BETHEL
UNIVERSITY

Large-scale analysis

OPEN DATA WORKSHOP 2024

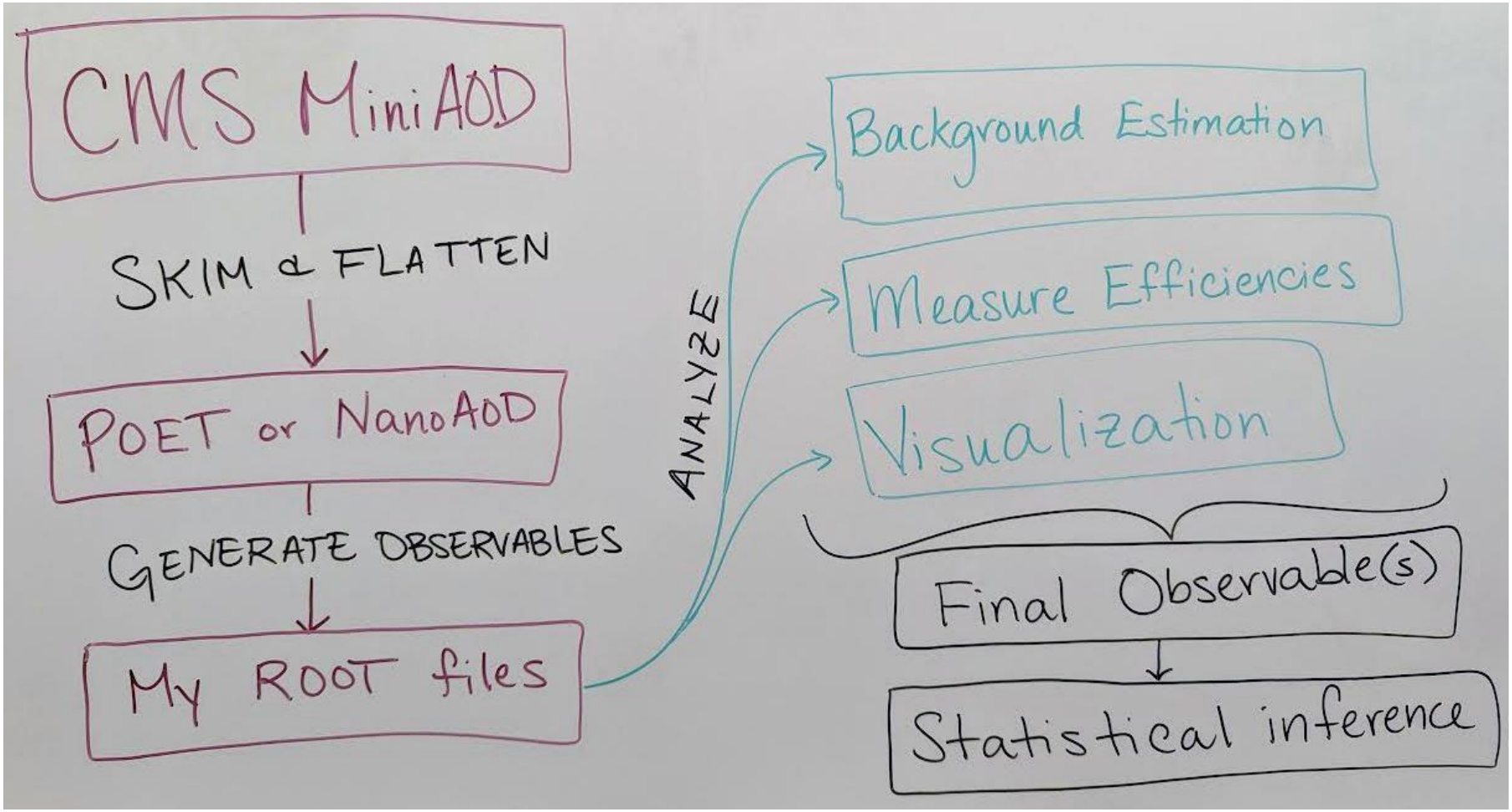
WHEPP CONFERENCE

JULIE HOGAN

1/10/2024

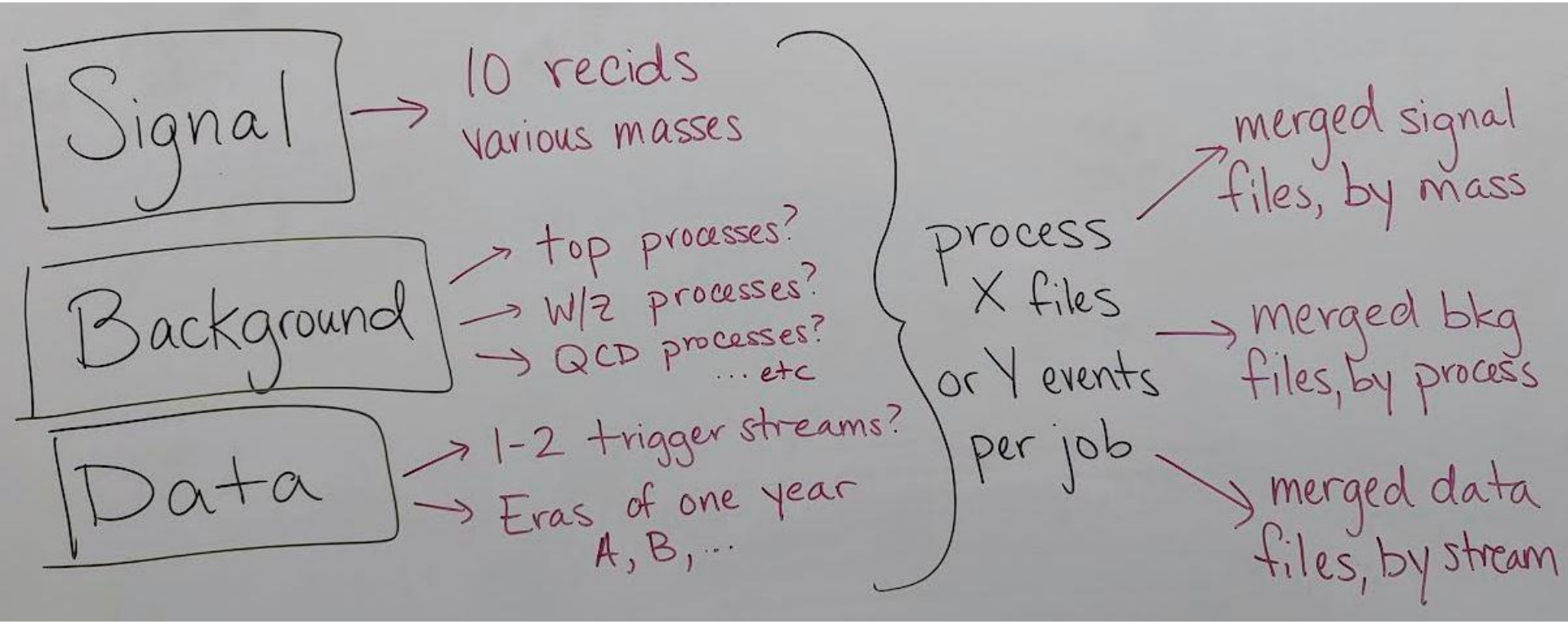
Processing at scale

- Episode 2 is my mini-dissertation on analysis workflow



Processing at scale

- Yesterday's lesson focused on methods to do local analysis of POET files for different physics processes
- Creating the POET files from AOD or MiniAOD tends to take the most resources



Processing at scale

- Local machine: one example analysis of H → 4 leptons cited 1 month to process simulations and 2011-2012 data on a laptop!

- Google cloud: anyone can pay for computing!
  - Requires modern workflow tools to configure
  - Kubernetes used to facilitate parallel processing using containers
  - See our tutorial from summer 2023

- Local clusters: put university resources to work!
  - Most analysis tools can be installed on Linux clusters (ROOT, python)
  - Condor batch systems are common
  - Main task is to access the CMSSW container for AOD processing

HTCondor at TIFR

BETHEL UNIVERSITY

- Your local student facilitators have prepared scripts to submit POET jobs via condor – a great template for your own use!
- We will look through the basics of the scripts very quickly – come back to the lesson webpage to read more thoroughly
- Then we'll practice submitting test jobs and see the results
- Follow episodes 3 and 4 on the webpage: https://cms-opendata-workshop.github.io/workshopwhepp-lesson-condor/03-condorscripts/index.html

More opportunities

- Please share your feedback!
  - https://cern.ch/odwswheppform

- Watch for a workshop in summer 2024!
- Mattermost will remain open for further discussion
- CMS Open Data Guide in the process of completion
- We watch the Open Data Forum actively for questions and issues! Happy to help you develop your Open Data analysis