Strategy for Computing Support for RHIC Experiments

James Dunlop

A Reminder from Last Year

- Priority choice, beginning in FY2016, to focus on experimental upgrades at higher priority than RCF capacity
  - Resources needed for iTPC and sPHENIX
  - Decision: Constrain RCF to constant capacity replacement level
  - Impacted computing for experiments
- Led to two recommendations in 2016 PAC report
  - BNL Management should immediately commission a panel of local computing experts to evaluate the STAR needs for computing resources, review the current practices, investigate possible optimization of the use of the local RCF resources, and report within three months.
  - BNL Management must look into the potential availability and use of additional computing resources within the laboratory, such as the Institutional Cluster within the BNL Computational Science Initiative, or seek additional computing resources, to enable timely reconstruction of STAR data.

Recommendation 1

Panel formed in July 2016, met in late August.

Members: E. Lancon (RACF), M. Begel (ATLAS), J. Lauret (STAR), J. Porter (NERSC/ALICE), C. Pinkenburg (PHENIX), J. Velkovska (PHENIX)

Charge:

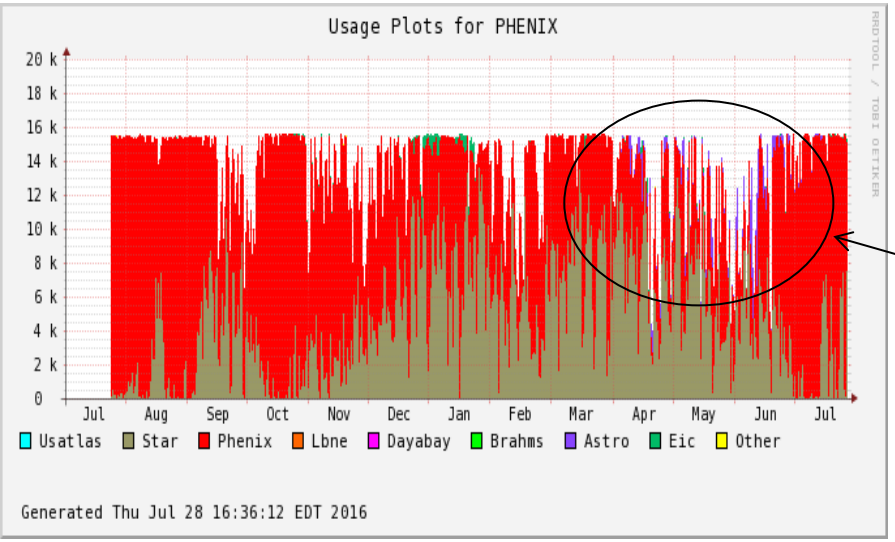
1. What are the STAR needs for computing resources from FY16-FY18 under the latest recommendations by the BNL NPP PAC for RHIC Runs 17 and 18?
2. What are the current practices for data reconstruction and analysis, both within STAR and in the larger context of the RCF, and how do these impact the reconstruction of STAR data?
3. What are the current practices for data reconstruction and analysis, both within STAR and in the larger context of the RCF, and how do these impact the reconstruction of STAR data?

3 action items identified to make progress:

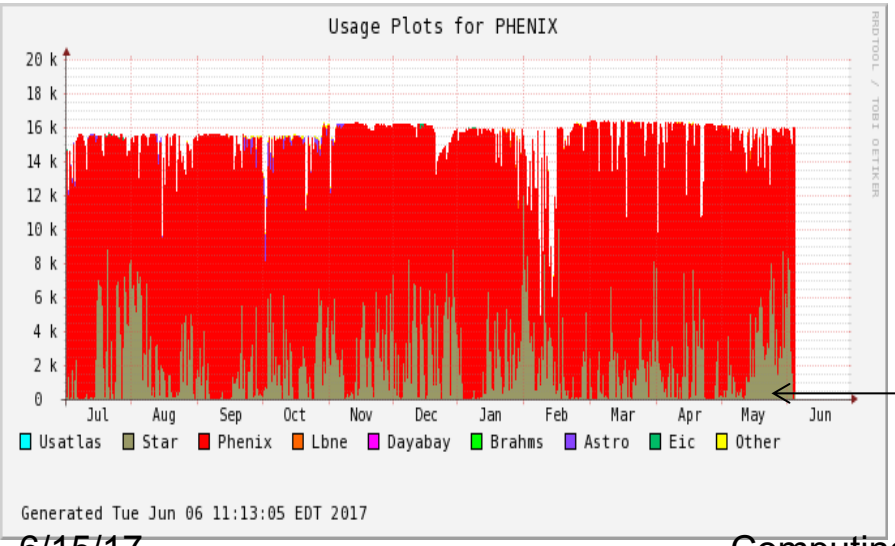
1. Resolve technical issues preventing running of STAR production jobs with PHENIX resources
2. Expeditiously implement the STAR picoDST, to reduce storage needs
3. Fully investigate new possibilities for external resources that were not clear at the time of the PAC

