User-level Grid monitoring with Inca 2

Shava Smallen
ssmallen@sdsc.edu

June 25, 2007

SDSC

TeraGrid

- Origins: national supercomputer centers, funded by the NSF
- 9 TeraGrid sites, 18 resources
- Mix of Architectures:
  - ia64, ia32: LINUX
  - Cray XT3
  - Alpha: True 64
  - SGI SMPs

- Connected via dedicated multi-Gbps links
- 1000s of CPUs, > 250 teraflops
- > 30 petabytes of online and archival data storage
- Coordinated user environment across heterogeneous resources
  - CTSS (Coordinated TeraGrid Software & Services)

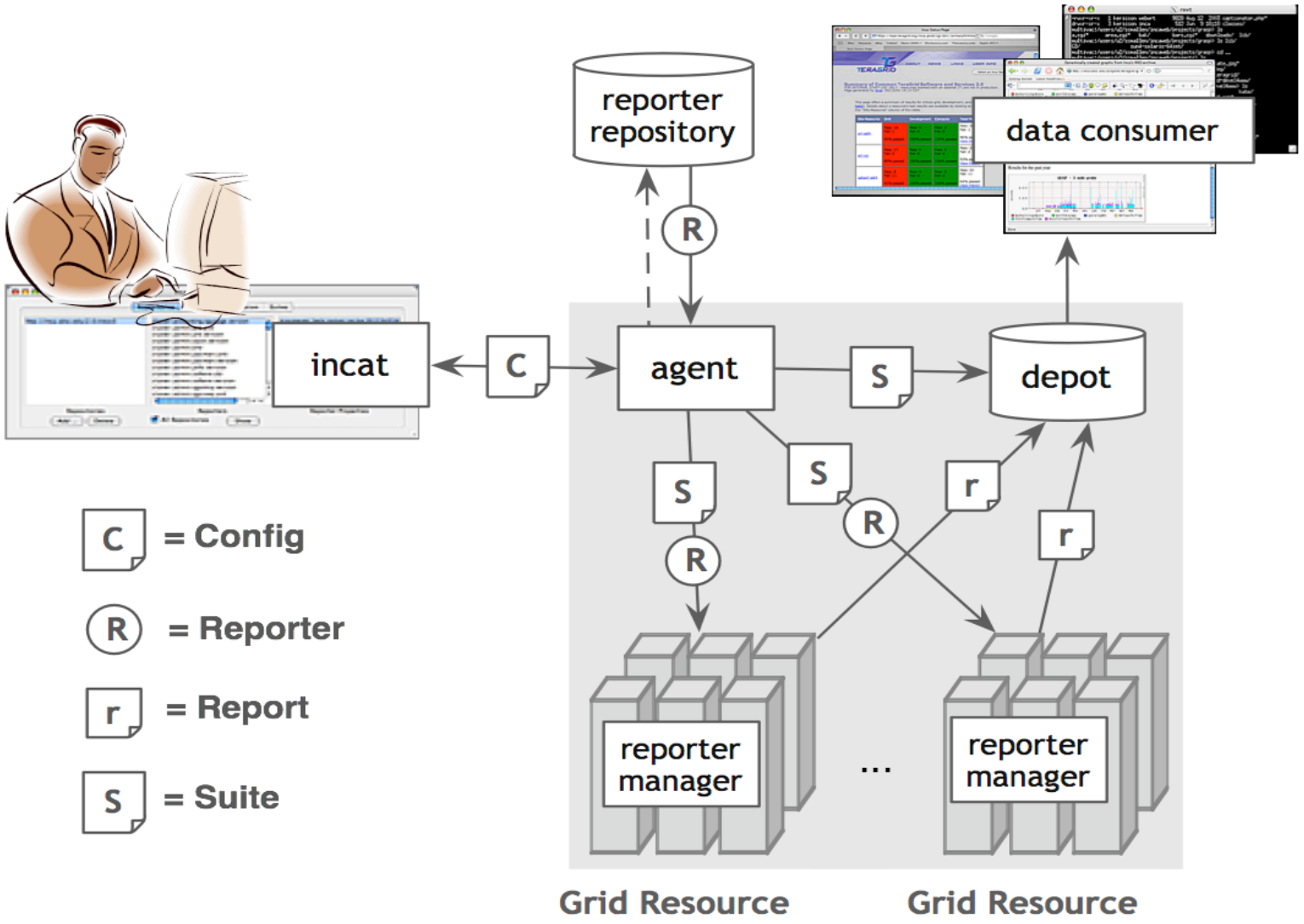
User-level Grid monitoring

- Testing and performance measurement from a generic, impartial user's perspective in order to detect and fix Grid infrastructure problems before the user's notice them.
- User-level Grid monitoring system:
  - Runs from a standard user account
  - Executes using a standard GSI credential
  - Uses tests that are developed and configured based on user documentation
  - Verifies user-accessible Grid access points
  - Centrally manages monitoring configuration
  - Automates periodic execution of tests
  - Easily updates and maintains monitoring deployment

