Welcome to the 3rd

Inca Workshop

Sponsored by the NSF

August 26-27, 2010

Presenters:

Shava Smallen ssmallen@sdsc.edu

Kate Ericson kericson@sdsc.edu

Paul Hoover phoover@sdsc.edu

Workshop Goals

- Introduce features and benefits of Inca to new or interested users.
- Help existing users to better utilize Inca for their Grid.
- Gather any feedback on new features, improvements to features, etc.

Agenda -- Day 1

9:00 - 10:00	Inca overview
10:00 - 11:00	Working with Inca Reporters
11:15 - 12:00	Hands-on: Reporter API and Repository
1:00 - 2:00	Inca Control Infrastructure
2:00 - 3:00	Administering Inca with incat
3:15 - 4:00	Hands-on: Inca deployment

Agenda -- Day 2

9:00 - 9:30	Inside the Inca Depot
9:30 – 10:10	Data display (data consumers)
10:20 - 11:00	Writing data consumers
11:00 - 12:00	Hands-on: Data display (data consumers)

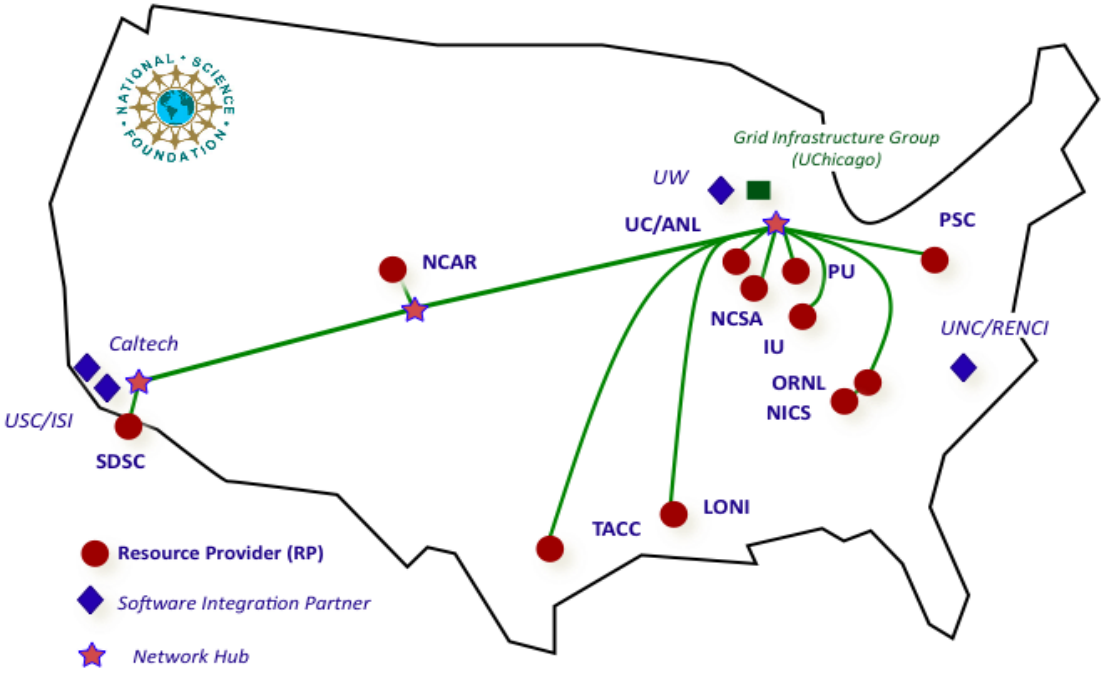
Inca Information

- Announcements: inca-users@sdsc.edu
- Email: inca@sdsc.edu
- Website: http://inca.sdsc.edu

- Supported by:

