



# **AOB: A New Tool For HTCondor Submission - pyHTC**

A. Poyet, G. Sterbini



Beam-Beam and Luminosity Meeting – 12<sup>th</sup> July 2019

# Context and Motivations

- To run a script a huge number of times, or needing a heavy computing time, one needs in general to send them to a cluster of nodes to run every scripts in parallel.
- At CERN, the batch service is based on HTCondor (previously, it was LSF).
- Some codes like SixTrack come with a submission environment that hide the submission from the user → Makes your life easier, but what's really going on in this black box?
- Useful reference to read: <http://batchdocs.web.cern.ch/batchdocs/index.html>
- **HTCONDOR TUTORIAL** : <http://batchdocs.web.cern.ch/batchdocs/tutorial/introduction.html>
  - See also 'HTCondor training' by F. Protopsalti : <https://indico.cern.ch/event/635217/attachments/1458551/2252370/HTCondor-Presentation.pdf>
- Submitting a job requires the creation of a proper submit file that can be understood by HTCondor
- In general, it would be useful to be able **to prepare and to submit jobs from your python script/notebook** → This is the aim of **pyHTC** 😊

# pyHTC

- pyHTC is python package that helps the user to prepare and submit jobs to HTCondor
  - <https://github.com/apoyet/pyHTC>
- Very preliminary at the moment : feel free to contribute 😊

apoyet / pyHTC

Code Issues Pull requests Projects Wiki Security Insights Settings

No description, website, or topics provided.

Manage topics

27 commits 1 branch 0 releases 1 contributor

Branch: master	New pull request	Create new file	Upload files	Find File	Clone or download
apoyet Merge branch 'master' of https://github.com/apoyet/pyHTC Latest commit a6b1ba 14 days ago					
example	Change the name of the package				14 days ago
pyHTC	Merge branch 'master' of https://github.com/apoyet/pyHTC				14 days ago
README.md	Update README.md				14 days ago
setup.py	First commit for setup and init				22 days ago

### pyHTC

pyHTC is a python package to ease the submission of jobs to the HTCondor cluster. It also provides functions to help in the organization of the parameters of each job, storing the data in convenient pandas DataFrames, allowing the user to find out which job corresponds to what study.

#### Installation

To install the package, you might use the command:

```
pip install --user git+https://github.com/apoyet/pyHTC.git
```

If you want to upgrade it, please issue the following command:

```
pip install --upgrade --user git+https://github.com/apoyet/pyHTC.git
```

**PRELIMINARY**

# Starting with pyHTC

- EOS Submission is not supported by the HTCondor service ([http://batchdocs.web.cern.ch/batchdocs/troubleshooting/eos\\_submission.html](http://batchdocs.web.cern.ch/batchdocs/troubleshooting/eos_submission.html)). Therefore, this example will be made assuming a notebook **located in AFS**.

- To install the package:

```
pip install --user git+https://github.com/apoyet/pyHTC.git
```

- Then import it in your favourite notebook/script ☺

```
In [1]: from pyHTC.Study import *
```

- pyHTC.Study? What's that? Hmmmm....

→ Better to see a real example!

# Conclusions

- This is only the beginning.... But it is already possible from a dictionary of parameters to produce the corresponding submission file and to submit through python notebooks.
- Outputs files paths are then recovered in a nice Pandas Dataframe, where all the information is summarized.
- The next step will be to implement the flexibility of the input, typically to allow the re-submission of a finer grid of parameters, without re-submitting the jobs already done.



***Thanks for your attention !***

