

# Interactive analysis for the ATS sector with SWAN



**Rodrigo Sobral**, Enric Tejedor  
On behalf of the SWAN team

<https://cern.ch/swan>

December 10<sup>th</sup>, 2024

Big Data User Forum





# 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

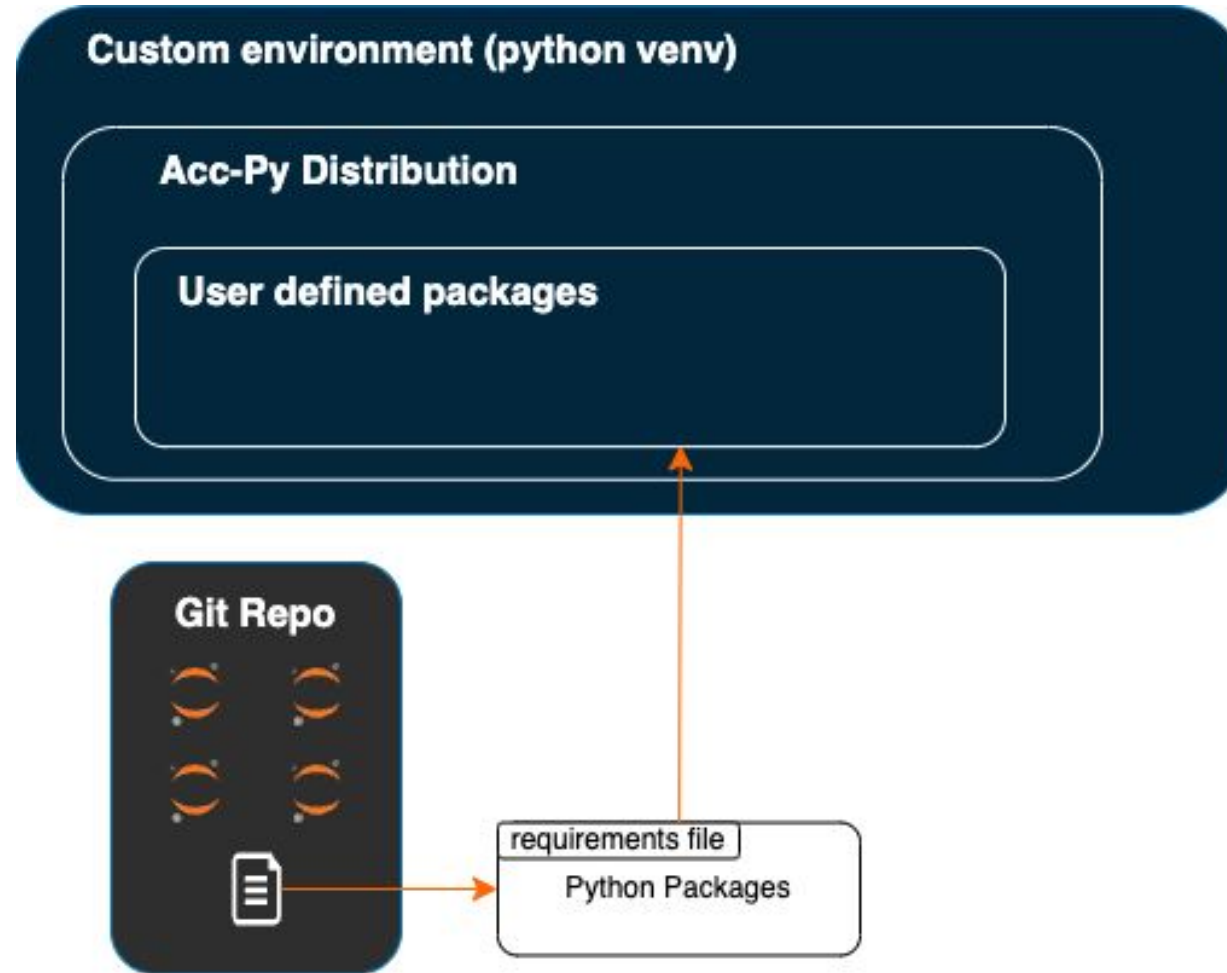


# The “SWAN on the TN” project

- > 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

The screenshot shows a GitHub repository named 'swan-on-tn-demo'. The repository is at the 'master' branch. It shows an 'Initial commit' by 'RodrigoSobral2000' with the commit hash '5b7da023'. Below the commit information is a table of files:

Name	Last commit	Last update
.gitignore	Initial commit	just now
NXCALs-Demo-PRO1.ipynb	Initial commit	just now
requirements.in	Initial commit	just now

- > matplotlib
- > pandas
- > pyarrow
- > ipyml
- > **nxcals**





# Selecting a custom environment

### Configure Environment ✕

Specify the parameters that will be used to contextualise the container which is created for you. See [SWAN service website](#) for more details and contact to administrators.

