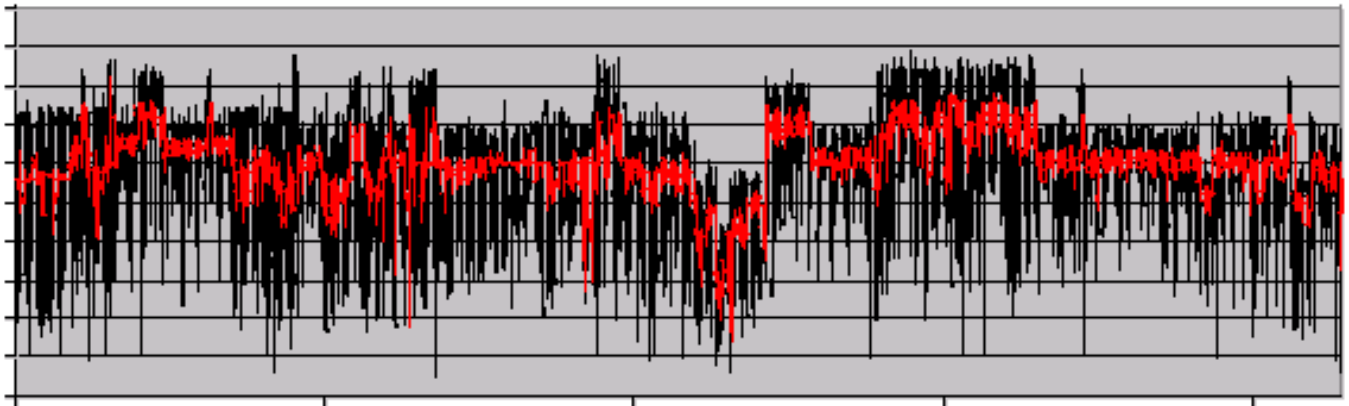


Grid Performance Workshop 2005



NeSC

June 22-23, 2005

www.mcs.anl.gov/~jms/GPW2005

Grid Performance Workshop 2005

<http://www.mcs.anl.gov/~jms/GPW2005>

The International Grid Performance Workshop - 2005 will address the science of performance and the Grid. It will be a forum for identifying and discussing the latest research and current application needs for performance and monitoring data for Grid systems, both in terms of design and use. The meeting will take place on June 22-23 at the NeSC in Edinburgh, UK. This event is part of the eScience Performance and Information Services for Smart Decision Making Event Theme at Edinburgh's eSI (www.mcs.anl.gov/~jms/ISTheme).

Applications are slowly being adapted to run over multiple administrative domains in a coordinated way, but they rarely achieve even a fraction of the possible performance of the underlying systems. In part this is because users very rarely know what performance they could achieve. There is no current data or infrastructure to formulate estimates of baseline performance with which to determine the difference between how an application is currently running and what is possible with some tuning. Following the success of last years workshop, this two-day working meeting will focus on the needs of three applications as use cases, and tools to help understand their problems, with the aim to extrapolate to other application use in the closing sessions. The result of the workshop will be a white paper on the state of the field and recommendations for future work.

Topics to be addressed include:

- What performance criteria are most important to users?
- How are the resources monitored – i.e. is the Grid up?
- How are failures detected – from job submission to file transfers?
- How do application scientists currently manage failures?
- What support tools are available for managing the performance of a Grid application, including
 - off-the-shelf tools
 - application-specific tools
 - what additional tools are needed?
- What role does simulation and modeling play in a Grid environment, and how can we verify its appropriateness?
- And the underlying themes of reproducibility, scalability, and usability

GPW'05 Steering Committee:

Chair, Jennifer Schopf, ANL and NeSC
Bill Gropp, ANL
Stephen Jarvis, University of Warwick
Mark Leese, Daresbury Laboratory
Brian Tierney, LBNL

GPW Advisory Committee:

Fran Berman, SDSC
Jack Dongarra, University of Tennessee
Ian Foster, ANL
Bill Gropp, ANL
Tony Hey, UK e-Science Centre
UK Finances, Mark Baker

Thanks

We gratefully thank NSF Division of Shared CyberInfrastructure and the Division of Computer and Network Systems, Microsoft, and the Joint Information Systems Committee (JSIC) for their help in supporting this meeting. We also thank Microsoft for their sponsorship of the conference dinner. We are also tremendously grateful to the whole NeSC crew for their assistance in local arrangements and planning.