Inca

- Provides user-level monitoring of Grid functionality and performance
- Features:
  - Collects wide variety of monitoring results
  - Captures context of monitoring result as it executes
  - Eases the writing and deploying of new tests or benchmarks
  - Supports sharing of tests and benchmarks
  - Stores and archives monitoring results
  - Securely manages short-term proxies
  - Measures system impact of tests and benchmarks

Inca Architecture

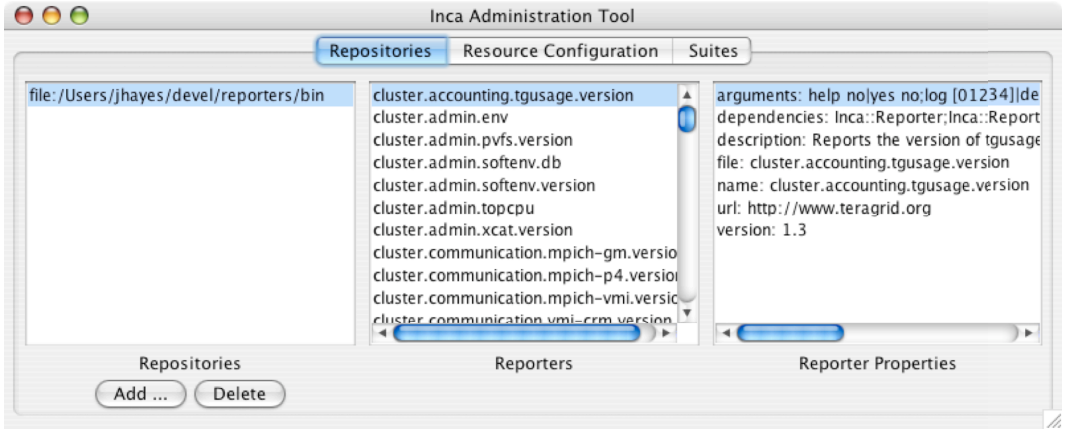


Collecting Monitoring Data

- Reporters
  - Executable program that measures some aspect of the system or installed software
  - Requirements:
    - Supports specific command-line options
    - Writes XML (Inca Reporter schema) to stdout
  - Supports multiple types of data
  - Extensive library support for perl scripts
  - Most reporters < 30 lines of code
  - Independent of other Inca components

Sharing Reporters

- Repositories: collection of reporters available via a URL
  - Supports package dependencies (Perl modules, Makefile, autoconf)



Screenshot of a repository using Inca GUI tool

  - Packages versioned to allow for automatic updates
  - Inca repository contains 157 reporters
    - Version, unit test, performance benchmark reporters
    - Grid middleware and tools, compilers, math libraries, data tools, and viz tool

Centralized configuration and deployment

- Incat
  - GUI interface to enable a large number of monitoring results to be collected with a minimum of effort
  - Configure the reporters to execute on a set of resources
  - Configuration stored in a XML file and sent to Agent
- Agent
  - Implements the configuration specified by Inca administrator
  - Stages and launches a reporter manager on each resource
  - Sends package and configuration updates

Storing data

- Depot
  - Stores configuration information and monitoring results
  - Uses relational database backend via Hibernate
  - Provides full archiving of reporter output
  - Supports SQL queries and provides predefined queries for latest monitoring results, report instance, and report history
  - Supports notifications

