GLUE Schema:
status and future directions

Sergio Andreozzi
INFN CNAF
Bologna, Italy
sergio.andreozzi@cnaf.infn.it

OSG Consortium Meeting
Seattle, WA - 21-23 August 2006

OUTLINE

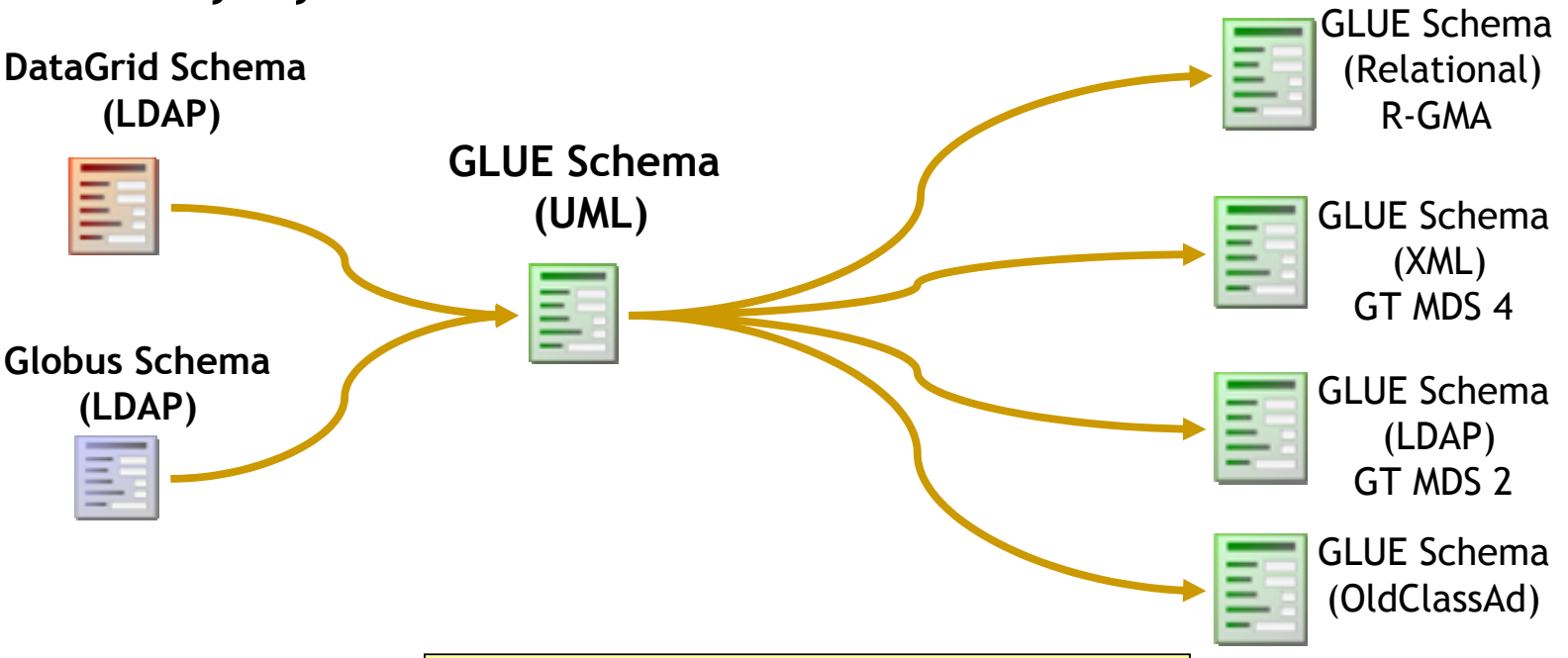
- Problem Statement
- GLUE Schema
  - History
  - Approach
  - Open Issues and Future Directions
- Conclusions

Problem Statement

- Resources available in Grid systems must be described in a precise and systematic manner if they are to be able to be discovered for subsequent management or use

GLUE Schema: a Bit of History

- Activity started in April 2002 by the EU-DataTAG and US-iVDGL projects; contributions from DataGrid, Globus, PPDG, GryPhyn

