



# LHC Big Data Project: Kickoff Part 1

Matteo, Lindsey

# Items for Discussion

The LHC Big Data Project aims at designing the most efficient HEP **analysis facility**:

- Front-end:

- Columnar analysis: coffea

- Back-end:

- Big data technologies

- Systems: spark (including laurelin, spark-ttree library), striped, parsl

- Services:

- Existing: VC3
- Future: slate, ssl, serviceX, funcX

- Hardware:

- Existing infrastructure, HPC, GPU (or combinations, like NERSC-9)

# People

- Groups:

- CERN-IT (Canali)
- Intel
- Nebraska (Bloom)
- UChicago (Gardner, Weinberg, Woodard, Wang)
- Notre Dame (Lannon, Tovar, Thain)
- FNAL (Kowalkowski, Paterno, Sehrish, Mandrichenko, Gray, Cremonesi, Hall, Smith)
- Vanderbilt (Melo)
- Purdue (Piperov)
- Caltech (Pata)
- Princeton (Pivarski)
- Urbana (Galewsky)

- Activities:

- Cafes (FNAL, CERN-IT, Vanderbilt, Intel)
- Coffea (FNAL, Princeton)
- ServiceX (UChicago, Urbana)
- FuncX (UChicago)
- Parsl (UChicago, Wisconsin, Nebraska)
- VC3 (UChicago, Notre Dame)
- HPC/Spark/HDF5 (FNAL)
- Striped (FNAL, Princeton)
- GPU (Caltech)