Outline for Hall D Document for the 2021 Computing Review

1. Reconstruction

- Facilities
    - JLab farm
    - HPC Centers
        - Pittsburgh
        - NERSC
        - IU
    - Usage history since the last review
        - List run period = campaign, dates of processing, number of events at each site, CPU used at each site in millions of core-hours (MCH) at each site.
    - History of implementation
        - Legacy Python scripts for JLab farm
            - Interacts with SWIF
            - Job-tracking database
        - SWIF2
            - Adds capability of staging input files to remote sites
                - Manage disk space at HPC site
                - Enhancement to work on multiple files in on a single node to get better throughput[?]
            - Job-tracking database

2. Analysis

- Analysis Launch: take reconstructed data and create root trees for individuals to do physics analysis
- Structure:
    - All reconstructed data for given run period for each launch
    - Selected reactions per launch
        - Reaction specified by a text string
        - Show examples from web page
    - One-to-many: one run period, several reactions
    - Variation from launch-to-launch in a 3-D space
        - Run period
        - Set of reactions
        - Software version
    - Root trees produced, one per run
        - Processing jobs are by data file (about 200 files per run)
        - Trees merged in post processing

- Same set of scripts as for reconstruction to drive process
- Processing done on JLab farm
    - One to two weeks
    - It is I/O intensive[?]
    - Tape latency can dominate elapsed time[?]
- Statistics
    - Number of launches
    - Number of reactions
    - Data volume

3. Simulation

- Geant4 simulation in productions use
- Bulk of simulation runs on OSG
    - Large input files are not needed
    - Package to manage submissions: MCwrapper
        - Web page interface
        - Command line interface
        - Job-tracking database
        - Python/javascript/php/shell-script
- Statistics
    - Number of jobs
    - CPU time used

4. Future Computing Needs

- List the run periods
- List the CPU resource need per run period
- List disk resource need per run period
- List tape resource need per run period

5. Ideas for the Future

- data catalog
    - data validation
    - Needs to be coupled to workflow manager
- reconstruction on osg
- better continuous integration
- better comprehensive test
- unit testing
- Documentation
    - impacts students
- workflow management

    - swif is big improvement over what we had before
    - need workflow manager that is coupled to data catalog
- Reconstruction on the OSG
    - Effort started
    - Must be managed by expert
    - Raw data must be staged from university cluster