Interactive analysis for the ATS sector with SWAN

Rodrigo Sobral, Enric Tejedor
On behalf of the SWAN team

https://cern.ch/swan

December 10th, 2024

Big Data User Forum

CERN

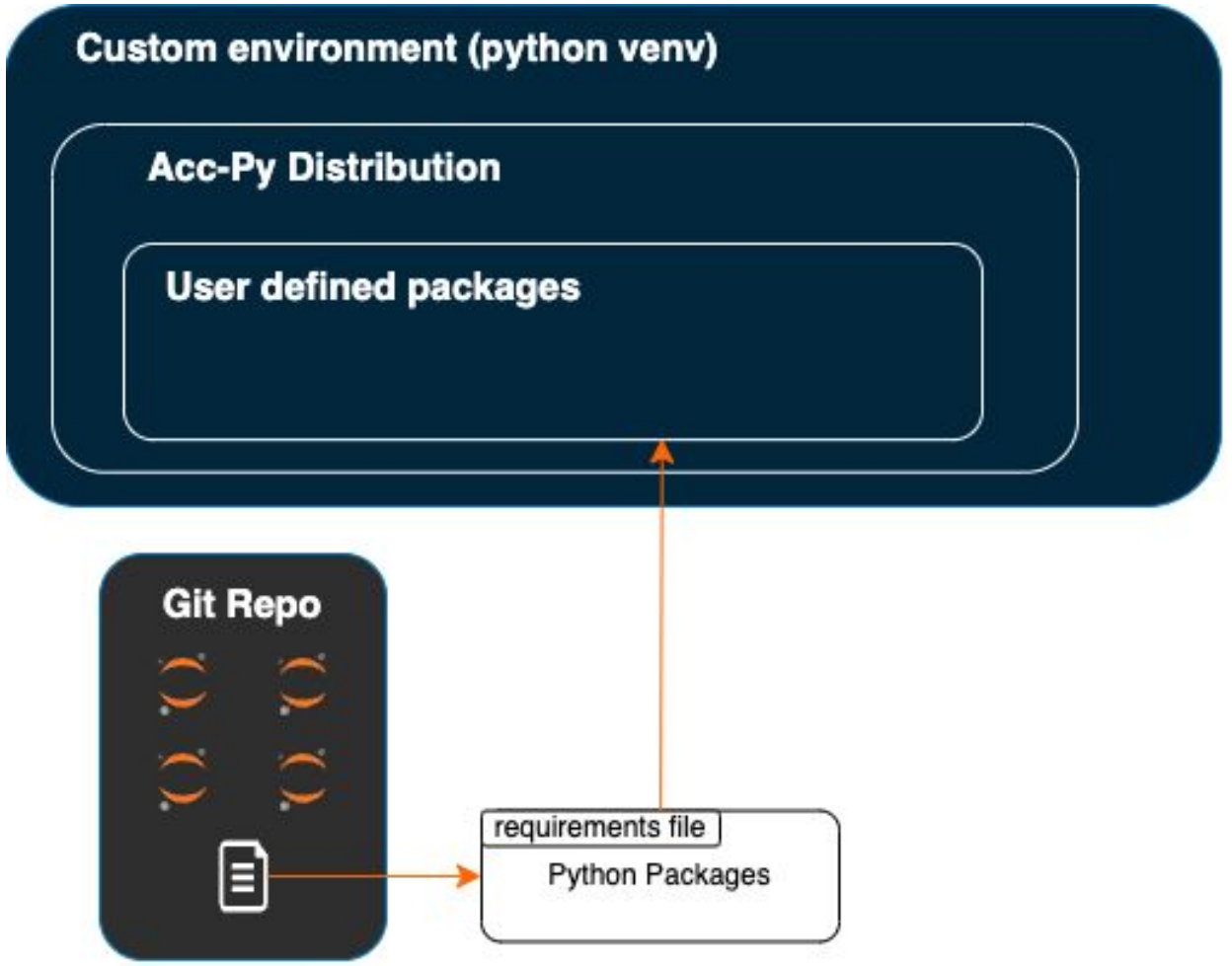
SWAN in a nutshell

- Interactive analysis with a web browser
  - No local installation is needed
  - Based on Jupyter Notebooks
  - Calculations, input data and results “in the Cloud”
- Good for data analysis and exploration, but also for teaching
- Easy sharing of scientific results: plots, data, code
- Added value: **integration with CERN infrastructure and services**!
  - Storage: EOS, CERNBox
  - Software: CVMFS
  - Computing: GPU, Hadoop, HTCondor