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

TeraGrid™

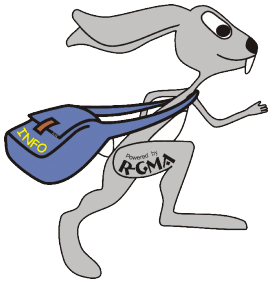
Related Grid monitoring tools

BIG BROTHER™

GridICE
the eyes of the Grid

MonALISA
MONitoring Agents using a Large Integrated Services Architecture

Nagios®



Service Availabilty Monitoring

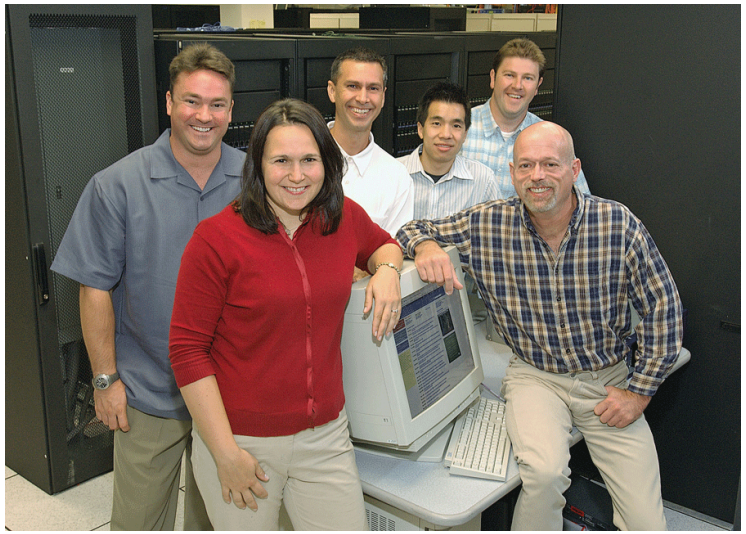
Inca's primary objective: user-level Grid monitoring

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

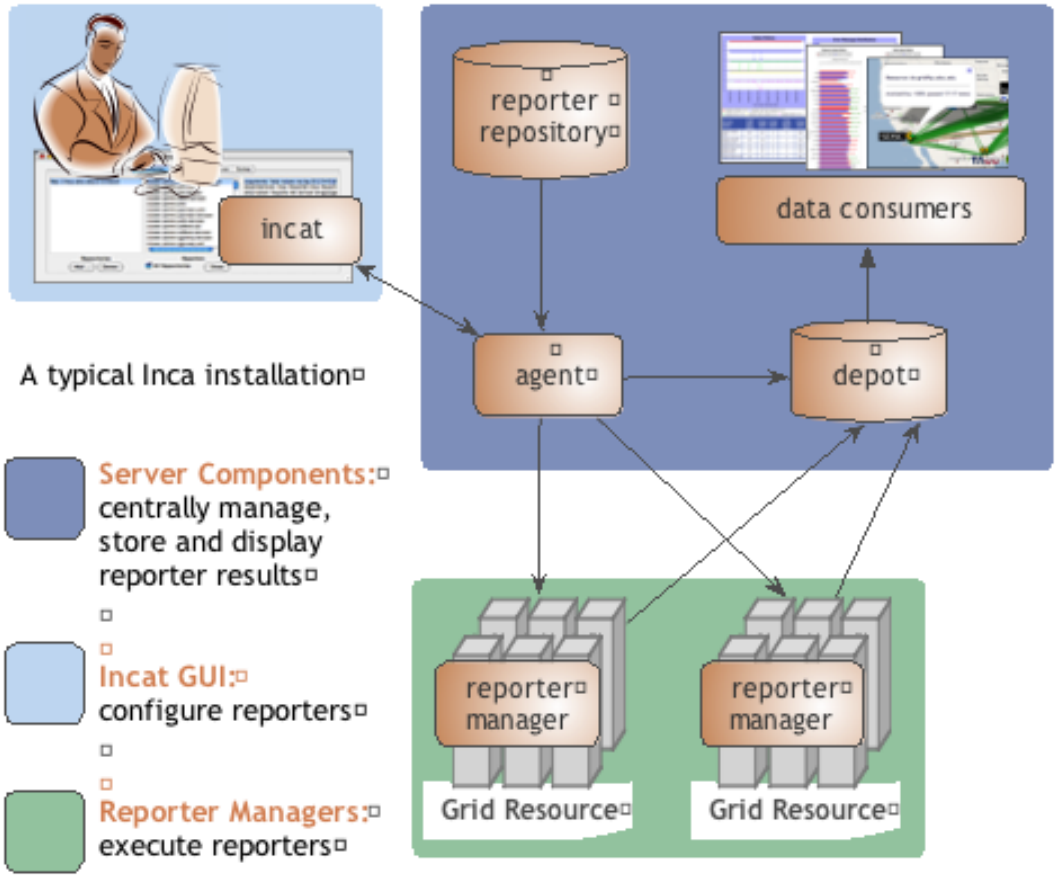
Who benefits from user-level grid monitoring?

