

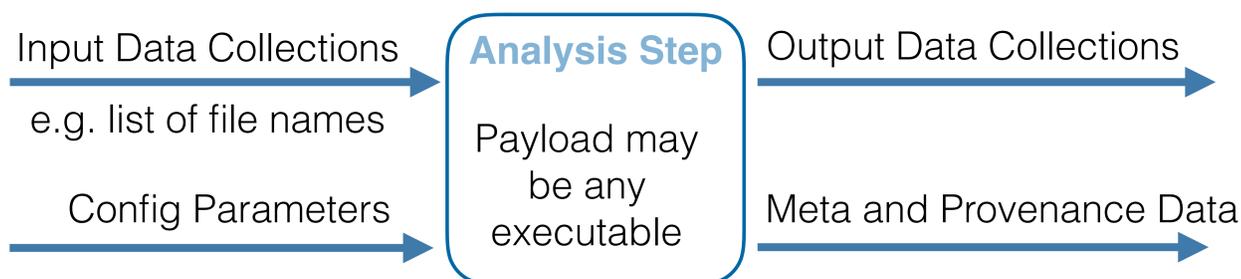
Workflow Management for Complex HEP Analyses

Motivation

- Modern HEP data analyses are
 - complex: need diverse techniques and software packages
 - big: many large datasets
- Results in large overhead for bookkeeping and data managing
- Time missing from doing actual analysis

