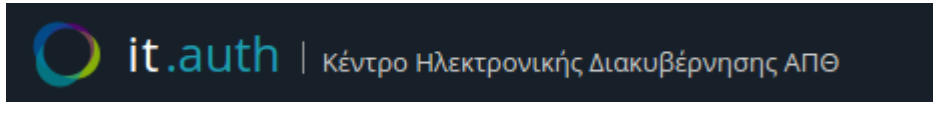
HPC WORKFLOWS USING SLURM

Machine Learning examples on Aristotle Cluster



IN THIS PRESENTATION

- Connect to Aristotle Web Interface
- Run a Jupyter Notebook on Aristotle cluster
- Submit a batch job to use additional computing resources

EXAMPLE JUPYTER NOTEBOOK

The Extreme Gradient Boosting (`XGBoost`) open-source library is used for this simple Regression example.

`XGBoost` implements machine learning algorithms under the Gradient Boosting framework.

PREPARATION FOR THIS SESSION

Please go through a `Unix Command cheat sheet` as the following:

- https://hpc.it.auth.gr/cheat-sheet/

A few unix commands can be useful to run the examples that follow.

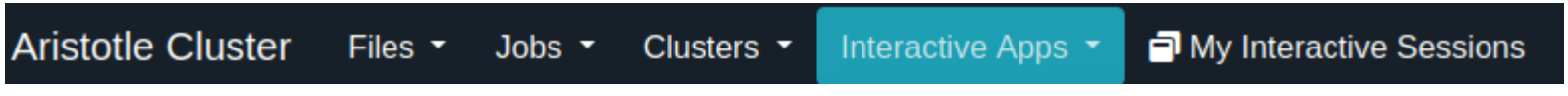
ACCESS ARISTOTLE HPC CLUSTER

from your browser: https://hpc.auth.gr/

1. Start Jupyter Server on the cluster
2. Use a custom virtual enviroment on Jupyter
3. Download results to your local machine

START JUPYTER SERVER

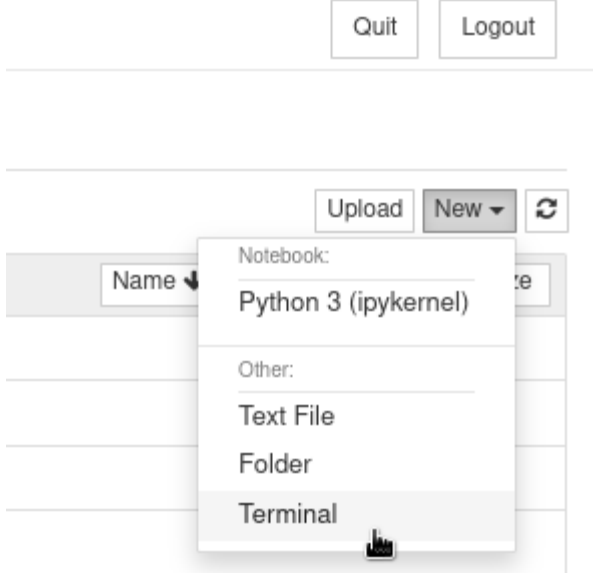
`Interactive Apps` -> `Jupyter Server`



... and launch!

START A NEW TERMINAL

on the Jupyter Server



Use `cp` command to copy the example jupyter notebook:

```
$ cp /mnt/apps/custom/jupyter/nb/xgboost_example.ipynb .
```

Source the prebuilt python virtual environment:

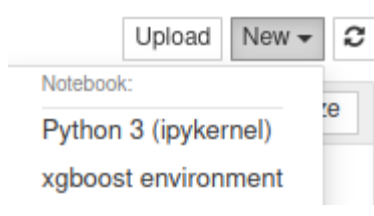
```
$ source /mnt/apps/custom/python-envs/xgboost-env/bin/activat
```

Install the `IPython kernel` in this environment for your user account:

```
$ python -m ipykernel install --user --name xgboost-env \
   --display "xgboost environment"
```

START NEW NOTEBOOK

Using the custom environment



On Jupyter menu select `File -> Open`

to load the `xboost example notebook`.

EXPORT PYTHON SCRIPT

At the notebook menu select:

`Download as` -> `Python (.py)`

PYTHON VIRTUAL ENVIROMENT

(+ Jupyter IPython Kernel)

To create a new custom python venv on your account the following process can be used:

```
module load gcc/9.4.0-eewq4j6 python/3.9.10-ve54vyn
python -m venv xgboost-env
source xgboost-env/bin/activate
pip install --upgrade pip
pip install jupyter xgboost matplotlib scikit-learn
python -m ipykernel install --user --name xgboost-env \
       --display "xgboost environment"
```

USING SLURM

to Access HPC Resources

SLURM WORKLOAD MANAGER

- Allocates and manages exclusive users access to cluster resources
- Provides a framework for job tracking and parallel job execution
  - Quick Start User Guide
  - Slurm Directives

SLURM USER COMMANDS (1)

- Submit a job to the cluster

```
$ sbatch <job_script>
```

- Show status of running and queued jobs

```
$ squeue
# Filter results for one user
$ squeue -u <username>
# Filter results for one partition
$ squeue -p <partition>
```

- Cancel a submitted job

```
$ scancel
```

SLURM USER COMMANDS (2)

- Show status of available partitions

```
$ sinfo
$ sinfo -N --long # how node status
```

- Show resources and efficiency of completed job

```
$ seff <jobid>
```

- Report job accounting information

```
$ sacct
```

BATCH JOB EXAMPLES

EXAMPLE 1: A TEST JOB

Steps:

1. Create a submission script
2. Submit job to Slurm
3. Monitor job execution
4. Get job results

Related docs:

- https://hpc.it.auth.gr/jobs/serial-slurm/

EXAMPLE 1: A TEST JOB

Submission script

```
#!/bin/bash
#SBATCH --time=10:00
#SBATCH --partition=testing
echo "Hello from $(hostname)"
sleep 30
echo Bye
```

EXAMPLE 2: MORE CPUS

```
#!/bin/bash
#SBATCH --partition=rome
#SBATCH --time=10:00
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=16

stress --cpu ${SLURM_NTASKS} --timeout 60
```

CPU Efficiency: `seff <jobid>`

EXAMPLE 3: MORE MEMORY

```
Memory Per Task = Total Memory on Node / #CPUs on Node
```

To allocate more memory use `--mem` directive:

```
#!/bin/bash
#SBATCH --partition=rome
#SBATCH --job-name=memory
#SBATCH --time=4:00
#SBATCH --mem=11G

./allocate-10gb
```

EXAMPLE 4: GPU JOBS

- Partitions:
  - **gpu** : 2 nodes with a NVIDIA Tesla P100
  - **ampere**: 1 node with 8 NVIDIA A100

```
#!/bin/bash
#SBATCH --partition=gpu
#SBATCH --gres=gpu:1
#SBATCH --cpus-per-task=20
#SBATCH --time=10:00

nvidia-smi
```

RUN XGBOOST EXAMPLE PYTHON SCRIPT

as a batch job on the cluster

```
#!/bin/bash
#SBATCH --job-name=xgboost-example
#SBATCH --partition=rome
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=8
#SBATCH --time=1:00:00

source /mnt/apps/custom/python-envs/xgboost-env/bin/activate

python example.py
```

THANK YOU !!