

SWAN: interactive data analysis on the web



Prasanth Kothuri

On behalf of the SWAN team

<https://cern.ch/swan>

Oct 11th, 2019

SWAN Users' Workshop



Introduction

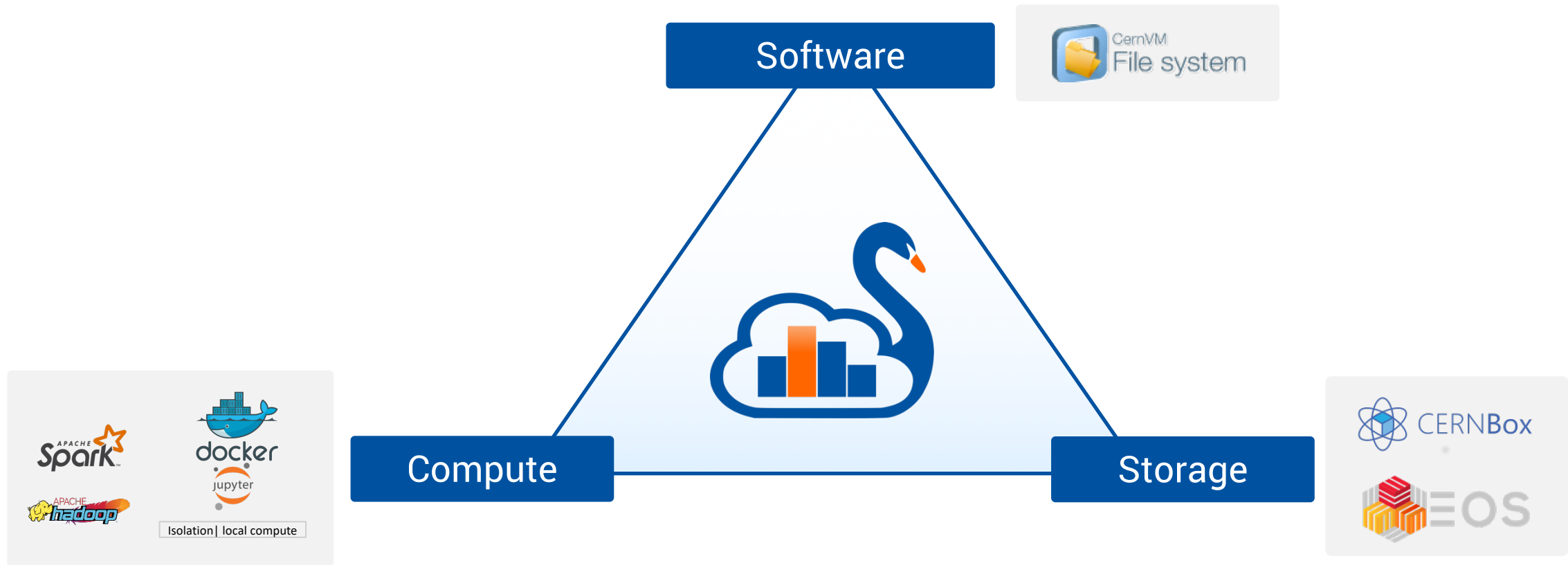


SWAN in a Nutshell

- > Analysis only with a web browser
 - No local installation needed
 - Based on Jupyter Notebooks
 - Calculations, input data and results “in the Cloud”
- > Support for multiple analysis ecosystems and languages
 - Python, ROOT C++, R and Octave
- > Easy sharing of scientific results: plots, data, code
- > Integration with CERN resources
 - software, storage, mass processing power



Integrating services





Jupyter - The Notebook as Interface

- > A web-based interactive interface and platform that combines code, equations, text and visualizations
 - Ideal for sharing/collaboration
 - A “shell opened within the browser”
- > Interactive, usually lightweight computations
 - And distributed parallel processing capability with the integration of mass processing system (Apache Spark)
- > Very useful for some use cases at CERN
 - Final steps of an Analysis, Exploration, Teaching, Documentation and Reproducibility





User Interface



Home Token Admin

SWAN © Copyright CERN 2016-2019. All rights reserved.
Home | Community | Support | Report a bug

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...](#)

Platform [more...](#)

Environment script [more...](#)


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

Memory [more...](#)

Spark cluster [more...](#)

Always start with this configuration











Start my Session




Projects Share CERNBox

SWAN > My Projects

My Projects +

<input type="checkbox"/> NAME ▲	STATUS	MODIFIED
 Proj1		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

