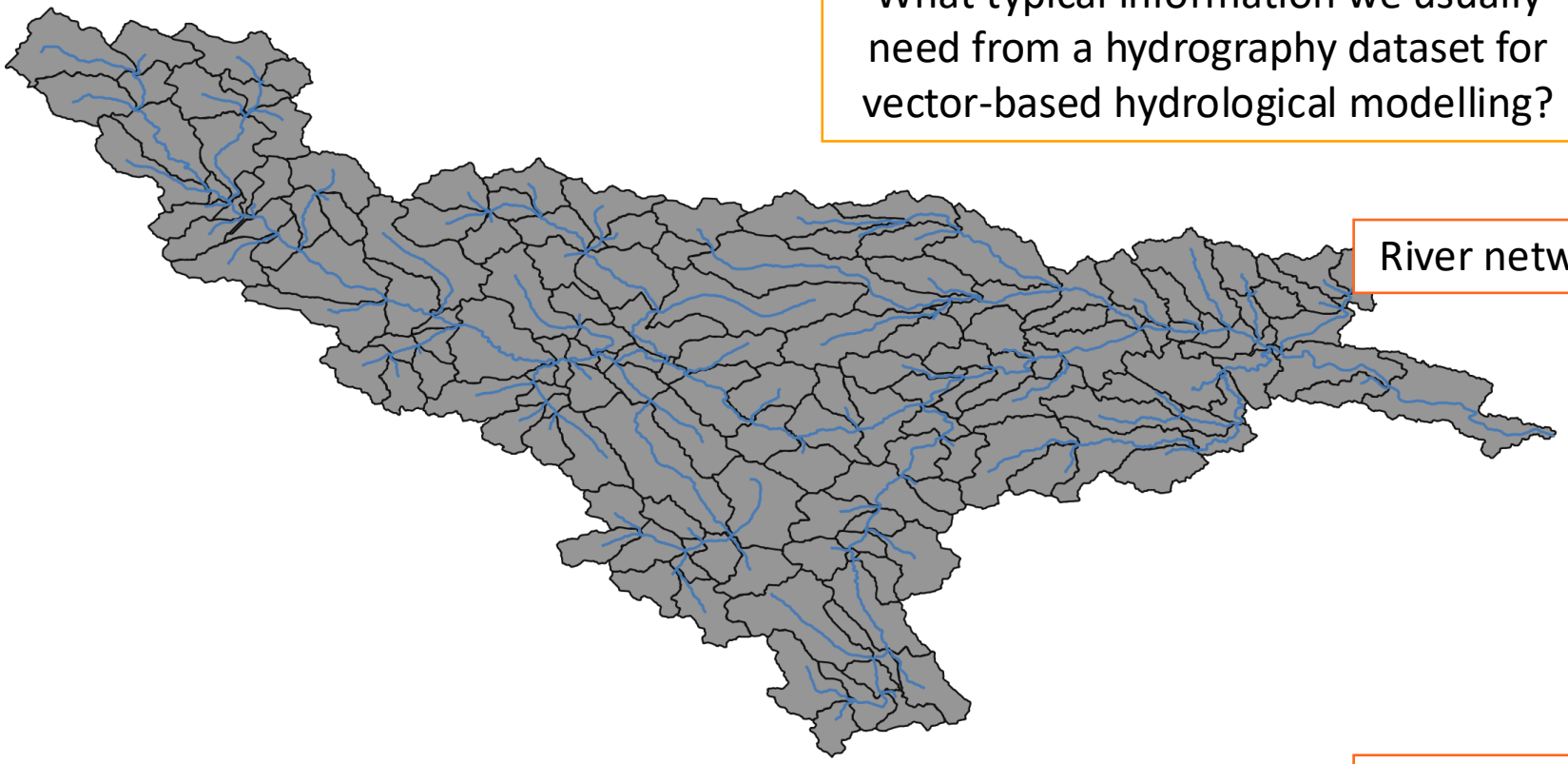
Non-contributing areas:
- Non-contributing area IDs (`FID`)
- Non-contributing area geometries

`MERIT-Basins` webpage: https://www.reachhydro.org/home/params/merit-basins

UNIVERSITY OF CALGARY

Setting up ECCC's National Water Model—MESH

What typical information we usually need from a hydrography dataset for vector-based hydrological modelling?



Bow River at Calgary

Sub-basins and river network extracted from `MERIT-Basins` using `Hydrant v0.1.0-dev0`

River network:

- River segment ID values
- Immediate downstream segment ID values
- Length

Depending on your routing scheme you may need:

- Slope [optional]
- Width [optional]
- etc.

Subbasins:

- Subbasin ID values
- Correspondence Mapping Table between Subbasin and River Segment IDs (could be identical value)
- Subbasin Area [optional]
- etc.

Let’s prepare our geospatial fabric for the study area

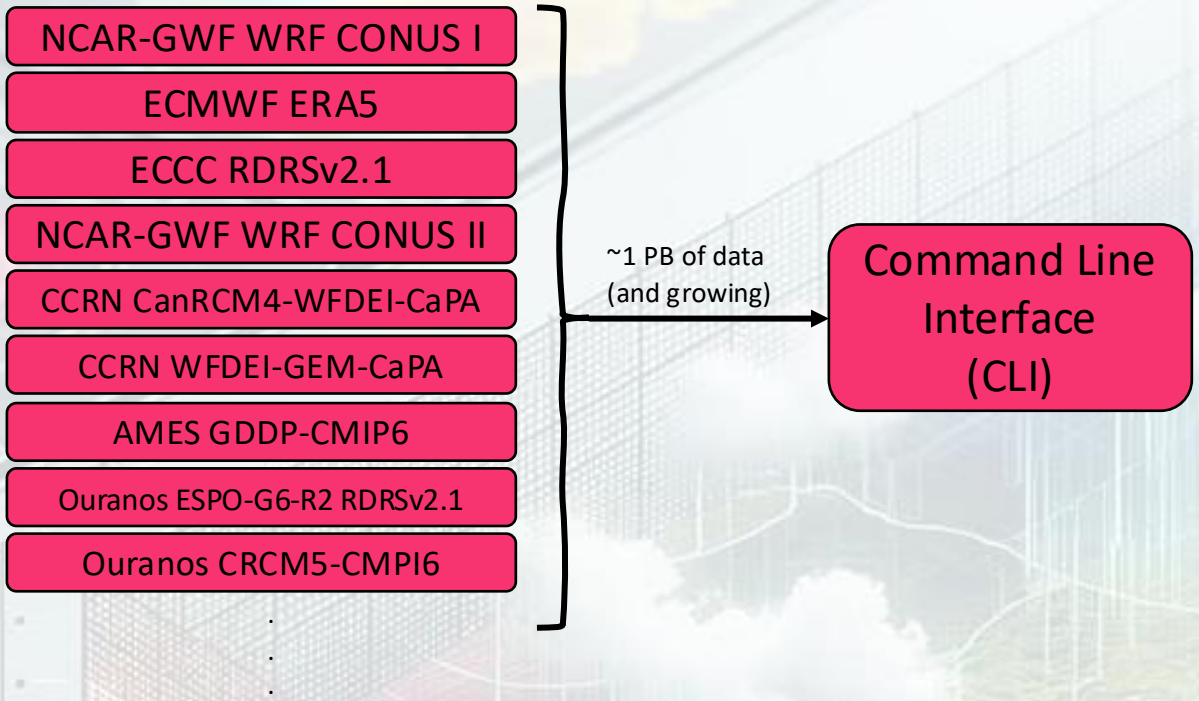
We use MERIT-Basins as the source dataset and extract sub-basins and river segments for the area

Setting up ECCC's National Water Model—MESH

Second step is to run the model-agnostic processes

DALLE generated image

datatool




```
./extract-dataset.sh  --dataset=CONUS1 \
  --dataset-dir="/project/rpp-kshook/Model_Output/WRF/CONUS/CTRL" \
  --output-dir="$HOME/scratch/conus_i_output/" \
  --start-date="2000-10-01 00:00:00" \
  --end-date="2013-09-30 23:00:00" \
  --lat-lims=49,54  \
  --lon-lims=-120,-98 \
  --variable="T2,PREC_ACC_NC,Q2,ACSWDNB,ACLWDNB,U10,V10,PSFC" \
  --prefix="conus_i" \
  --email="youremail@company.ca" \
  -j;
```

datatool

Steps:

