

Workload Management



Massimo Sgaravatto

INFN Padova



Overview

- ✍ Goal: define and implement a suitable architecture for distributed scheduling and resource management in a GRID environment
 - ✍ Large heterogeneous environment
 - ✍ PC farms and not supercomputers used in HEP
 - ✍ Large numbers (thousands) of independent users in many different sites
 - ✍ Different applications with different requirements
 - ✍ HEP Monte Carlo productions, reconstructions and production analyses
 - ✍ "Scheduled" activities
 - ✍ Goal: throughput maximization
 - ✍ HEP individual physics analyses
 - ✍ "Chaotic", non-predictable activities
 - ✍ Goal: latency minimization
 - ✍ ...



Overview

- ✍ Many challenging issues :
 - ✍ Optimizing the choice of execution location based on the availability of data, computation and network resources
 - ✍ Optimal co-allocation and advance reservation of CPU, data, network
 - ✍ Uniform interface to different local resource management systems
 - ✍ Priorities, policies on resource usage
 - ✍ Reliability
 - ✍ Fault tolerance
 - ✍ Scalability
 - ✍ ...
- ✍ INFN responsibility in DataGrid






Tasks

- ✍ Job resource specification and job description
 - ✍ Method to define and publish the resources required by a job
 - ✍ Job control language (command line tool, API, GUI)
- ✍ Partitioning programs for parallel execution
 - ✍ “Decomposition” of single jobs in multiple, “smaller” jobs that can be executed in parallel
 - ✍ Exploitation of task and data parallelism




Tasks

Scheduling

-  Definition and implementation of scheduling policies to find the best match between job requirements and available resources
-  Co-allocation and advance reservation
-  Resource management

Services

-  Authentication, authorization, bookkeeping, accounting, logging,



Effort breakdown (mm)

	Funded	Unfunded	
INFN	216	184	400
DATAMAT	108	0	108
CESnet	72	72	144
PPARC	0	18	18
	396	274	670