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



Text

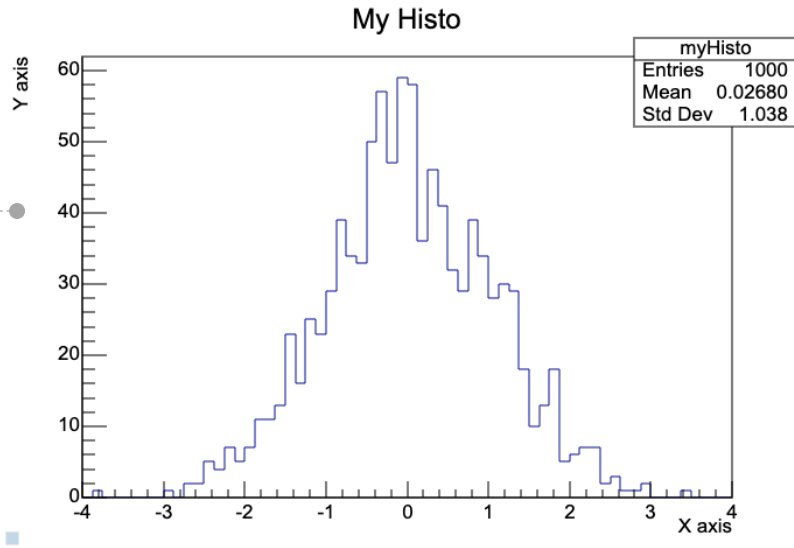
Code

Graphics

2 Displaying graphics

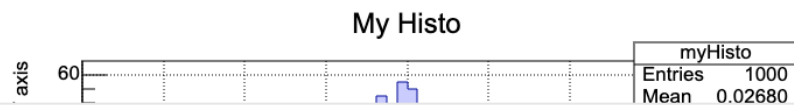
We can now draw the histogram. We will at first create a [canvas](#), the entity which in ROOT holds graphics primitives. Note that thanks to [JSROOT](#), this is not a static plot but an interactive visualisation. Try to play with it and save it as image when you are satisfied!

```
In [5]: c = ROOT.TCanvas()
        h.Draw()
        c.Draw()
```



We'll try now to beautify the plot a bit, for example filling the histogram with a colour and setting a grid on the canvas.

```
In [6]: h.SetFillColor(ROOT.kBlue-10)
        c.SetGrid()
        h.Draw()
        c.Draw()
```



FILE EDIT VIEW INSERT CELL KERNEL HELP Trusted Python 2

Do the heavylifting in spark and collect aggregated view to panda DF

```
In [11]: df_loadAvg_pandas = spark.sql("SELECT submitter_host, \
    avg(body.LoadAvg) as avg, \
    hour(from_unixtime(timestamp / 1000, 'yyyy-MM-dd HH:mm:ss')) as hr \
    FROM loadAvg \
    WHERE submitter_hostgroup = 'hadoop/itdb/datanode' \
    AND dayofmonth(from_unixtime(timestamp / 1000, 'yyyy-MM-dd HH:mm:ss')) = 15 \
    GROUP BY hour(from_unixtime(timestamp / 1000, 'yyyy-MM-dd HH:mm:ss'), submitter_host")\
    .toPandas()
```

Apache Spark: 90 EXECUTORS 180 CORES Jobs: 1 COMPLETED

Job ID	Job Name	Status	Stages	Tasks	Submission Time	Duration
3	toPandas	COMPLETED	2/2	388 / 388	4 minutes ago	36s

Visualize with seaborn

```
In [19]: # heatmap of service availability
plt.figure(figsize=(10, 6))
ax = sns.heatmap(df_loadAvg_pandas.pivot(index='submitter_host', columns='hr', values='avg'), cmap="Blues")
ax.set_title("Heatmap of loadAvg")
```

Out[19]: Text(0.5,1,u'Heatmap of loadAvg')

