Report on the
sPHENIX Software and Computing Review

BNL, September 5-6 2019

Committee members: Alexei Klimentov (BNL, co-hair), Jérôme Lauret (BNL, co-chair), Christoph Paus (MIT), Brett Viren (BNL), Tony Wong (BNL)
Jamie Dunlop (ex-officio)

September 20, 2019

Revision delivered October 7th 2019

Executive Summary

A BNL review of the sPHENIX experiment DAQ, Software and Computing was held on Sep.5 - Sep.6, 2019 at Brookhaven National Laboratory, using an independent review committee of 5 members. This was the second review of sPHENIX DAQ, SW & Computing. The purpose of this review was to evaluate the progress to-date on the sPHENIX for data acquisition, data storage, computing and software development plans and provide advice and guidance on future planning, addressing in particular three points. We summarize here our conclusions on these points.

Are the resources required to transmit and store the data adequately understood?

Committee response : Yes

The technical plans and resource needs for transmitting and storing the data appear well understood and under control, drawing on deep experience, commodity components, and well established collaboration with RACF.

- Data rate estimates at its peak is 1.4 PB/day and the plan has moved to relying on offline storage transferred over LAN (managed by the RACF).
- Network bandwidth is well provisioned - 72 fibers (capable of 100 Gbit/sec each) will provide well over the required 135 Gbits/sec (1.4 PB/day) needed.
- Storage needs have been estimated

Are the resources required to process the data to a form suitable for physics analyses adequately understood?

Committee response : No

The work to develop technical plans and resource estimates for processing the data to a form suitable for physics analyses is in progress. The estimation that core SW&C team of 14.2 FTE will be needed appears correct and reasonable, but the staffing of SW&C team is largely incomplete. Overall, a resourced computing model developed with cost/benefit in mind, and a timeline of deliverables including confronting the system before commissioning with at-scale operation (e.g. via data challenges), will be required in order to have confidence that the resource requirements are understood and under control. Alternate computing models (data processing and data storing scenarios) should be further evaluated.

Is the plan for developing software processes and framework adequately understood?

Committee response : No

The plans for developing the software processes and framework appear well founded and fairly understood, The sPHENIX software team should be commended for their work to reduce memory usage. It is one of the vital topics for SW in the next year. At the same time a more detailed labor plan, priorities and risk assessments should be developed.

Given that the committee identified areas in which plans and resource needs are not yet adequately understood, and are subject to large uncertainties that should be clarified over the next year, we would recommend a follow-up review in summer 2020.

The panel thanks the sPHENIX project for excellent presentations, documentation and for being responsive to our requests for information. While area of clarifications were identified, the committee was pleased to see much progress and follow-ups from the past review recommendations and an incredible amount of work being done to design and commission sPHENIX Software & Computing.

Committee Report

Here follows the committee's comments and recommendations across the areas covered by the review, organized as Findings, Comments and Recommendations. ***Findings*** are significant points of information highlighted by the committee as relevant to the charge. ***Comments*** are inferences and conclusions drawn. ***Recommendations*** suggest a particular action or direction, often in light of a finding and comments.

General, planning

Findings

1. sPHENIX has identified 9 physics topics, many of them are complementary to the LHC Heavy Ion program.
2. sPHENIX has had several successful reviews in the past year.
3. Several Task Forces (TF) were organized to address critical topics. Some of the TFs were organized together with ALICE
4. New SW group is set up in BNL PH department (NPPS) to support NP and HEP experiments. sPHENIX is part of the NPPS leadership team.
5. sPHENIX has initiated discussions with CSI experts related to new computing architectures and HPCs.

6. Much of the success stories (analysis train, tutorials, seamless framework) from PHENIX are being reused in the sPHENIX era.
7. The present computing model envisions to store and process all sPHENIX raw data at BNL.
8. Discussions with the RACF on feasibility of the approach considered has begun.
9. The committee was informed that the RACF will migrate to a new data center in 2021 due to infrastructure constraints (floor space, power and cooling) in the existing data center.
10. Cost/benefit estimates for a reasonable number of scenarios were not presented.
11. The estimated core software effort is 14.2 FTE, staffing of SW&Computing team is largely incomplete. It was not clear to the committee which FTE's would come from the collaboration (what growth or support model the collaboration has in place) and which would come from BNL (sPHENIX, NPPS, CSI, …).

Comments

1. The sPHENIX team has successfully addressed the bulk of earlier recommendations.
2. There is good synergy between the DAQ and the offline SW teams.
3. sPHENIX has approached CSI but the effort to create collaborative efforts seem modest and peripheral (participation of workshops, initial discussions but no comprehensive survey of where and how CSI could contribute).
4. sPHENIX will benefit from NPPS expertise, a staffing plan for S&C is needed.
5. 2020 MC simulation campaign is well planned and rely on RACF resources.
6. Resource needs for simulation campaign were presented in a comprehensive manner. MC resources request is relatively small in comparison with overall computing budget.
7. We are pleased sPHENIX has taken a forward looking approach to address diversity in the collaboration.
8. As noted at the first review, growth and wide participation could also be drawn and secured with funding from remote institutions (for example via MoUs or an LHC-like model where participation is proportional to the number of authors). It was still unclear if a plan is in place but we understood that a review of the collaboration participation and effort is underway and that built-in mechanism to draw efforts exists. The statement was that the ongoing survey will serve as the basis for a gradual evolution toward a more formal system.

