The Off-line Computing System for the PHENIX Experiment at RHIC

David P. Morrison, Han Geng, Qihang Liu, Li Qun, Mark Pollack, Kyle Pope, Victor Perevoztchikov, Soren P. Sorensen, and Yan Zhao

University of Tennessee, Knoxville, TN 37996-1200, USA

Phenix is one of four experiments currently under construction at the RHIC accelerator complex at Brookhaven National Laboratory. The experiment will study very high energy collisions of heavy-ions with the aim of demonstrating the existence of a quark-gluon plasma and investigating its properties. The experiment will collect 350 Terabytes of raw data annually and will require close to 150 GFlops of continuous CPU power to perform all needed computational functions. The Phenix computing model specifies that the bulk of the data will be stored in a hierarchical storage system at the central RHIC Computing Facility (RCF). Strong emphasis will be put on a storage architecture that will permit database-like queries of the physics quantities. The RCF will also be the site of most of the initial event reconstruction, later physics analysis, and data mining. Plans are also underway for the creation of a major regional computing center in Japan which will be responsible for all simulation tasks and performing CPU intensive event generation calculations.

Key words: Off-line Computing System, Computing Model,

1 Introduction

The PHENIX experiment is currently under construction at the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory. Data taking is scheduled to begin in the Fall of 1999. The main physics goals of PHENIX is to study hot and dense nuclear matter and in particular demonstrate the existence of the Quark-Gluon Plasma (QGP) and study its properties. PHENIX is focusing on the use of penetrating probes like electrons, muons and photons in order to be able to study the early phases of the nuclear collision, where

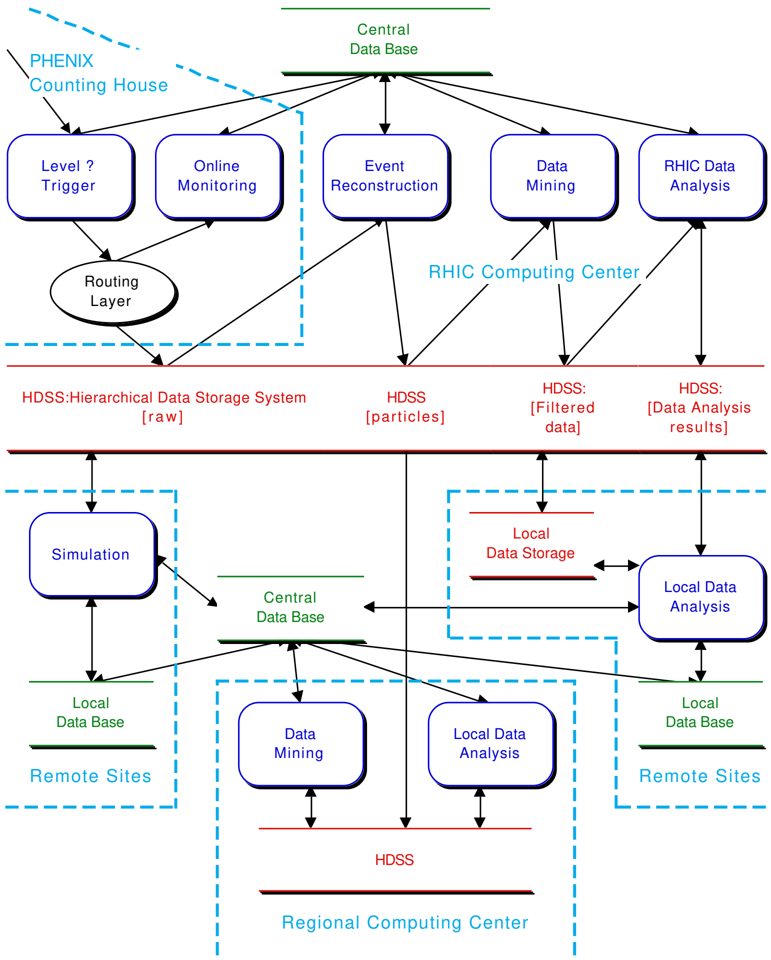

Fig. 1. A pictorial view of the PHENIX Computing Model. Please refer to the text for further details.

the likelihood of creating the QGP is presumed to be highest. The PHENIX collaboration consists of more than 400 physicists from USA, Japan, Russia, China, Canada, Germany, Sweden, India and several other countries.