AGENDA

June 22

9:00-9:30 Registration

9:30-10:00 Introduction and Overview
Jennifer Schopf, Welcome

10:00-12:00 Applications: Three talks by application scientists, focusing on performance
Andrew Porter, Reality Grid
Justin Burruss, Fusion Grid
Birger Koblitiz, LCG2

12:00-12:30 Applications Panel: An opportunity for tool developers to ask questions of a panel of applications scientists

12:30-1:15 Lunch

1:15-2:45 Themed Tool Talks: Is the Grid up?
Shava Smallen, Inca
Nick LeRoy, Hawkeye
David Colling, LCG

2:45-3:15 Coffee break

3:15-4:15 Themed Tool Talks: Why did my job/file transfer fail?
Brian Tierney, NetLogger
John Gurd, Manchester

4:15-5:30 Tool Panel: An opportunity for application scientists to ask questions of a panel of tool developers

OR

Breakout sessions looking at issues identified earlier in the day

AND

Start of collection of issues for final panel to address

6:30 Banquet

June 23

9:00-10:00 KeyNote: Simulation - it's role with applications and Grid performance
Henri Casanova

10:00-10:30 Coffee Break

10:30-11:30 Themed Tool Talks: How long will my file transfer take?
Matthew Allen, NWS
Les Cotrell, PingER

11:30-1:00 Themed Tool Talks: How long will my job take to run?
Seung-Hye Jang, TAMU, Prophesy
Daniel Rueckert , Imperial College
I-hsin Chung, Active Harmony

1:00-2:00 Lunch

2:00-3:30 Breakout sessions on specific performance topics:
Are there any disadvantages to the approaches used by the tools for measuring/predicting performance?
If so, which of those issues do the applications people need addressing, and what is the best way forward?

3:30-4:00 Workshop Summary

ABSTRACTS

Andrew Porter, Reality Grid

Title: Performance Considerations within RealityGrid

Abstract:

The focus of the RealityGrid project is the use of the Grid to facilitate the simulation of condensed-matter systems such as material surfaces, miscible fluids and macromolecules. The scientists doing such work typically have existing codes (written in a variety of languages) for doing the calculations and require access to powerful, parallel computing resources.

The RealityGrid project has provided a set of tools for use by the application scientists. These tools include functionality for launching calculations on GT2-based Grids, monitoring and interacting with a running job (including the provision of on-line visualization) and managing the checkpoints produced by a job or set of jobs. The latter functionality has been used to aid tasks such as parameter space exploration and job migration.

In this talk I will discuss the aspects of the project where performance is critical to the application scientists, the level of functionality currently offered and what additional features we would consider desirable.

Justin Burruss, Fusion Grid

Title: Grid Performance and Fusion Science

Abstract:

Fusion research seeks the development of an environmentally and economically attractive power plant. Fusion science is advanced through experiments carried out on fusion devices called “tokamaks” such as the DIII-D tokamak in San Diego, California. Fusion experiments are interactive—not “batch mode”—and require rapid data analysis. Furthermore, because fusion researchers are geographically dispersed, fusion experiments require scientists to work remotely, both to lead and otherwise participate in experiments. For these reasons, managing data analysis throughput and remote collaboration reliability are the chief performance concerns. As fusion data analysis moves off of the local network and onto distributed resources on the WAN, performance problems become trickier to diagnose and repair. Solutions to grid performance management are required.

