Big data science Day 1 - Hands on

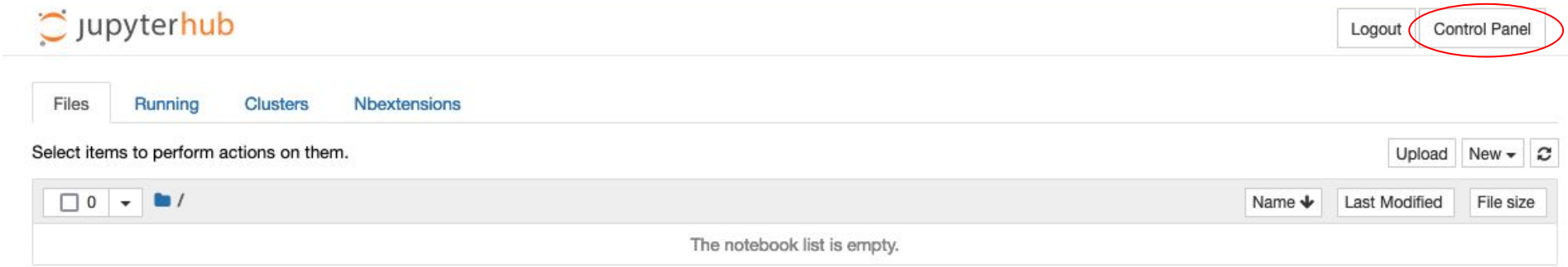
F. Legger - INFN Torino

https://github.com/Course-bigDataAndML/MLCourse-INFN-2022

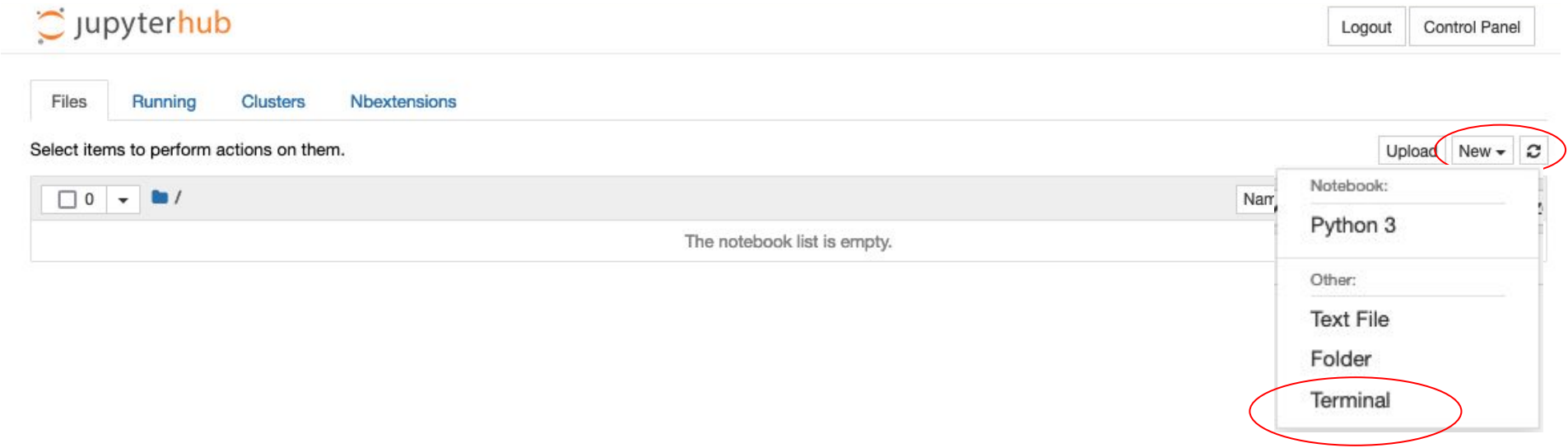
How to start

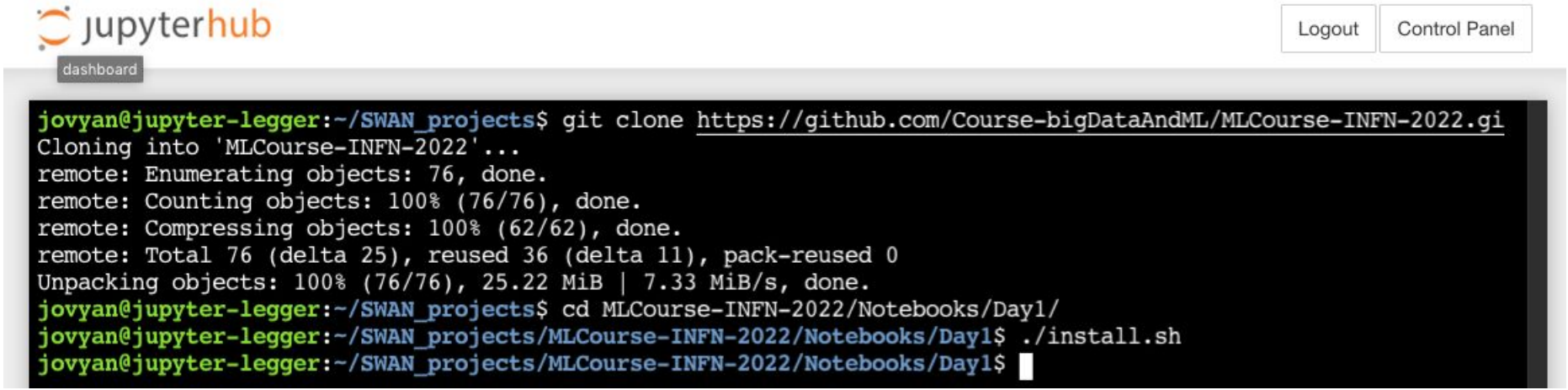
- Point your browser to the JHub link you received by email
- Authenticate
- It should look like this:

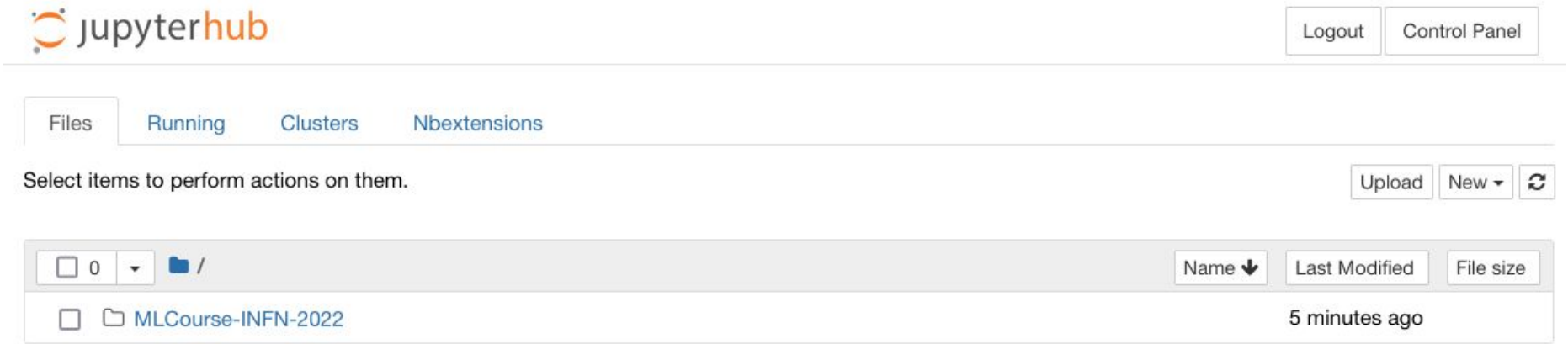
Start/stop jupyterHub



- **Open a terminal**
  - git clone https://github.com/Course-bigDataAndML/MLCourse-INFN-2022.git
  - cd MLCourse-INFN-2022/Notebooks/Day1/
  - ./install.sh




jupyterhub
dashboard
Logout
Control Panel
jovyan@jupyter-legger:~/SWAN_projects$ git clone https://github.com/Course-bigDataAndML/MLCourse-INFN-2022.gi
Cloning into 'MLCourse-INFN-2022'...
remote: Enumerating objects: 76, done.
remote: Counting objects: 100% (76/76), done.
remote: Compressing objects: 100% (62/62), done.
remote: Total 76 (delta 25), reused 36 (delta 11), pack-reused 0
Unpacking objects: 100% (76/76), 25.22 MiB | 7.33 MiB/s, done.
jovyan@jupyter-legger:~/SWAN_projects$ cd MLCourse-INFN-2022/Notebooks/Day1/
jovyan@jupyter-legger:~/SWAN_projects/MLCourse-INFN-2022/Notebooks/Day1$ ./install.sh
jovyan@jupyter-legger:~/SWAN_projects/MLCourse-INFN-2022/Notebooks/Day1$


jupyterhub
Logout
Control Panel
Files
Running
Clusters
Nbextensions
Select items to perform actions on them.
Upload
New
0
/
Name
Last Modified
File size
MLCourse-INFN-2022
5 minutes ago