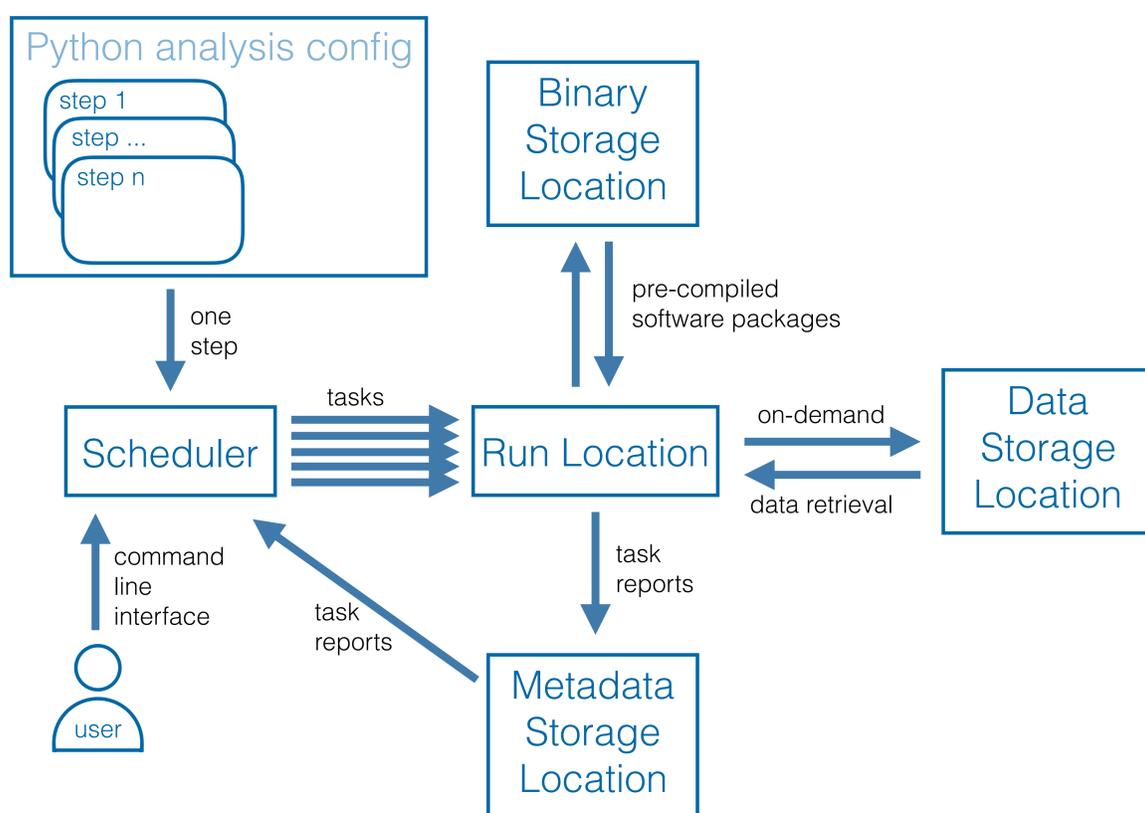
Concept

- Analysis comprises interconnected steps
- Implemented in Python
- As independent from specific infrastructures as possible
- Abstraction of computing resources:
 - Run location
 - Storage location
 - Software environment



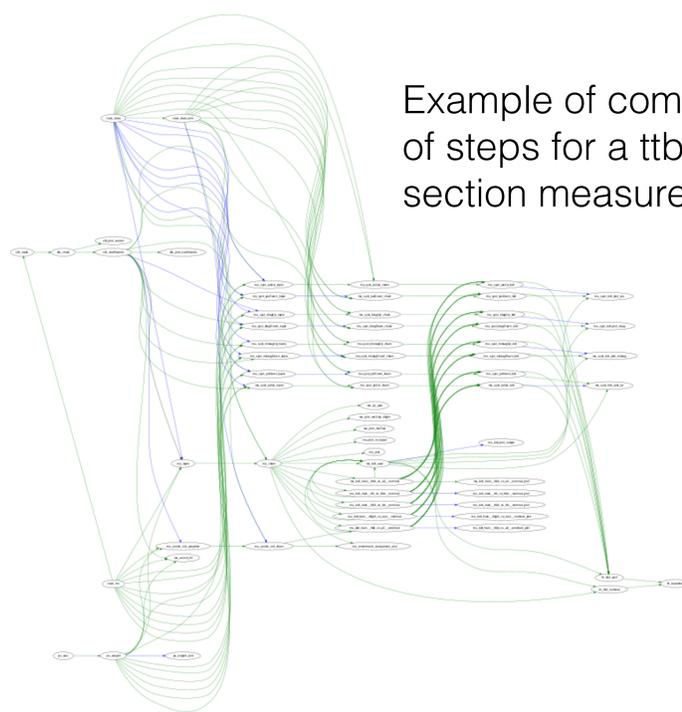
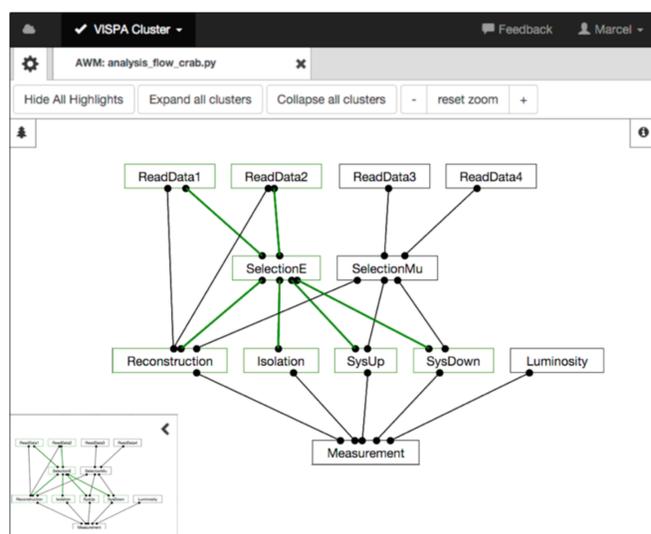
Execution Scheme

- Each step is executed in one or multiple tasks
- Scheduler determines unfinished tasks
- Unfinished tasks are send to run location, e.g. WLCG site or local computer
- Software environment is setup specific to step and run location
- Tasks store metadata and provenance info in JSON task reports
- IO data is retrieved from and stored to reachable storage locations, e.g. dCache



Workflow Visualization

Visualization of steps and their connections on the VISPA Internet Platform (Poster 315)



Example of complex graph of steps for a ttbb cross section measurement