- Grid operators
  - Verify requirements are fulfilled by resource providers
  - Identify failure trends
- System administrators
  - Email notification
  - Debugging support
- End users
  - Debug user account/environment issues
  - Advanced users: feedback to Grid/VO



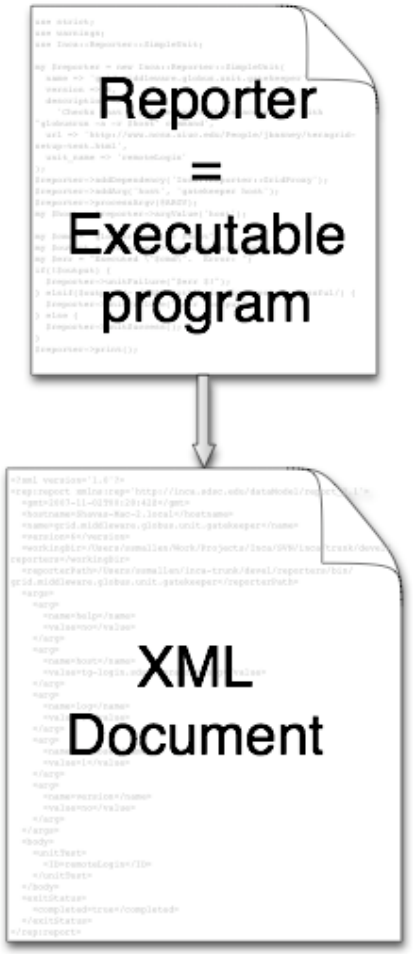
Inca provides user-level grid monitoring