The “SWAN on the TN” project

- SWAN on the TN: ATS-IT joint project
- Goal: **deploy a SWAN instance for the ATS sector**
- Main project requirements
  - Access to TN devices, to use their data in the analysis
  - Creation of custom software environments, based on acc-py software
  - Startup of SWAN sessions via URLs with predefined arguments
  - Support of version control for notebooks

Custom software environments

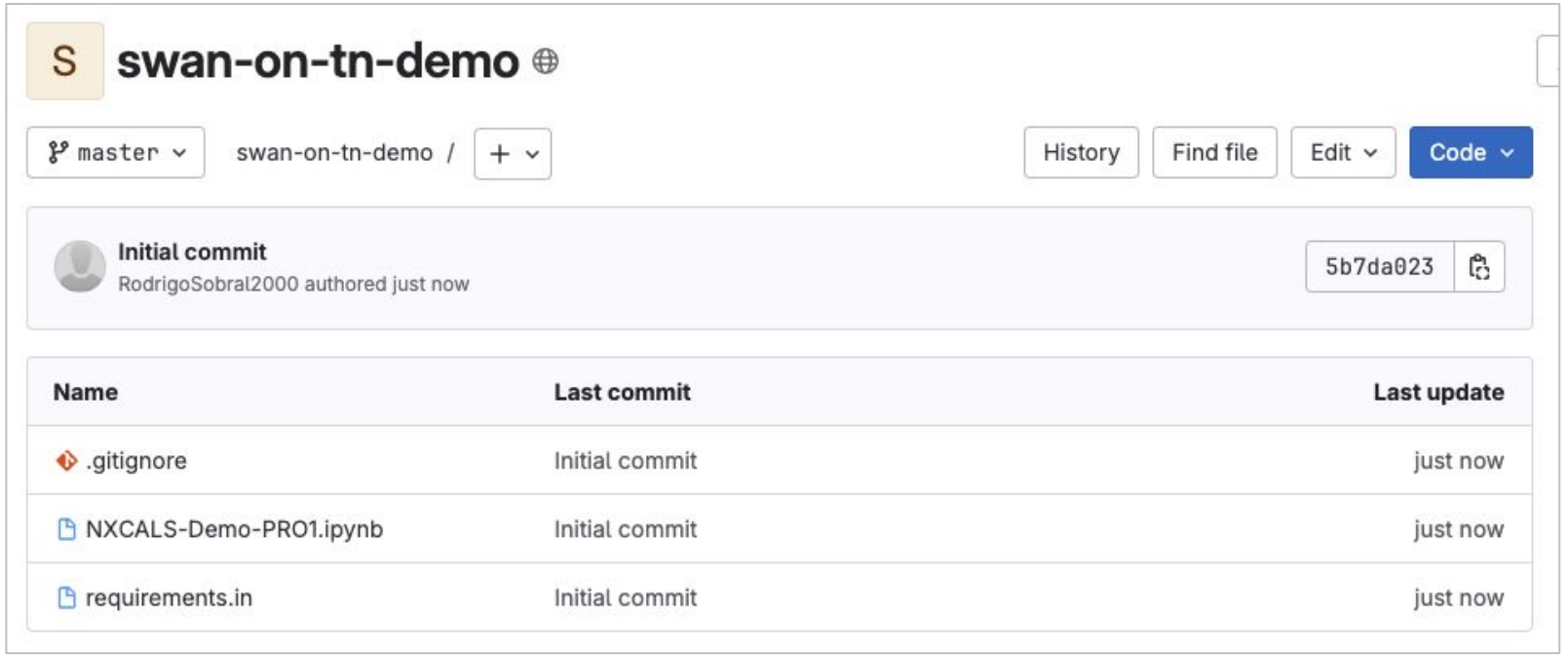


Note: Independent of LCG releases!