Workload Management in the INFN-GRID project

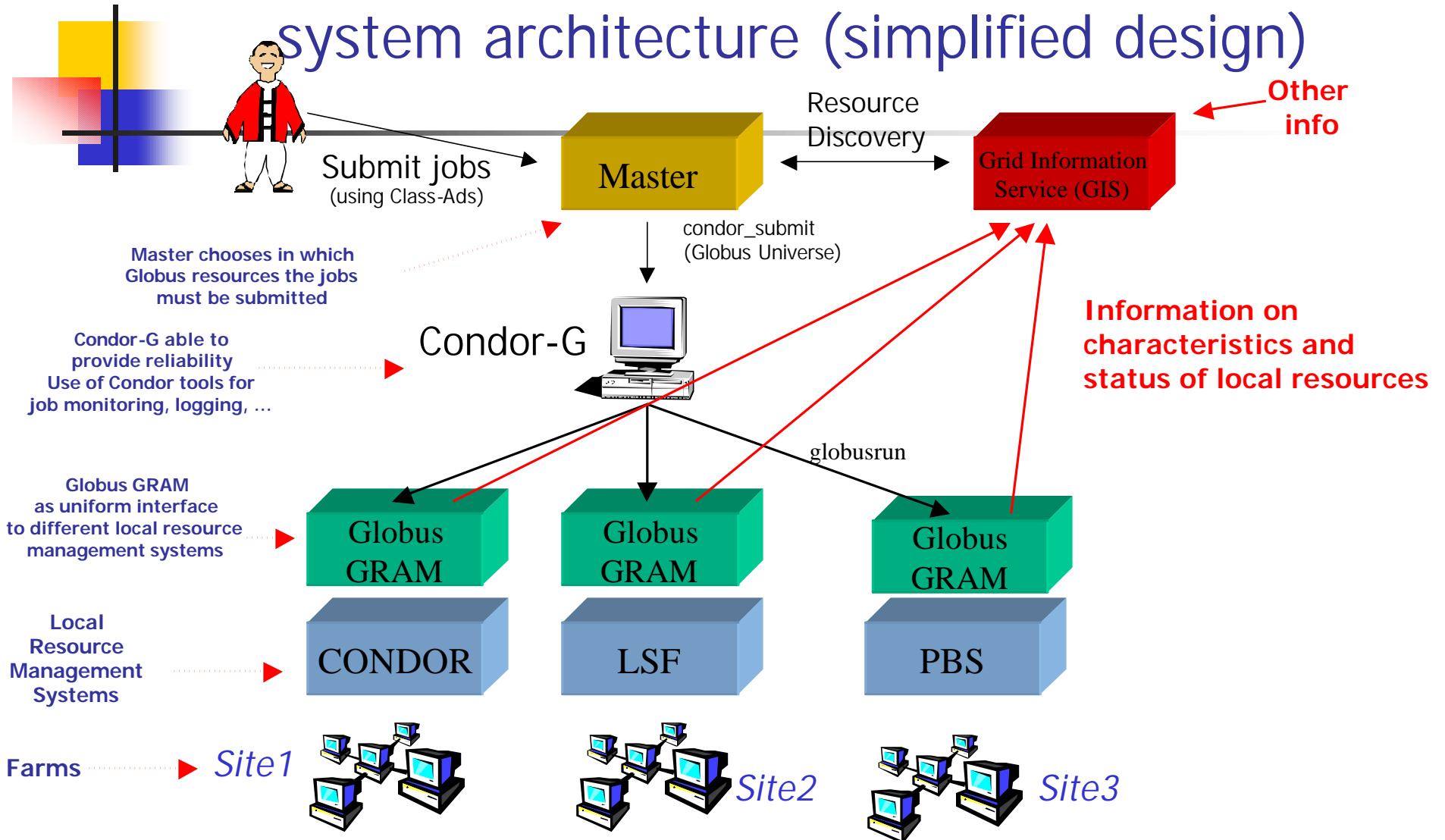
- ✦ Integration, adaptation and deployment of middleware developed within the DataGrid project
 - ✦ GRID software must enable physicists to run their jobs using all the available GRID resources in a “transparent” way
- ✦ HEP applications classified in 3 different “classes”, with incremental level of complexity
 - ✦ Workload management system for Monte Carlo productions
 - ✦ Goal: throughput maximization
 - ✦ Implementation strategy: code migration (moving the application where the processing will be performed)
 - ✦ Workload management system for data reconstruction and production analysis
 - ✦ Goal: throughput maximization
 - ✦ Implementation strategy: code migration + data migration (moving the data where the processing will be performed, and collecting the outputs in a central repository)
 - ✦ Workload management system for individual physics analysis
 - ✦ “Chaotic” processing
 - ✦ Goal: latency minimization
 - ✦ Implementation strategy: code migration + data migration + remote data access (accessing data remotely) for client/server applications



First Activities and Results

- ✍ CMS-HLT use case (Monte Carlo production and reconstruction) analyzed in terms of GRID requirements and GRID tools availability
 - ✍ Discussions with Globus team and Condor team
 - ✍ Good and productive collaborations already in place
 - ✍ Definition of a possible high throughput workload management system architecture
 - ✍ Use of Globus and Condor mechanisms
 - ✍ But major developments needed

High throughput workload management system architecture (simplified design)





First Activities and Results

- ✈ On going activities in putting together the various building blocks
 - ✈ Globus deployment
 - ✈ INFN GRID distribution toolkit to make Globus deployment easier and more automatic
 - ✈ INFN customizations
 - ✈ Evaluation of Globus GRAM
 - ✈ Tests with job submissions on remote resources
 - ✈ Globus GRAM as uniform interface to different underlying resource management systems (LSF, Condor, PBS)
 - ✈ Evaluation of Globus RSL as uniform language to describe resources
 - ✈ "Cooperation" between GRAM and GIS