Recommendations

A.1 We encourage to continue close collaboration with other teams like ALICE, CBM and other collaborations to benefit from past experience and create synergy.

A.2 The RACF and the department should make it a priority to make collaborative tools available based on inputs and requests from sPHENIX and other groups.

A.3 Investigation of possible synergies with CSI should be re-evaluated and increased.

A.4 Work together with the RACF on cost-benefit analysis (for the computing resources) is needed at this stage of maturity of the plan. sPHENIX and RACF need to carefully coordinate timeline of the new data center and availability of compute and storage resources.

A.5 sPHENIX needs to formalize the process of how and to what level and when collaborating institutions contribute so that external resources - hardware and workforce - can be included in the planning process and leveraged. This will strengthen the collaboration planning and negotiations with BNL. Understanding the process may be dynamic, a more mature support model should be presented at the next review.

A.6 The Computing planning needs to be further developed and delivered by January 2020.

a. Priorities, risks and time required for each identified task should be done.
b. Develop (together with NPPS and Physics Management) a S&C staffing plan.

A.7 We recommend again to BNL management that the laboratory encourage and catalyze collaboration between CSI and experiments in the Physics Department. We quote our past recommendation in that regard: "*Cross-department project teams that aim to boost the capabilities and functionalities of experimental software (projects of broad interest and impact) are highly desirable."* - we suggested to leverage LDRDs though PDF (Program Development Funds) should also be considered.

A.8 We recommend to schedule the next review in the *summer of 2020*

Data acquisition, data transfer, data storage and data processing

Findings

1. The TPC dominates the data volume and CPU requirements, tracking dominates reconstruction time.
2. C-AD has been engaged regarding the beam crossing angle. A model has shown guidance for optimization. Tuning and validation is planned.
3. Data streaming has been tested in the 2019 test beam as a proof of principle.
4. RAW data are collected, stored and archived in separate files for a given fraction of the sub-detectors. Event building will be done on the fly during each pass of the reconstruction.
5. With the current planned resources, the raw data cannot be streamed simultaneously into and out from HPSS. It is assumed that HPSS performance will be a limitation for simultaneous raw data writing and reading.

Comments

1. In our opinion there are potential risks related to the way how RAW data are stored.
2. The TPC distortion corrections (calibration) is a critical point with large risk and

uncertainty.

Recommendations

B.1 Perform a cost/benefit and risk analyses for three event-building scenarios: (a) online, (b) offline before tape storage and (c) on-the-fly for each reconstruction pass.

B.2 We recommend to consider an approach where the first pass reconstruction (asynchronous) is later efficiently deleted from HPSS.

Simulations overview & calorimeter performance

Findings

1. QA procedure for simulation to detect changes and divergence from reference are in place.
2. Simulation of space charge is being worked on with ALICE but not yet implemented in sPHENIX simulation workflows.
3. Calorimeter is the driver in CPU usage for simulation.
4. The need for fast simulation (using approaches like caloGAN) is being considered but not for the immediate future.
5. Simulation CPU and storage requirements are modest when compared to real data.

Comments

1. Plans, numbers and priorities in this area seem adequate and well thought out. Resource needs for simulation campaign were presented in a comprehensive manner. MC resources request is relatively small in comparison with overall computing budget.
2. The 2020 MC simulation campaigns rely on RACF resources though at first glance, the initial target of 60k CPUs needed for 10 weeks seemed to not be fitting within BNL's planning. Later requirements precisions relaxed to 30k spanning a longer timeline (12k cores coming from the PHENIX resources would serve as the base) seemed more realistically achievable.
3. External resources (NERSC, ORNL, LLNL, OSG, and XSEDE) were thought of and noted but no firm plan or commitment seemed established. It was our understanding that an effort to evaluate the use of external resources will be soon underway.
4. Workforce has been identified as requiring specific Physics expertise and the committee felt those FTEs (for example, TPC calibrations and slow simulator requiring Physicists with domain-specific knowledge) were not well suited for a generic workforce pool. We

would not recommend a NPPS hire for knowledge typically held by Physicists but to identify or hire or retain physicist to carry those tasks within the collaboration.

5. However, the committee also felt that expert computing knowledge for activities such as code optimization, vectorization, multi-threading, data model (not yet addressed) are natural targets for NPPS' help (where knowledge likely already exists).
6. We understood that at least one of the two FTEs noted for the TPC specific tasks has been identified from within the collaboration.

Recommendations

C.1 Revisit the TPC distortions plans and requirements together with teams having a TPC based experiment and long expertise. Example: experts from BNL (STAR) and CERN (ALICE) should review and comment on sPHENIX approach.

