

www.eu-egee.org

JRA1 All Hands Meeting, 15-17 Nov 2004

GLUE Schema

Sergio Andreozzi JRA1 IT/CZ Cluster sergio.andreozzi@cnaf.infn.it



EGEE is a project funded by the European Union under contract INFSO-RI-508833 Copyright (c) Members of the EGEE Collaboration. 2004.





- Problem Statement
- GLUE Schema
 - From the beginning to specification version 1.1
 - Current Revision Process
 - What's new in the current draft for spec version 1.2



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
- A shared description allows multiple experts to contribute to the problem and serves as a communication mean between different knowledge domains

INFORMATION MODEL

- Abstraction of real world into constructs that can be represented in computer systems (e.g., objects, properties, behavior, and relationships)
 - Not tied to any particular implementation
 - Used to exchange information among different domains

Problem Statement



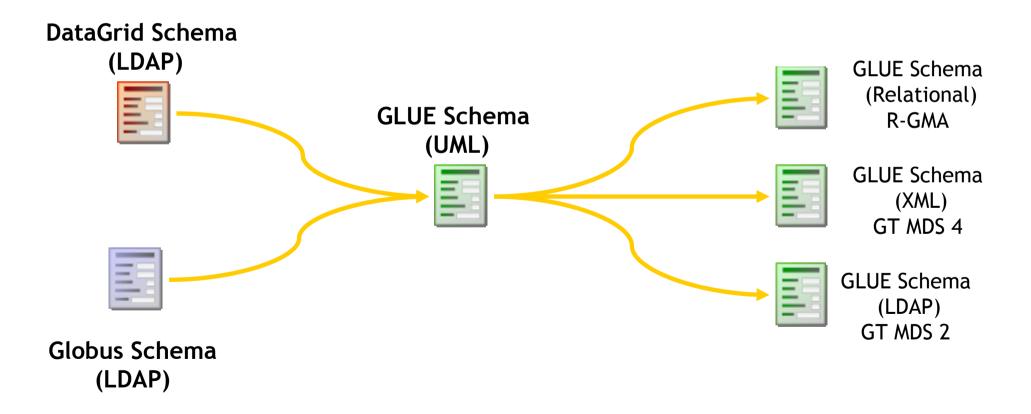
• Main Use Cases:

- Discovery for brokering and access:
 - "what are the Computing Elements available to the VO CMS and that offer the SL3 operating system with installed the CMKIN software package?"
 - "what are the Storage Elements that offer 20 gigabytes of disk space for the VO ATLAS?"
- Discovery for monitoring
 - "how many CPUs the site XYZ is offering to the EGEE Grid?"
 - "what is the success rate of job submitted per site?"

GLUE Schema



- approach to the information modeling of Grid resources started in April 2002 by the DataTAG and iVDGL projects
- Contributions from DataGrid, Globus, PPDG, GryPhyn



GLUE Schema - current spec

- 31 March 2003: GLUE Schema version 1.1
 - Computing Resources
 - Computing Element
 - Cluster
 - SubCluster
 - Host
 - Storage Resources
 - Storage Element
 - Storage Space
 - Data Access Protocol
 - Storage Library
 - Computing/Storage Relationship
 - CESEBind



[1]



GLUE Schema Current Revision Process

GLUE Schema - next steps



- There is a number of issues to be addressed:
 - Simplification
 - "bug fixing"
 - better documentation
 - new use cases
 - Extensions:
 - Per-VO view of computing resources (e.g., Estimated Traversal Time, Free Job Slots) (Jeff Templon)
 - Grid3 (some already in the new draft) [4]
 - GridICE (Grid monitoring)
 - INFN (Monitoring the Connectivity of a Grid) [5]
 - JRA1 Advance Reservation (being defined with T. Ferrari and E. Ronchieri)

GLUE Schema - current revision process



Communication:

- Mailing list: <u>glue-schema@hicb.org</u>
- Tracker: http://infnforge.cnaf.infn.it/tracker/?atid=118&group_id=9&func=browse
- Phone Calls or face-to-face meeting are <u>not scheduled</u> since one year

Revision Process:

- Items for schema revision can be added in the tracker by anyone
- People can post their comments in the tracker
 - Each tracker update is forwarded to the mailing list
- When an agreement is reached, the change is inserted in the next schema revision

GLUE Schema - current revision process



- Even though the <u>active participation</u> in this activity is <u>still a problem</u>, there is a periodical expression of interest in this work and its evolution
- In order to go ahead, each involved project should refresh who are the representatives that should actively participate and make decision on behalf of the project
- Area of interests vs. resouce cateories
 - we should propose names for EGEE

	Computing	Storage	Network
Brokering	V	V	
Data Management	Ø	Ø	?
Advance Reservation			\checkmark
Monitoring			

GLUE Schema - next steps

- Possible approach:
 - Minor revision taking into account only <u>bug fixing</u> and extensions that maintain <u>backwards compatibility</u> (e.g., GLUE Schema 1.2 draft)
 - Major revision including <u>refactoring for simplification</u> (e.g., GLUE Schema 2.0 to be written)
 [3]
 - gLite should consider the major one





GLUE Schema - The importance of Use Cases

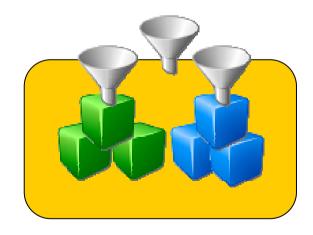


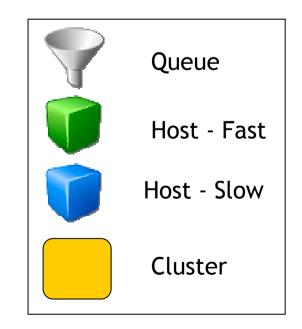
- Use Cases are important
- So far, we lack of a Use Case document
- Use cases are for information models what Unit Tests are for software
- People participating in the deployment should help in collecting a number of significant use cases

GLUE Schema - The importance of Use Cases an example



- Site A has 6 worker nodes (3 fresh new and fast, 3 old and slow)
- The farm is configured as follows:
 - a high-end queue to the 3 fast WN's
 - a slow queue to the 3 slow WN's
 - a background queue to the 6 WN's (lower priority)

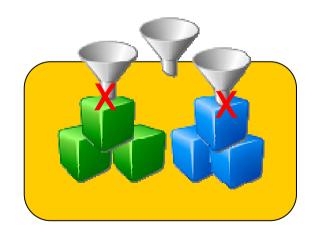


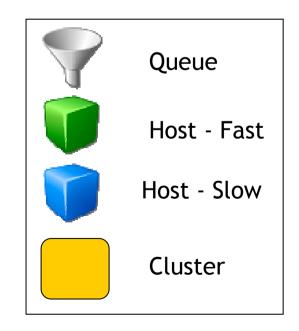


GLUE Schema - The importance of Use Cases an example



- Representation in the GLUE Schema v. 1.1
- Queue -> <u>Computing Element</u> (CE_highend, CE_slow, CE_background)
- Characteristics of new/fast WN's -> <u>Subcluster</u> A
- Characteristics of old/slow WN's -> <u>Subcluster</u> B
- A+B=<u>Cluster</u>
- **Problem:** there is no explicit relationship between CE and SubCluster; <u>CE are</u> <u>associated</u> to Cluster, hence <u>to all underlying SubCluster</u>