- 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



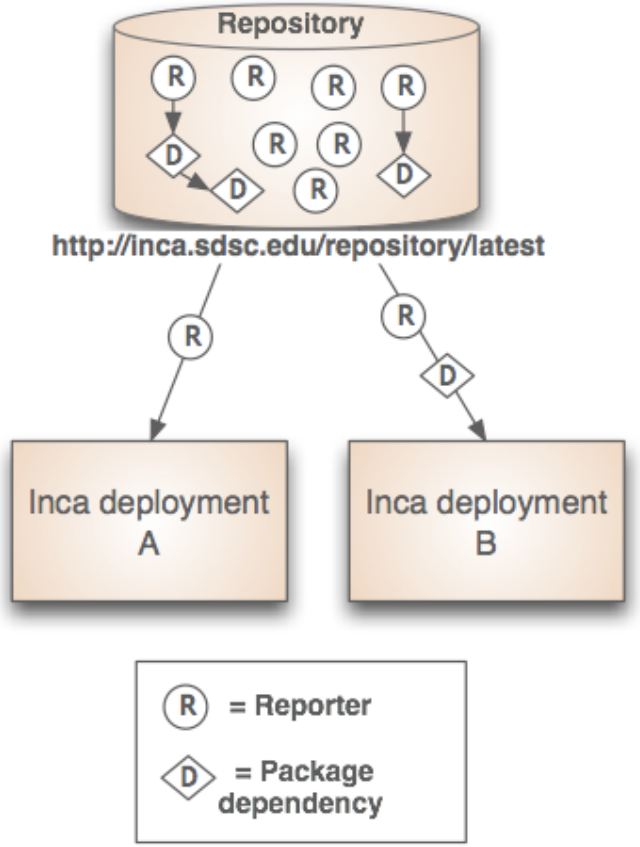
Reporters collect monitoring data

- 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



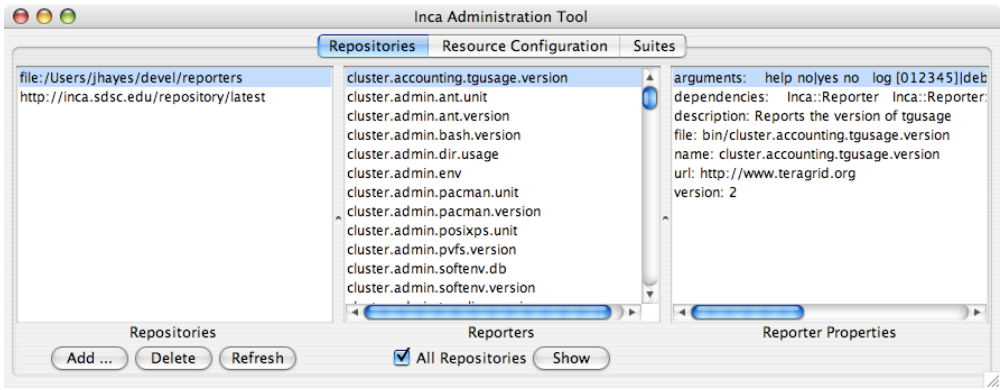
Repositories support sharing

- 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

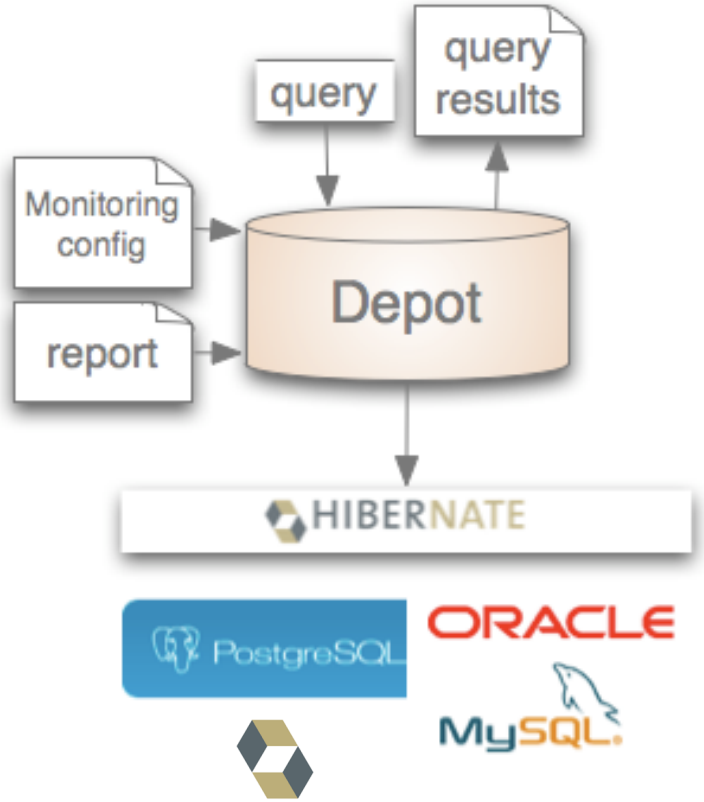
- Implements the configuration specified by Inca administrator
- Stages and launches a reporter manager on each resource
- Sends package and configuration updates
- Manages proxy information
- Administration via GUI interface (incat)



Screenshot of Inca GUI tool, incat, showing the reporters that are available from a local repository

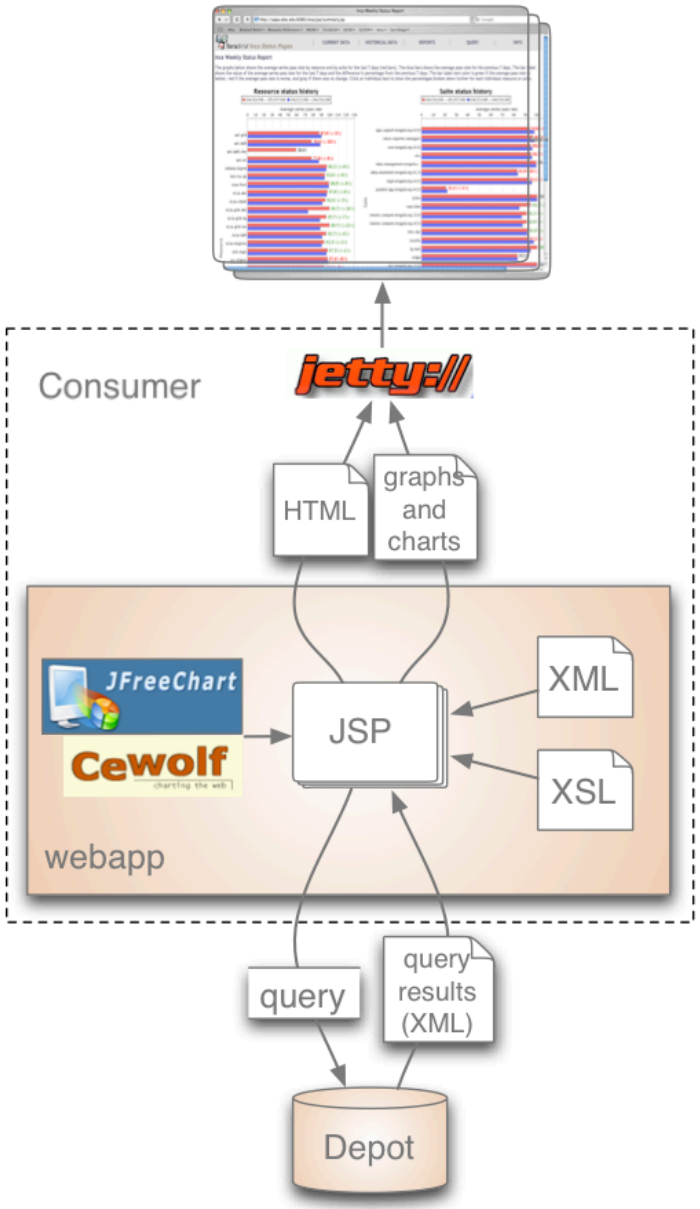
Depot stores and publishes data

- 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)
- Supports fault tolerance
- Web services - Query data from depot and return as XML