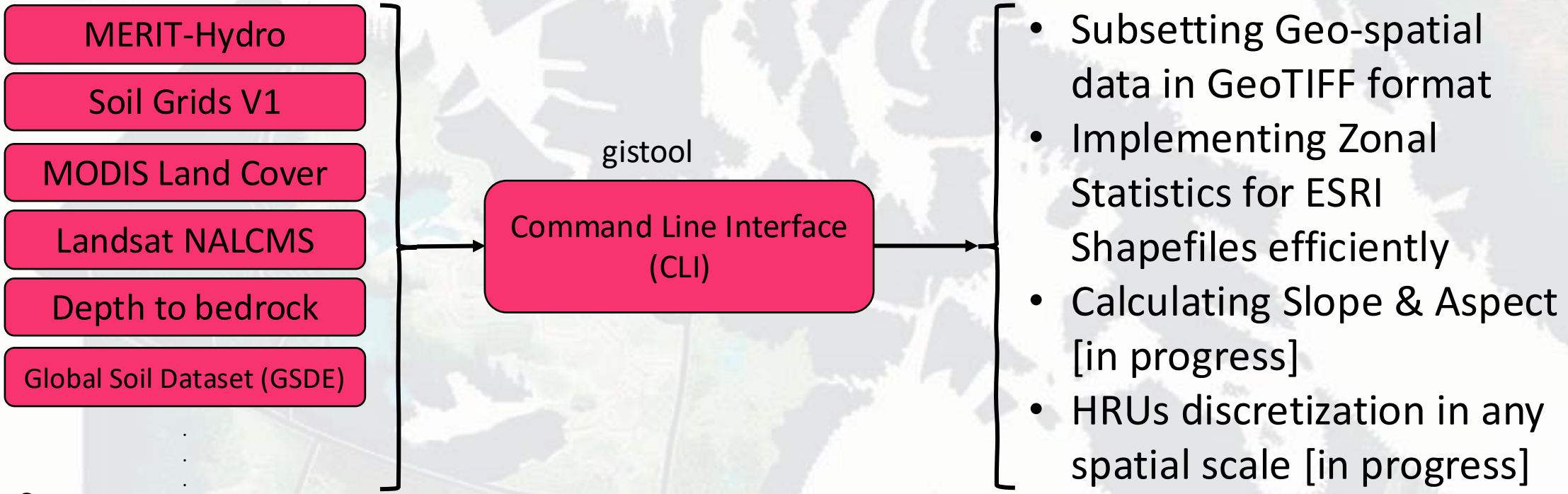
1. Login to cluster of choice,
2. Make sure you have access to the designated project space and allocation,
3. Clone the remote GitHub repository,
4. Subset any spatial and temporal extents of interest for any of the available datasets with only **one line of code**.

https://github.com/CH-Earth/datatool.git

UNIVERSITY OF CALGARY

DALL-E generate image

gistool



Steps:

1. Login to your cluster of choice,
2. Make sure you have access to the designated project space and allocation,
3. Clone the repository,
4. Subset and/or implement zonal statistics for any spatial and temporal extents of interest for any of the included datasets with only **one line of code**.

https://github.com/CH-Earth/gistool.git

UNIVERSITY OF CALGARY

Setting up ECCC's National Water Model—MESH

datatool
(HPC-agnostic)

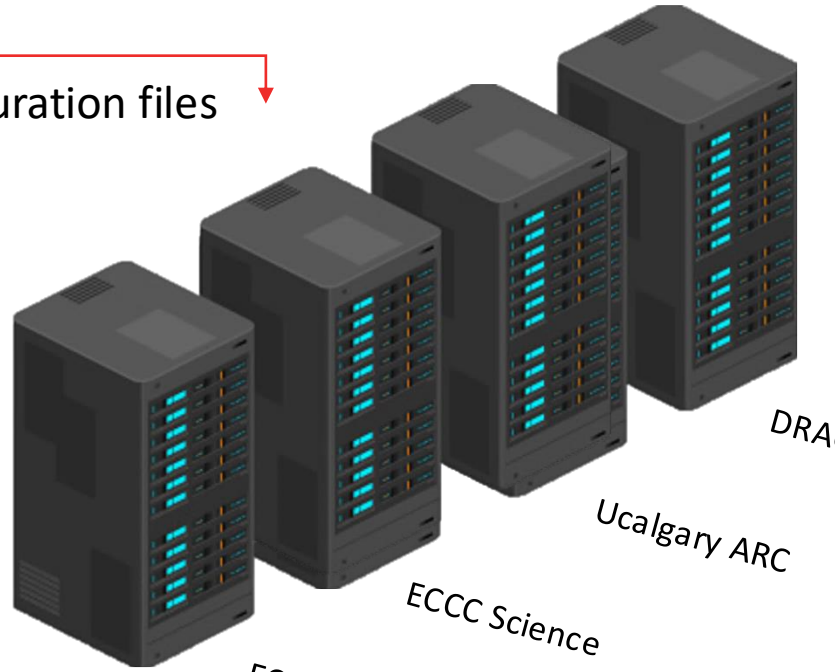
Hardware dependent
job settings

HPC-specific settings

Scheduler dependent
settings

JSON
files

JSON configuration files



DRAC Graham

Ucalgary ARC

ECCC Science

ECCC Collab

… any other HPC

slurm workload manager

Altair | PBS Works™

IBM Spectrum LSF

UNIVERSITY OF CALGARY

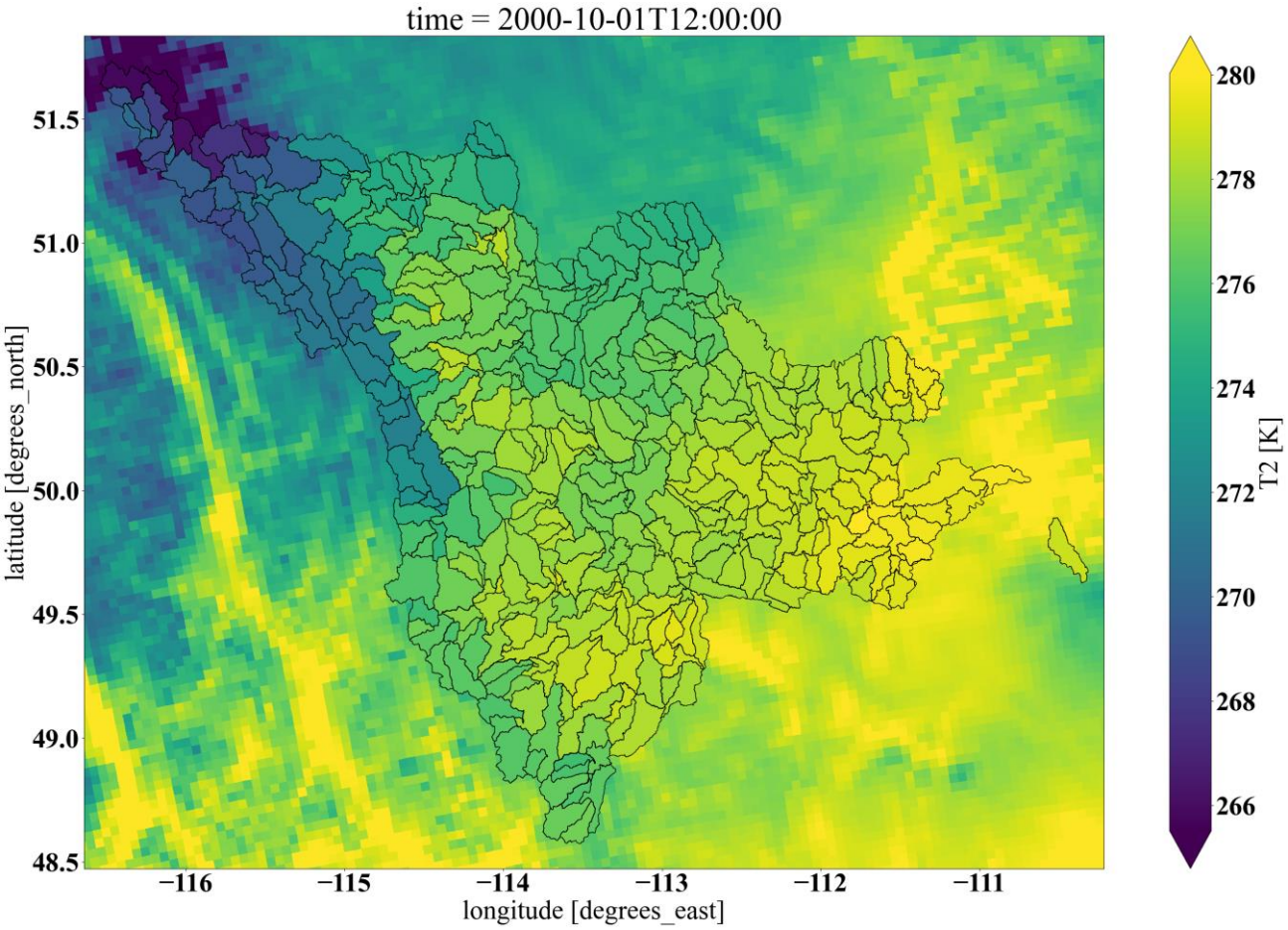
https://datatool.readthedocs.io/

Let’s clone *datatool* and *gistool* repositories and see if everything is alright with the HPC-specific JSON file.