**Try out our new experimental interface based on JupyterLab and let us know your feedback!**

**User Interface** [more...](#)

Try the new JupyterLab interface (experimental)

**Software**

**Source** [more...](#)

LCG  Custom Environment

**Software stack** [more...](#)

105a ▼

Use Python packages installed on CERNBox

**Platform** [more...](#)

AlmaLinux 9 (gcc13) ▼

**Environment script** [more...](#)

e.g. `$(CERNBOX_HOME)/MySWAN/myscript.sh`

**Session resources**

**Number of cores** [more...](#)

2 ▼

**Memory** [more...](#)

8 GB ▼

**External computing resources**

**Spark cluster** [more...](#)

None ▼

**HTCondor pool** [more...](#)

None ▼

**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

### Configure Environment

Specify the parameters that will be used to contextualise the container which is created for you. See [SWAN service website](#) for more details and contact to administrators.

Try out our new experimental interface based on **JupyterLab** and let us know your feedback!

**User Interface** [more...](#)

Try the new JupyterLab interface (experimental)

**Software**

**Source** [more...](#)

LCG  Custom Environment

**Repository** [more...](#)

**Builder** [more...](#)

**Session resources**

**Number of cores** [more...](#)

**Memory** [more...](#)

**External computing resources**

**Spark cluster** [more...](#)

None  
 BE NXCALS (NXCals)

Acc-Py

- ✓ 2023.06
- 2021.12
- 2020.11

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







# Creating the environment



Creating environment...

```
wire ~/k...  
Requirement already satisfied: asttokens>=2.1.0 in /opt/acc-py/base/2023.06/lib/python3.11/site-packages (from stack-data->ipython-c9->ipyml->-r /tmp/swan-on-tn-demo/requirements.in  
(line 4)) (2.4.1)  
Requirement already satisfied: pure-eval in /opt/acc-py/base/2023.06/lib/python3.11/site-packages (from stack-data->ipython-c9->ipyml->-r /tmp/swan-on-tn-demo/requirements.in (line 4))  
(0.2.3)  
Downloading https://acc-py-repo.cern.ch/repository/vr-py-releases/resources/matplotlib/matplotlib-3.9.3-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (8.3 MB)  
----- 8.3/8.3 MB 137.3 MB/s eta 0:00:00  
Downloading https://acc-py-repo.cern.ch/repository/vr-py-releases/resources/pandas/pandas-2.2.3-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (13.1 MB)  
----- 13.1/13.1 MB 133.3 MB/s eta 0:00:00  
Downloading https://acc-py-repo.cern.ch/repository/vr-py-releases/resources/pyarrow/pyarrow-18.1.0-cp311-cp311-manylinux_2_28_x86_64.whl (40.1 MB)  
----- 40.1/40.1 MB 68.9 MB/s eta 0:00:00  
Downloading https://acc-py-repo.cern.ch/repository/vr-py-releases/resources/ipyml/ipyml-0.9.4-py3-none-any.whl (516 kB)  
----- 516.3/516.3 kB 83.4 MB/s eta 0:00:00  
Downloading https://acc-py-repo.cern.ch/repository/vr-py-releases/resources/nxcals/nxcals-1.5.30-py3-none-any.whl (2.3 kB)  
----- 2.3/2.3 kB 100.0 MB/s eta 0:00:00  
Downloading https://acc-py-repo.cern.ch/repository/vr-py-releases/resources/nxcals-extraction-api-python3/nxcals_extraction_api_python3-1.5.30-py3-none-any.whl (15 kB)  
----- 15.0/15.0 kB 100.0 MB/s eta 0:00:00  
Downloading https://acc-py-repo.cern.ch/repository/vr-py-releases/resources/nxcals-spark-session-builder/nxcals_spark_session_builder-1.5.30-py3-none-any.whl (772.8 MB)  
----- 772.8/772.8 MB 100.0 MB/s eta 0:00:00
```