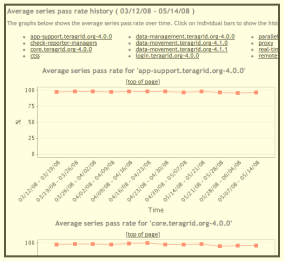


Consumer displays data

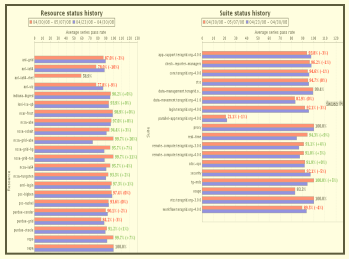
- Current and historical views
- Web application packaged with Jetty
- JSP 2.0 pages/tags to query data and format using XSLT
- CeWolf/JFreeChart to graph data
- Ability to fetch Inca data in HTML or XML format via REST URLs *new*
- Allow “run nows” from the Inca web status pages *new*



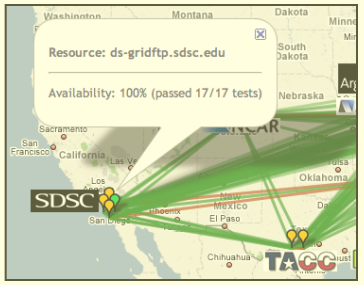
Tests Summary



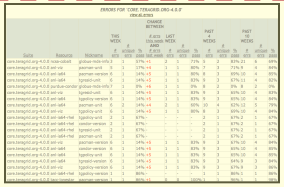
Resource status history



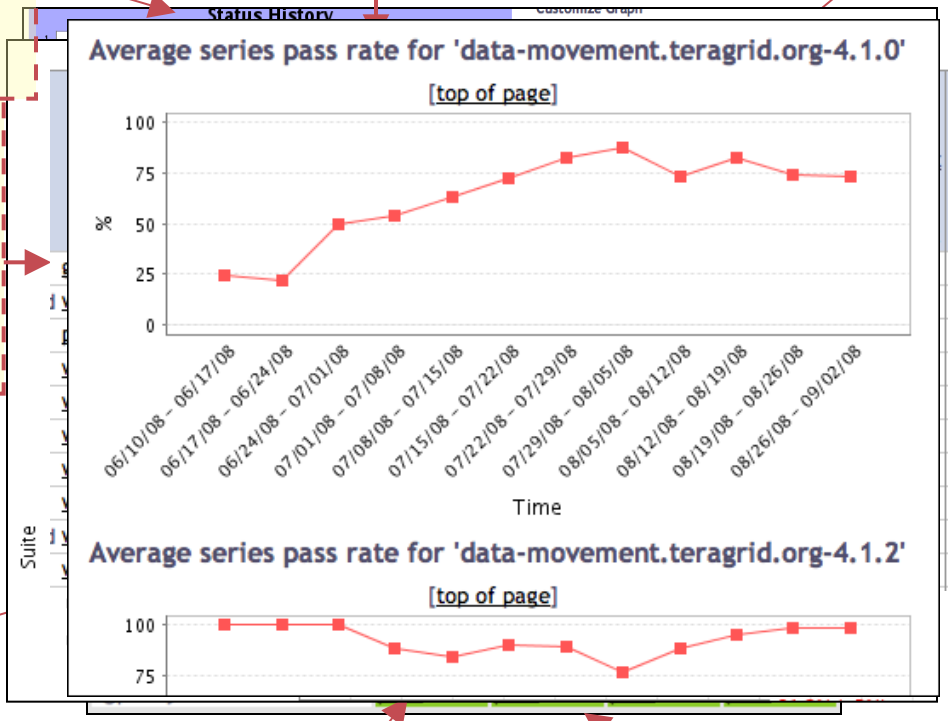
Weekly status report

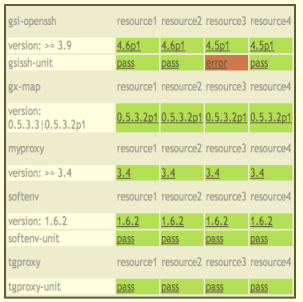


Cumulative test status by resource



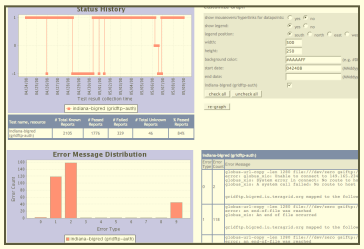
Error history summary



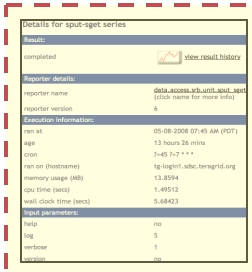


Test status by package and resource

Related test histories



Individual test history



Individual test result details

Test Details

Historical

Current status

Inca components communicate using SSL

- Provides credential based authentication for all communication