Demo

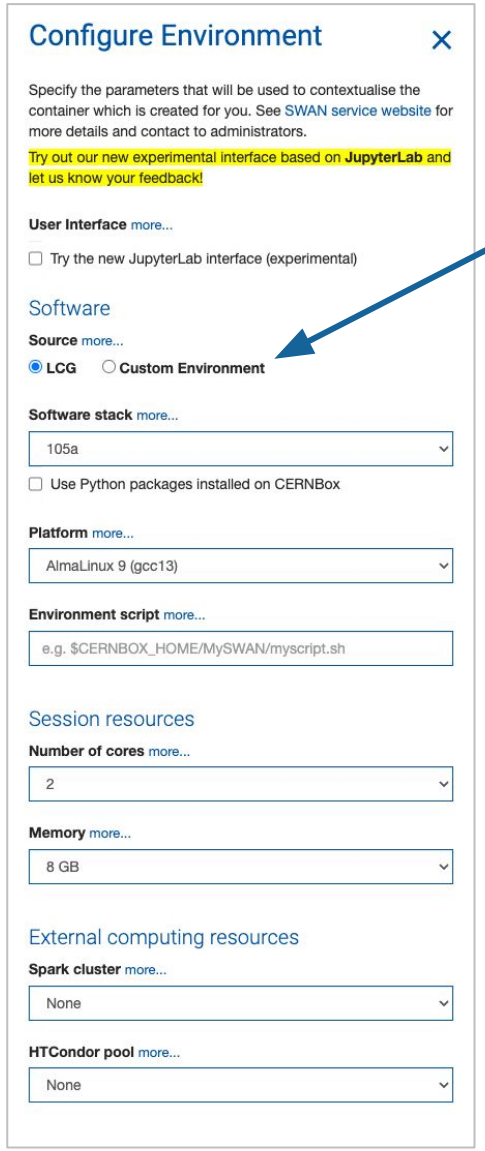
Source of truth: git repository

- Custom environments can be created from **git repositories** that contain:
  - A **requirements file**: describes what packages should be included in the environment
  - Notebooks (and other files) that work with such environment



- matplotlib
- pandas
- pyarrow
- ipympl
- **nxcals**

Selecting a custom environment



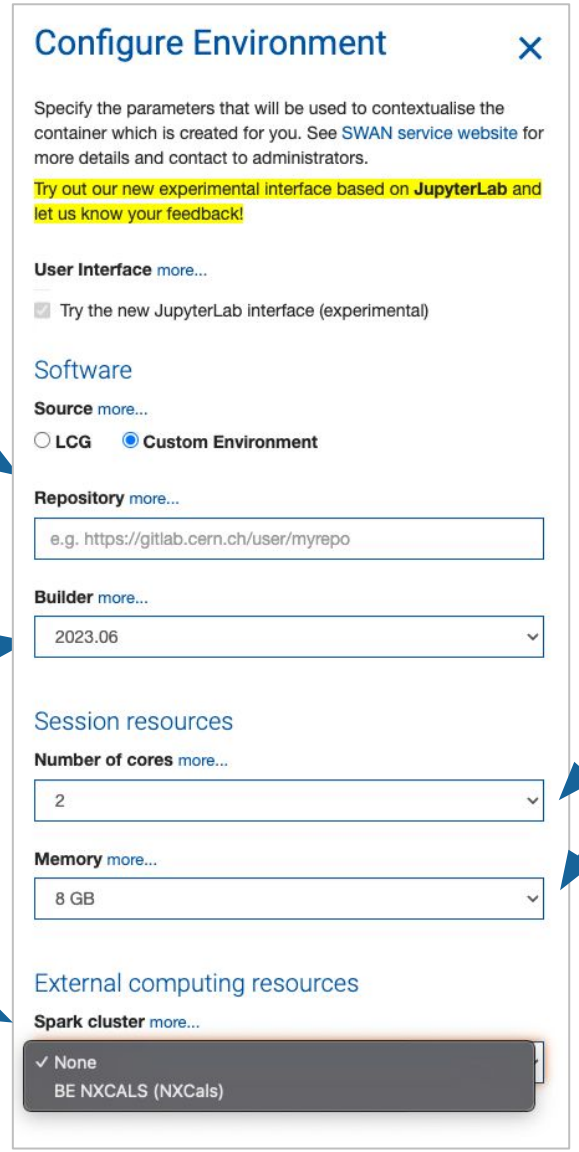
Source: Select Custom Environment to provision software via such environment

