Current and Future Generation Grid Technology

Summer School 2006

24th - 28th July, Bonn (Germany)

Version: 1.1 Date: 06.07.06
Organization: Rosa Badia, Stephan Springstubbe, Wolfgang Ziegler

Sponsors

IBM STG Infrastructure Solutions

Sun Microsystems GmbH

Abstracts of Lectures

1) Introduction to Grid Technologies

"Session: Next Generation Grid"

a) Domenico Laforenza: *"Grid Technologies and their evolution"*

Abstract:
This talk is an introduction to Grid technologies and on their evolution. The term Grid Computing was introduced at the end of 90s by Foster and Kesselman; it was envisioned as "*an important new field, distinguished from conventional distributed computing by its focus on large-scale resource sharing, innovative applications, and, in some cases, high-performance orientation*". Since then a lot of water passed under the bridge, and perhaps this definition needs to be a bit revised. Recently the CoreGrid Executive Committee reached an agreement on the following definition: a Grid is "*a fully distributed, dynamically reconfigurable, scalable and autonomous infrastructure to provide location independent, pervasive, reliable, secure and efficient access to a coordinated set of services encapsulating and virtualizing resources (computing power, storage, instruments, data, etc.) in order to generate knowledge*". This is a more modern service-oriented vision of the Grid that stems from the conviction that the great majority of complex software applications will, in the mid-long term, be dynamically built by composing services which will be available in an open market of services and resources. In this sense, the Grid will be conceived as a "world-wide cyber-utility" populated by self-∗ cooperating services which interact as in a complex and gigantic software ecosystem.

b) Stephen Davey: *"NextGRID – Architecture for Next Generation Grids"*

Abstract:
NextGRID is part of a major initiative to ensure that Europe is a world leader in Grid technology, central to the new knowledge-based economy. NextGRID's Objectives are to develop Grid architecture to support mainstream use and meet the needs of business users by addressing in particular security and economically viable business models.

The project will last for three years and will consist of integrating activities on architecture, development activities on Grid foundation services and protocols, dynamic federation, user interaction and representative applications, and the consolidation and standardisation of these enhancements.

The overall result of the project will be a collection of new architectural designs, key middleware components and application support mechanisms that will underpin the Next Generation Grid.

NextGRID's vision is an architecture for Next Generation Grids that will enable their widespread use by research, industry and the ordinary citizen thus creating a dynamic marketplace for new services and products.

"Session: Grid Principles I"

c) Hiro Kishimoto: *"Open Grid Services Architecture"*

Abstract:
The Global Grid Forum (GGF) has embraced the Open Grid Services Architecture (OGSA) as the blueprint for standards-based grid computing. "Open" refers to the process used to develop standards that achieve interoperability. "Grid" is concerned with the integration, virtualization, and management of services and resources in a distributed, heterogeneous environment. It is "service-oriented" because it delivers functionality as loosely coupled, interacting services aligned with industry-accepted Web service standards. The "architecture" defines the components, their organizations and interactions, and the design philosophy used. In this talk I will explain how a service-oriented grid fits with other leading-edge enterprise technologies and provide details of OGSA and its core capabilities, including progress and plans.

d) Alvaro Arenas: *"Grid Security"*

Abstract:
This talk introduces the different concepts and technologies relevant to trust and security in Grid systems. We analyse the principal security requirements following the traditional lifecycle of Virtual Organisations, including particular requirements such as authentication, authorisation, secure communication, confidentiality and distributed trust. We then examine the available technologies that meet these requirements.

2) Grid information, resource management and monitoring

"Session: Grid RMS and Monitoring"

a) Ramin Yahyapour: *"RMS and Scheduling for Future Generation Grids"*

Abstract:
While many application scenarios for Grids originally stemmed from High Performance Computing, future Grid systems will have to provide generic solutions for a much broader use. This includes, for instance, the consideration of more and different types of resources as well as the need for coordination and orchestration of resource usage for Grid jobs.
Here, many existing solutions in resource management and scheduling may not be sufficient. This presentation discusses current concepts in scheduling and highlights new requirements for future Grid

systems. In addition, an outlook will be given for possible solutions for future resource management and scheduling in Grids.