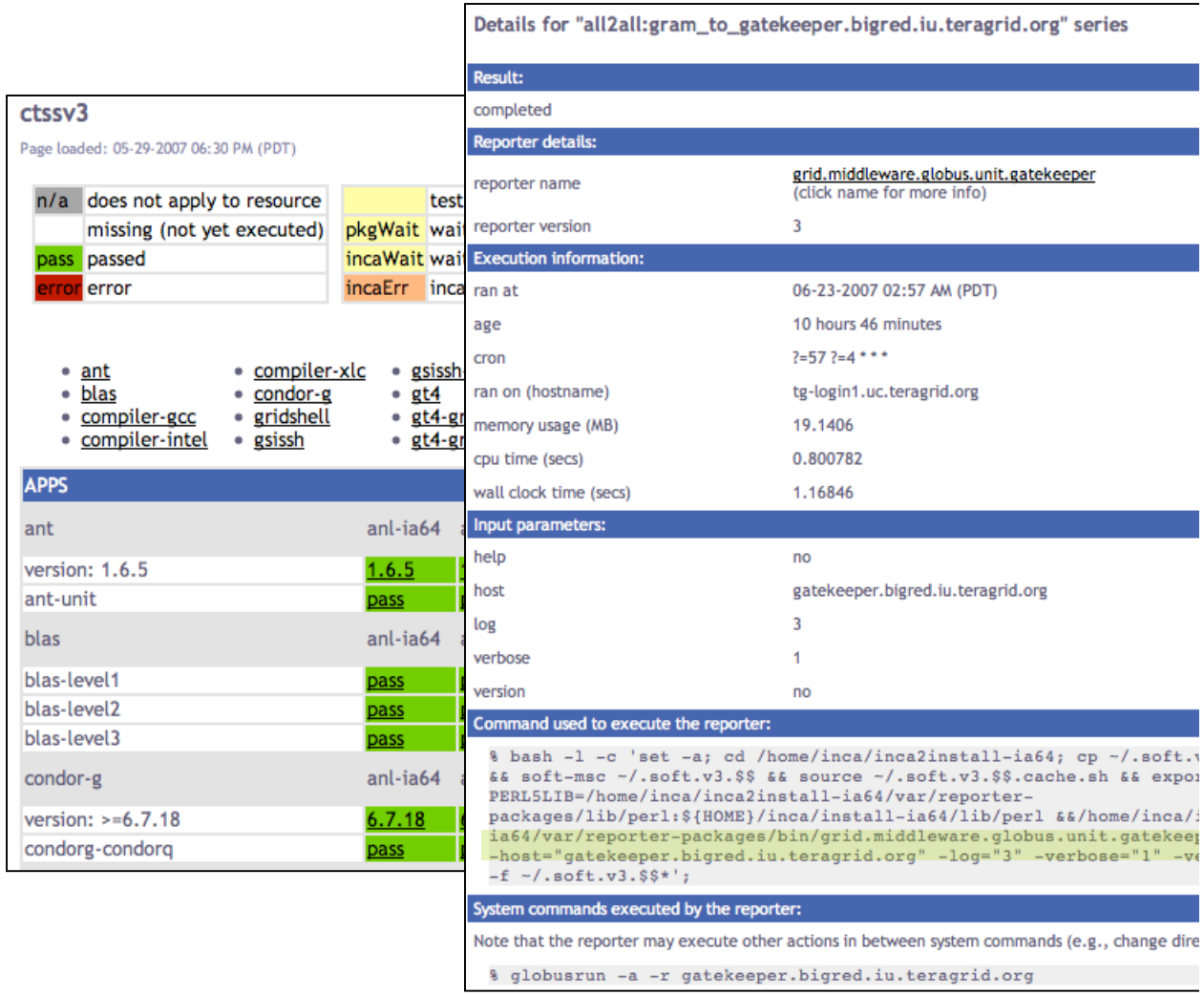
Displaying and publishing data

- Data Consumer
  - Web application that queries and displays monitoring data
  - Packaged with Jetty
  - JSP tags to query data and format using XSL