First Activities and Results

✂ Evaluation of Condor-G

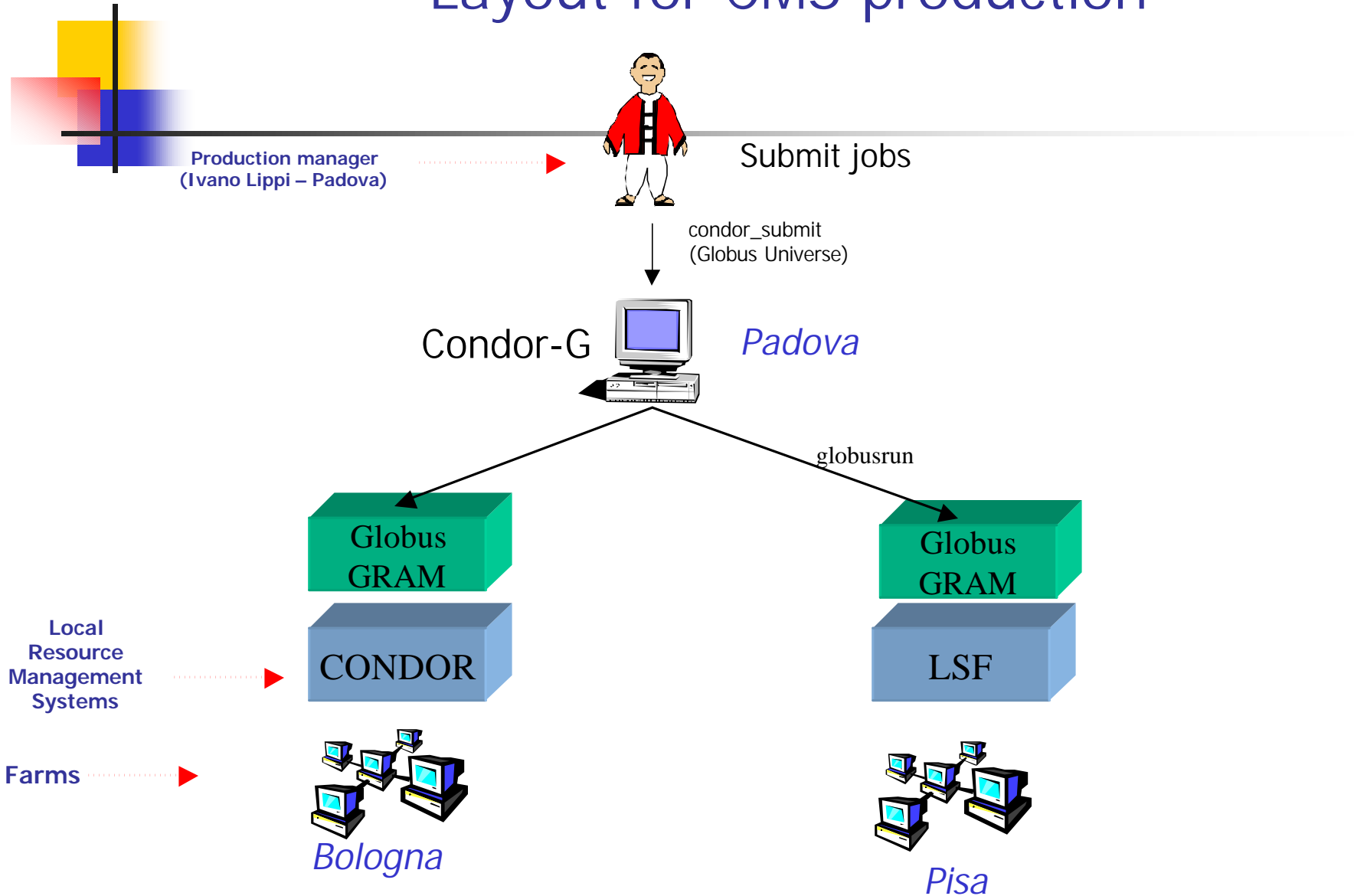
- ✂ It works, but some problems must be fixed:
 - ✂ Very difficult to understand about errors
 - ✂ Problems with log files
 - ✂ Problems with scalability in the submitting machine
 - ✂ Condor-G is not able to provide fault tolerance and robustness (because Globus doesn't provide these features)
 - ✂ Fault tolerance only in the submitting side
- ✂ Condor team is already working to fix some of these problems
 - ✂ They are also implementing a new Globus jobmanager



First activities and results

- ✍ Tests with a real CMS MC production
 - ✍ Real applications (Pythia)
 - ✍ Real production environments
 - ✍ Jobs submitted from Padova using Condor-G and executed in Bologna and Pisa
 - ✍ Many many memory leaks found in the Globus jobmanager !!!
 - ✍ Fixes provided by Francesco Prelz

Layout for CMS production





Some next steps

- ✍ Evaluation of the new Globus jobmanager and the new Condor-G implementations (when ready)
- ✍ Master development !!!



Other info

 <http://www.infn.it/grid>