

Job and Data Accounting on the Open Science Grid

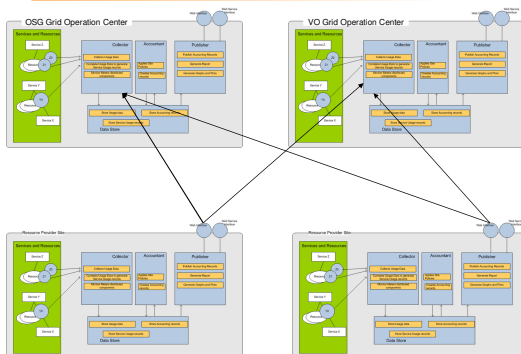
Ruth Pordes, Fermilab
with thanks to Brian Bockelman, Philippe Canal, Chris Green, Rob Quick



Open Science Grid Accounting

- Overview
- Sustainability
- The Future

Overview – a Distributed Interconnected Accounting System



OSG Accounting Gratia

..designs and deploys robust, scalable, trustable and dependable grid accounting, publishes an interface to the services and provides a reference implementation.

- In production use for >3 years.
- Ongoing requests for upgrades and extensions as utility and scale grows.
- Lot of data now available for “mining”
- Main OSG Database now >100 Gigabytes in size.





Technology Snapshot

- Architecture
probe->collector->collector
- Implementation
basic schema is extension of OGF Usage Record
python library, JMS web server, java code,
hibernate, mysql/innodb, birt, graphtools
- Development environment
svn, make, Metronome, OSG software process (see
Rob Quicks talk)

5



Evolving Team

- **Gratia** started as a joint project between the Fermilab Computing Division, US CMS, as an external software development project to meet **Requirements for Job and Data Accounting** for Fermilab distributed systems locally, the US LHC experiments reporting requirements to the WLCG, and the OSG.
- Since started **extended with contributions** from US ATLAS, OSG itself, Oklahoma University, University of Nebraska, Condor project (for condor and boinc probes) and now UTA (testing).
- Software installation and configuration scripts distributed, to OSG and its partners, as part of the **OSG Virtual Data Toolkit**

6



Capabilities

Collects a record per job that uses a batch system (condor, pbs, sge, psf) locally, including end of job status conditions to the Grid interface (Gram 2 or 4).

Records data transfer from instrumented storage systems – Bestman, dCache, GridFTP.

Provides for Linux based accounting with psAcct probe.

Summarizes records per Site, User, VO for automated reports and for selection through the web interfaces.

Collects availability test results (OSG RSV). Interfaces to BDII/SAM for availability and reliability information

Interfaces to APEL for EGEE and WLCG accounting.

7



Reliability (from Philippe)

1. The Probe library caches the XML messages locally when the communication with the Collector fails
 - Allowed seamless server upgrades
 - Recovery from probe misconfiguration.
2. The Collector caches the XML messages locally when the communication with the back-end database fails.
3. The Collector keep locally copy of the process messages for a (configurable) while.
4. The Back-end database is regularly backed-up. "expired" Records are archived
5. data chain has several points where data are buffered in the event that the upstream receiver is offline



Accounting Repositories

OSG Accounting Central Database

- MYSQL+ INNODB for scaling and transactions.
- Centrally collected OSG Job records kept in DB for 3months then archived to tape.
- Automated replication for data warehousing/mining.
- Streamed for security officer – SPLUNK analysis.

~6 other "OSG site" Repositories at large resources

Fermilab and Nebraska campus repositories for extended local accounting.

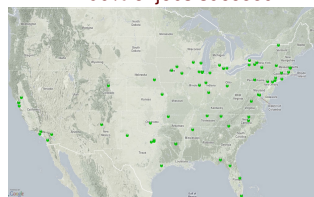
OSG Reliability Database

9



Scale of Resource Usage on OSG

- ~10,000 jobs/ hour
- >20,000 files/hour
- ~30,000 CPU days/day
- ~75 sites used per day
- ~ 20 VOs
- ~90% of jobs succeed

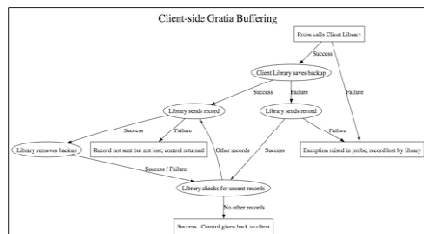


10



Completeness of Information

- Gratia attempts to ensure no data is lost.
- Data is buffered at each stage.
- "Catch up" of data is supported even when this takes several weeks.