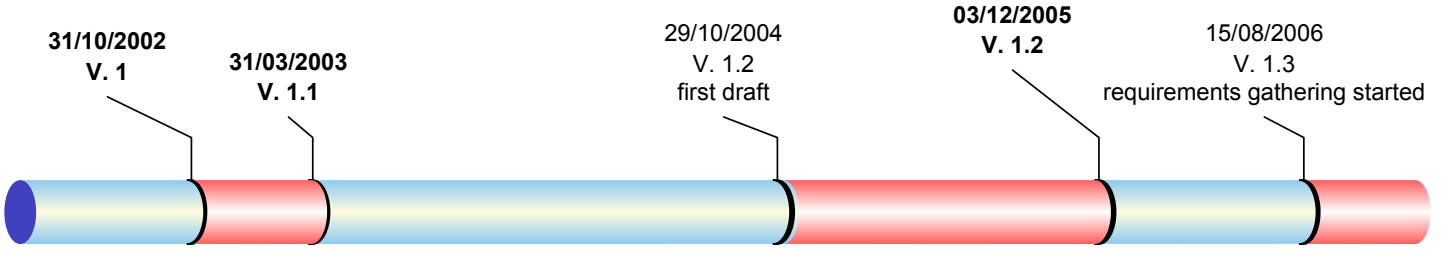


http://glueschema.forge.cnaf.infn.it

Modeling Guidelines

- Focus on the abstractions at the virtual level of a Grid
- Aggregate descriptions of virtual pools of resources
- Generalization
  - capture common aspects for different entities providing the same functionality (e.g., uniform view over different batch services)
- Supporting discovery for:
  - **Brokering and Access**: concerns those attributes that are meaningful for locate resources on the base of a set of preferences/constraints
  - **Monitoring**: concerns those attributes that are meaningful to describe the status of resources

GLUE Schema Timeline



Next steps:

- Short Term: minor revision v. 1.3
  - Backwards-compatible
  - To include requirements from different projects
- Medium Term: major revision v. 2.0
  - No need for backwards-compatibility
  - Refactoring for simplification
  - Convergence to CIM ?

GLUE Schema Main Concepts

- Site
  - Service
  - Cluster
    - Computing Element
    - SubCluster
    - Host
  - Storage Element
    - Storage Area
    - Control Protocol
    - Access Protocol
  - Computing Element – Storage Element relationship

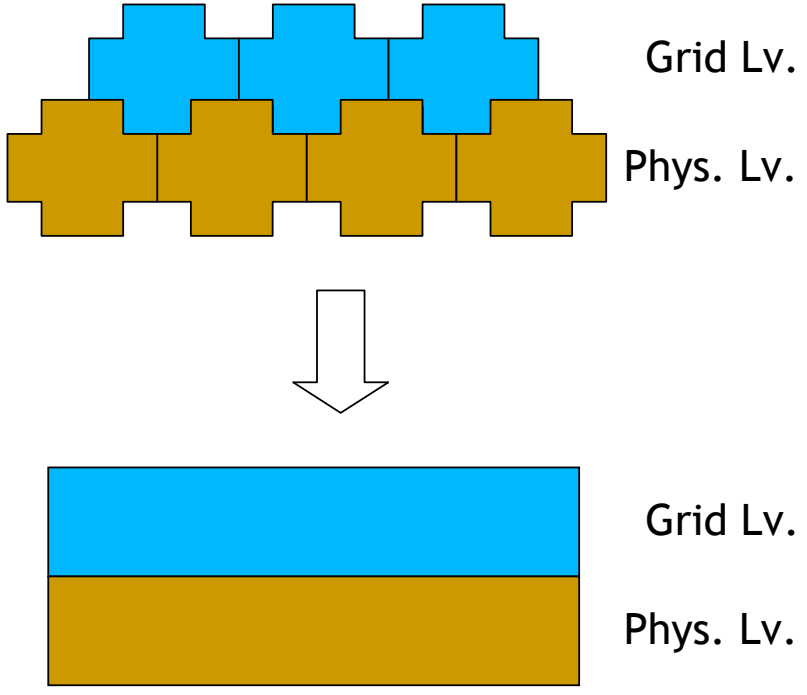
(Some) Open Issues for Future Revision

Computing Element /1

- The virtual concept of Computing Element is not well isolated from the physical one
  - Direct relationship with the underlying batch queue
  - Batch queues
    - can be used/configured in many different ways to provide the same service
    - metrics related to a batch queue may assume a different value depending on the submitting user
      - led to the introduction of VOView in GLUE Schema 1.2
    - may not exist on some physical system (e.g. MAUI does not have a concept of queue)