Text

Code

Monitoring

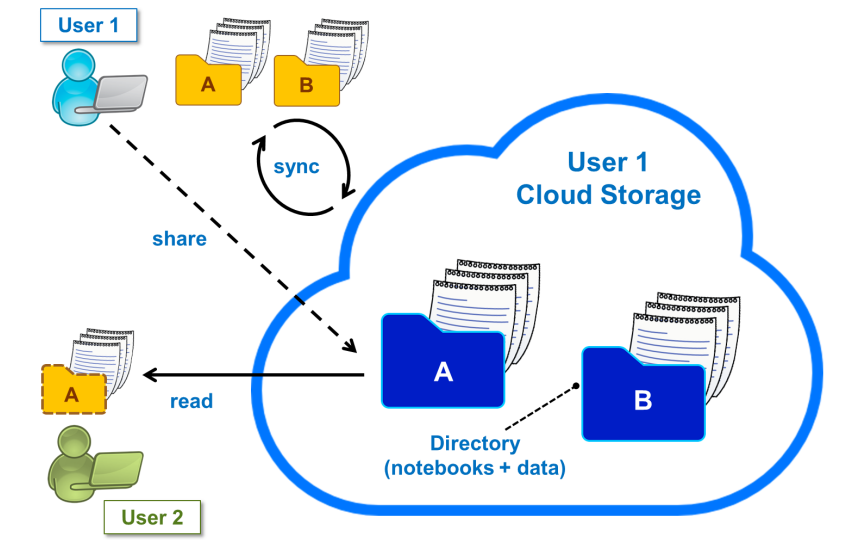
Visualizations





Cloud storage as your Home

- > CERNBox is SWAN's home directory
 - Storage for your notebooks and data
 - 16k users and 6PB of user data
- > Uses EOS disk storage system
 - All experiment data potentially available
 - 250PB of experimental data at CERN (LHC and others)
- > Sync & Share
 - Files synced across devices and the Cloud
 - Collaborative analysis





Sharing made easy

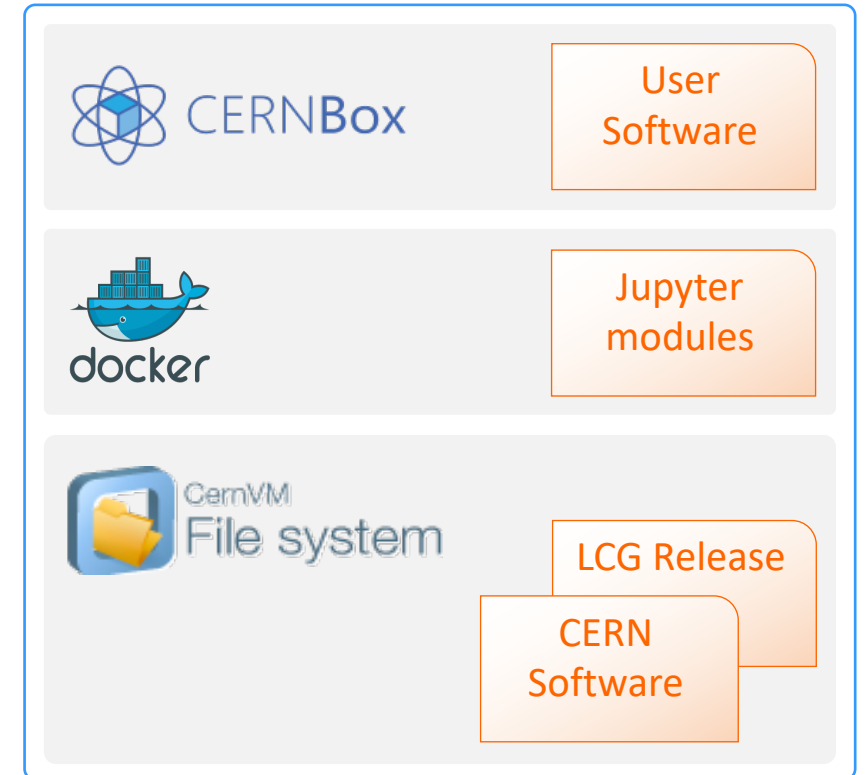
- > Sharing from inside SWAN interface
 - Integration with CERNBox
 - List shares from other users
- > Users can share “Projects”
 - Special kind of folder that contains notebooks and other files, like input data
 - Self contained

The screenshot displays the SWAN interface with a 'Share Project' modal open. The main view shows a breadcrumb trail: SWAN > My Projects > Super Real Analysis with TOTEM data. Below this, the project title 'Super Real Analysis with TOTEM data' is shown with an upward arrow. A list of files follows: 'DistillDistribution.ipynb' (with a notebook icon) and 'dataset.root' (with a folder icon). The 'Share Project' dialog on the right has a close button (X) in the top right. It contains the text 'You are sharing: Super Real Analysis with TOTEM data'. Below this is a search instruction: 'Search by name or username. Use "a:" for secondary accounts.' and an input field with the placeholder 'Start typing to add names...'. Underneath is a section titled 'Shared with' containing two entries: 'Danilo Piparo (danilo)' and 'Enric Tejedor Saavedra (enric)'. At the bottom of the dialog are two buttons: 'Stop Sharing' (red) and 'Update' (blue). The footer of the main interface reads: 'SWAN © Copyright CERN 2016-2018. All rights reserved. Home | Contact | Support | Report a bug'.