Configuring a custom environment

Repository: git repository with a requirements file containing the packages to install

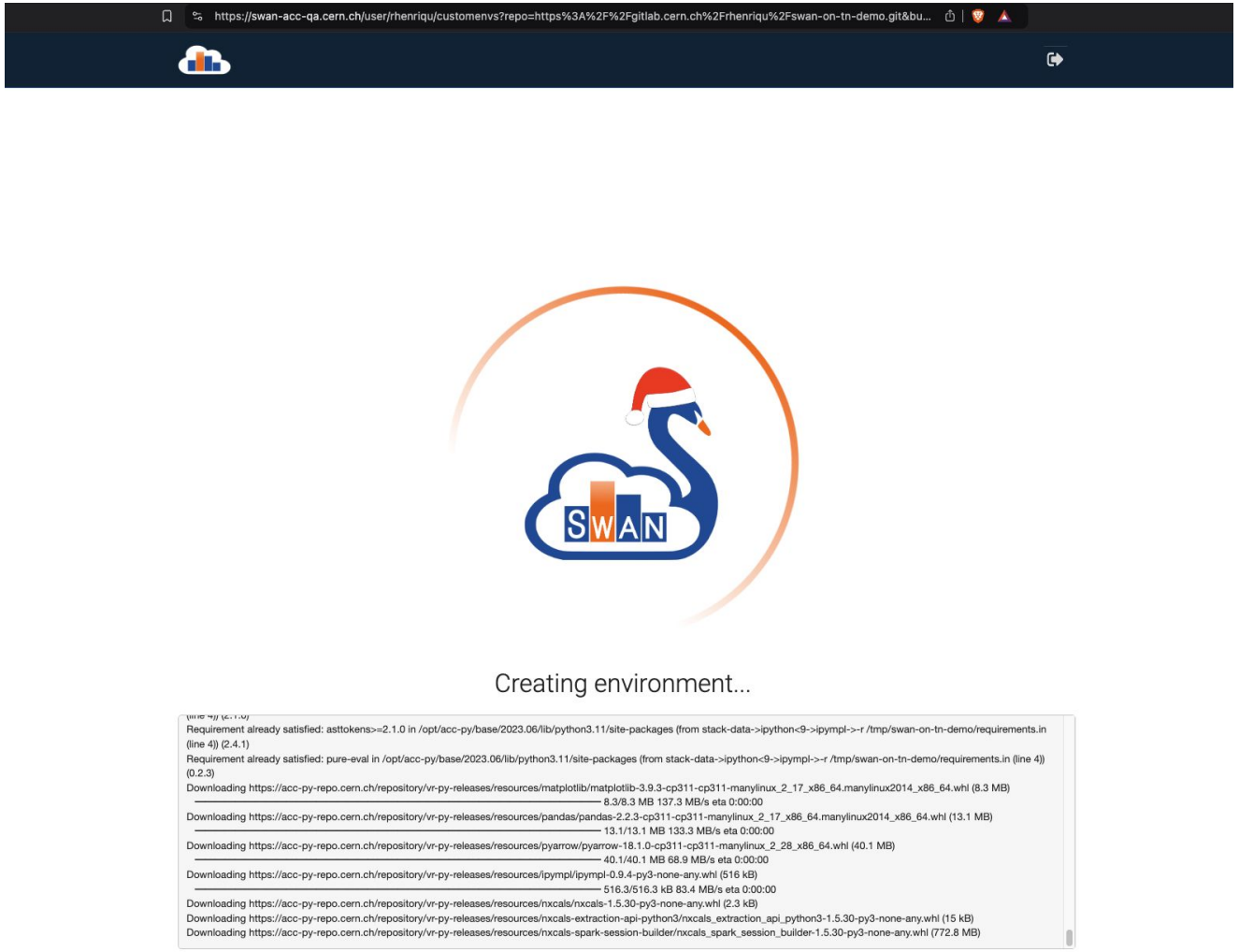
Builder: Choose acc-py distribution to use as base for your environment