b) Simon Alexandre: *"RMS - HPC4U"*

Abstract:
Next Generation Grids will demand for a flexible negotiation mechanisms supporting various ways of Quality-of-Service (QoS) guarantees. In this context, a QoS guarantee covers simultaneous allocations of various kinds of different resources, such as processor runtime, storage capacity, or network bandwidth, which are specified in the form of Service Level Agreements(SLA). Currently, a gap exists between the capabilities of Grid middleware and the underlying resource management systems concerning their support for QoS and SLA negotiation. The goal of the HPC4U project is to create a software-only solution for a fault- tolerant cluster environment, providing reliability and the negotiated level of QoS. This talk will highlight the main architecture elements of the envisaged cluster middleware and present the current state of development as well as latest achievements.

c) Robert Piotter: *"Mobile networks and RMS – Akogrimo"*

Abstract:
The integration of Mobility into Grid scenarios imposes new challenges on resource management. The changing context of resource users and providers such as bandwidth of the access network or device capabilities requires new solutions. The virtualisation of network resources, the communication of context changes to application allow a more dynamic solution and enable the use of Service Level Agreements (SLAs) also in this volatile environments by a close co-operation between SLA Management, Execution Management Service and QoS Broker interfaces. This talk will provide an overview of these additional challenges, their impact on existing solutions and how Akogrimo has structured its end-to-end QoS framework.

d) Norbert Meyer: *"Grid middleware services in CoreGRID"*

Abstract:
Grid infrastructures are subject to frequent changes and disruptions in the service of a huge number of services they are built of. To manage such a dynamic system and its resources, online monitoring of the resources to determine their availability is required. Monitoring is an essential, even critical, part of any Grid infrastructure, helping to hide its complexity. Other important topics for supporting dynamic systems is checkpointing, required to minimize site effects of hardware and software failures. Complex job workflows represent another challenge, as the monitoring information must be almost synchronously gathered from many different sources and appropriately processed to provide a coherent view (state information) of the whole workflow and its components. The problem of managing user accounts becomes a non-trivial one in a distributed environment (especially in a Grid environment) that includes many independent sites and virtual organizations with hundreds or even thousands of nodes with allocated user accounts. Therefore the complexity rises significantly in dynamic systems, where are many of them in production environments.
The presentation will focus on results produced by the CoreGRID Institute on Grid Information, Resource and Workflow Monitoring Services, especially network monitoring, checkpointing, workflow, user account management and accounting.

3) Web and Grid Services

"Session: Grid Principles II"

a) Roger Menday: *"Web Services"*

Abstract:
Web services refer to a loosely coupled infrastructure of services accessed over the Internet for engaging in distributed computing on a global scale. The foundation of XML and the SOAP protocol promotes a highly interoperable framework. At the root the ideas are simple, although when viewing the collection of specifications and technologies forming the Web services stack, one could be forgiven for thinking otherwise. In this session we present an overview of important technologies for building Service Oriented Architectures, emphasising best-practices and real-world usage.

b) Roger Menday: *"Grid Services – WSRF"*

Abstract:
The Grid is modelled as a collection of resources which are created, addressed, inspected, discovered and managed. The Grid use-case was the main driver for the Web Service Resource Framework (WSRF) family of specifications. WSRF adds a 'stateful' characteristic to conventional Web services, by defining a WS-Resource. A WS-Resource has a name, properties, and lifetime management mechanisms, as do resources on the Grid. This session describes in detail the WSRF approach and specifications and is illustrated by examples of its usage within the UniGrids project.

4) Current Grid Platforms and Services

"Session: Grid Environments"

a) Thomas Rueter: *"IBM RMS and distributed Queuing in Grids"*

Abstract:
With the new offerings from like IBM Grid and Grow Express, Batch on Grid and IBM Virtualization Engine IBM is committed more than ever to enable clients to embrace flexible IT infrastructure solutions. This talk will cover the principles and some examples of flexible infrastructure solutions which uses Grid technology. The reference example will cover research engagements and commercial projects in the financial services sector.

b) Constantin Gonzalez: *"Distr. Resource Management with the Sun Grid Engine"*

Abstract:
A critical part of any grid infrastructure is its distributed resource management (DRM) middleware. A DRM matches compute job requirements to available resources, enforces policies set by administrators and manages the complete life cycle of a job in a grid. This session provides an introduction into the Sun Grid Engine DRM system. It covers basic installation, administration, policy setup as well as advanced scheduling features and accounting aspects of the software. See also: http://www.sun.com/gridware/ and http://gridengine.sunsource.net/

c) Wolfgang Ziegler: *"Resource Orchestration in Grids"*

Abstract:
This lecture will discuss why we need resource orchestration in Grids and the added value of this approach with respect to commonly used best effort approaches. Based on this we will identify the capabilities local resource management systems (schedulers, queuing systems) must bring along in order to efficiently support the orchestration process. The state of the art technology available for resource orchestration, both commercial and public domain, will be discussed and a number of examples will be presented. Finally, we will look in more detail into one orchestration service developed in a German project, which can be used to orchestrate arbitrary resources. Co-allocation of

compute and network resources will be presented as an example. We will discuss the negotiation process and the use of Service Level Agreements for reservation of resources in this approach. Remarks on the ongoing standardisation work in this area will conclude the talk.