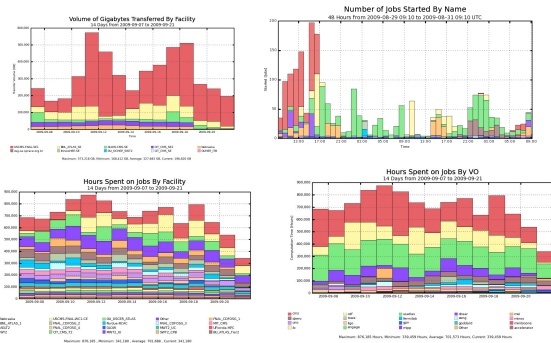
11



Text Based Reporting and Validation

- Cron jobs report job, data, VO, efficiency usage daily and weekly.
- Reports sent to VO managers and Site Administrators listing usage by user.
- "New user" report alerts security team to watch.
- Gratia checks VOs and Sites reporting against OSG registration databases.
 - Gratia is used by partner Grids such as NYSGrid so some sites are not registered with the OSG itself.
 - Some sites use Gratia internally to report local jobs. Checks identify misconfigurations.

12

[illegible]

- This has been done for more than 2 years. Basis for WLCG reports to the funding agencies. Checked monthly against direct OSG reports
- Units of HEPspec2006 ready for use.
- Sites that report controlled by VO management through the OSG registration process.
- Full User DN publishing for WLCG under test (CN available already).



- Attention to support of software in long-term production use.



- The software source is available **open source on SourceForge**.
- Modicum of “developer independence” provided by the Project lead (Philippe Canal, developer, HEP domain knowledgeable) being Fermilab and the OSG Software Project Liaison (Brian Bockelman, CS/Maths PhD) being University of Nebraska.
- **OSG external project liaisons provide written requirements & priorities. They are expected to provide a knowledgeable conduit** between external projects and the OSG Consortium.
- Accept other sources for contributions to date UTA, OU, BNL. Software **contributions are coordinated** through the weekly project meetings.
- Releases are managed and release notes written and s/w is built and tested using Metronome before being put into VDT.

17



Gratia Development and Test Environment

- Define and follow Testing Procedures for each new version of the software.

- [Description](#)
- [Software](#)
- [Gratia Daily Builds](#)
- [Daily installs and upgrades](#)
- [Gratia Upgrade Script](#)
- [General Development](#)
- [Reports and Services \(Collectors\)](#)
- [Test Site Grid](#)
- [VOMS Test Data](#)
- [Gratia ITB Test Environment](#)
- [Databases](#)

18

[illegible]

Table 1. The number of subjects and percentage of the total data base (2003 data) March 2003. The numbers are taken from the job opening and a number of later data from the 2003 data base survey.

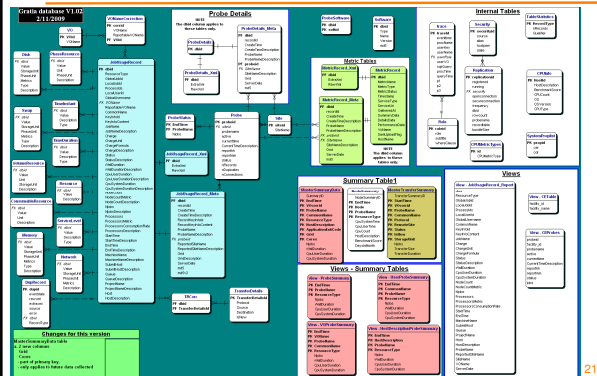
19



Test Description	Probe	Possible results	Result assumed
Check Windows event logs for errors	ALL	CRITICAL WARNING INFO	Y
Check Windows event logs for warnings	ALL	CRITICAL WARNING INFO	Y
Check performance on parameters (if found a problem, then check hardware)	ALL	CRITICAL WARNING INFO	N
Check for software in copyright	ALL	WARNING INFO	N
Check for service status in parameters	ALL	WARNING INFO	N
Verify if the system is running as administrator	Console	WARNING WARNING (can't run as administrator)	N
Can Windows event monitor be loaded?	FILE, LSP	CRITICAL WARNING INFO	N
Is the system configuration file loaded?	FILE, LSP	CRITICAL WARNING INFO	N
Is not a 386MS (type setting in the system configuration is not recognized)	FILE, LSP	CRITICAL WARNING INFO	N
Check LSPMS type setting matches probe type	FILE, LSP	CRITICAL WARNING INFO	N
Check FILE log directory exists	FILE	CRITICAL WARNING INFO	Y
Check LSP log directory and log contents	LSP	CRITICAL WARNING INFO	Y
Check files written to the sent from the probe	FILE, LSP	CRITICAL WARNING INFO	Y

20

Entity Relationship Diagram Maintained



21

Interoperability

- Gratia framework will allow transform of information to report to other databases - given requirements and acceptance of policies by the VOs and Sites affected.
- Initial discussions with TeraGrid imply that publishing Gratia data to central Amie database might be useful.
- Additionally distributed Collector framework has potential utility for TeraGrid sites.

22

Future Plans

- Continue use and scalability
 - Improvements in data transport have enabled use at large data sites, BNL as well as Fermilab.
 - Information collected showing increased utility for validation and metrics both locally and gathered from multiple source.
- Continue extensions based on Customer requests
 - Will forward information to TeraGrid accounting/allocation databases when needed.
 - Want to complete work for validation between VO and Grid accounting layers. Initial work done with ATLAS and LIGO-Boinc jobs.
 - CS researchers want to provide a web data mining interface for analysis and understanding.
- Further interoperability with Campus, Regional, EGI, NGIs

23