Check the accounts you have access to and adjust the JSON file:

$ sacctmgr show associations user=$USER

Easymore




```
kasras-mbp:easymore ShervanGharari$ easymore
Usage: easymore [OPTIONS] COMMAND [ARGS]...

  EASYMORE is a collection of functions that allows extraction of the
  data from a NetCDF file for a given shapefile such as a basin,
  catchment, points or lines. It can map gridded data or model output to
  any given shapefile and provide area average for a target variable.

Options:
  --version  Show the version and exit.
  --help     Show this message and exit.

Commands:
  cli   Run Easymore using CLI
  conf  Run Easymore using a JSON configuration file

  For bug reports, questions, and discussions open an issue at
  https://github.com/ShervanGharari/EASYMORE.git
```

Command Line Interface (CLI)
and
Python Library

https://github.com/ShervanGharari/EASYMORE.git

UNIVERSITY OF CALGARY

Any questions so far?

Setting up ECCC's National Water Model—MESH

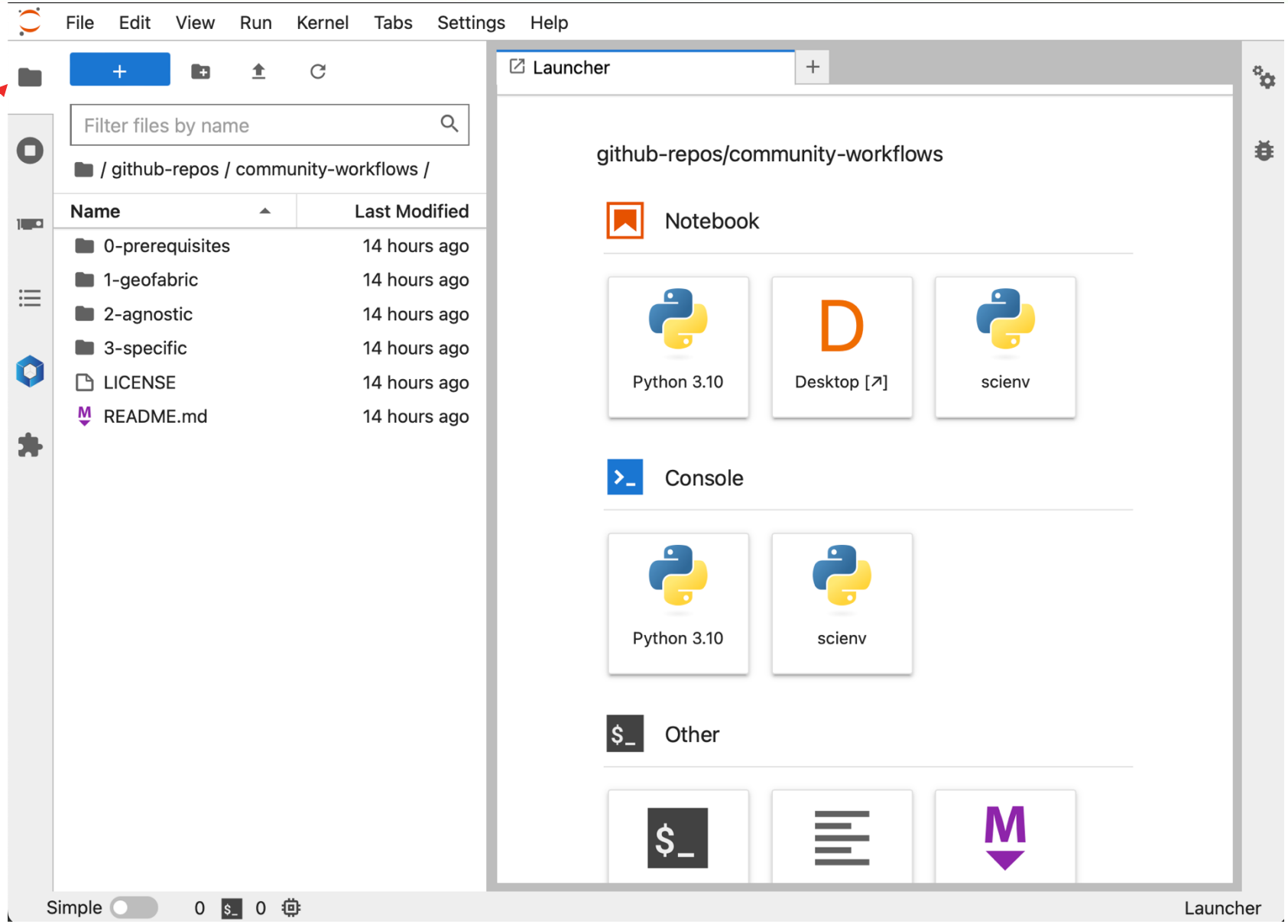
Third and final step is to run the model-specific processes

UNIVERSITY OF CALGARY

DALLE generated image

Setting up ECCC's National Water Model—MESH

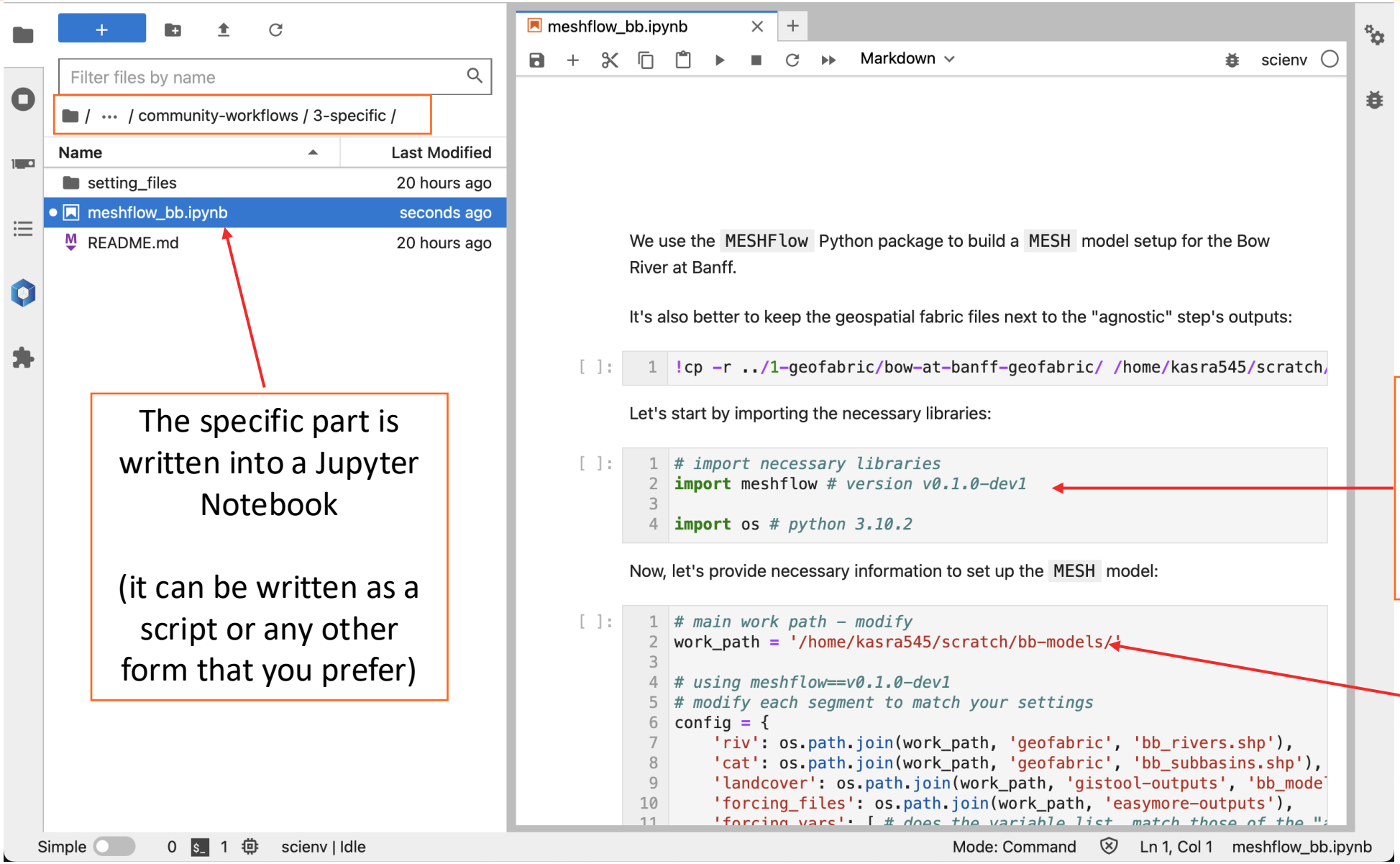
Navigate using JupyterLab file browser



UNIVERSITY OF CALGARY

Let’s switch back to the Jupyter environment for the final step.