# Using the environment in JupyterLab

The screenshot displays the JupyterLab web interface. At the top, the browser address bar shows the URL: `https://swan-acc-qa.cern.ch/user/henriqu/lab/tree/SWAN_projects/swan-on-tn-demo`. The interface features a top menu bar with options: File, Edit, View, Run, Kernel, Spark, Git, Tabs, Settings, and Help. On the left, a sidebar contains a file browser with a search bar labeled "Filter files by name" and a table listing files in the `/SWAN_projects / swan-on-tn-demo /` directory:

Name	Last Modified
NXCALS-D...	2 minutes ago
requireme...	2 minutes ago

The main workspace area is titled "SWAN\_projects/swan-on-tn-demo" and contains a "Launcher" view. This view offers several options for starting a new environment or tool:

- Notebook:** A section with a Python icon and a button labeled "Python (swan-on-tn-demo\_env)".
- Console:** A section with a terminal icon and a button labeled "Python (swan-on-tn-demo\_env)".
- Other:** A section containing icons for "Terminal", "Text File", "Markdown File", "Examples Gallery", "Python File", and "Show Contextual Help".

At the bottom of the interface, a status bar shows "Simple" mode, a memory usage indicator "Mem: 236.97 / 8192.00 MB", and a "Launcher" tab indicator.





# URLs with default arguments

[https://swan-acc-ga.cern.ch/hub/spawn?software\\_source=customenv&repository=https://gitlab.cern.ch/rhenriqu/swan-on-tn-demo.git&builder=accpy:2023.06&file=NXCAL S-Demo-PRO1.ipynb&clusters=hadoop-nxcals](https://swan-acc-ga.cern.ch/hub/spawn?software_source=customenv&repository=https://gitlab.cern.ch/rhenriqu/swan-on-tn-demo.git&builder=accpy:2023.06&file=NXCAL S-Demo-PRO1.ipynb&clusters=hadoop-nxcals)

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

### Configure Environment ×

Specify the parameters that will be used to contextualise the container which is created for you. See [SWAN service website](#) for more details and contact to administrators.

**Try out our new experimental interface based on JupyterLab and let us know your feedback!**

**User Interface** more...

Try the new JupyterLab interface (experimental)

**Software**

**Source** more...

LCG  Custom Environment

**Repository** more...

**Builder** more...

**Session resources**

**Number of cores** more...

**Memory** more...

**External computing resources**

**Spark cluster** more...

**Open File**

**File path** more...

# 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](mailto:rodrigo.sobral@cern.ch))

Enric Tejedor ([etejedor@cern.ch](mailto:etejedor@cern.ch))



# How to get started?

History Find file Edit Code

**Clone with SSH**  
ssh://git@gitlab.cern.ch:7999/rh

**Clone with HTTPS**  
https://gitlab.cern.ch/rhenriqu/

**Clone with KRB5**  
https://:@gitlab.cern.ch:8443/rh

**Open in your IDE**

- Visual Studio Code (SSH)
- Visual Studio Code (HTTPS)
- IntelliJ IDEA (SSH)
- IntelliJ IDEA (HTTPS)

**Download source code**

- zip
- tar.gz
- tar.bz2
- tar

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

## Configure Environment

Specify the parameters that will be used to contextualise the container which is created for you. See [SWAN service website](#) for more details and contact to administrators.

[Try out our new experimental interface based on JupyterLab and let us know your feedback!](#)

User Interface [more...](#)

Try the new JupyterLab interface (experimental)

Software

Source [more...](#)

LCG  Custom Environment

Repository [more...](#)

https://gitlab.cern.ch/rhenriqu/swan-on-tn-demo.git

Builder [more...](#)

2023.06

Session resources

Number of cores [more...](#)

2

Memory [more...](#)

8 GB

External computing resources

Spark cluster [more...](#)

BE NXCALLS (NXCALLS)



# 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=NXCALSDemo-PRO1.ipynb](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=NXCALSDemo-PRO1.ipynb)

Also applicable for LCG sessions

## Configure Environment ×

Specify the parameters that will be used to contextualise the container which is created for you. See SWAN service website for more details and contact to administrators.

Try out our new experimental interface based on JupyterLab and let us know your feedback!