Action item 1: RCF Utilization

2015-2016: PHENIX utilization



2016-2017: PHENIX Utilization



Issue from last year:

Opportunistic use of PHENIX resources by STAR ***analysis*** jobs was in place

BUT: Not fully efficient. No opportunistic STAR ***production*** jobs due to STAR job size and inability of CRS software to stage multiple times Conflict:

- PHENIX: rapidly insert jobs
- STAR: long-running production jobs

Technical barriers overcome in Summer 2016 after discussions between RCF, STAR, and PHENIX.

In the end, only small increase in resources to STAR possible, as PHENIX utilization was high in the past year

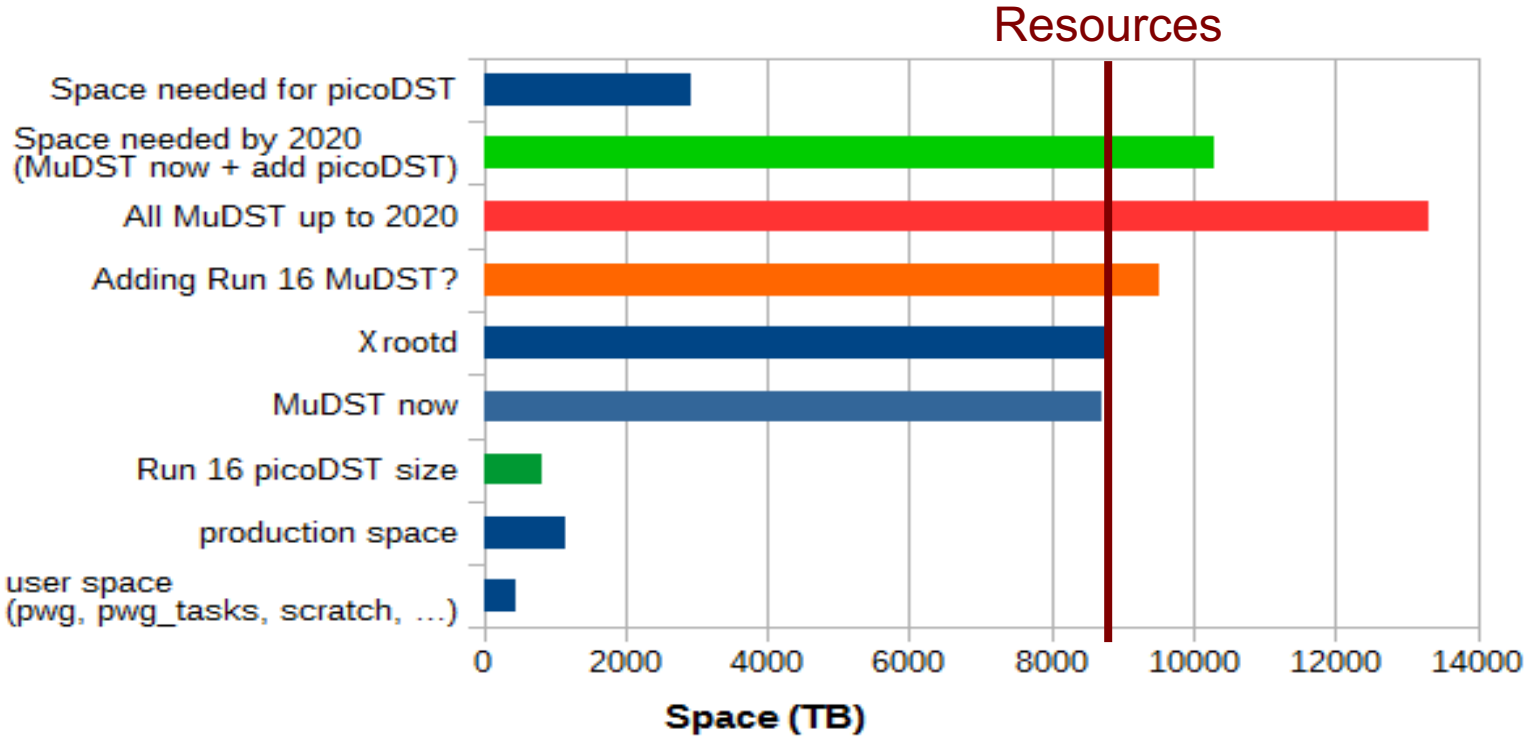
- Developed, tested, but not used
  - Workforce more effective elsewhere
- Analysis jobs harvested 17.25%
- Other short jobs, such as data format conversion, can also work

PHENIX Needs

- PHENIX is actively producing the Au+Au VTX data
  - Storage and staging issues have led to a shift from the “analysis taxi” approach towards a more coordinated approach, due to constrained storage resources

- While not a direct subject of last year’s recommendations, Hippocratic policy taken with regard to PHENIX
  - No reduction in dedicated resources
  - No lengthening of queue length, due to the need for rapid insertion of jobs in the PHENIX computing model.