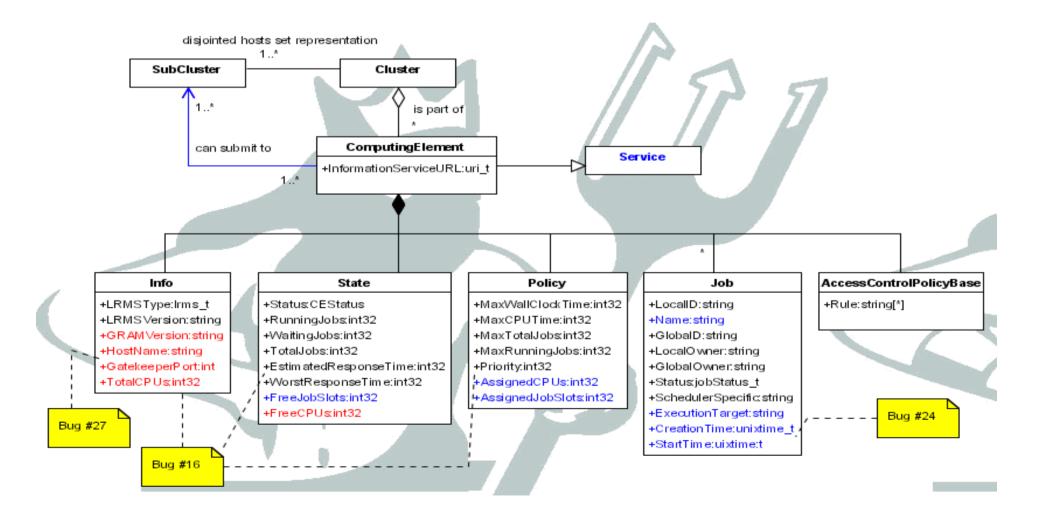




GLUE Schema - The importance of Use Cases an example

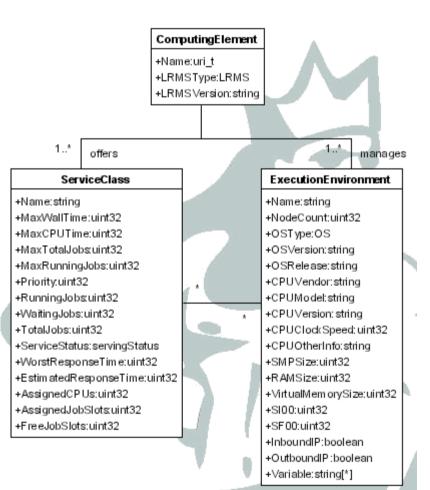


- Possible Fix in GLUE Schema v. 1.2 (Additive approach)
- Add a new relationship: <u>CE (*)</u> ---can submit to--- (*) SubCluster



GLUE Schema - The importance of Use Cases

- Possible Major Refactoring in GLUE Schema v.2.0
 - <u>CE is a site cluster</u> (EGEE Architecture document)
 - Queues are used to differentiate the service
 - The service offers access and management of available execution environments
- ServiceClass:
 - HighEnd, Slow, Background
- ExecutionEnvironment:
 - <u>charact node A, charact node B</u>