Setting up ECCC's National Water Model—MESH



The specific part is written into a Jupyter Notebook

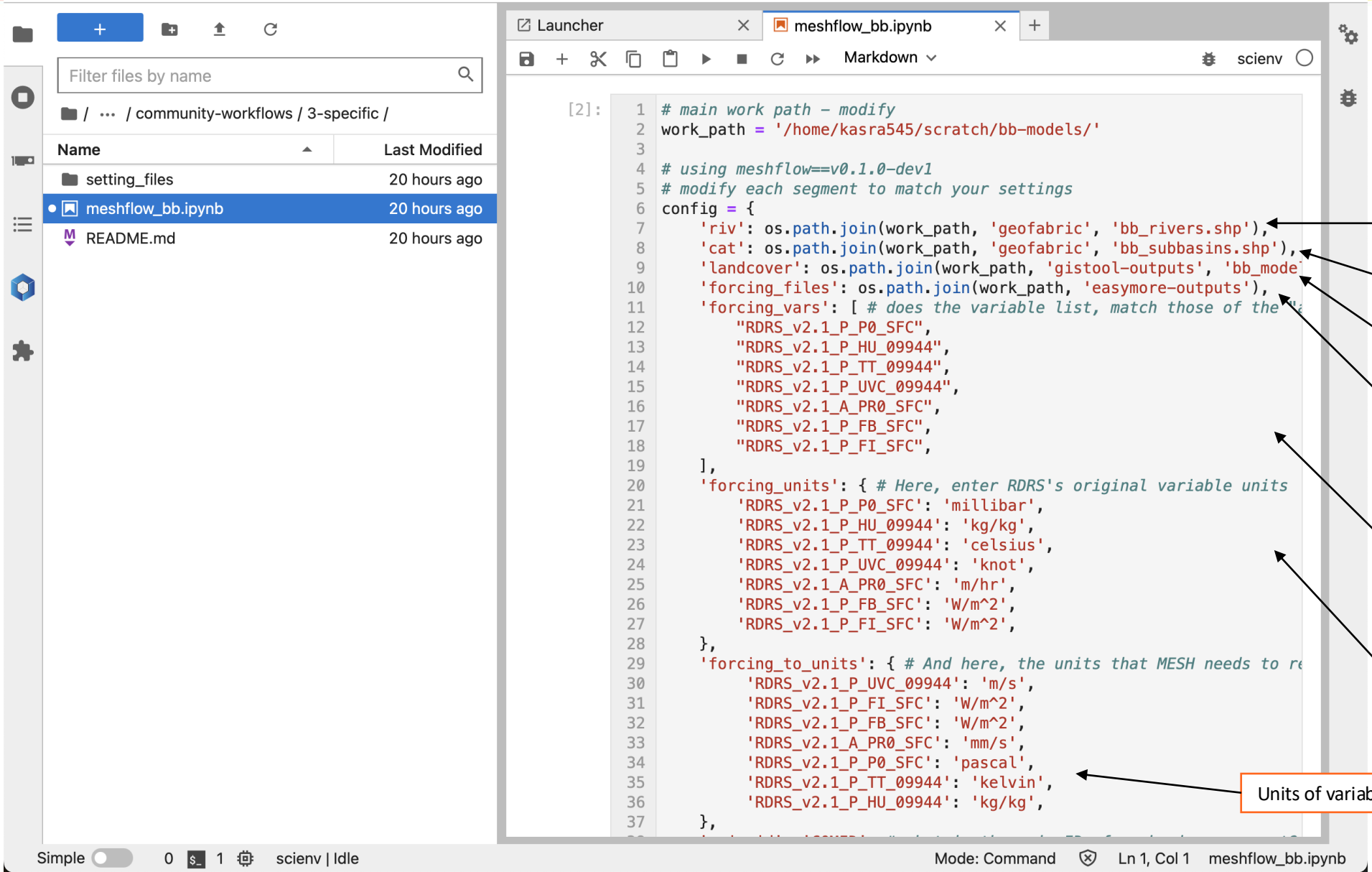
(it can be written as a script or any other form that you prefer)

`Meshflow` is the package to build a MESH model out of all the data we processed so far

The root address of where the data is stored

UNIVERSITY OF CALGARY

Setting up ECCC's National Water Model—MESH



River network geometry file

Subbasin geometry file

Landcover fractions file(s)

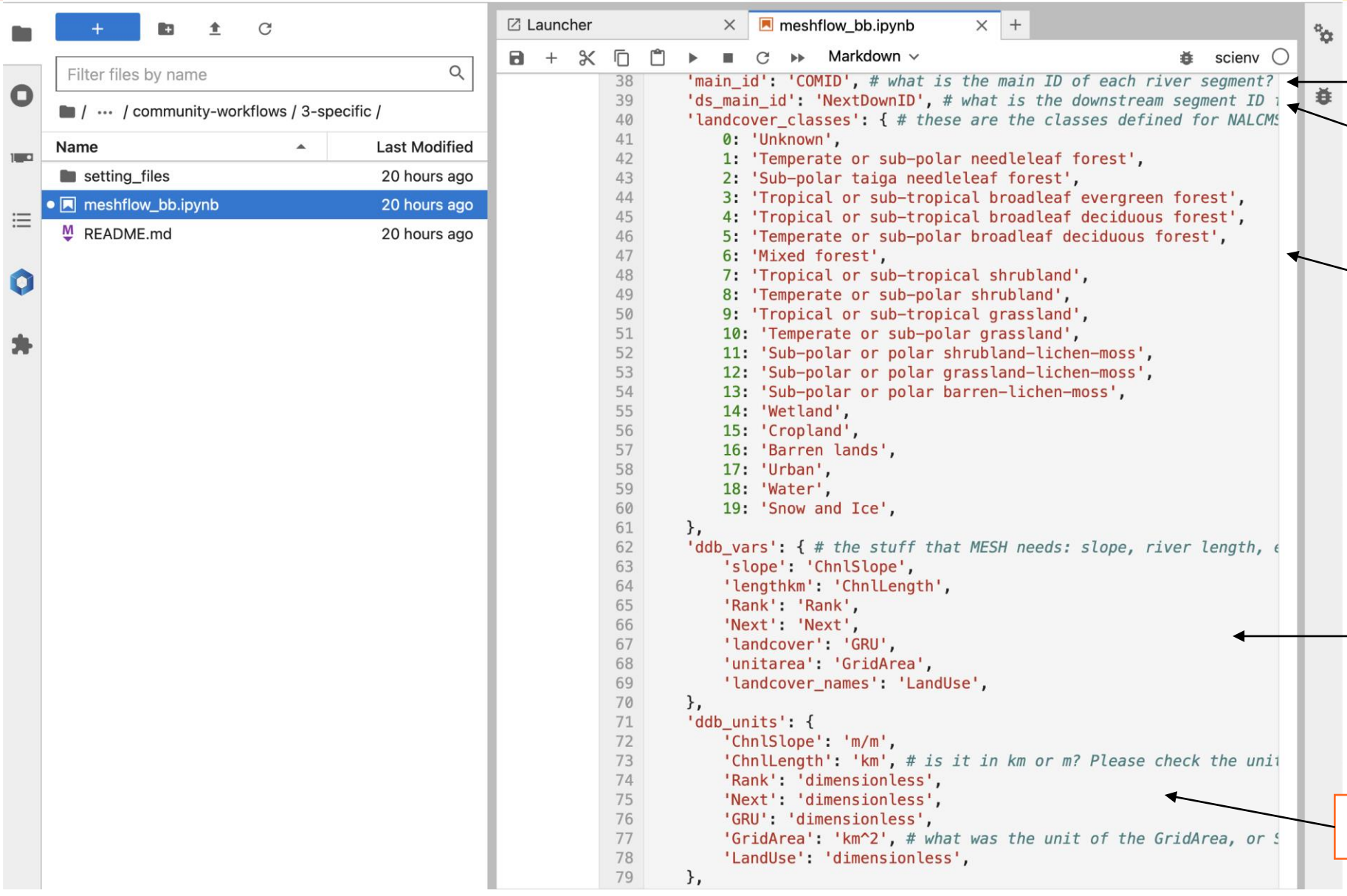
Prepared forcing file(s)

