Monitoring User-Level Grid Functionality and Performance using inca

Shava Smallen
ssmallen@sdsc.edu
May 15, 2008
Goal: reliable grid software and services for users

- Over 750 TF
- Over 30 PB of online and archival data storage
- Connected via dedicated multi-Gbps links
- 30-63 software packages and 6-23 services per resource

11 TeraGrid sites, 21 resources
User-level grid monitoring

- Runs from a standard user account
- Executes using a standard GSI credential
- Uses tests that are developed and configured based on user documentation
- Centrally manages monitoring configuration
- Automates periodic execution of tests
- Verifies user-accessible Grid access points
- Easily updates and maintains monitoring deployment
**Inca features and architecture**

- Stores and archives a wide variety of monitoring results
- Captures context of monitoring result as it is collected
- Eases the writing, deploying, and sharing of new tests or benchmarks
- Flexible and comprehensive web status pages
- Secure
Reporters collect monitoring data

- Executable programs that measure some aspect of the system or installed software
- Supports a set of command-line options and writes XML to stdout
- Schema supports multiple types of data
- Extensive library support for perl and python scripts (most reporters < 30 lines of code)
- Independent of other Inca components
Repositories support sharing

- Collection of reporters available via a URL
- Supports package dependencies
- Packages versioned to allow for automatic updates
- Inca project repository contains 150+ reporters
  - Version, unit test, performance benchmark reporters
  - Grid middleware and tools, compilers, math libraries, data tools, and viz tool
Agent provides centralized configuration and management

• Implements the configuration specified by Inca administrator

• Stages and launches a reporter manager on each resource

• Sends package and configuration updates

• Manages proxy information

• Administration via GUI interface (incat)

Screenshot of Inca GUI tool, incat, showing the reporters that are available from a local repository
Depot stores and publishes data

- Stores configuration information and monitoring results
- Provides full archiving of reports
- Uses relational database backend via Hibernate
- Supports HQL and predefined queries
- Supports plug-in customization (e.g., email notifications, downtimes)
- Web services - Query data from depot and return as XML
**Consumer displays data**

- Current and historical views
- Web application packaged with Jetty
- JSP 2.0 pages/tags to query data and format using XSLT
- CeWolf/JFreeChart to graph data
Summary of errors this week

Cumulative test status by resource

Average test pass rate

Test status by package and resource

Related test histories

Individual test result details

Individual test history

Current status

Test Details

Historical
Software status and deployments

Current software version: 2.3

(available from Inca website)

http://inca.sdsc.edu
**Inca TeraGrid deployment**

- Running since 2003
- Testing for CTSS
- Cross-site tests
- GRAM usage
- CA certificate and CRL checking
- Resource registration in MDS

Screenshot of Inca status pages for TeraGrid
http://inca.teragrid.org/
Inca GLEON deployment

- Sensors in lake: dissolved oxygen level, temperature, velocity (some), etc.
- Monitoring Data Turbine deployments since Oct. 24
- Currently deployed for Lake Sunapee and Lake Erken
Benefits of using Inca

• Detect problems before the users notice them
• Easy to write and share tests and benchmarks
• Easy to deploy and maintain
• Flexible and comprehensive displays
More information

Website:  
http://inca.sdsc.edu

Announcements:  
inca-users@sdsc.edu

Email:  
inca@sdsc.edu

Funded by:

NSF  
TeraGrid™