eGee

Enabling Grids for E-science in Europe

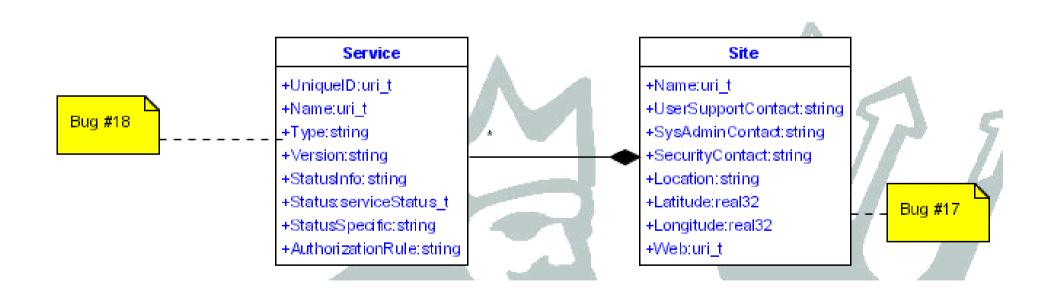


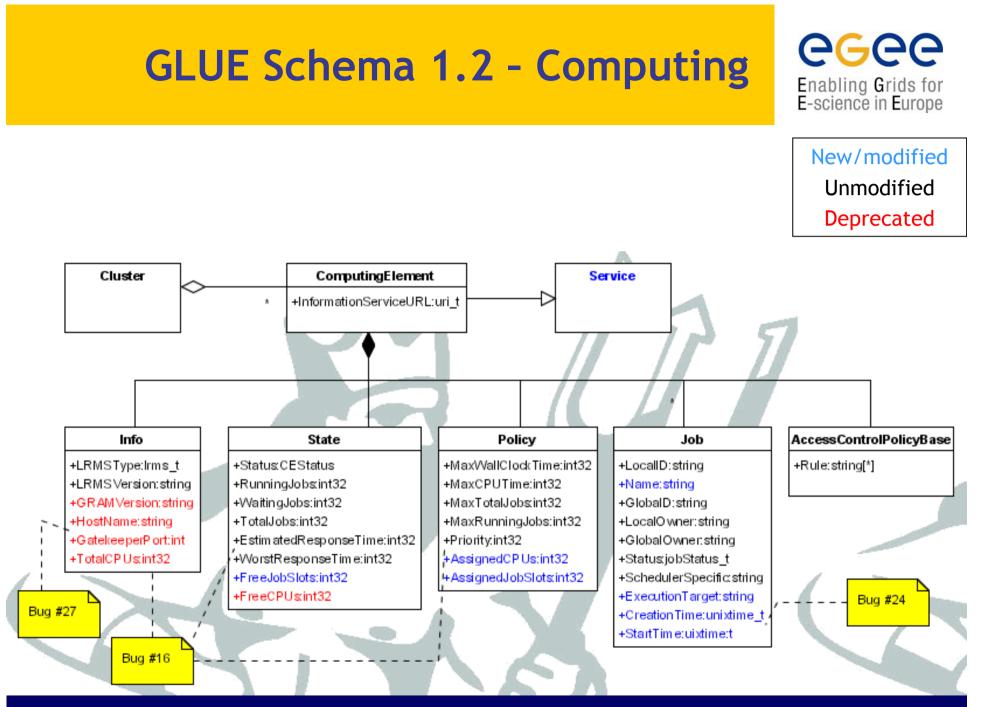
GLUE Schema Version 1.2 Draft proposal

GLUE Schema 1.2 - Core

Enabling Grids for E-science in Europe

New/modified Unmodified Deprecated





GLUE Schema 1.2 - Cluster/SubCluster/Host



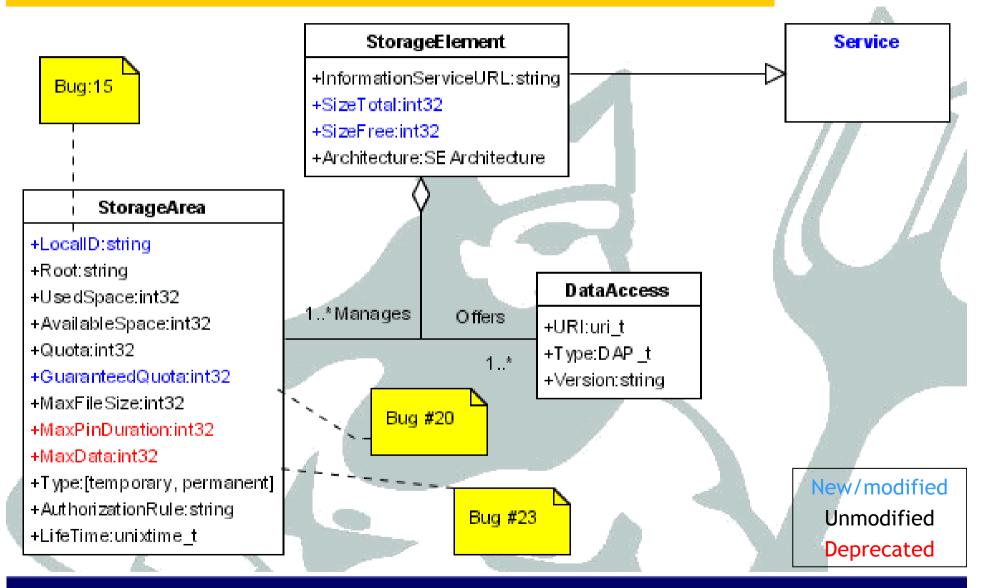
Cluster +UniqueID:uri t +Name:urit disjointed hosts set representation 1 SubCluster further attributes that can be part of participating host +UniqueID:uri t the subcluster are a subset of the +Name:urit host attributes that are meaningful 1..* to be used to as a criterion to partition hosts in homogeneous subsets Host +UniqueID:uri_t +Name:urit homogeneity +UpTime:unixtime_t OperatingSystem Architecture +Name:string MainMemory +PlatformType:string Benchmark +Release:string +SMPSize:int32 +RAMSize:int32 +Version:string +SI00:int32 +SMTSize:int32 +RAMAvailable:int32 +SF00:int32 +VirtualSize:int32 +VirtualAvailable:int32 NetworkAdapter Applic ation Software File Bug #26 +Name:string +RunTimeEnvironment:string[*] +Name:string +IP Address: string +Size:int32 Load +CreationDate:unixtime_t +MTU:int32 FileSystem +LastModified:unixtime_t +InboundIP:boolean +Last1Min:int32 +OutboundIP:boolean +Last5Min:int32 LastAccessed:unixtime_t StorageD evice +Name:string Latency: int32 +Last15Min:int32 +Root:string +Name:string LifeTime:int32 +Size:int32 +Type:string +Owner:string file storage +AvailableSpace:int32 Processor +TransferRate:int32 +ReadOnly:boolean +Size:int32 +Vendor:string 0..1 +Type:string Bug #22 AvailableSpace:int32 +Model:string +Version:string resides on Directory 0..1 +ClockSpeed:int32 +InstructionSet:string 0..1 mount +OtherProcessorDescription:string StoragePartition +CacheL1:int32 +Name:string +CacheL1I:int32 Size:string +CacheL1D:int32 RemoteFileSystem LocalFileSystem 0..1 export +ReadRate:int32 +CacheL2:int32 +WriteRate:int32 ProcessorLoad SMPLoad +Last1Min:int32 +Last1Min:int32 +Last5Min:int32 +Last5Min:int32 Last15Min:int32 Last15Min:int32 Bug #25