Birger Koblitz, LCG2

Title: Catalogue access on the grid

Abstract:

I present the studies of the ARDA project on the access to metadata and file catalogues. Both services rely on databases as a backend and a fast access is very important to user applications, while security needs to be taken into account as well as to enable the user to make use of the capabilities of a database. I will present performance comparisons on the LFC and Fireman EGEE file-catalogues as well as the ARDA metadata service, which allows to directly compare SOAP and text-streaming access to a database. Finally some design considerations of the ARDA/gLite metadata interface are presented.

Shava Smallen, Inca

Title: Inca Test Harness and Reporting Framework

Abstract:

Running applications on the Grid remains challenging for users. This difficulty is in part due to knowing when Grid services and resources are up and available and possible inconsistencies in the available Grid software environment. Inca is a flexible framework for the automated testing, benchmarking and monitoring of Grid systems. It includes mechanisms to schedule the execution of information gathering scripts, and to collect, archive, publish, and display the results. This talk covers how Inca can be utilized to detect installation, configuration, user environment, and/or performance problems on Grids and how this information benefits users. I will also discuss how Inca is currently being used by the TeraGrid project, a NSF-funded Grid effort composed of nine sites across the United States. Finally, I'll conclude with a status summary of our next version of Inca.

Nick LeRoy, Hawkeye

Title: Keeping a Hawkeye on the Grid

Abstract:

With the increasing availability of grid computing technologies, we provide researchers with unprecedented access to computing resources. As the size of individual clusters and the number of these clusters that comprise these grids continues to grow, so do the associated maintenance costs. Advanced and automated tools are increasingly required to keep these resources functioning properly. In this talk I will discuss Hawkeye and how it can be used to monitor various aspects of computing grids.

David Colling, LCG

Title: Performance and the LCG

Abstract:

The LCG is an operational Grid currently running at 136 sites in 36 countries, offering its users access to nearly 14,000 CPUs and approximately 8PB of storage. Monitoring the state and performance of such a system is challenging but vital to successful operation. The standard monitoring tools are described and an analysis of the performance of this LCG presented.

Brian Tierney, LBL, NetLogger

Title: Techniques for Monitoring Large Loosely-coupled Cluster Jobs

Abstract:

Science and commerce have both experienced an explosion in the sheer amount of data that must be analysed. More and more compute clusters are being used to analysing these huge data sets. However, debugging and tuning clusters requires specialized tools. Current cluster performance tools are more oriented towards tightly coupled parallel applications. We describe how the NetLogger Toolkit methodology is more appropriate for loosely coupled cluster computing, and describe our new automatic workflow anomaly detection component. We also describe how this methodology is being used in the Nearby Supernova Factory (SNfactory) project at Lawrence Berkeley National Laboratory.

John Gurd, Manchester

Title: Use case scenarios for performance control of Grid-based metacomputing

Abstract:

A Grid-based performance control system, PerCo, has been developed as part of the "deep track" activity in the RealityGrid project. PerCo is capable of controlling the behaviour of component-based Grid applications at two levels. Behaviour in a single component can be altered by changing actuating parameters that affect the way that the component executes in its allocated environment; this is controlled by a component performance steerer (CPS) to which the component code is attached. Behaviour of the application as a whole can be altered by moving components from one execution environment to another; this is controlled by a central application performance steerer (APS). The PerCo infrastructure is fully implemented, and is capable of cleanly closing down components, transferring them to a new platform, and restarting them correctly in the new environment.

The question addressed by this presentation is: what can the PerCo system do that will improve the situation faced by a scientist using high end computing to pursue some

scientific objective by means of large-scale simulation? We use our contacts with the scientists in the RealityGrid (EPSRC) and IntBioSim (BBSRC) e-Science projects to develop practically useful use case scenarios that can help them to progress more rapidly with their scientific investigations.

