SWAN: service for web based analysis



CERN

D. Castro, E. Tejedor, D. Piparo, P. Mato E. Bocchi, J. Moscicki, M. Lamanna

https://swan.web.cern.ch

Mar 13th, 2018 Jupyter notebooks at CERN – EPFL meeting







> Analysis only with a web browser

- Available everywhere and at anytime
- Integrated with other analysis ecosystems : ROOT, R, Python, ...

> Easy to use (but powerful)

- No local installation and configuration needed
- > Create easily sharable scientific results: plots, data, code
- Integration with CERN resources
 - Access software, user/experiments data, mass processing power













Integrating Jupyter

> 4 kernels

- Python (2 or 3), ROOT C++, R and Octave
- Configurable environments through user defined scripts
- Jupyterhub to allow multiple Jupyter instances
 - Single instance of Jupyter per user
 - Proxy the container to the outside
- User sessions spawned as Docker containers
 - Enforces resource limits per user
 - Sessions culled after 6h of inactivity







> Uses EOS mass storage system

- All experiment data potentially available
- Installed on host, mounted on container
- User personal space, synchronized through CERNBox
 - All files synced across devices, the cloud and other users
 - Based on OwnCloud







Software distributed through CVMFS

- "LCG Releases" pack a series of compatible packages
- Software used by researchers is available
- Strong caching mechanisms
- Possibility to install other libraries in user storage (CERNBox)









CERN



> Containerized version of all the infrastructure

- Includes EOS, CERNBox, CVMFS and all Swan services (Jupyter Docker image, JupyterHub)
- Available in <u>https://github.com/cernbox/uboxed</u>
- > Easily deployable on premises
 - Installable in Linux systems
 - Based on Docker Compose

Recent developments





Configure Environment

Specify the parameters that will be used to contextualise the container which is created for you. See the online SWAN guide for more details.

÷

Software stack more...

91

2

Platform more
x86_64-slc6-gcc62-opt

Environment script more...

e.g. \$CERNBOX_HOME/MySWAN/myscript.sh

Number of cores more...

Memory	more

8 GB			
0.00			

Spark cluster more...

Hadalytic

Always start with this configuration

Start my Session



Waiting for swan-ga004.cern.ch...



Starting your session

)

CERN

•



a	Projects	Share	CERNBox		>_ ••• 🕞
SWAN > My Projects					
My Projects					(+)
□ NAME .				STATUS	MODIFIED
🗑 Proj1				4	5 days ago
Proj2					15 days ago
Project					21 days ago
Project 1					2 months ago
Project 2					4 months ago
ProjTest					15 days ago
Spark					7 days ago
SWAN-Spark_NXCALS_Example					20 days ago
🗑 teste					19 days ago





Sharing made easy

- Sharing from inside
 SWAN interface
 - Integration with CERNBox API
- Users can share "Projects"
 - Special kind of folder that contains notebooks and other files, like input data

&	Projects	CERI	Share Project	×
SWAN > Share			You are sharing: Proj1	
Projects shared with me	^		You can share with people or groups. Your contacts will be able to see your project, including all the fil inside it, and clone it. You can prefix the search by "a:" to also look into secondary and service accounts, or "g:" to only search for unix groups.	es.
ProjTest		5	Start typing to add names	
			Shared with	
Projects shared by me			deiparo	
NAME				
🗑 Proj1				
SWAN © Copyright CERN 2017. All rights reserved. Home Contacts Support Report a bug Imprint			Stop Sharing Update	

• Sharing made easy

- Users can clone a shared Project directly from the interface
 - Jupyter doesn't allow concurrent editing

-					
•••	Projects	< Share	CERNBox		>_ ••• 🗭
SWAN > Share					
Projects shared wi	ith me 🔿				
NAME			SIZE	SHARED BY	DATE
ProjTest			5.64 MB	diocas	25 days ago
Projects shared by	/ me \land				
NAME				SHARED WITH	DATE
🗑 Proj1				2 people/groups	5 days ago

SWAN Copyright CERN 2017. All rights reserved. Home | Contacts | Support | Report a bug | Imprint



Integration with Spark

- Connection to CERN
 Spark Clusters
- User data accessed through EOS
- Graphical Jupyter extensions developed
 - Spark Connector
 - Spark Monitor



Integration with Spark

×



Spark > Spark_Simple Spark clusters connection FILE EDIT VIEW INSERT CELL KERNEL HELP 🖹 🕂 🛠 🖓 🖪 🛧 🔸 射 🔳 C' Markdown 💠 📼 📰 Trying to connect to Spark Clusters. This may take a while Registering MapOutputTracker Registering BlockManagerMaster Using org.apache.spark.storage.DefaultTopologyMapp er for getting topology information Simple example with Spark BlockManagerMasterEndpoint up Created local directory at /tmp/blockmgr-e3b5b0d9-82ec-4e7d-a190-b76cf7c87015 This notebook illustrates the use of Spark in SWAN. MemoryStore started with capacity 912.3 MB The current setup allows to execute PySpark operations on a local small datasets. Registering OutputCommitCoordinator In the future, SWAN users will be able to attach external Spark clu Successfully started service 'SparkUI' on port 901 Moreover, a Scala Jupyter kernel will be added to use Spark from 7. Bound SparkUI to 172.17.0.9, and started at htt Import the necessary modules p://swan006.cern.ch:9017 The pyspark module is available to perform the necessary impor In []: from pyspark import SparkContext Cancel







In [5]: sc.parallelize(range(0,10)).count() sc.parallelize(range(0,20)).count() Apache Spark: 1 EXECUTORS 4 CORES Jobs: 2 COMPLETED 12 N 11. Job ID Job Name Stages Tasks Submission Time Status Duration 1/1 3 count a few seconds ago 0s 4 1/1 a few seconds ago 0s count Out[5]: 20

×









SWAN and the community



CERN user community

- SWAN development is guided by our user community
 - New features (libs, kernels, ...) are requested by users from their real usage needs
- Gallery of examples
 - Made in collaboration with our users
 - Almost 50 notebooks in 7 categories

Example notebooks at swan.web.cern.ch SWAN Interactive Data Analysis, in the Cloud. Home Gallerier FAQ Talks and Publications Baix ROT Primer Accelerator Complex Machine Learning Apache Spark

Basic Examples

This is a gallery of basic example notebooks: click on the images to inspect the underlying document, open in SWAN the single notebooks or the full git repository!

pen in 🛃 SWAN

Many of the notebooks are ROOTbooks, based on the ROOT framework. To know more about ROOT, visit root.cern.ch.



Other collaborations

Building block in UP2University European Project

- Bridge the gap between secondary schools, higher education, and the research domain
- SWAN used by students to learn Physics and other Natural Sciences
- Let the kids use the very same tools & services used by real researchers doing Big Science at CERN
- Using Boxed distribution

Future research





> Apache Guacamole

- Docker container available as remote desktop in a browser
- Access to already configured software environment
- More complex use cases not possible with Jupyter
 - i.e. access to ROOT framework

> Jupyterlab

- More complex interface
- Access to different editors
- Concurrent editing











- > SWAN is a CERN service that provides Jupyter Notebooks on demand
- SWAN promotes a cloud based analysis model where users can do analysis only with their browser
- SWAN federates CERN services for software, storage and infrastructure so that users can find what they need in the service
- > SWAN fosters collaboration and results sharing between scientists
- SWAN is an Interface for Mass Processing Resources (Spark)



SWAN: service for web based analysis

Thank you