2 The PHENIX Computing Model

All the main ingredients in the PHENIX off-line computing system and their relations to each other is summarized in our computing model as shown in Fig. 1.

The raw data from the DAQ will be send through a routing layer (most likely a powerful workstation) directly via a high-speed network to the hierarchical data storage system (HDSS) at the RHIC Computing Facility (RCF). The routing layer will also send data to a multitude of monitoring tasks running on workstations in the counting house. These monitoring tasks will use the same

Table 1
Overview of PHENIX estimated requirements for CPU.

	CPU		Location
	GFlops	kSPECint92	
Event Reconstruction	70	175	RCF
Data Mining	10	25	RCF
Data Analysis	20	60	Primarily RCF
Simulation	25	75	RCF & Japan
Models	25	75	Japan
Total On-site	115	305	RCF
Total Off-site	35	105	Japan
Total	150	410	

software, the PHENIX Event Processor (PEP), as will be used for all off-line tasks. All simulated data, both model calculations and detector hit generation, will be generated at a regional PHENIX computing center in Japan.

The event reconstruction will be performed on the main PHENIX Event Reconstruction Cluster at the RCF by reading events from the HDSS after calibrations have been performed. It is assumed, that after an initial start-up phase, that event reconstruction will be done with only a few hours time delay after the data collection. The event reconstruction layer will output calibrated data and higher order objects like hits, track-stubs, tracks, calorimeter clusters, RICH rings, identified particles etc. to the HDSS in a format suitable for the subsequent data mining.

The data mining stage will permit any user to search through all the data and create datasets suitable for their use. This stage corresponds to the production of micro-DSTs in other collaborations, but we find the data mining concept interesting, since it allows each user to basicly create their own micro-DSTs. The datasets from the data mining will either be stored in the central HDSS or be transferred to remote sites for further analysis.

The final data analysis will be performed both at RCF and at remote sites. It is currently assumed, that 50-75% of the PHENIX users will use the central I/O intensive cluster for this analysis and will only send X-window traffic over the network.

Table 2
Overview of PHENIX estimated annual requirements for data storage and distribution among storage media. It is assumed, that a tape vault will be the final storage medium for the data and that data will not accumulate on the disks or in the robots.

Type of data	Annual requirement TBytes	Storage Medium Disk	Robot
Raw data	350	1	12
Reconstructed data	175	1	8
Analysis data sets	100	5	10
Simulated data	150	1	5
Model data	10	1	5
Data base	0.1	0.1	
Total	≈ 800	≈ 10	≈ 40

3 Computing Requirements

Even if the PHENIX detector is of a similar size as several current HEP detector the computing requirements are much larger than any current HEP detector. This is caused by the very high event multiplicities, which can reach 10,000 charged particles in the most central Au+Au collisions. In this environment tracking becomes extremely complicated, forcing us to use very CPU intensive algorithms. The large event sizes (>300kBytes) combined with the expected high annual duty cycle (> 40 weeks) of the RHIC accelerator

The CPU requirements, as shown in Table 1, are totally dominated by the difficult pattern recognition during the event reconstruction. Tests on a variety of platforms have shown that our tracking code scales very well with the SPECint performance of the platforms. The estimated uncertainties in these values are at least a factor of 2.

Following the PHENIX computing model, we will need a substantial amount of CPU power in order to do the data mining after the event reconstruction. Here we have assumed, that a powerful shared memory machine around 10 GFlops will be able to serve the 5-10 concurrent users, who are performing data searches. This estimate is based on an extrapolation of the machines currently used by large private corporations for doing data mining.