List of variables to be used

Variables' default units in the NetCDF files

Units of variables that MESH requires

Setting up ECCC's National Water Model—MESH



River Segment/Subbasin ID

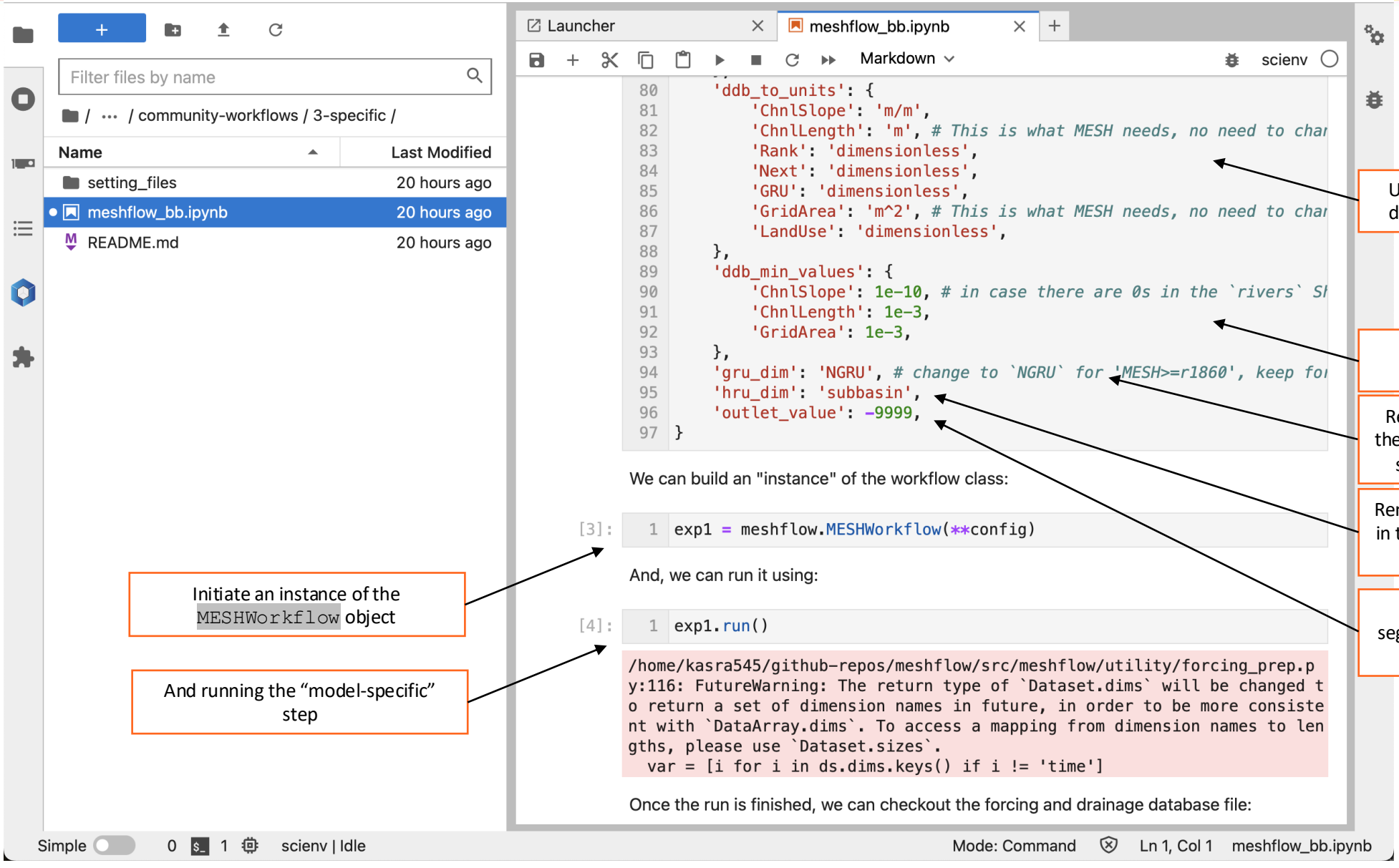
Downstream river segment/subbasin ID

Landcover classes of the landcover dataset used

All the variables to be used in the "drainage database" MESH file

Units of variables to be included in the "drainage database" MESH file

Setting up ECCC's National Water Model—MESH



Units of variables in the "drainage database" file that MESH requires

Minimum values of "drainage database" MESH file

Renaming GRU dimension name in the MESH "drainage database" file to satisfy diverse versions of MESH

Renaming sub-basin dimension name in the MESH "drainage database" file if needed

The value of the outlet river segments created using "`Hydrant`" (`Hydrant` default is `-9999`)

Initiate an instance of the `MESHWorkflow` object

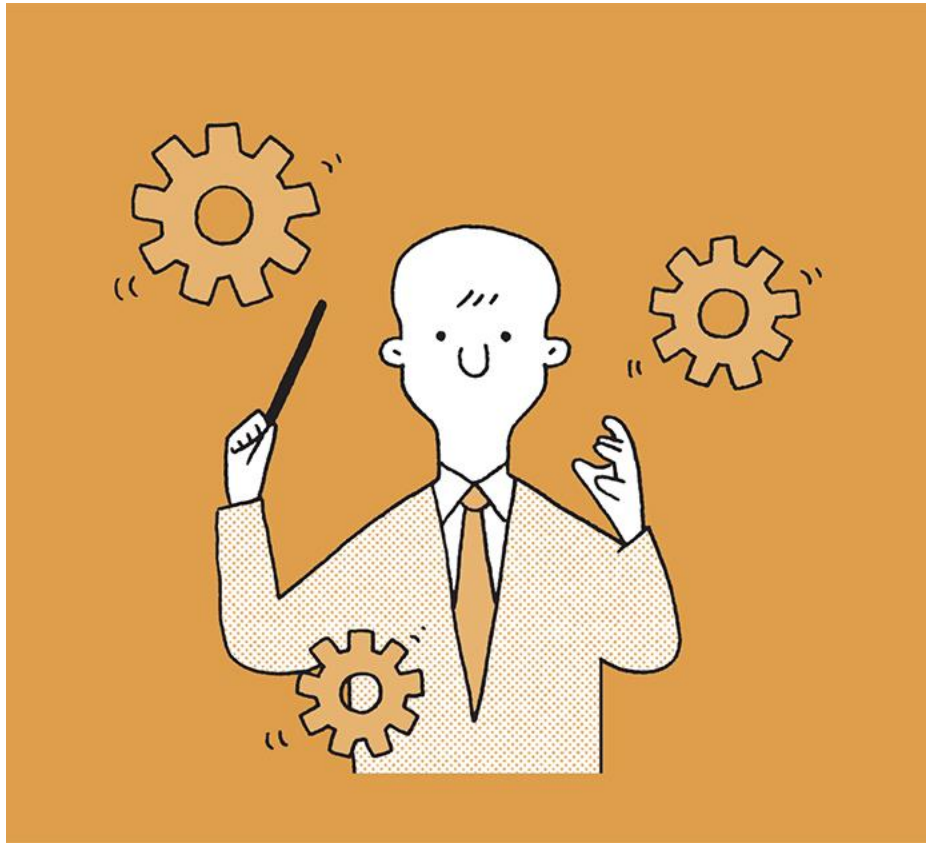
And running the "model-specific" step

UNIVERSITY OF CALGARY

Any questions so far?

Before finishing, let’s look at one last tool.