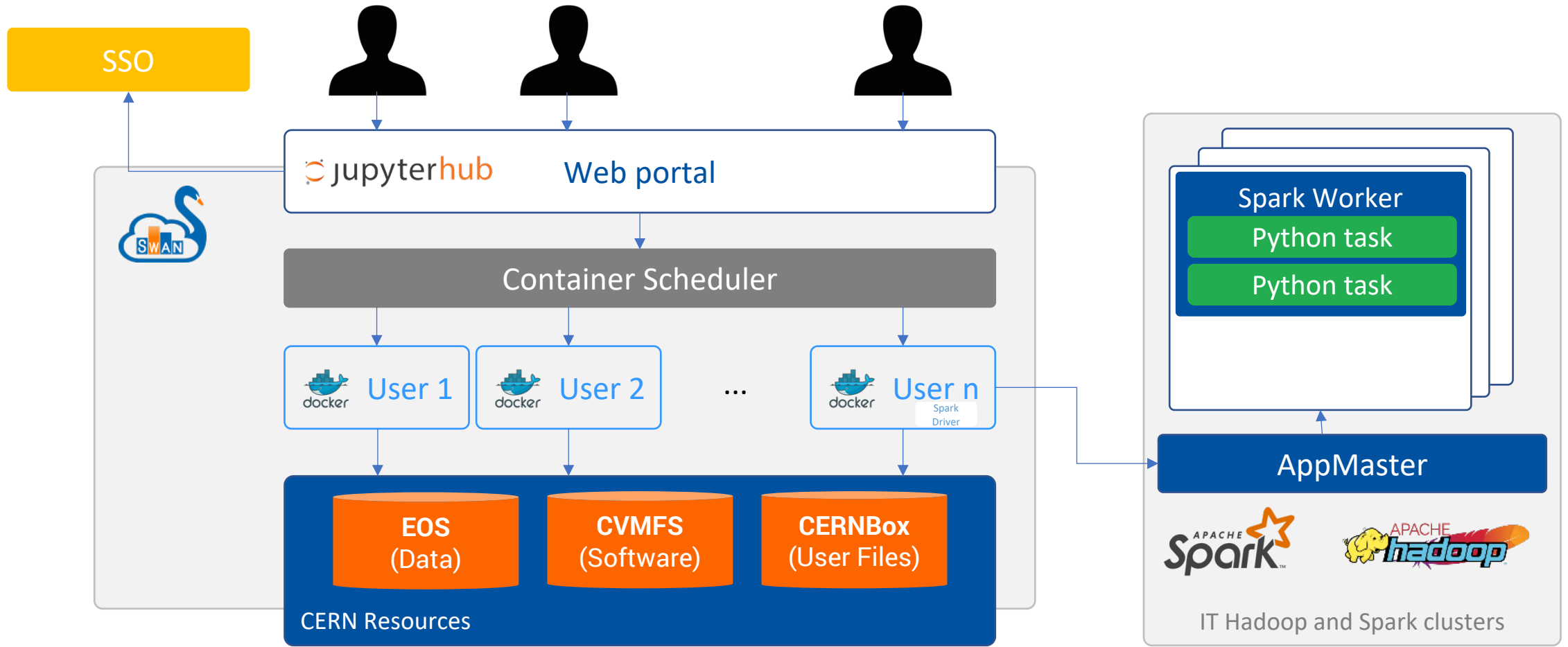
Software

- > Software distributed through CVMFS
 - Distributed read-only filesystem
 - "LCG Releases" - pack a series of compatible packages
 - Reduced Docker Images size
 - Lazy fetching of software
- > Possibility to install libraries in user cloud storage
 - Good way to use custom/not mainstream packages
 - Configurable environment





Architecture



* User sessions are terminated after 6h of no activity to optimize usage of compute resources