Spark Cluster: NXCALS cluster available

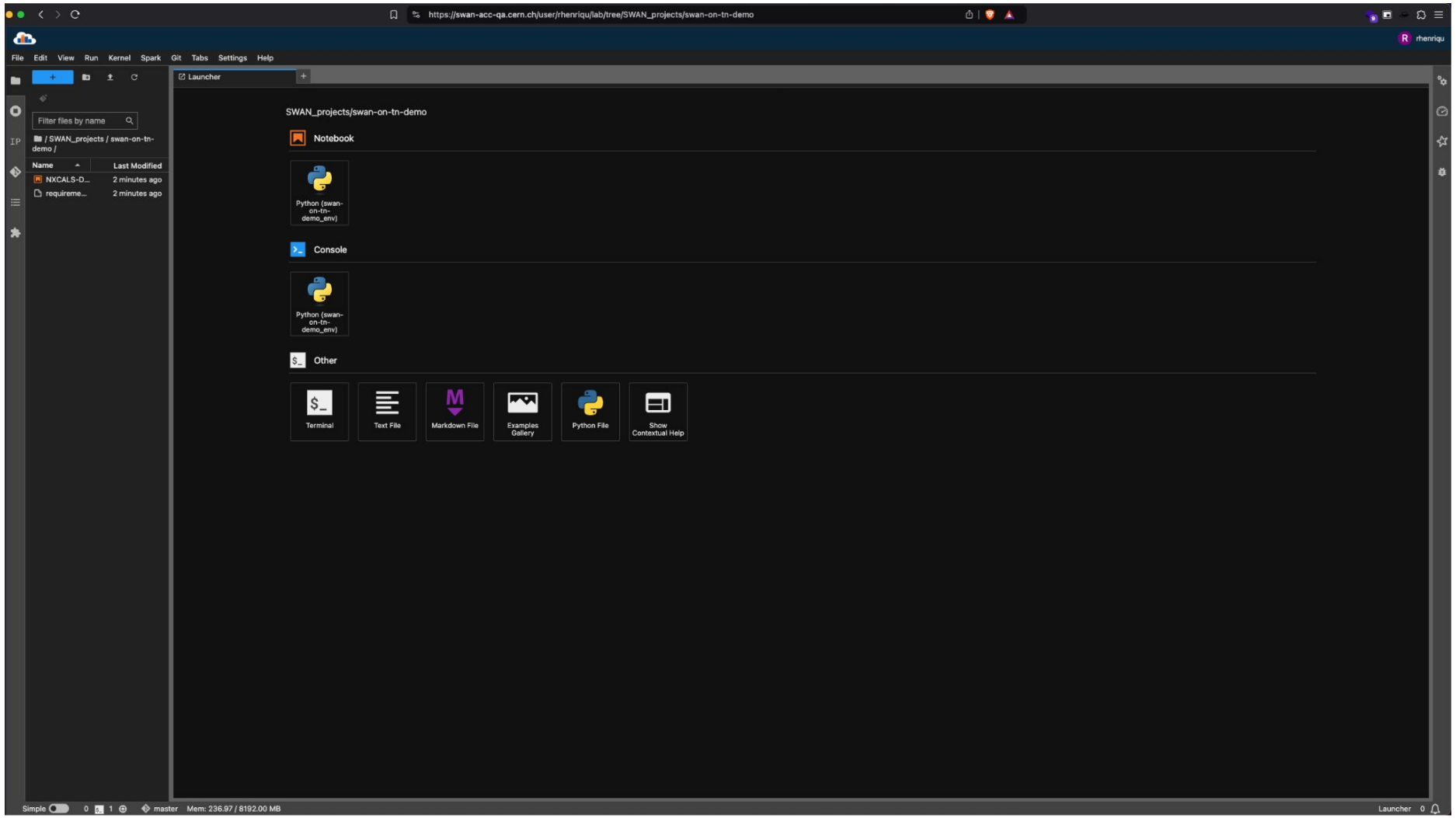


How much **memory**? And **cores**?