Henri Casanova, UCSD

Title: Simulation for Grid Computing

Abstract:

In this presentation we discuss the need for well-established simulation practice and technologies for the purpose of conducting research in the area of Grid computing. After identifying the need for Grid simulation and placing today's state-of-the-art in perspective with that in other areas of computer science, we discuss two main issues: the generation of synthetic grid platforms for simulation purposes, and the simulation of applications on these platforms. We review today's main approaches and discuss the advantages and shortcomings of current technology. We discuss several open questions, such as those regarding the trade-offs between simulation speed and simulation accuracy, and we conclude with a description of recent efforts that attempt to answer these open questions.

Matthew Allen, NWS

Title: Comparing Performance Measurement Time Series

Abstract:

Testing link conditions is standard practice for active Internet users. Network administrators monitor link performance to detect disruptive behaviour, determine performance bottlenecks, and locate faulty routers. Wide area distributed applications developers use measurements to predict performance or minimize communication time between nodes. To this end, a large number of tools have been developed to measure link bandwidth. Iperf, netperf, NTTCP, and the Network Weather Service (NWS) all measure network performance through active network probing. These techniques, however, all differ slightly in their methodology, causing them to report different results. In this talk, we attempt to understand whether these different methodologies effect the measurements produced. In particular, we try to develop techniques to compare measurements produced by different tools.

Les Cotrell, PingER

Title: Forecasting Network Performance

Abstract: Predicting how long a file transfer will take, requires forecasting network and application performance. However, such forecasting is beset with problems. These include seasonal (e.g. diurnal) variations in the measurements, the increasing difficulty of making accurate active low network intrusiveness measurements especially on high speed

(>1 Gbits/s) networks and with Network Interface Card (NIC) offloading, the intrusiveness of making more realistic active measurements on the network, the differences in network and large file transfer performance, and the difficulty of getting sufficient relevant passive measurements to enable forecasting. We will discuss each of these problems, compare and contrast the effectiveness of various solutions, look at how some of the methods may be combined, and identify practical ways to move forward.

Seung-Hye Jang, TAMU, Prophecy

Title: Performance-Directed Resource Allocation

Abstract:

Grid systems provide vast compute and data resources to users for large-scale applications such as cosmology, ocean modelling and gravitational-wave physics. One of the major issues to be addressed with grids is that of resource selection. In this talk, we will quantify the advantages of using performance prediction versus load information for resource selection. Our work uses the Prophecy infrastructure to predict application performance on different sites. Prophecy uses historical data to generate analytical performance models for predictions. The quantification is based upon two case studies. The first case study involves a large-scale scientific application, called GEO LIGO, for which the experimental results indicate an average of 33% performance improvement as compared to a load-based method. The second case study involves a web-based, educational application, called AADMLSS, for which the results indicate an average of 10% performance improvement as compared to a load-based method.

Daniel Rueckert , Imperial College

Title: Performance in Medical Image Computing

Abstract:

In this talk we will describe why performance modelling and predication is crucial for many applications in medical image computing. We will focus on grid applications using image registration which are employed in computer-aided diagnosis and computer-assisted surgery.

I-hsin Chung, Active Harmony

Title: Towards Automatic Performance Tuning

Abstract:

Software today makes extensive use of libraries and re-usable components in order to speed development. However, libraries used by an application may not be performance tuned to the application's need. To address this issue, we developed the Active Harmony automated runtime tuning system. I will describe the interface used by programs to make applications tunable. I will also present the optimisation algorithm used to adjust

application parameters together with a library to expose multiple variations of the same API using different algorithms. In order to speed up the tuning process, our tuning server utilizes historical data. When the system to be tuned has numerous parameters, our system uses techniques to prioritise parameters and identify the relations among parameters to avoid trying unnecessary configurations. In addition, for homogeneous compute nodes, we use parameters replication to tune all nodes as one. We have successfully applied the Active Harmony system to commercial and scientific applications. The experimental results show that frequently there is no single configuration performs well for all kinds of workloads. The performance improvement cannot easily be achieved by tuning individual components for such a system. For an e-commerce web site, Active Harmony helps the system adapt to different workloads and improve the performance up to 16% (70% with reconfiguring node roles) and for a parallel plasma simulation code, execution time is reduced up to 70%.