

# SparkSQL lab

# Zoe login

- <https://cloud-platform.eurecom.fr/zoe/login>
- Click on EURECOM gitlab and login with EURECOM creds



Username:

Password:

Login

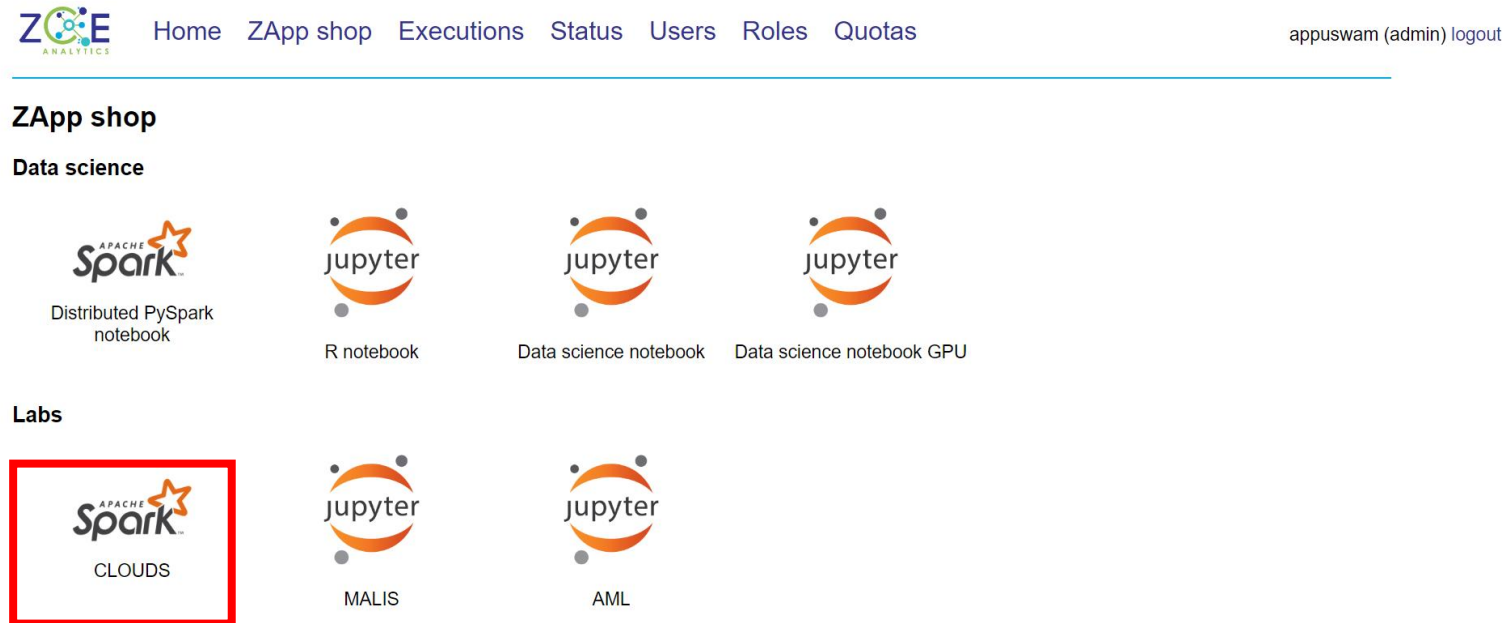
Or login with:

Eurecom GitLab

Zoe Analytics v. 2018.12 — Give us some feedback! — © Copyright 2017 by DSG


# Clouds ZApp

- Click on Zapp Shop option and click on CLOUDS



# Spark cluster

- Leave default options and click on “Start”



## CLOUDS Lab ZApp

This ZApp must be used during the laboratory activities of the CLOUDS course at Eurecom.

Combine the full power of a distributed Apache Spark cluster with Python Jupyter Notebooks.

Spark is configured in stand-alone, distributed mode. This ZApp contains Spark version 2.2.2.

To start clone the repository at <https://github.com/EURECOM-CLOUDS-COURSE> and open the notebook corresponding to the laboratory session you are in.

### ZApp details

This ZApp is composed by the following services:

- 1 spark-master (1 essential)
  - Suggested memory allocation: 1.00 GiB
  - Suggested CPU core allocation limit: 1
- Image: `zapps/spark-master:10396`
- 2 spark-worker (1 essential)
  - Suggested memory allocation: 12.0 GiB
  - Suggested CPU core allocation limit: 3
- Image: `zapps/spark-worker:10396`
- 1 spark-jupyter (1 essential)
  - Suggested memory allocation: 12.0 GiB
  - Suggested CPU core allocation limit: 2
- Image: `zapps/spark-jupyter-notebook:10396`

The following persistent volumes will be available:

- `/mnt/workspace`: User workspace (RW)
- `/mnt/datasets` (RO)

### Start-up parameters:

EXECUTION NAME:

spark-master

MEMORY ALLOCATION (GiB)

1.0

MINIMUM CORE ALLOCATION:

1

spark-worker

MEMORY ALLOCATION (GiB)

12.0

MINIMUM CORE ALLOCATION:

3

spark-jupyter

MEMORY ALLOCATION (GiB)

12.0

MINIMUM CORE ALLOCATION:

2

☐ Download JSON with these parameters for command-line execution

[Back to the ZApp shop](#)

# Launch Jupyter

- Wait until “status” turns to running from queued
- Click “[Jupyter Notebook interface](#)” (internal link in EURECOM)
  - You can get to Spark master web i/f later from here