- Web services
  - Query data from depot and return as XML

Inca in Use: TeraGrid

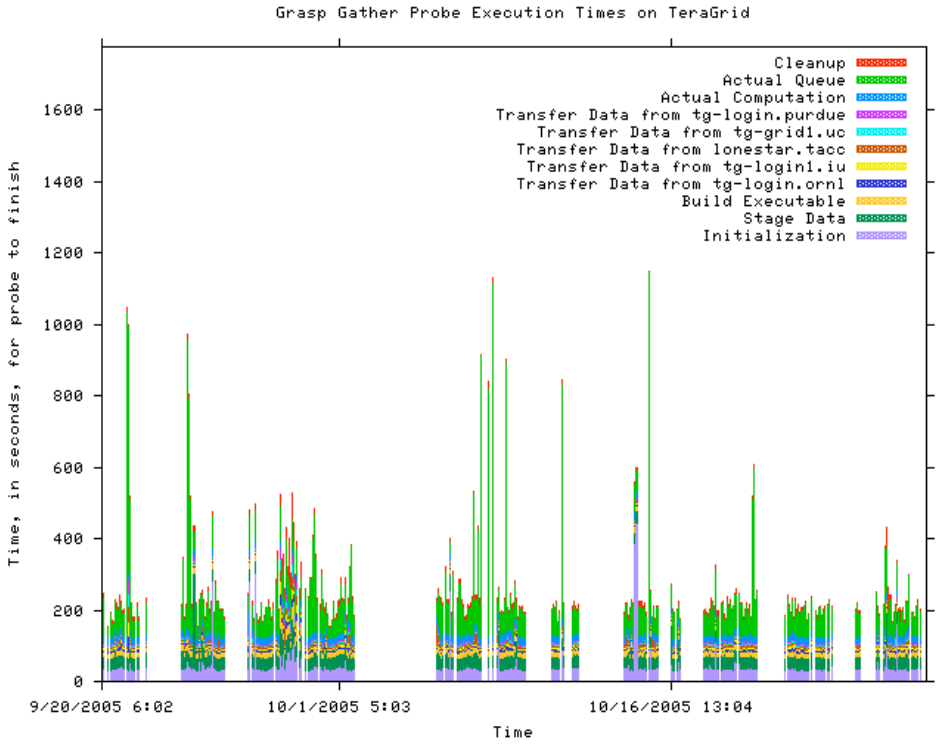
- Currently monitoring all 18 allocated TeraGrid resources
  - Monitoring of CTSSv3
  - Monitoring of CTSSv4 (in progress)
  - Grid jobs (Globus gatekeeper logs)
  - CA certificate and CRL checking (notify if 2 weeks from expiration)
  - Resource registration in MDS

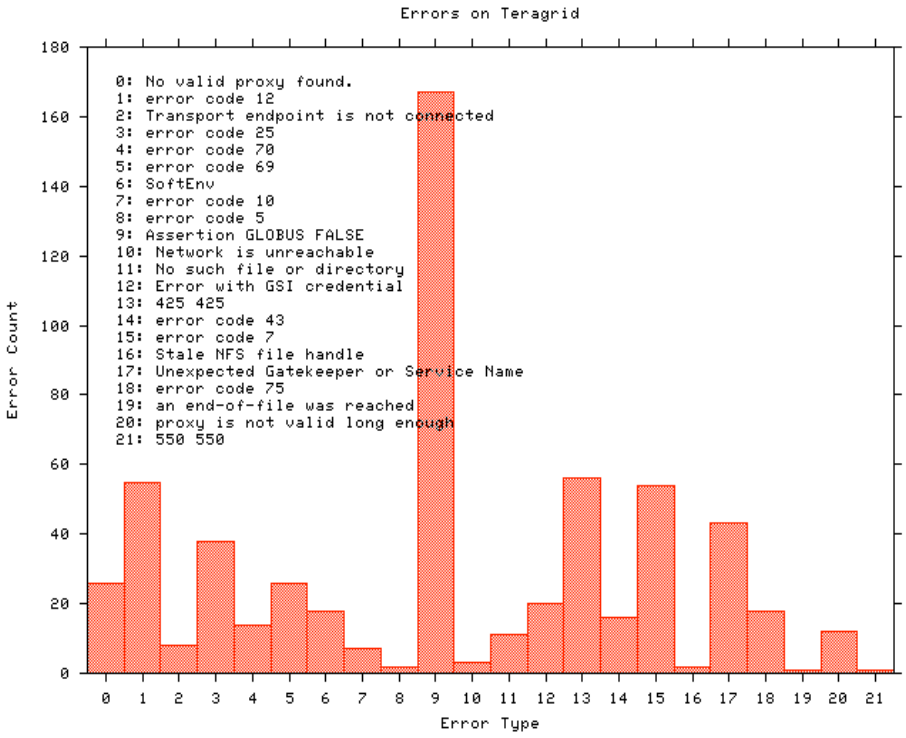


SDSC TeraGrid

Inca in use: Grid Assessment Probes

- Set of probes designed to emulate Grid applications
- Deployed using Inca to GEON and TeraGrid





Software Status

Current software version: 2.03
(available from Inca website)
http://inca.sdsc.edu

Other Inca deployments:

Summary

- User-level Grid monitoring: Testing and performance measurement from an impartial user perspective to detect problems before the users notice them
- Standalone reporter APIs and repositories make it easy to write and share tests and benchmarks (reporters)
- Centralized configuration enables uniform monitoring and makes it easy to deploy Inca monitoring to a set of resources
- Data consumer and web services interface enable publishing and displaying of Inca monitoring data

More Information

Website:
http://inca.sdsc.edu

Announcements:
inca-users@sdsc.edu

Email:
inca@sdsc.edu

Supported by: