Inference as a service for particle tracking



Xiangyang Ju
Computing System
Engineer@LBNL



Andrew Naylor

NESAP for Data

postdoctoral

fellow@NERSC

TrkX











Inference as a service for particle tracking EvtGen Client Server **GPU-enabled Node** Digitization **Triton Inference Server Network CPU Nodes NERSC Perlmutter Login Node GPU-enabled Node** Clustering **AWS Instance Triton Inference Server Google Cloud CPU GPU** GPU TrackFinding SP Embedding new SP FRNN --Edges_ Edges-Edge Scores > applyFiltering Filtering WCC -Tracks → ● **TrackFitting** Edge Scores