- Credentials created during setup
  % inca createauth
- Configure via inca.properties

```
inca.consumer.auth = true | false
inca. consumer.cert=componentcert.pem
inca. consumer.key=componentkey.pem
inca. consumer.trusted=trusted
inca. consumer.password=stdin:password>
# inca.consumer.depot=inca://localhost:6324
inca.consumer.depot=incas://localhost:6324
```

Software status and deployments

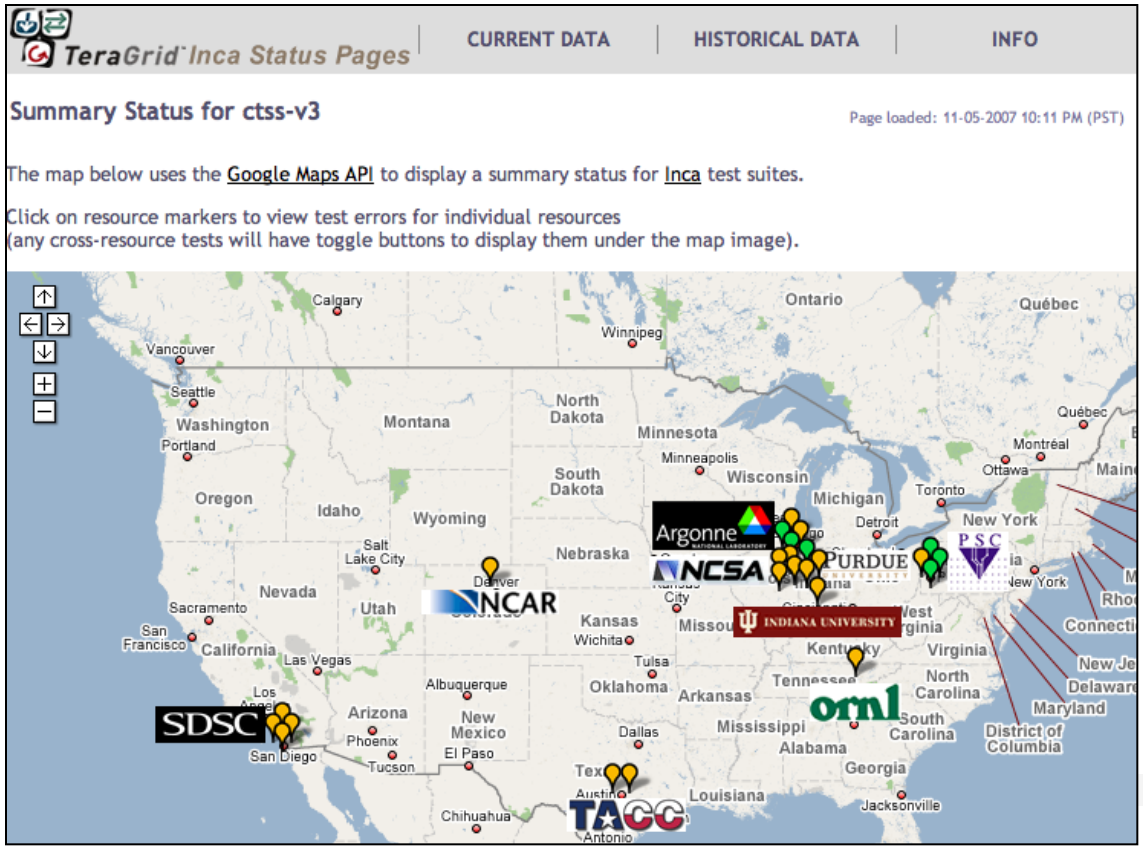
Current software version: 2.5

(final 2.6 release within a month)

http://inca.sdsc.edu

Inca TeraGrid deployment

- Running since 2003
- Total of 2660 tests running on 20 login nodes, 3 grid nodes, and 3 servers
- Coordinated software and services
- Cross-site tests
- GRAM usage
- CA certificate and CRL checking
- Resource registration in information services



Screenshot of Inca status pages for TeraGrid

http://inca.teragrid.org/

Inca monitoring benefits TeraGrid end users

"Inca reported errors mirror failures we've observed and as they are addressed we've noticed an improvement in TeraGrid's stability."