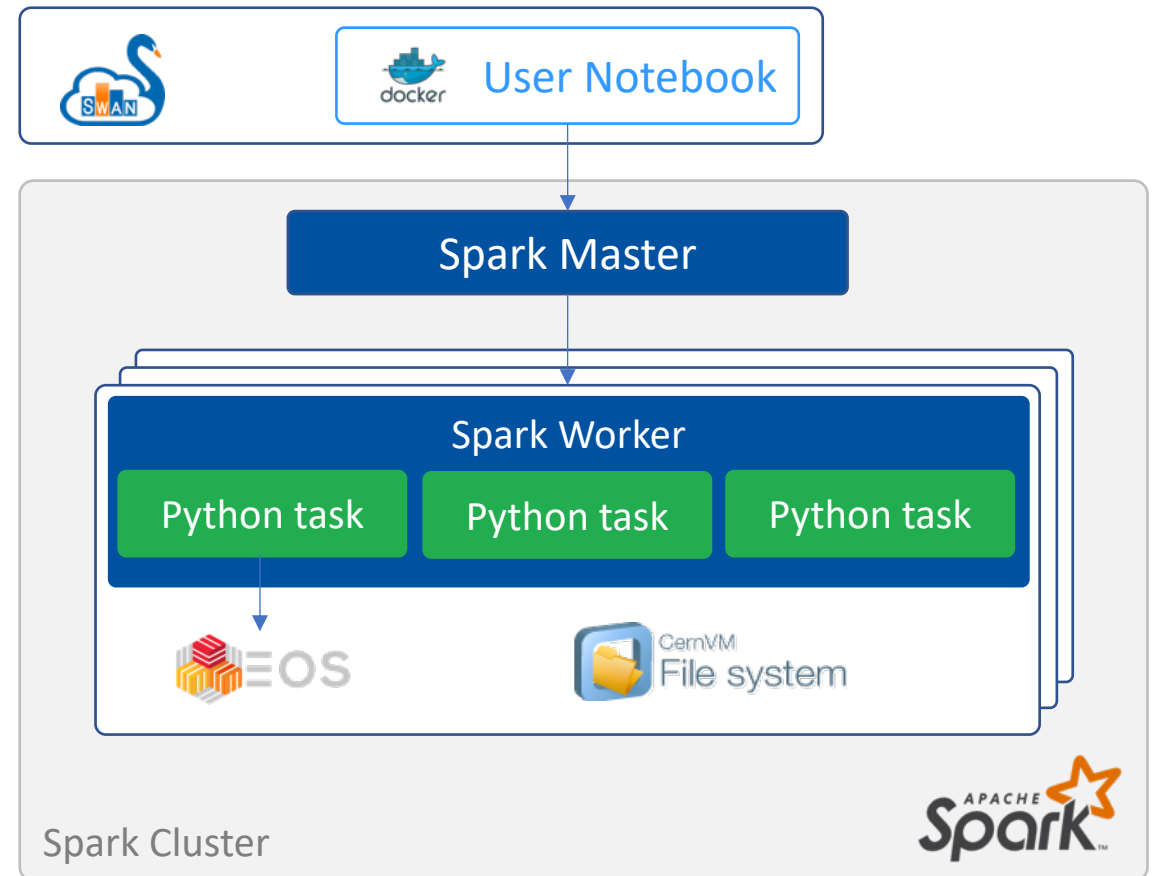


Access to Computing Resources



Integration with Spark

- > Connection to CERN Spark Clusters
 - Spark: general purpose distributed computing framework
- > Same environment across platforms (local/remote)
 - Software - CVMFS
- > Graphical Jupyter extensions developed
 - Spark Connector
 - Spark Monitor
- > Spark Clusters
 - NXCals: – Dedicated cluster for accelerator logging
 - Analytix: – General purpose YARN cluster
 - Cloud Containers: – General purpose Kubernetes cluster





Spark Connector

Spark > Spark_Simple (autosaved)

FILE EDIT VIEW INSERT CELL KERNEL HELP

Markdown

Simple example with Spark

This notebook illustrates the use of [Spark](#) in [SWAN](#).

The current setup allows to execute [PySpark](#) operations on a local small datasets.

In the future, SWAN users will be able to attach external Spark clusters. Moreover, a Scala Jupyter kernel will be added to use Spark from...

Import the necessary modules

The `pyspark` module is available to perform the necessary imports

```
In [ ]: from pyspark import SparkContext
```

Spark clusters connection

You are going to connect to: **hadalytic**

You can configure the following [options](#).
Environment variables can be used via {ENV_VAR_NAME}.