C.2 Consolidate the simulation requirements and implementation needs in regards to treatment of distortions, misalignment and other considerations that may introduce uncertainties in the estimate of efficiencies. The timeline for such implementation should be clearly justified.

C.3 Fast simulations should be further investigated. While simulation constitutes a small portion of the total CPU budget, they may provide a path to fast evaluation / analysis.

C.4 The sPHENIX collaboration should identify from within the collaboration the individuals with the knowledge (or hire & grow individuals to have such knowledge) needed to support TPC domain-specific tasks (calibrations, detailed simulations) and furthermore, have their retention in mind.

Framework, software development, calibration and reconstruction

Findings

1. Much of the success stories (analysis train, tutorials, seamless framework) from PHENIX are being reused in the sPHENIX era.
2. sPHENIX uses one common online, offline and analysis computing framework which conserves developer effort.
3. Containers and CVMFS are now in use in the collaboration and is seen as a step toward a possible distributed computing workflow.
4. Average memory usage has been reduced from 10 GB/event to 4.5 GB/event with a wide spread.
5. Calibration aspects are better understood and folded into the development plan and calibration strategies.
6. New seeding schemes were evaluated and implemented. They reduced the CPU needs.

7. The RAW data reconstruction plan relies on a cycling buffer of 20 PB size disk cache (full copy still in HPSS) which roughly corresponds to 2 weeks depth at full data rate (10% of the annual data).
8. Tracking efficiency is 70% @ 9 sec, with the goal to reach 90% @ 5 sec.
9. The reconstruction passes consist of one synchronous pass (for TPC calibration) and 2 (asynchronous passes) with TPC reconstruction and TPC+calo, respectively.
10. With the current planned resources, the raw data cannot be streamed simultaneously into and out from HPSS. It is assumed that HPSS performance will be a limitation for simultaneous raw data writing and reading.
11. The DST format is still under discussion and postDST formats are only envisioned.

Comments

1. All CPU estimates presented were based on current CPU technologies - the projections to outer years did not apply Moore's law corrections.
2. Running in a container is an important initial step towards using non-BNL resources. However, the approach to distributed computing still appears not well thought out.
3. It is not clear if synergies are possible with the EIC Software Development efforts ongoing at BNL and beyond. Framework components and approaches as well as experience and best practices from that community could be yet another resource for effort saving.
4. While tools such as Rucio was noted as targeted for adoption, it is not always clear if workforce is thought of for testing and integration within the sPHENIX context. Adoption will not come free of effort.
5. The implementation of a multi-threaded framework does not appear to be on the horizon for sPHENIX. Discussion and initial investigation revealed that there is little identified shared memory and the collaboration is likely to make MT a low priority.
6. The envisioned reconstruction workflow, relying on multiple input files, raised operational concerns from the committee.
7. The calibration efforts seems well on track and sPHENIX have clearly benefited from the ALICE experience and processes. We were pleased with the rapid progress in this area.
8. Track finding efficiency has room for improvement (90%).
9. The committee inferred that the RACF will provide the bulk of the analysis resources. Analysis is seen as a small fraction of the total resources needed.
10. The committee recommended in the past a cost/benefit analysis under several scenarios. Cache versus more HPSS resources for example. It did not sound like such optimization was carried out and supported the statements we were presented.

Recommendations

D.1 At this stage of maturity of the plan, we highly recommend to work together with the RACF on computing resources plans that would fold in cost decrease and CPU performance increase (Moore's and Kryder's law scaling) - projections should be updated accordingly.

D.2 Continue work related to DST and postDST format definitions - this will soon be needed.

D.3 The work to improve track finding performance to satisfy the goal of 90% should continue.

Appendices

Appendix A - Charge

Memorandum from James Dunlop

sPHENIX Software & Computing review September 5-6, 2019

Charge to the Review Committee

The sPHENIX detector, currently under development, is designed to facilitate large acceptance, ultra-high rate measurements of fully reconstructed jets and high resolution spectroscopy of Upsilon states at the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory (BNL). The experiment is aimed at addressing scientific questions prioritized in the 2015 NSAC Long Range Plan and generally enhancing the physics reach afforded by the RHIC complex prior to the possible construction of an Electron Ion Collider (EIC).

The plans for DAQ software, data storage and other computing aspects are at present at an early stage. The committee is charged to review the software and computing aspects of the project and provide advice and guidance on future planning:

1. Are the resources required to transmit and store the data adequately understood?

2. Are the resources required to process the data to a form suitable for physics analyses

adequately understood?

3. Is the plan for developing software processes and framework adequately understood?

A report should be submitted to my office by close of business on Monday Sept. 16, 2019.

I very much appreciate your willingness to lend your time and expertise in this important process and look forward to receiving your assessment.

James Dunlop
Associate Chair for Nuclear Physics,
Physics Department Brookhaven National Laboratory

Appendix B - Review schedule and materials

Review materials provided by the sPHENIX team for the review were available at :
https://indico.bnl.gov/event/6702/