## Detailed information for execution clouds (86711)

- Application name: clouds
- Owner: appuswam
- Time submitted: 11/13/2019, 9:28:50 PM
- Time started: 11/13/2019, 9:28:54 PM
- Time finished: not yet
- Status: **running**
- Will be killed at: 11/20/2019, 9:28:50 PM
- Actions: [Terminate](#)
- [Resource usage plots](#)

## Endpoints

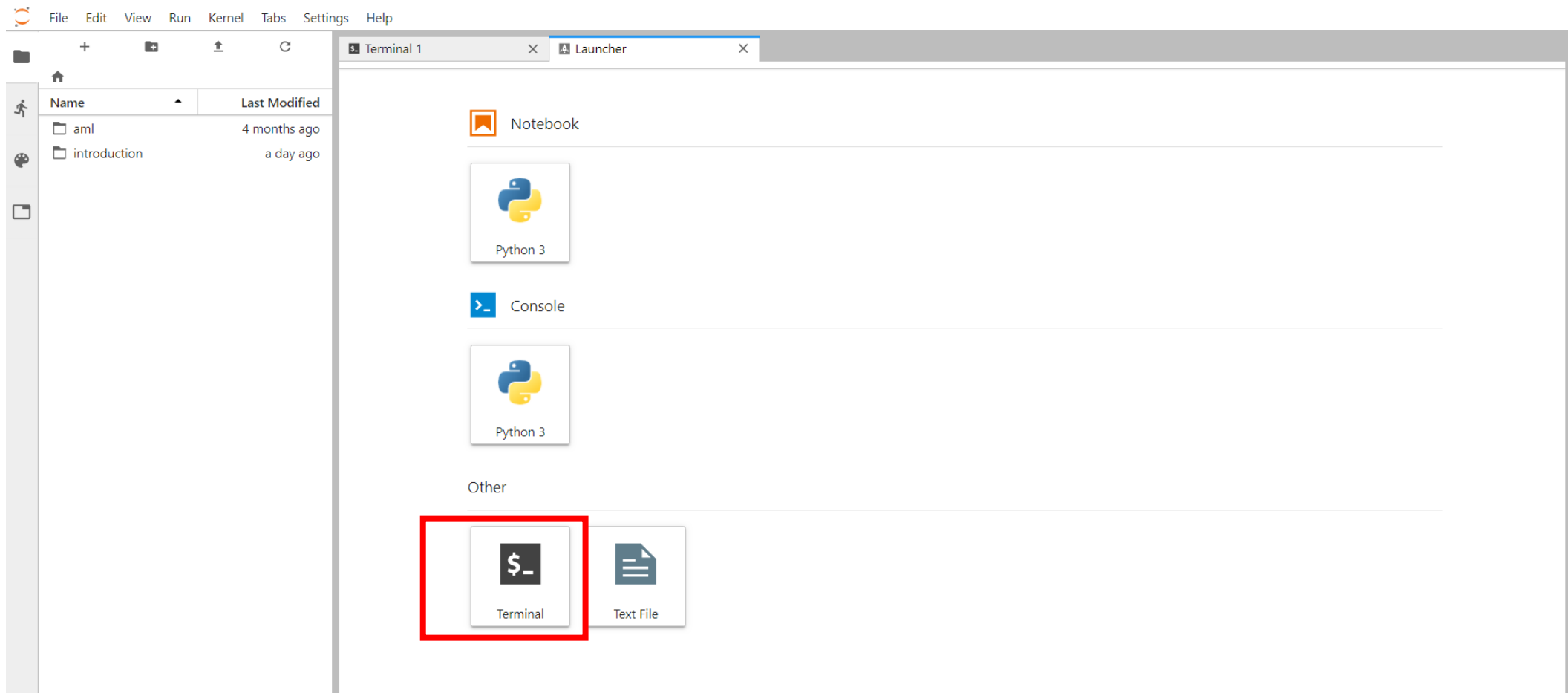
- [Jupyter Notebook interface \(internal link\)](#)
- [Spark master web interface \(http://192.168.47.8:34496/\)](#)

## Services:

ID	Name	Zoe status	Backend status	Host	Labels	Output logs	Errors
93953	spark-master0	active	started	bf8	labs	<a href="#">open</a>	None
93956	spark-jupyter0	active	started	bf16	labs	<a href="#">open</a>	None
93954	spark-worker0	active	started	bf20	labs	<a href="#">open</a>	None
93955	spark-worker1	active	started	bf22	labs	<a href="#">open</a>	None

# Launch terminal

- Click on “Terminal” in Jupyter launcher



# Clone and launch notebook

- Clone git repository that contains first lab
  - “git clone <https://github.com/EURECOM-CLOUDS-COURSE/SparkSQL.git>” in the terminal
- You should see directory “SparkSQL” appear. Double click to enter the directory.
- Double click on file “SPARKSQL.ipynb” within directory to open up the notebook

# Admin (same as before)

- Team up (upto 2 per group) or solo
- All labs based on Zoe – our cloud platform
  - Remote work possible: Zoe, notebooks accessible outside EURECOM
  - Beware: But spark admin UI accessible only in EURECOM network
- Submission deadline: **1 week** (Following Thursday end of day)
  - If you cannot finish on the same day
- To submit
  - Download the ipython notebook locally
  - Email notebook with subject line: “***groupname* CLOUDS lab submission**”
    - Groupname: 4 chars of partner1’s lastname - 4 chars of partner2’s lastname
    - Ex: Alan Turing & John von Neumann = turi-neum