-- Suresh Marru (LEAD developer)

- Tests resources and services used by LEAD. E.g.
  - Pings service every 3 mins
  - Verifies batch job submission every hour
- Automatically notifies admins of failures
- Show week of history in custom status pages

Inca GEON deployment

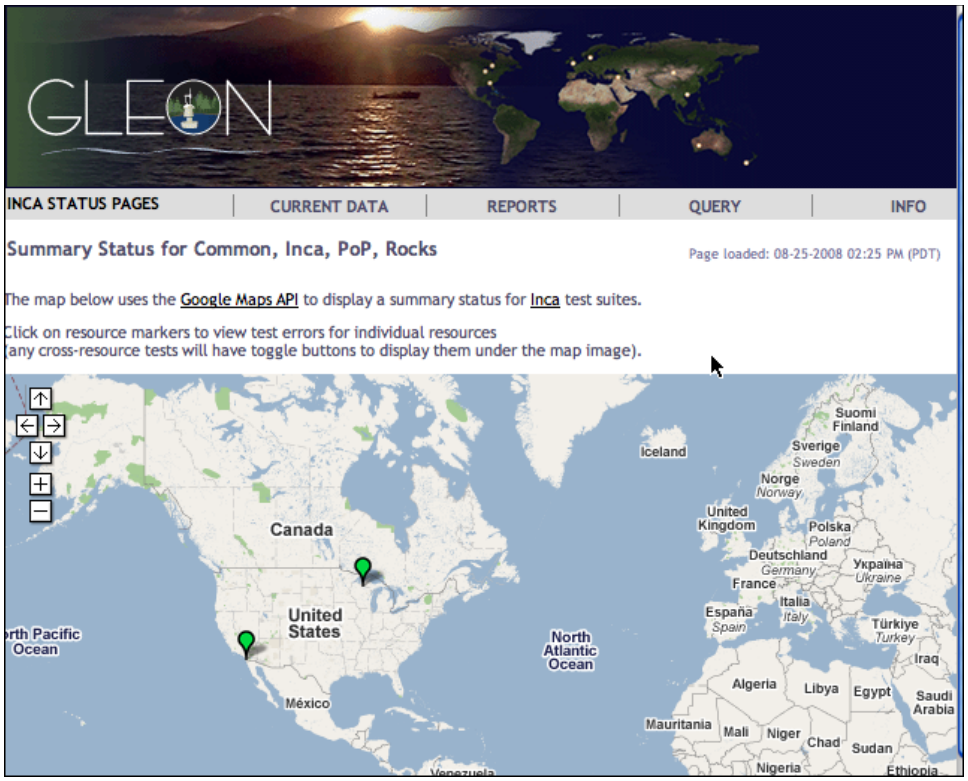
- Running since Feb 2008
- Total of 206 tests running on 5 login nodes and 6 servers
- LiDAR workflow services
- Web servers
- Ssh connectivity
- Base system information (Rocks, Gcc, Java, etc.)

Screenshot of Inca status map for GEON

http://inca-geon.sdsc.edu

Inca GLEON deployment

- Sensors in lake: dissolved oxygen level, temperature, velocity (some), etc.
- Monitoring Data Turbine deployments since Oct 2007
- Total of 26 tests running on data server at SDSC and windows box in Northern Temperate Lakes in Wisconsin