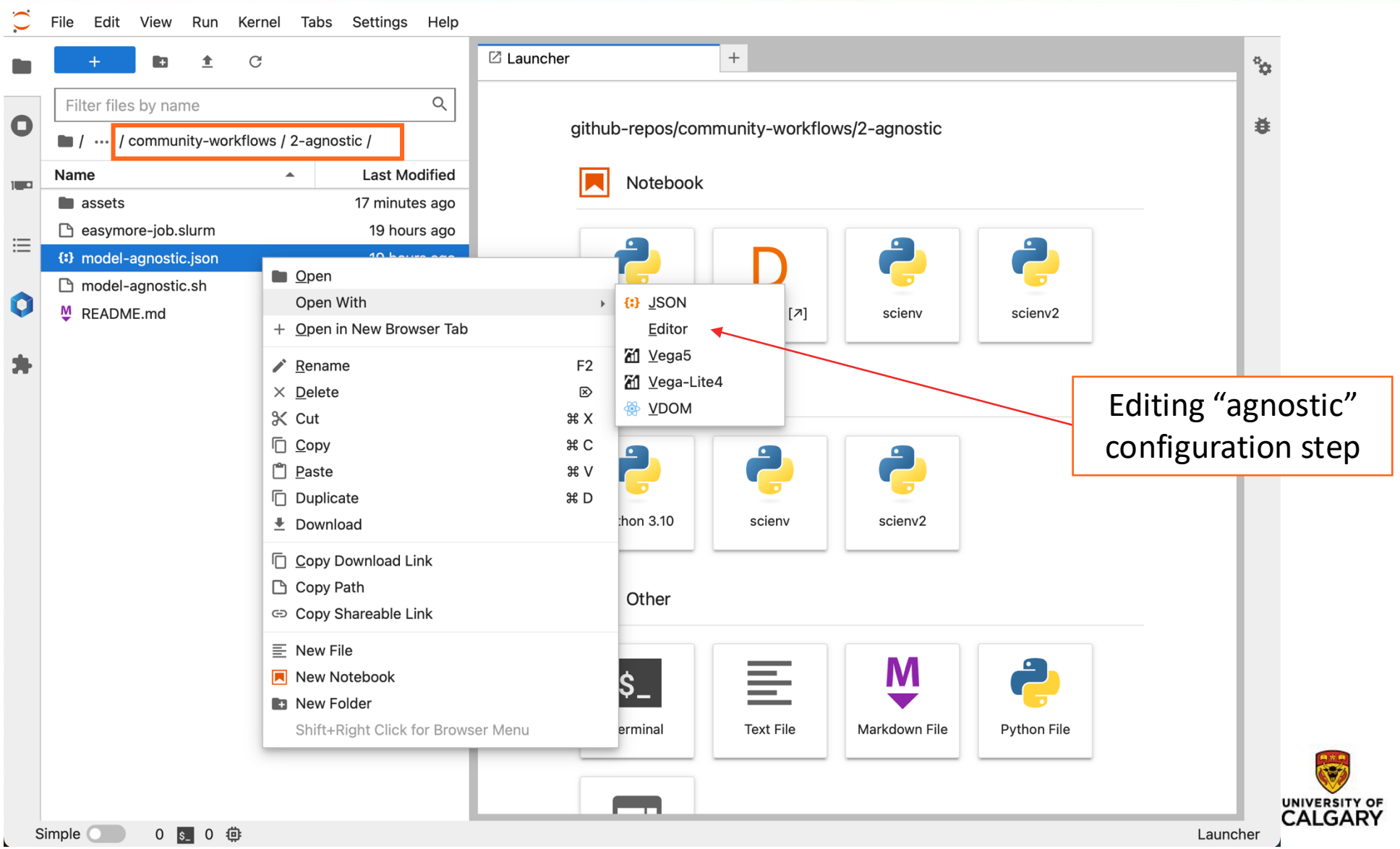
Setting up ECCC's National Water Model—MESH



- Through a tool called "agnostic orchestrator" all the relevant data needed for this setup is processed by the Graham HPC
- We have previously downloaded all the relevant tools for the "agnostic" step
- We use the "agnostic orchestrator" configuration file instruct the file processing hierarchy

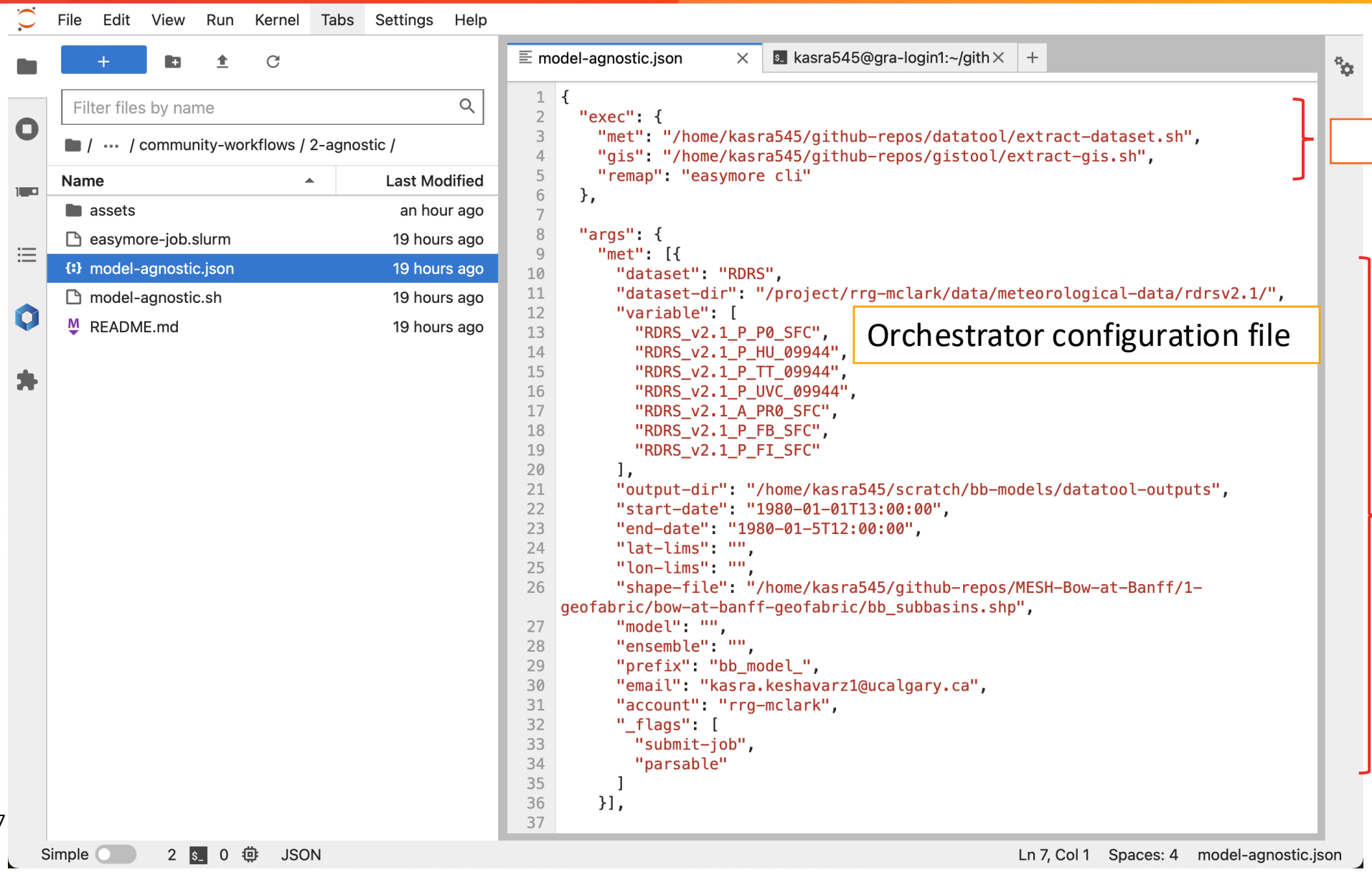
Setting up ECCC's National Water Model—MESH

Agnostic workflow execution using its "Orchestrator"



UNIVERSITY OF CALGARY

Setting up ECCC's National Water Model—MESH



Orchestrator configuration file

Executable paths

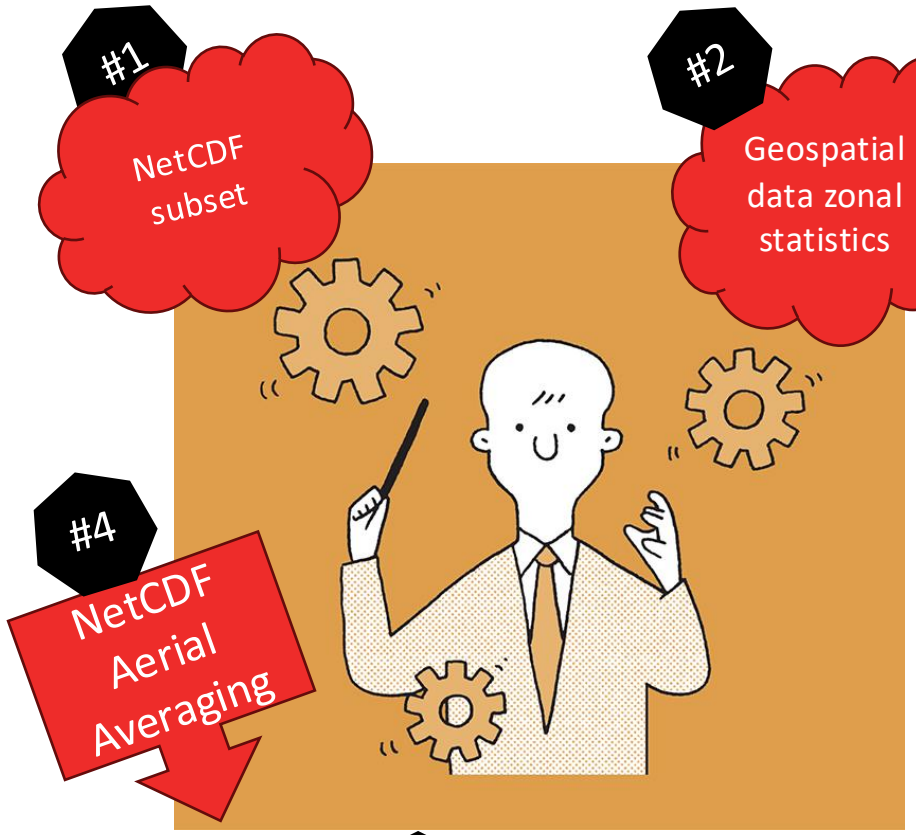
Executable arguments and options

(empty options are ignored)