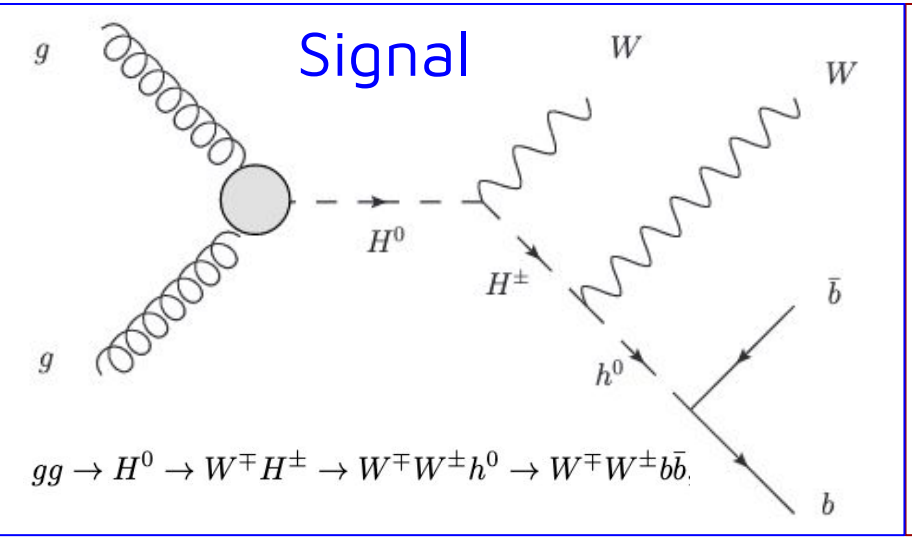
What we will use

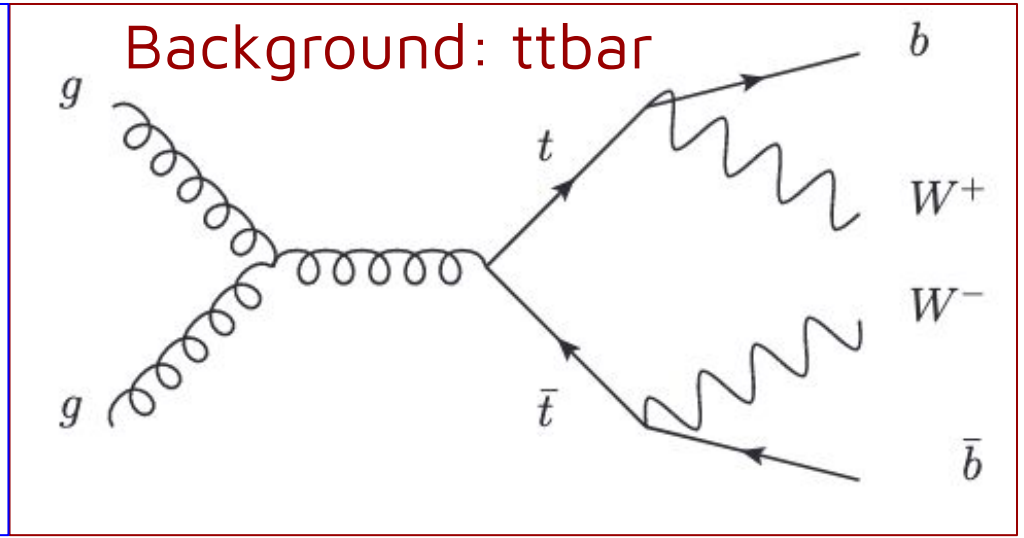
- Python with Jupyter notebooks
- **Day 1:**
  - familiarise with ML dataset, parquet files
- **Day 2:**
  - Gradient Boosting Trees GBT MLlib
  - Multilayer Perceptron Classifier MCP MLlib
- **Day 3:** Neural networks
  - Keras Sequential model

Input dataset for hands-on

https://archive.ics.uci.edu/ml/datasets/HIGGS

- Open HEP dataset @UCI
- Signal (heavy Higgs) + background (ttbar)





Baldi, Sadowski, and Whiteson. "Searching for Exotic Particles in High-energy Physics with Deep Learning." Nature Communications 5

Input dataset for hands-on

- Monte Carlo events
  - **21 low level features**
    - pt's, angles, MET, b-tag, ...
  - **7 high level features**
    - Invariant masses (m(jj), m(jjj), ...)