Creating the environment



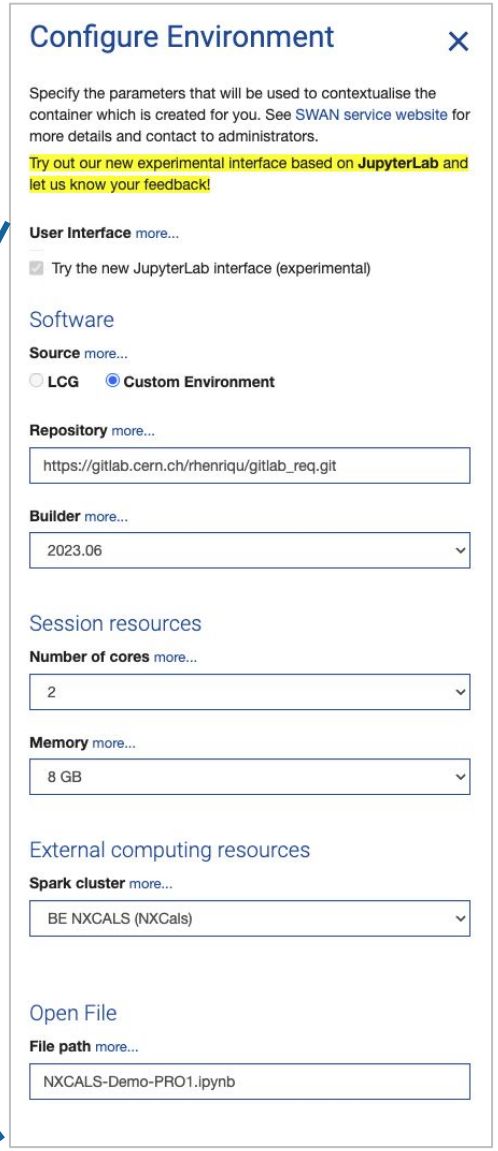
Using the environment in JupyterLab



URLs with default arguments

https://swan-acc-qa.cern.ch/hub/spawn?software_source=customenv&repository=https://gitlab.cern.ch/rhenriqu/swan-on-tn-demo.git&builder=accpy:2023.06&file=NXCALS-Demo-PRO1.ipynb&clusters=hadoop-nxcals

You can build URLs to automatically fill up the form and open a given file



Future work

Future work

- Provide access to NFS servers in the TN
- Deploy a tool for tracking user activity
- ATS SWAN instance to be in production end of June 2025

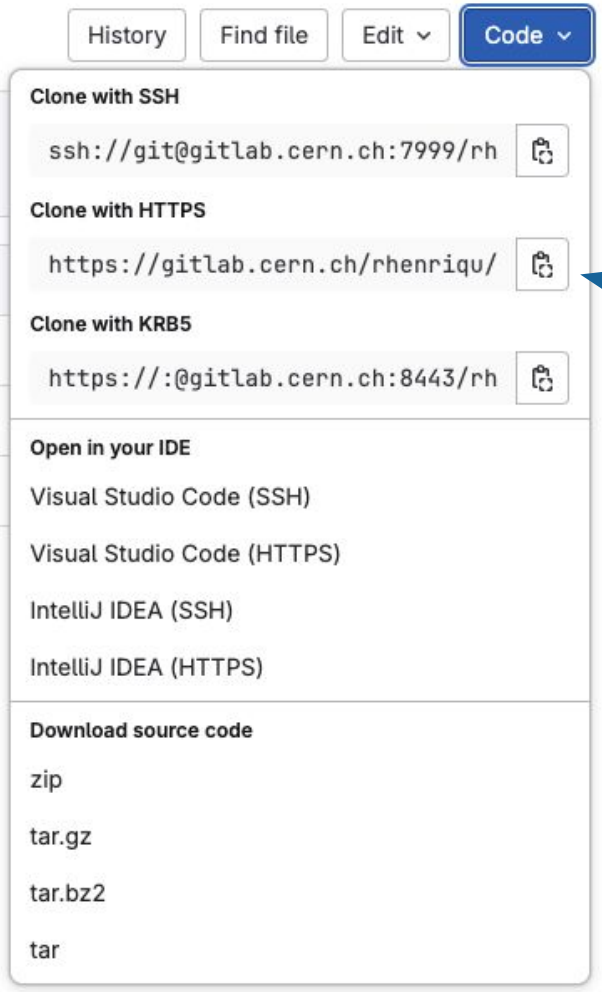
- Expose features implemented in the project to the general SWAN users
  - E.g custom software environments