Add a new option

Bundled configurations

Include NXCALs options

Selected configuration

- spark.shuffle.service.enabled: false
- spark.driver.memory: 2g
- spark.executor.instances: 4

Connect

- > Spark Connector – handling the spark configuration complexity
 - User is presented with Spark Session (Spark) and Spark Context (sc)
 - Ability to bundle configurations specific to user communities
 - Ability to specify additional configuration

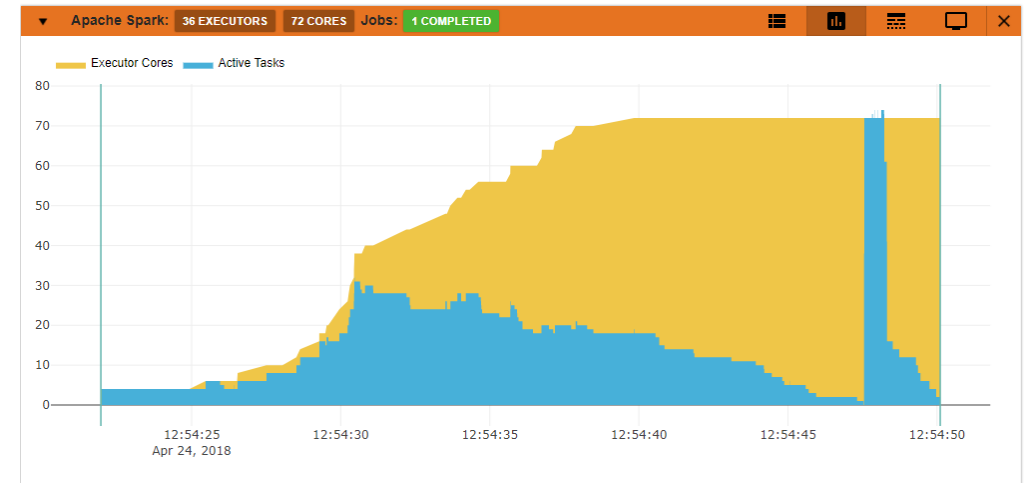




Spark Monitor

> Spark Monitor – jupyter notebook extension

- For live monitoring of spark jobs spawned from the notebook
- Access to Spark WEB UI from the notebook
- Several other features to debug and troubleshoot Spark application



Job ID	Job Name	Status	Stages	Tasks	Submission Time	Duration
11	toPandas	RUNNING	0/2 (1 active)	4 + 4 / 281	a few seconds ago	-





How to get help?

> SWAN Community

- <https://cern.ch/swan-community>
- Find solution to the commonly encountered issues / questions on the usage of Jupyter notebooks, LCG releases, storage and spark
- Report improvements / new features to the service
- E.g: How to install custom user packages

> Service Now

- Report issues to the service
- E.g: Unable to start a session

> Help on various functionality of the tool

Help

1. Introduction

- > What is SWAN
- > Jupyter notebooks
- > Cloud storage: CERNBox and EOS
- > Software: CVMFS

2. Create and manage a SWAN session

- > Select a configuration
- > Set a configuration as default
- > Switch to a new configuration
- > Terminate a session

3. Working with SWAN

- > Create a Project
- > Create a Notebook
- > Create a Folder
- > Open a Terminal



How to get started?

> Gallery of sample notebooks for varied usage of SWAN

- Quick way to be productive
- Also accessible from cern.ch/swan

Gallery

- > Basic Examples
- > ROOT Primer
- > Accelerator Complex
- > FCC
- > LHC Signal Monitoring
- > Beam Dynamics
- > Machine Learning
- > Apache Spark
- > Outreach
- > Awake

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!

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

Simple ROOTbook (Python)



Simple ROOTbook (C++)