New/modified Unmodified Deprecated



GLUE Schema 1.2 - Storage

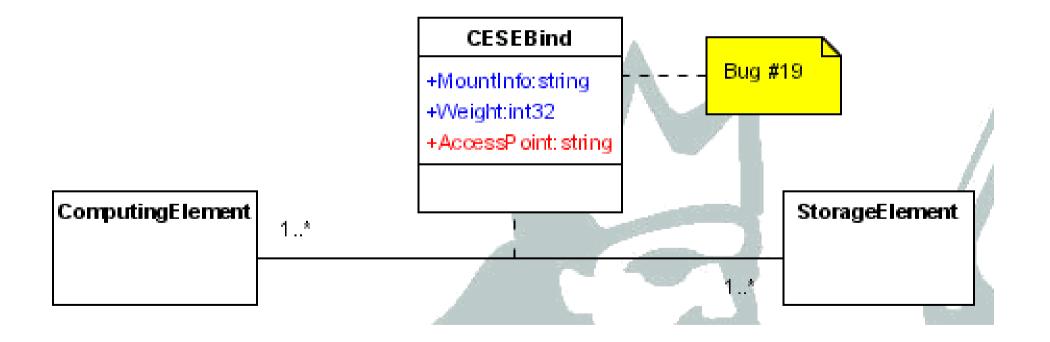
CGCC Enabling Grids for E-science in Europe



GLUE Schema 1.2 - CE/SE relationship

Enabling Grids for E-science in Europe

New/modified Unmodified Deprecated







[1] GLUE Schema Resources: http://www.cnaf.infn.it/~sergio/glue

- [2] GLUE Schema open issues: http://infnforge.cnaf.infn.it/docman/view.php/9/65/GLUEInfoModel_V_1_2_draft_1.pdf
- [3] GLUE Schema Version 1.2 First Draft <u>http://infnforge.cnaf.infn.it/tracker/?atid=118&group_id=9&func=browse</u>
- [4] Grid3 metrics <u>http://grid.uchicago.edu/metrics/metrics-table.html</u> <u>http://griddev.uchicago.edu/download/grid3/doc.pkg/monitoring-metrics/Grid3-metrics.doc</u>
- [5] Monitoring the Connectivity of a Grid. S. Andreozzi, A.Ciuffoletti, A. Ghiselli, C. Vistoli. In Proc. of the 2nd International Workshop on Middleware for Grid Computing (MGC 2004) in conjunction with the 5th ACM/IFIP/USENIX International Middleware Conference, Toronto, Canada, October 2004.