### User Interface more...

Try the new JupyterLab interface (experimental)

### Software

#### Source more...

LCG  Custom Environment

#### Software stack more...

105a

Use Python packages installed on CERNBox

#### Platform more...

AlmaLinux 9 (gcc13)

#### Environment script more...

e.g. \$CERNBOX\_HOME/MySWAN/myscript.sh

### Session resources

#### Number of cores more...

4

#### Memory more...

16 GB

### External computing resources

#### Spark cluster more...

General Purpose (Analytix)

#### HTCondor pool more...

None

### Open File

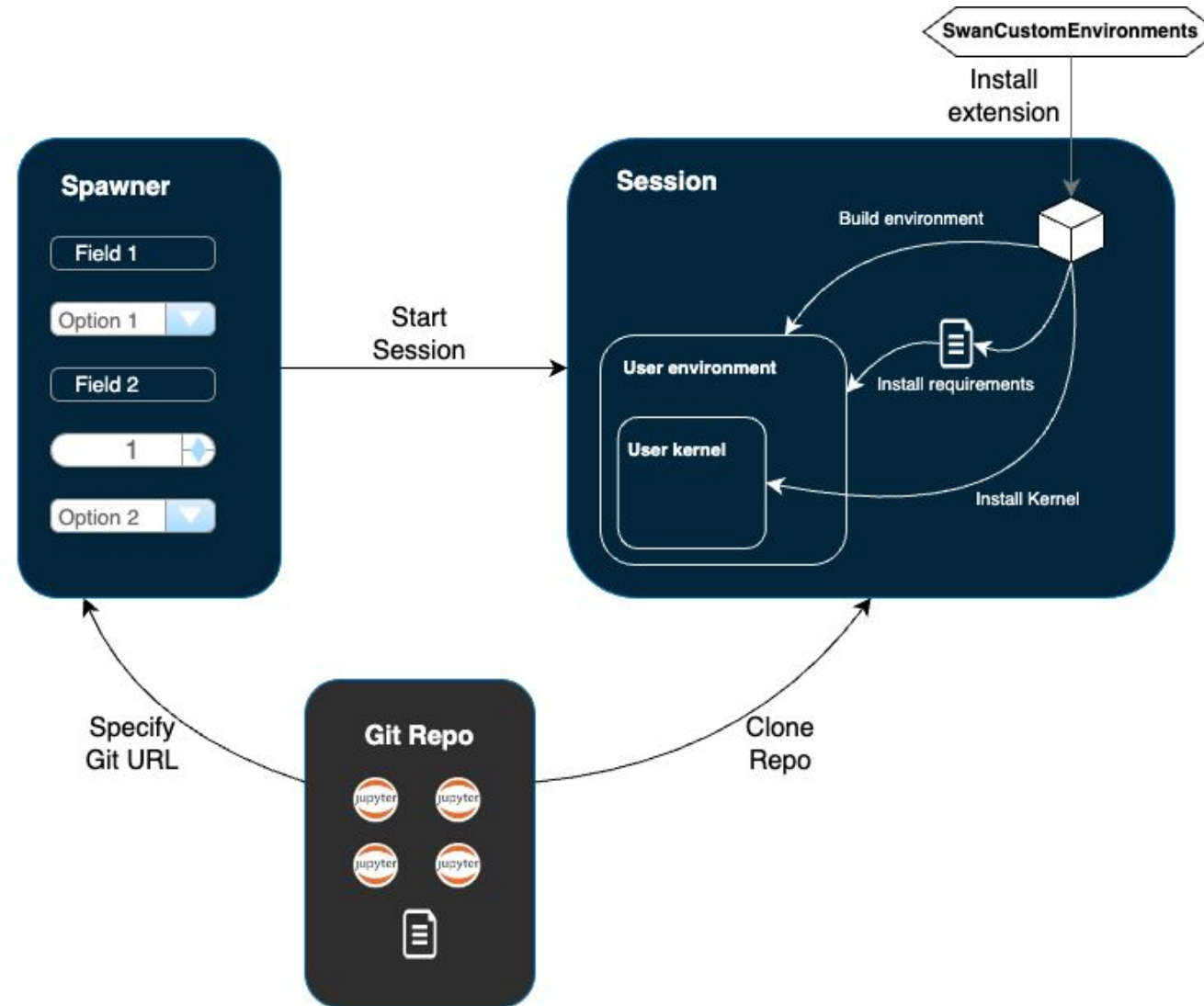
#### File path more...

NXCALS-Demo-PRO1.ipynb





# Custom software environments





# Custom software environments