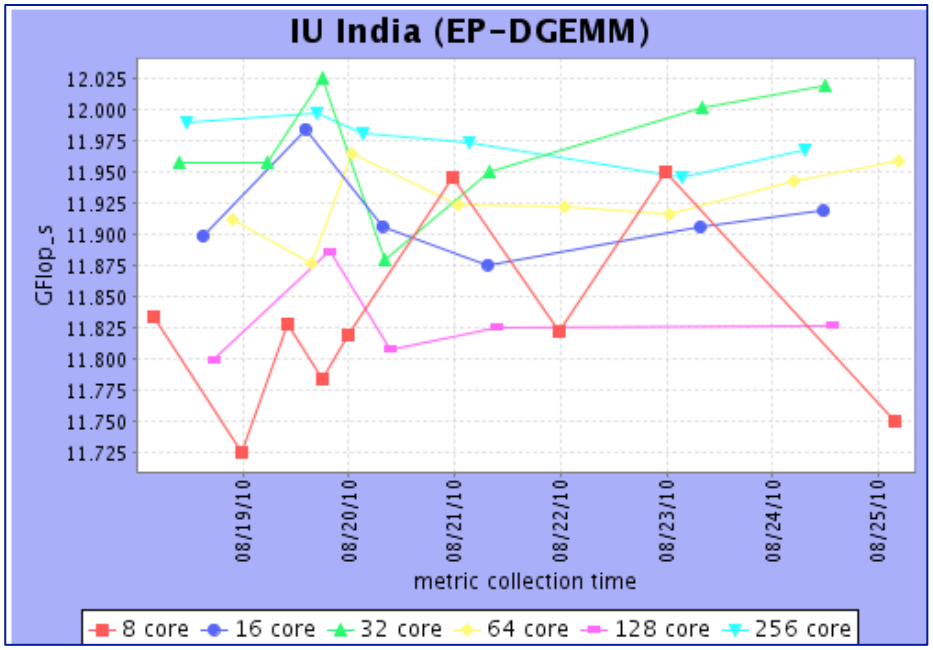
OPEN SOURCE DATA TURBINE INITIATIVE

Empowering the Scientific Community with Streaming Data Middleware

http://inca-gleon.sdsc.edu

Inca performance deployments

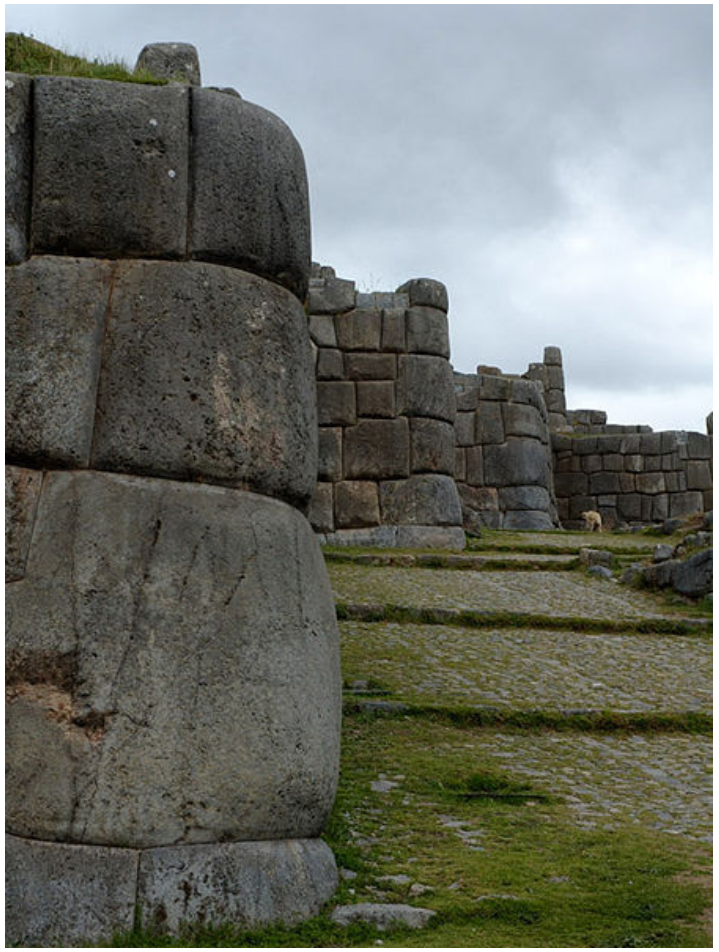
- Past
  - GrASP performance measurements in 2006
  - Deployed IPM instrumented MPI applications to TeraGrid in 2009
- New for 2.6
  - Deployed HPCC and other planned benchmarks for FutureGrid
  - Deployed HPCC for SDSC's Dash
  - Deployed MADbench, hycomm, and others to TeraGrid



Partial HPCC results on FutureGrid machine at IU

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



Future work

- More support for benchmarks
- Consolidate user interfaces
- Publish suites in Inca repository
- More data management of Inca report data (e.g., limit on stored histories)

Agenda -- Day 1

9:00 - 10:00	Inca overview
10:00 - 11:00	Working with Inca Reporters
11:15 - 12:00	Hands-on: Reporter API and Repository
1:00 - 2:00	Inca Control Infrastructure
2:00 - 3:00	Administering Inca with incat
3:15 - 4:00	Hands-on: Inca deployment (part 1)