d) Roberto Barbera: *"gLite and GILDA – EGEE"*

Abstract:
The Enabling Grids for E-sciencE (EGEE) project is operating the largest production grid infrastructure for e-Science existing in the world and is developing gLite, the next generation of grid middleware. During the lecture, EGEE in general and gLite in particular will be presented and some space will be reserved to describe how new interested users can test gLite by themselves using the GILDA t-Infrastructure, the EGEE training and dissemination service.

5) Grid Applications in Life Sciences

"Session: Grid Applications"

a) Marc Zimmermann: *"The WISDOM project – Grid Computing and neglected diseases"*

Abstract:
In the biomedical community, grid computing has initiated several projects on large scale in silico drug screening approaches. The project WISDOM (Wide In Silico Docking On Malaria, http://wisdom.eu-egee.fr/) was amongst the first projects in the public domain that made use of grid enabled in silico docking to simulate the interaction of potential drugs with target proteins. In silico docking enables researchers to simulate the interaction of potential drugs with a target protein. It is the first step of virtual screening, which is one of the most promising approaches to speed-up and to reduce the costs of the development of new drugs to treat diseases such as malaria. The first biomedical data challenge for drug discovery ran on the EGEE grid production service from 11 July 2005 until 31 August 2005. Entitled WISDOM, the challenge saw over 46 million docked ligands during that period. The WISDOM data challenge demonstrated how grid computing can help drug discovery research by speeding up the identification of potentially active compounds.

b) Jose R. Valverde: *"Life Science applications"*

Abstract:
Life Sciences have seen an exponential growth in data collection (and hence in their needs for data analysis) in the last decades. After the completion of the Human Genome and subsequent development of Genomics and related technologies the panorama is growing even faster, with hyperexponential growth calling at the door. Coping with this data has very specific requirements stemming both from its high-throughput nature and the historical methodology applied to Life Sciences problem solving. In this scenario, the Grid provides an affordable entry point for HPC that is allowing us to readily attack the most urgent problems, but we still need to go forward several steps to meet the upcoming demands for the foreseeable future: multidisciplinary combined approaches integrating 'classical' Grid technologies (like EGEE), web services based distributed development (like BioMOBY), tight security and encryption mechanisms and workflow technologies are just the first steps towards a fully functional Life Sciences infrastructure. In this talk we will start by laying out the needs arising in the Life Sciences, then we will daydream together about an ideal solution and finally we will take a walk of current initiatives to grasp the state of the art.

6) Introduction to UNICORE

"Session: Grid Practices I"

a) Morris Riedel: *"UNICORE – From Project Results to Production Grids"*
b) Morris Riedel: *"UNICORE RMS, Workflow"*
c) Morris Riedel: *"UNICORE Security, Interoperability"*
d) Morris Riedel: *"UNICORE Installation"*
e) Morris Riedel: *"UNICORE Configuration "*
f) Morris Riedel: *"UNICORE Command Line Interfaces & Plugins"*
g) Morris Riedel: *"Introduction to Unicore/GS"*

Abstract:
The UNICORE Grid-technology provides a seamless, secure and intuitive access to distributed Grid resources. It is used in daily production worldwide and serves as a solid basis in many European and International projects. UNICORE is available as open source under BSD licence at SourceForge. The CoreGrid Summer School is an excellent opportunity for Grid users, developers, administrators, and researchers to get an inside view of UNICORE and its future developments. The interaction of UNICORE with resource management systems, its workflow and interoperability capabilities, command line interfaces, and security issues will be explained in presentations. Furthermore, participants are able to make hands-on experiences with UNICORE by installing the server and client components as well as developing simple plugins through the UNICORE plugin API. The day closes with discussions, share of experiences and a sneak preview of the upcoming Web services-based UNICORE 6.0.