SWAN-On-TN

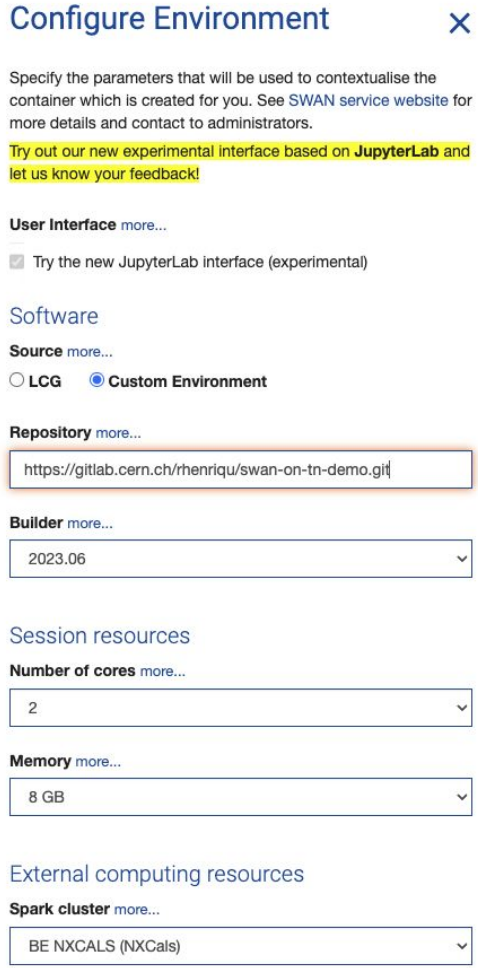
Thank you

Rodrigo Sobral (rodrigo.sobral@cern.ch)
Enric Tejedor (etejedor@cern.ch)

How to get started?



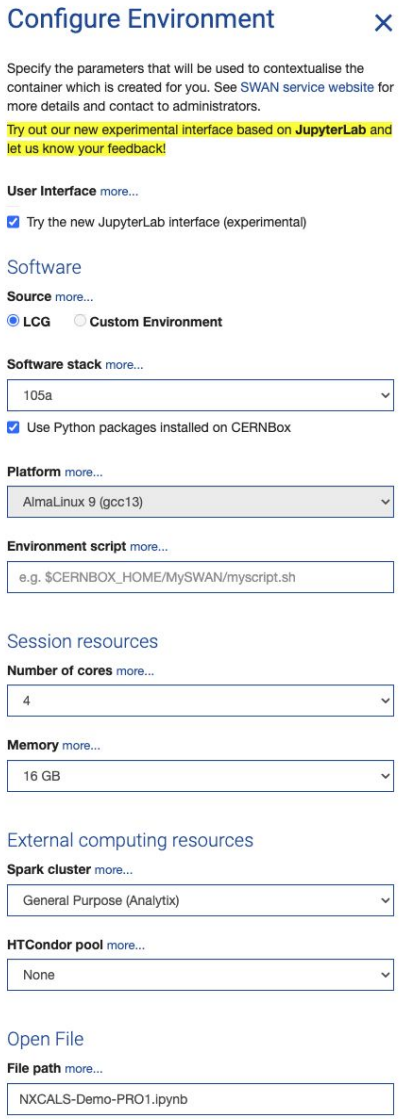
Copy the Git/HTTPS URL to the right repository field and fulfill the remaining fields, according to it



