



# pandas + HDF5 event selection

Marc Paterno

HEP Analytics on HPC Collaboration Meeting

6 Nov 2018

# Status of containers

- We are currently doing development in Docker containers
  - convert to Shifter for NERSC (Cori)
  - convert to Singularity for ALCF (Theta)
  - Build using numpy optimized for KNL (through MKL, from Anaconda Python)
  - Special builds of HDF5 (most recent, MPI enabled)
  - MPI suitable for replacement by Cray MPI @NERSC and @ALCF
- Dockerfiles and makefiles for building and testing in git repository on BitBucket
  - `git clone git@bitbucket.org:ssehrish/pyhpc-docker.git` ; public repository
  - Images available on Docker Hub
    - `docker pull heponhpc/pyhpc:0.5.0`

# Status of code, next steps

- Event selection code is in the repository  
[git@bitbucket.org:fnalscdcomputationalscience/hep\\_hpc\\_python.git](https://git@bitbucket.org:fnalscdcomputationalscience/hep_hpc_python.git)
- Task list is on the BaseCamp site
  - it is not complete; please help complete it
- Major elements
  - refactoring toward production-quality *framework* code
  - write ROOT macro equivalent of currently-completed event selection
  - creation of large file for verification and performance measurement