7) Grid Computing and the Globus Toolkit

"Session: Grid Practices II"

a) Jennifer Schopf: *"Globus Toolkit Introduction"*
b) Jennifer Schopf: *„GT4 Components I, II"*
c) Hamza Mehammed: *"GT4 Installation I, II"*
d) Hamza Mehammed: *"GT4 Configuration"*
e) Hamza Mehammed: *"GT4 Tests"*

Abstract:
The Globus Project is developing fundamental technologies needed to build Grids --persistent environments that enable software applications to integrate instruments, displays, computational and information. This paradigm has become common in many application communities that use distributed resources in a coordinated manner. This talk gives a brief overview of Grid and web services, and how those standards have grown to address current needs.
We then describe the Globus Toolkit, an open source software toolkit that includes services and libraries for security, resource management, monitoring and discovery, file transfer and data management. The current release, GT4, which is a reference implementation of the WSRF standard, is discussed in great detail. Future plans and open issues conclude the talk.

8) Grid Applications in Industry – "Life Sciences"

"Session: Industrial Experiences"

N.N.: ***„Oracle's Grid Technology"***

Abstract: coming soon!

a) Tassilo Eckerle: *"Anterio's MOAB – Clustering Grids"*

Abstract:
Moab Grid Suite, developed by Cluster Resources, Inc, is unlike other grid software products in that it is a feature-abundant metascheduler and a grid workload manager, that eliminates the political barriers associated with grids. Moab lets sites unify multiple clusters of heterogeneous hardware, operating systems, and resource managers into a single grid environment. Moab intelligently schedules all grid resources to automatically migrate jobs and data to the optimal cluster on the grid, while taking into account all needed variables such as network, political policies, job priority, licenses, etc. Additionally, Moab virtually eliminates political struggles by allowing each cluster in the grid to maintain autonomy, while still sharing resources within the grid.
Moab is extensively used in every industry including the government, academic, pharmaceutical, energy, weather, etc. Moab is the scheduling software trusted by many of the world's largest grids such as TeraGrid (the world's largest grid), CMA (China's largest grid), CDAC (projected to be India's largest grid), etc. Moab has been tested on grid environments of over 80,000 nodes running 50,000 jobs.

9) Visions of Future Grids

"Session: Visions and Discussion"

a) Daniel Mallmann: *"Vision of UNICORE – Unigrids and beyond"*

Abstract:
The UniGrids project developed a Grid Service infrastructure that is compliant with the Web Service Resource Framework standard and based on the UNICORE Grid software and the Open Grid Services Architecture. This infrastructure will be the base of the upcoming UNICORE 6. The UniGrids project influenced and adopted several standards and proposed atomic services that provide interoperability between different Grid systems, demonstrated for UNICORE and Globus.

b) Jennifer Schopf: *"Globus – Where Are We Going For the Next Ten Years?"*

Abstract:
The Globus Project will soon celebrate the Tenth birthday of it's initial use, and is currently the underpinning to many of the largest Grid projects internationally. This talk will briefly discuss the state of Globus, it's new open source contribution environment, dev.globus.org, and future plans for development, standards, and users.

c) Keith Jeffery: *"Vision of Next Generation Grids"*

Abstract:
In the first half of 2003 a group of high level experts was convened by the European Commission, Unit INFSO/F2, in order to produce a report entitled "Next Generation Grids, European Grid Research 2005-2010" [NGG1]. In this report, known as the NGG report, the experts pioneered the vision of the 'Invisible Grid', whereby the complexity of the Grid is fully hidden to users and developers through the complete virtualisation of resources, and sketched the research priorities underpinning the realisation of the Next Generation Grids. The NGG report 2004 [NGG2] identified additional requirements that arose in the light of one more year of experience by the experts working in the Grids domain. In particular, the shortcomings of existing Grids middleware were much better understood, and despite the development directions concerning OGSA providing some integration with service-oriented architectures, it was becoming clearer that applications in the Grids environment require a greater range of services than can be provided by the combination of currently evolving Grids middleware and existing operating systems. The report concluded that a new operating system environment was required to assure security, trust, performance and self-* properties , and indicated that the foundationware and middleware layers required more intelligence in order to provide the dynamics of the self properties required.

In the last quarter of 2005, Unit F2 "Grid Technologies" has reconvened the Next Generation Grids expert group in order to identify the gaps between the leading-edge of Grid technologies and the scientific and technological requirements necessary to evolve Grids towards the wider and more ambitious vision of Service Oriented Knowledge Utilities (SOKU). The work of the expert group can be seen both as a follow-up of the former NGG consultations on the future of Grid research, as well as a main contribution to the wider consultation process on the FP7-ICE Pillar on "Software, Grids, Security and Dependability".