Action Item 2: Storage Issue



STAR analysis model emphasizes flexibility for the analyzer

- Any user can look at any data at any time, if on disk
- Past attempts to buffer from HPSS unsuccessful: too much stress on tapes
- Data is either chosen by collaboration to be on disk or not available

Requirements under this model far outstrip storage resources

- Active prioritization: can lead to some analyses delayed

Action Item 2: picoDST Solution

- Principle: keep flexibility of STAR's analysis approach

- Solution: reduce amount of information stored
  - STAR's analysis storage format "MuDST" >~10 years old
  - Opportunity to reduce requirements by reducing information stored
  - Choice of what information to include must be driven by the collaboration and those doing analysis
  - Global exercise within the collaboration in Fall 2016, converged by Spring 2017

- Result: Factor 6.5 reduction in needed space
  - Current storage can handle projected needs to FY20
    - Requires conversion of past data to picoDST
    - Discussion of conversion priorities have begun

Action Item 3, Rec. 2: External Resources

"or seek additional computing resources"

- Dubna: 1000 slots, +7% if all used
  - Proof of principle test successful up to 500 slots: 97% efficient
  - Plan to shift specific, self-contained datasets to Dubna
- CORI at NERSC: +25% this year
  - Novel use of High Performance Computing (HPC) resources for data-intensive analysis
  - Proof of principle run in the fall, reported at CHEP 2016
    - CPU efficiency 99%, end-to-end workflow efficiency of 95%
    - Tested complex simulation workflows (embedding): CPU efficiency 95%
  - Led to allocation of 25M CPU hours for large-scale production
    - ~ ¼ CPU*year of RCF resources
    - After some technical work to establish full scalability, production has begun

Resources vs. Need

Stream	Size [TB]	Comments
W	70	Sivers Asymmetry of W
fms	300	Drell-Yan. No TPC tracking (fast).
mtd	700	Charmonium heavy ion baseline
rp	1000	Diffractive physics with Roman Pots
physics	2800	Jet studies (transversity)

- All runs previous to Run 17 finished within 5 months
  - "STAR is unlikely to suffer major delays in delivering datasets from Run 17 flagged as high priority"
- Run 17 data well structured, enabling prioritization
  - Would enable external resources to focus on specific topics
    - e.g. 2011 W production in the cloud

    http://www.nersc.gov/news-publications/nersc-news/science-news/2011/magellan-tackles-mysterious-proton-spin/
- Runs 18-20 (through BES-2) well under control
  - Under current plans "the out years are light on resource demands and do not seem to raise significant computing resource challenges"

A note on Simulations

- Unlike LHC experiments, simulations for RHIC experiments have not historically driven computing needs
  - STAR and PHENIX: ~10% of total (external+local) resources used for simulations
  - STAR simulations rely heavily on external resources
    - Grid resources for pure simulations for years
    - Embedding has primarily run on external resources
    - STAR can run all simulation and embedding workflows on any resource, including CORI
  - sPHENIX has used Grid resources for specific campaigns
    - https://www.opensciencegrid.org/using-open-science-grid-to-prepare-for-the-next-big-thing-at-brookhaven/
- Design stage of detectors (sPHENIX, EIC)
  - Increased need for resources and flexibility for rapid resource spin-up
  - Design framework to make most flexible use of all resources
  - Likely need to continue to rely on external sources for large campaigns

Towards the Future

- Program after BES-II will have high resource demands
  - High luminosity and high datarate, ~20 PB/year
  - Future detectors likely follow trend of LHCb, ALICE, CBM in mixing traditional “online” and “offline” needs into just “computing” needs
- Major shift in architecture in the next 5 years
  - Shift towards massively multi-core underlying technology
  - Large external investments likely in High Performance Computing
    - Not optimized specifically for our needs (IO, memory): we need to adapt
- Physics department in the process of investigating optimal structure to address these needs
  - Some cross-group benefits already happening (e.g. event display, experiment record handling, phone books). Further possibilities.
  - BNL in-house experimental computing knowledge competitive in field
  - Committee formed, report expected this summer