Simple Fitting

```

1 // In this ROOT notebook, we will fit a histogram with a Gaussian.
2
3 // The histogram is created in the previous notebook.
4
5 // We define the Gaussian fit function.
6
7 // We fit the histogram with the Gaussian function.
8
9 // We print the fit parameters.
10
11 // We print the fit chi-square.
12
13 // We print the fit covariance matrix.
14
15 // We print the fit residuals.
16
17 // We print the fit pull.
18
19 // We print the fit quality.
20
21 // We print the fit status.
22
23 // We print the fit message.
24
25 // We print the fit error.
26
27 // We print the fit warning.
28
29 // We print the fit error message.
30
31 // We print the fit warning message.
32
33 // We print the fit error message.
34
35 // We print the fit warning message.
36
37 // We print the fit error message.
38
39 // We print the fit warning message.
40
41 // We print the fit error message.
42
43 // We print the fit warning message.
44
45 // We print the fit error message.
46
47 // We print the fit warning message.
48
49 // We print the fit error message.
50
51 // We print the fit warning message.
52
53 // We print the fit error message.
54
55 // We print the fit warning message.
56
57 // We print the fit error message.
58
59 // We print the fit warning message.
60
61 // We print the fit error message.
62
63 // We print the fit warning message.
64
65 // We print the fit error message.
66
67 // We print the fit warning message.
68
69 // We print the fit error message.
70
71 // We print the fit warning message.
72
73 // We print the fit error message.
74
75 // We print the fit warning message.
76
77 // We print the fit error message.
78
79 // We print the fit warning message.
80
81 // We print the fit error message.
82
83 // We print the fit warning message.
84
85 // We print the fit error message.
86
87 // We print the fit warning message.
88
89 // We print the fit error message.
90
91 // We print the fit warning message.
92
93 // We print the fit error message.
94
95 // We print the fit warning message.
96
97 // We print the fit error message.
98
99 // We print the fit warning message.
100
101 // We print the fit error message.
102
103 // We print the fit warning message.
104
105 // We print the fit error message.
106
107 // We print the fit warning message.
108
109 // We print the fit error message.
110
111 // We print the fit warning message.
112
113 // We print the fit error message.
114
115 // We print the fit warning message.
116
117 // We print the fit error message.
118
119 // We print the fit warning message.
120
121 // We print the fit error message.
122
123 // We print the fit warning message.
124
125 // We print the fit error message.
126
127 // We print the fit warning message.
128
129 // We print the fit error message.
130
131 // We print the fit warning message.
132
133 // We print the fit error message.
134
135 // We print the fit warning message.
136
137 // We print the fit error message.
138
139 // We print the fit warning message.
140
141 // We print the fit error message.
142
143 // We print the fit warning message.
144
145 // We print the fit error message.
146
147 // We print the fit warning message.
148
149 // We print the fit error message.
150
151 // We print the fit warning message.
152
153 // We print the fit error message.
154
155 // We print the fit warning message.
156
157 // We print the fit error message.
158
159 // We print the fit warning message.
160
161 // We print the fit error message.
162
163 // We print the fit warning message.
164
165 // We print the fit error message.
166
167 // We print the fit warning message.
168
169 // We print the fit error message.
170
171 // We print the fit warning message.
172
173 // We print the fit error message.
174
175 // We print the fit warning message.
176
177 // We print the fit error message.
178
179 // We print the fit warning message.
180
181 // We print the fit error message.
182
183 // We print the fit warning message.
184
185 // We print the fit error message.
186
187 // We print the fit warning message.
188
189 // We print the fit error message.
190
191 // We print the fit warning message.
192
193 // We print the fit error message.
194
195 // We print the fit warning message.
196
197 // We print the fit error message.
198
199 // We print the fit warning message.
200
201 // We print the fit error message.
202
203 // We print the fit warning message.
204
205 // We print the fit error message.
206
207 // We print the fit warning message.
208
209 // We print the fit error message.
210
211 // We print the fit warning message.
212
213 // We print the fit error message.
214
215 // We print the fit warning message.
216
217 // We print the fit error message.
218
219 // We print the fit warning message.
220
221 // We print the fit error message.
222
223 // We print the fit warning message.
224
225 // We print the fit error message.
226
227 // We print the fit warning message.
228
229 // We print the fit error message.
230
231 // We print the fit warning message.
232
233 // We print the fit error message.
234
235 // We print the fit warning message.
236
237 // We print the fit error message.
238
239 // We print the fit warning message.
240
241 // We print the fit error message.
242
243 // We print the fit warning message.
244
245 // We print the fit error message.
246
247 // We print the fit warning message.
248
249 // We print the fit error message.
250
251 // We print the fit warning message.
252
253 // We print the fit error message.
254
255 // We print the fit warning message.
256
257 // We print the fit error message.
258
259 // We print the fit warning message.
260
261 // We print the fit error message.
262
263 // We print the fit warning message.