This is the only file for the "agnostic" step

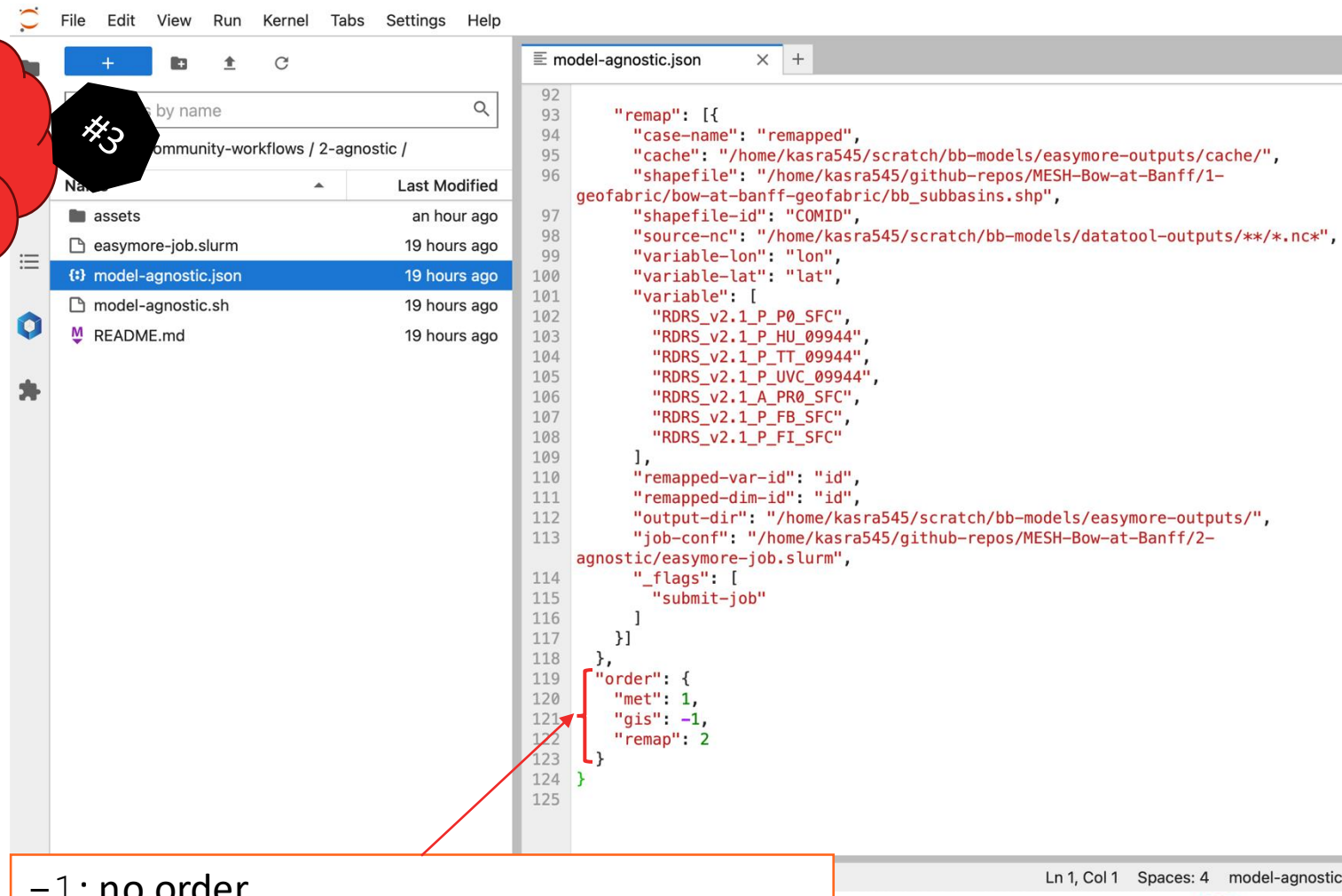
UNIVERSITY OF CALGARY

Setting up ECCC's National Water Model—MESH



: Order of process

Automated and ordered job submission to HPC



`-1`: no order
`ordered integers`: order of processes

UNIVERSITY OF CALGARY

Any questions so far?

Thank you for attending

https://ucalgary.ca/civil

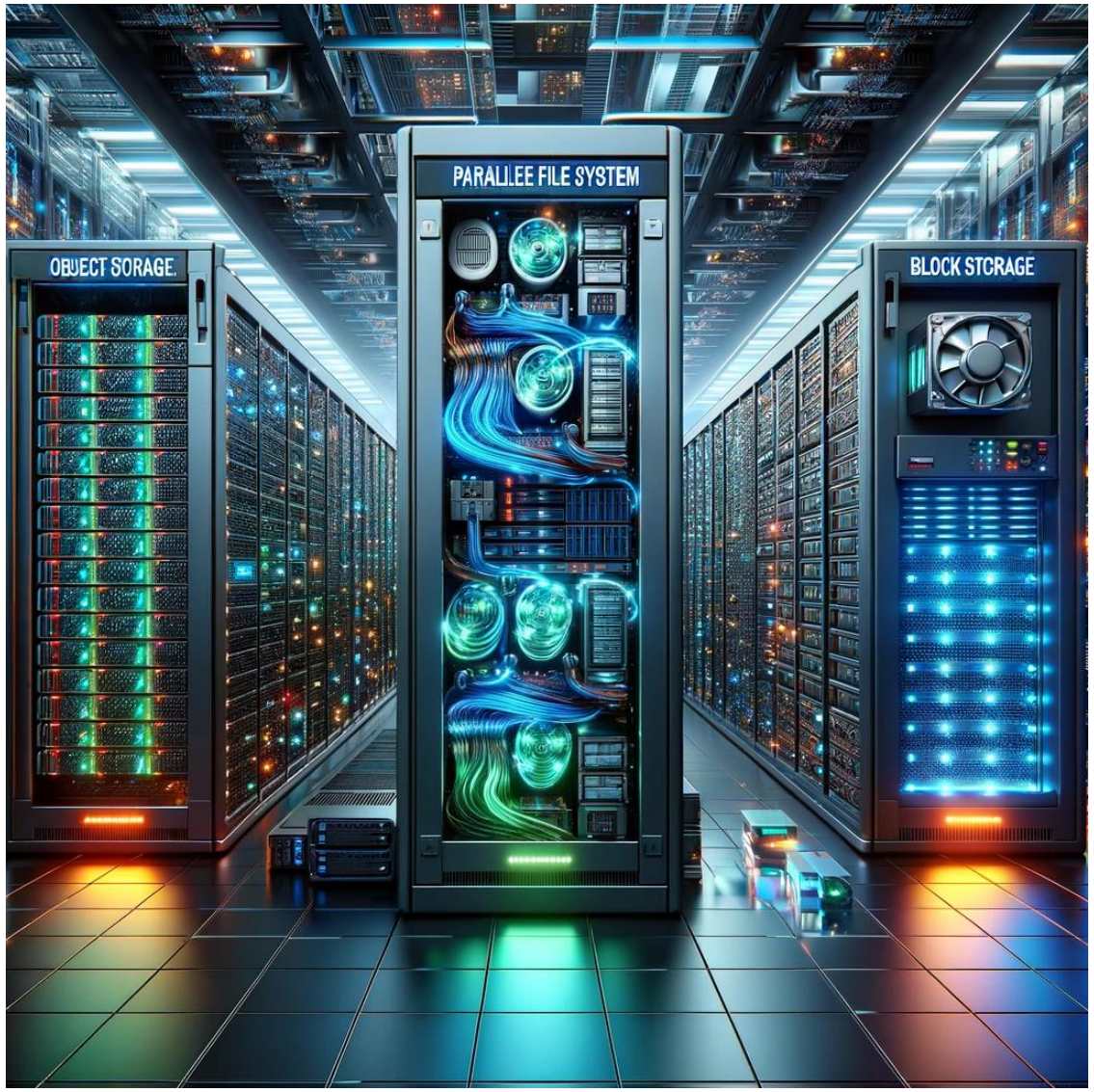
UNIVERSITY OF CALGARY

Backup slides

UNIVERSITY OF CALGARY

Computational Resources

1	Only for personal use	Home space 500 GB per user on ARC
2	Only for data processing	Scratch space 15 TB per experiment
3	NOT for personal use – only for shared resources	Project space comphyd_lab: 0.2 PB