Computing Element /2

- Refine concentring on:
  - Isolation of physical level from Grid level
  - Identification of the core concepts for the Grid level
    - What is the shared resource?
      - Execution environment
    - How the access is policied?
      - Service differentiation



Heterogeneous Clusters

- Since the beginning, the GLUE Schema enables to model subsets of homogenous nodes in a cluster
- This feature has been never used because from the Grid level it is not possible to submit a job providing requirements on the worker node characteristics (at least with GT2 GRAM)
- This implied the usage of only one subcluster instance related to the less powerful machines!!!
- How can this be improved?
  - One single description with min/avg/max values?
  - Waiting for more powerful job submission protocols?
  - …

Multiple Endpoints to the Same Service

- The current schema does not support multiple network endpoints to the same service instance
  - E.g.: a single cluster with two head-nodes/gatekeepers appears as two different services?

- Comments:
  - Evaluate the reasons for the need of multiple headnodes
    - Dedicated access point to a certain VO?
    - Load balancing?
  - An approach is to hide different headnodes using the DNS
    - You don't choose among different access point when you search on google
  - If you advertise multiple service endpoints as being related to the same service, how do you choose among them?

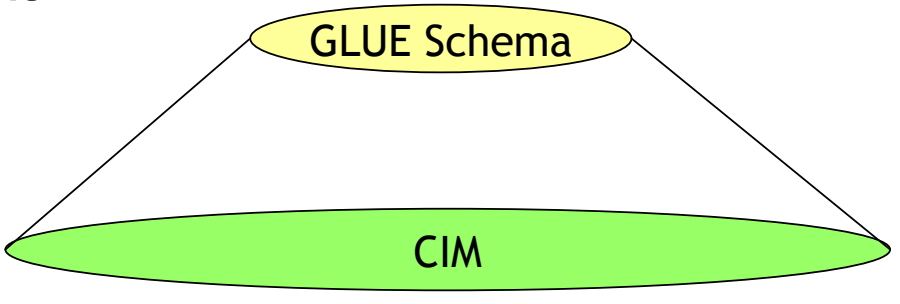
GLUE Schema and CIM /1

- GLUE Schema
  - tailored approach to the modeling of Grid resources for brokering, access and monitoring
  - aggregate description
  - description in UML, mapping into different data models
  - HEP-Grid community standard (expanding adoption)

- CIM
  - modeling of IT resources for their management
  - fine-grained description
  - description in UML, mapping into concrete data models (mainly XML)
  - DMTF standard
  - Part of WBEM to provide infrastructure for management comprising interface, protocol and query language to interact with a managed resource

GLUE Schema and CIM /2

- My view on a possible convergence path:
  - To be investigated for GLUE Schema 2.0
  - CIM to be extended with GLUE Schema concepts
  - GLUE Schema 2.0 concept as a restrictied view of CIM concepts
  - Need for committements from the involded parterns



Other Issues

- From yesterday discussion:
  - Problem about binary compatibility between execution environment and applications
    - OS name, type and version
      - not enough
      - sometimes misconfigured
    - My suggestion goes in two directions:
      - Improve data quality
      - Add application-level prologue that checks if the execution environment contains what expected, if not quit reporting application-level error
  - WallTime
    - Hard vs. soft deadline
- There are other standing issues from different projects
  - e.g.: EGEE, NorduGrid, OSG

see http://glueschema.forge.cnaf.infn.it/Spec/V13

CONCLUSIONS

- GLUE Schema is adopted mainly in HEP-related Grid and is expanding
- Revision process relies on communities contributions in terms of requirements and participation
- The involved projects can contribute a meaningful experience for:
  - Refininements in v.1.3
  - Refactoring for a major version stable for the next few years

EVERYTHING IS POSSIBLE IF PEOPLE PARTICIPATE AND CONTRIBUTE