264
265 // We print the fit error message.
266
267 // We print the fit warning message.
268
269 // We print the fit error message.
270
271 // We print the fit warning message.
272
273 // We print the fit error message.
274
275 // We print the fit warning message.
276
277 // We print the fit error message.
278
279 // We print the fit warning message.
280
281 // We print the fit error message.
282
283 // We print the fit warning message.
284
285 // We print the fit error message.
286
287 // We print the fit warning message.
288
289 // We print the fit error message.
290
291 // We print the fit warning message.
292
293 // We print the fit error message.
294
295 // We print the fit warning message.
296
297 // We print the fit error message.
298
299 // We print the fit warning message.
300
301 // We print the fit error message.
302
303 // We print the fit warning message.
304
305 // We print the fit error message.
306
307 // We print the fit warning message.
308
309 // We print the fit error message.
310
311 // We print the fit warning message.
312
313 // We print the fit error message.
314
315 // We print the fit warning message.
316
317 // We print the fit error message.
318
319 // We print the fit warning message.
320
321 // We print the fit error message.
322
323 // We print the fit warning message.
324
325 // We print the fit error message.
326
327 // We print the fit warning message.
328
329 // We print the fit error message.
330
331 // We print the fit warning message.
332
333 // We print the fit error message.
334
335 // We print the fit warning message.
336
337 // We print the fit error message.
338
339 // We print the fit warning message.
340
341 // We print the fit error message.
342
343 // We print the fit warning message.
344
345 // We print the fit error message.
346
347 // We print the fit warning message.
348
349 // We print the fit error message.
350
351 // We print the fit warning message.
352
353 // We print the fit error message.
354
355 // We print the fit warning message.
356
357 // We print the fit error message.
358
359 // We print the fit warning message.
360
361 // We print the fit error message.
362
363 // We print the fit warning message.
364
365 // We print the fit error message.
366
367 // We print the fit warning message.
368
369 // We print the fit error message.
370
371 // We print the fit warning message.
372
373 // We print the fit error message.
374
375 // We print the fit warning message.
376
377 // We print the fit error message.
378
379 // We print the fit warning message.
380
381 // We print the fit error message.
382
383 // We print the fit warning message.
384
385 // We print the fit error message.
386
387 // We print the fit warning message.
388
389 // We print the fit error message.
390
391 // We print the fit warning message.
392
393 // We print the fit error message.
394
395 // We print the fit warning message.
396
397 // We print the fit error message.
398
399 // We print the fit warning message.
400
401 // We print the fit error message.
402
403 // We print the fit warning message.
404
405 // We print the fit error message.
406
407 // We print the fit warning message.
408
409 // We print the fit error message.
410
411 // We print the fit warning message.
412
413 // We print the fit error message.
414
415 // We print the fit warning message.
416
417 // We print the fit error message.
418
419 // We print the fit warning message.
420
421 // We print the fit error message.
422
423 // We print the fit warning message.
424
425 // We print the fit error message.
426
427 // We print the fit warning message.
428
429 // We print the fit error message.
430
431 // We print the fit warning message.
432
433 // We print the fit error message.
434
435 // We print the fit warning message.
436
437 // We print the fit error message.
438
439 // We print the fit warning message.
440
441 // We print the fit error message.
442
443 // We print the fit warning message.
444
445 // We print the fit error message.
446
447 // We print the fit warning message.
448
449 // We print the fit error message.
450
451 // We print the fit warning message.
452
453 // We print the fit error message.
454
455 // We print the fit warning message.
456
457 // We print the fit error message.
458
459 // We print the fit warning message.
460
461 // We print the fit error message.
462
463 // We print the fit warning message.
464
465 // We print the fit error message.
466
467 // We print the fit warning message.
468
469 // We print the fit error message.
470
471 // We print the fit warning message.
472
473 // We print the fit error message.
474
475 // We print the fit warning message.
476
477 // We print the fit error message.
478
479 // We print the fit warning message.
480
481 // We print the fit error message.
482
483 // We print the fit warning message.
484
485 // We print the fit error message.
486
487 // We print the fit warning message.
488
489 // We print the fit error message.
490
491 // We print the fit warning message.
492
493 // We print the fit error message.
494
495 // We print the fit warning message.
496
497 // We print the fit error message.
498
499 // We print the fit warning message.
500
501 // We print the fit error message.
502
503 // We print the fit warning message.
504
505 // We print the fit error message.
506
507 // We print the fit warning message.
508
509 // We print the fit error message.
509

```



Where to find us



Where to find us

> Contacts

- swan-admins@cern.ch
- <http://cern.ch/swan>

> Repository

- <https://gitlab.cern.ch/swan>

> Science Box

- <https://cern.ch/sciencebox>

SWAN and its analysis ecosystem

Thank you

Prasanth Kothuri
prasanth.kothuri@cern.ch