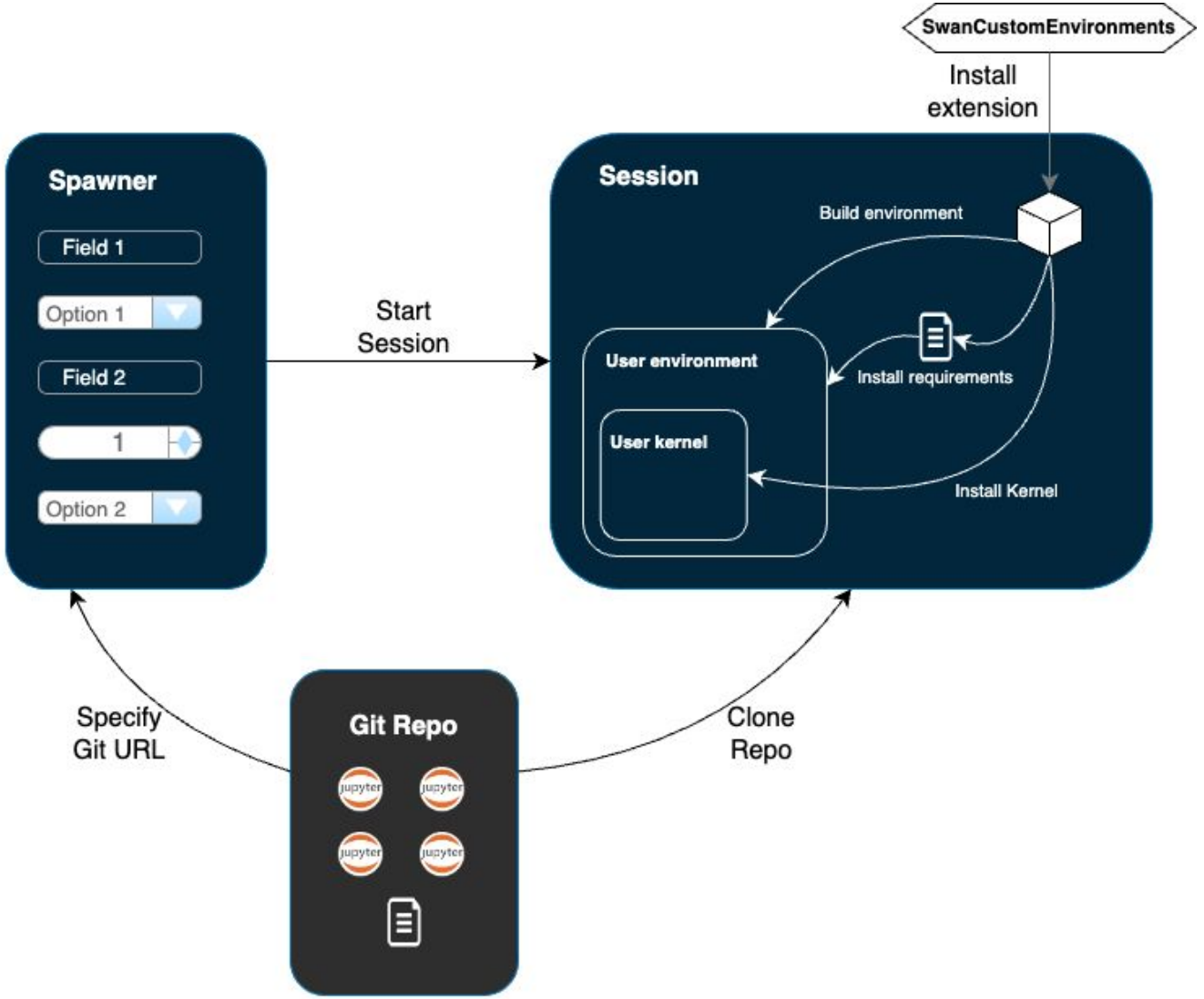
URLs with default arguments (II)

https://swan-acc-qa.cern.ch/hub/spawn?software_source=lcg&memory=16&cores=4&clusters=analytix&use-jupyterlab=true&use-local-packages=true&file=NXCALS-Demo-PRO1.ipynb

Also applicable for LCG sessions



Custom software environments



Custom software environments