The CPU requirements for the actual data analysis, where the results of the data mining is rescanned and investigated numerous times, is not expected

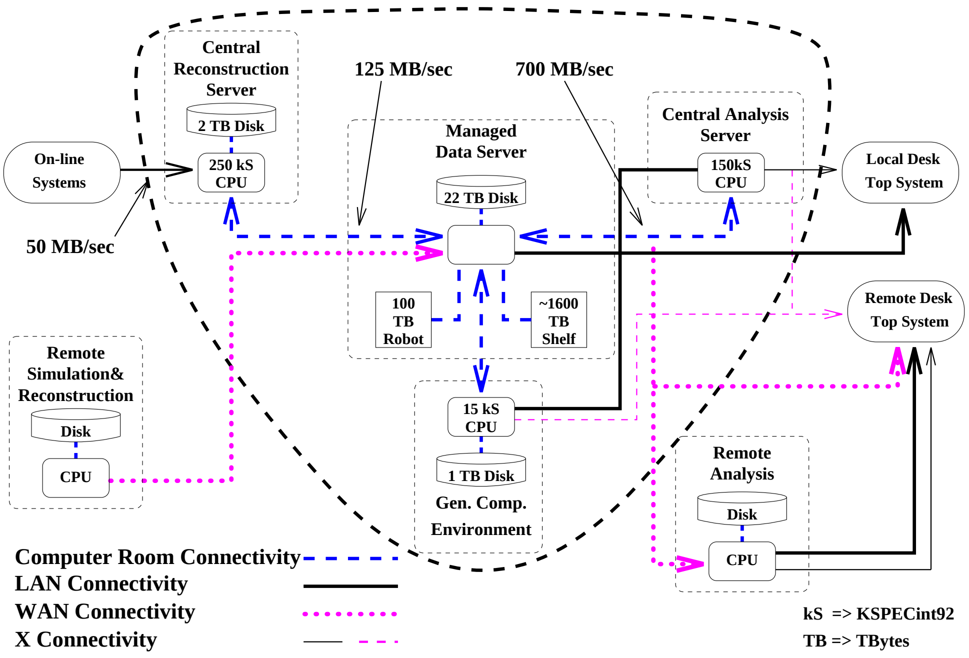

Fig. 2. The physical model of the RHIC Computing Facility at the time of turn-on of RHIC.

to surpass the requirements for event reconstruction. The main emphasis in PHENIX is on physics signals based on leptonic probes and since in general the number of detected leptons per event will be less than 10, the complexity of each event at this stage of analysis is limited. The most CPU intensive part of the data analysis will probably be the analysis of the hadronic sector, especially HBT and $\phi \to KK$.

Table 2 summarizes PHENIX data storage needs.

4 RHIC Computing Facility

As described above most of the actual hardware for the PHENIX off-line computing system will be part of the RHIC Computing Facility. RCF has recently received initial funding for FY97 and is now aggressively building up a large-scale system. Fig. 2 shows the layout of the system at the time of turn-on of RHIC [2]. This system will be shared between the two large RHIC experiments, PHENIX and STAR, and the two smaller, BRAHMS and PHOBOS. In analogy with the PHENIX Computing Model RCF will consists of 4 components:

The Central Reconstruction Server (CRS): is responsible for the reconstruction and calibration of all the event data.

The Managed Data Server (MDS): is responsible for storing and retrieving all data, both raw data, reconstructed data and various more advanced level of analyzed data.

The Central Analysis Server (CAS): will be used for analyzing most of the data by scanning and manipulating the data obtained from the MDS through data mining requests.

The General Computing Environment (GCE): will serve as an user interface to the other component and for the many more mundane tasks the user will have to perform: email, report writing, database browsing etc.

5 Current Technology Choices

Through the design phase we have chosen to make our software as modular as possible in order to be able to switch to new and better technologies in the future. Currently we have chosen to work within the following environment:

Operating Systems: 5 flavors of UNIX (SGI, HP, Sun, IBM and DEC) and Windows NT.

Programming languages: All future user code in C or C++, but we still will have to carry along a substantial amount of legacy F77.

CASE: The BridgePoint CASE tool based on the Shlaer-Mellor Object-Oriented Analysis. This tool is currently only being used by the members of the core group.

Interprocess communication: PVM, but we are considering a switch to a CORBA implementation.

GUI: Tcl/TK

Database: The Off-line database is based on ORACLE. The online group is pursuing an OODB solution and we following their efforts in order to see if a change is advantageous.

6 Acknowledgments

This work was supported by the Department of Energy through the RHIC project and through contract DOE DE-FG02-96ER40982.

References

[1] B. S. Kumar *et al.*, Off-line Computing at RHIC, http://www.rhic.bnl.gov/html/planning/rococo2/rococo2.html

[2] B. Gibbard *et al.*, RHIC Computing Facility, a proposal to DOE, BNL, June 24, 1996