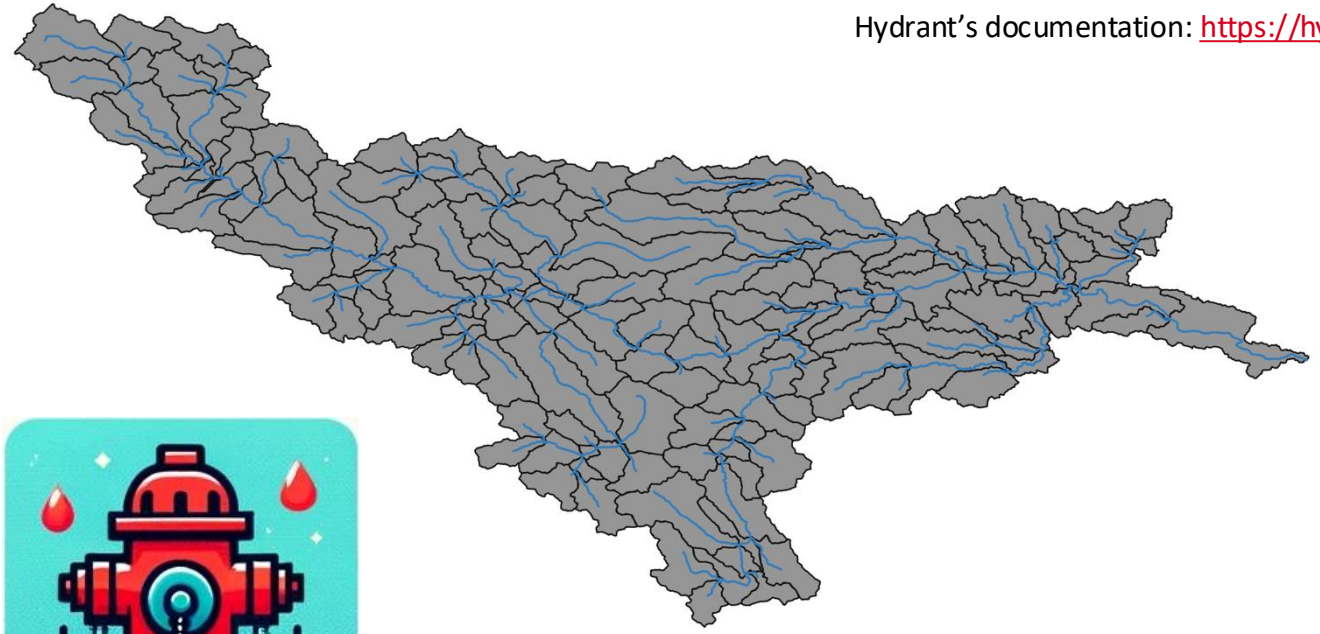
DALL-E 3 generated image

Setting up ECCC's National Water Model—MESH

How does `HydrAnT` help us?

Hydrant's documentation: https://hydrantpy.readthedocs.org

NetworkX
Network Analysis in Python



HYDRANT

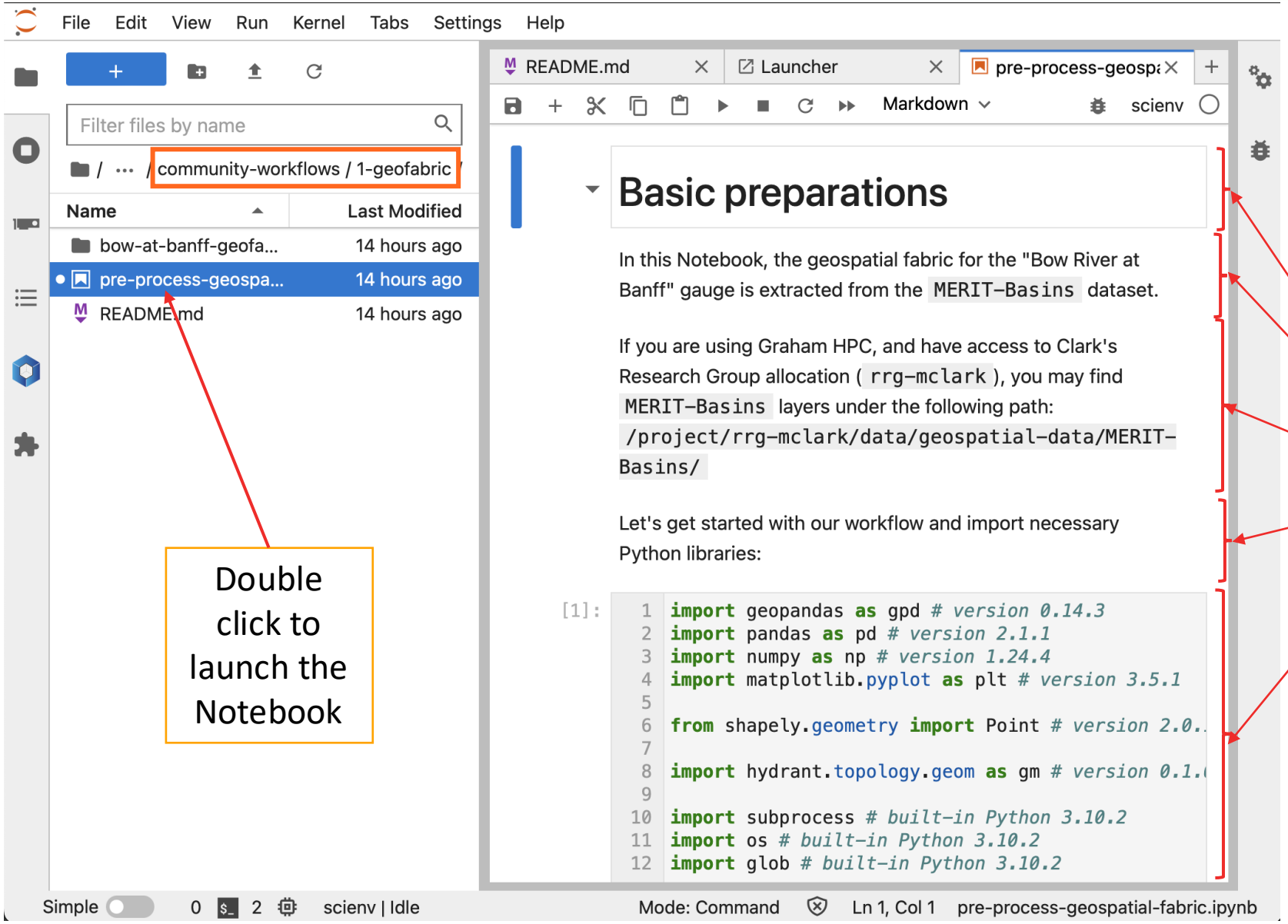
Bow River at Banff

Sub-basins and river network extracted from `MERIT-Basins` using `Hydrant v0.1.0-dev0`

- Extracting subsets of hydrography datasets given a point or a boundary area,
- Hydrant is data-agnostic; if the input hydrography is valid and standard, you can use `Hydrant`,
- Common sanity checks for datasets:
  - Is water going upstream?
  - Is there a cycle of river segments?
  - etc.
- Aggregation methods to dissolve elements of hydrography datasets
- Knowledge extraction:
  - What is the longest branch of river segments?
  - What are the upstream segment of a certain point in the river network?
  - What are the downstream segments of a river segment?
  - What are the Strahler order of river segment?

UNIVERSITY OF CALGARY

Setting up ECCC's National Water Model—MESH



Double click to launch the Notebook

Use "Shift + Enter" to execute Notebook cells

Notebook cells

Tip: review keyboard shortcuts!