Welcome to the
Inca 2.0 Workshop

Sponsored by the San Diego Supercomputer Center

Presenters:
Shava Smallen ssmallen@sdsc.edu
Jim Hayes jhayes@sdsc.edu
Kate Ericson kericson@sdsc.edu
Cathie Olschanowsky cmills@sdsc.edu

Thank you for attending
Workshop Goals

- Understand Inca goals and objectives
- Learn new features of v2
- Deploy or upgrade to an Inca v2 installation
- Give feedback to Inca project team

Agenda -- Day 1

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 - 10:00</td>
<td>Inca 2.0 overview</td>
</tr>
<tr>
<td>10:00 - 11:00</td>
<td>Working with Inca Reporters</td>
</tr>
<tr>
<td>11:15 - 12:00</td>
<td>Hands-on: Reporter API and Repository</td>
</tr>
<tr>
<td>1:00 - 2:00</td>
<td>Inca Control Infrastructure</td>
</tr>
<tr>
<td>2:00 - 3:00</td>
<td>Administering Inca with incat</td>
</tr>
<tr>
<td>3:15 - 4:00</td>
<td>Hands-on: Inca deployment (part 1)</td>
</tr>
</tbody>
</table>
### Agenda -- Day 2

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 - 10:00</td>
<td>Inside the Inca Depot</td>
</tr>
<tr>
<td>10:00 - 11:00</td>
<td>Data display (data consumers)</td>
</tr>
<tr>
<td>11:15 - 12:00</td>
<td>Hands-on: Data display (data consumers)</td>
</tr>
<tr>
<td>1:00 - 3:00</td>
<td>Hands-on: Inca deployment (part 2)</td>
</tr>
<tr>
<td>3:15 - 4:00</td>
<td>Wrap up</td>
</tr>
</tbody>
</table>

---

### Inca Information

- **Announcements:** [inca-users@sdsc.edu](mailto:inca-users@sdsc.edu)
- **Bugs/Feature Requests:** [http://inca.sdsc.edu/bugs](http://inca.sdsc.edu/bugs)
- **Email:** [inca@sdsc.edu](mailto:inca@sdsc.edu)
- **Website:** [http://inca.sdsc.edu](http://inca.sdsc.edu)

- **Supported by:**
  - SDSC
  - TeraGrid
  - NMI
  - PMaC

---

---
Inca 2 Architecture

1. Create a “suite”
2. Submit suite to Reporter Agent
3. Reporter Agent invokes Reporter Managers and distributes suite and reporters
4. Reporter Managers send data to Depot
5. GUIs can display collected data by querying Depot

Inca 2 Components

• Data Collection

A Reporter is an executable program that tests or measures some aspect of the system or installed software.

A Reporter Repository contains a collection of reporters and is available via an URL.

A Suite specifies a set of reporters to execute on selected resources, their configuration, and frequency of execution.
**Inca 2 Components (cont.)**

- **Administration and control**

  - **Report Manager**
  
  A **Report Manager** is responsible for managing the schedule and execution of reporters on a single resource.

  - **Report Agent**
  
  A **Report Agent** is a server that implements the configuration specified by the Inca Administrator.

  - **Incat**
  
  **Incat** is a GUI used by the Inca administrator to control and configure the Inca deployment on a set of resources.

- **Data Storage and Display**

  - **Depot**
  
  A **Depot** is a server that is responsible for storing the data produced by reporters.

  - **Data Consumer**
  
  A **Data Consumer** is typically a web page client that queries a Depot for data and displays it in a user-friendly format.
### v1/v2 Inca Components

<table>
<thead>
<tr>
<th>Components</th>
<th>Version 1</th>
<th>Version 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Consumers</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Depot</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Reporter</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Controller</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Reporter Manager</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Collector</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Incat</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Reporter Agent</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Reporter Repository</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

### Inca 2.0 Overview

Shava Smallen  
ssmallen@sdsc.edu

Inca 2.0 Workshop  
February 23, 2006
Grid Reliability

- **Grid computing**: The ability to dynamically link resources together as an ensemble to support the execution of large-scale, resource-intensive, and distributed applications

> “You know you have [a distributed system] when the crash of a computer you’ve never heard of stops you from getting any work done.” -- Leslie Lamport

Simple Grid application

Is the Grid up?

- Can user X run application[s] Y on Grid[s] Z?
  - Access dataset[s] N?
    - Can I login?
    - Are Grid services the application[s] use available? Compatible versions?
    - Are dataset[s] N accessible to user X? Credentials?
    - ...

[Diagram of grid reliability and grid up checks]
Testing a Grid

1. Iteratively define a set of concrete requirements
2. Write tests to verify requirements
3. Periodically run tests and collect data
4. Publish data

*Automate Steps 3 and 4*

What type of testing?

- Deployment testing
  - Automated, continuous checking of Grid services, software, and environment
  - E.g., gatekeeper ping or scaled down application

Software Package (unit, integrated)

Software Stack (interoperability)

Software Deployment
Who are the consumers?

- Grid/VO management
  - Responsible for designing & maintaining requirements
  - Verify requirements are fulfilled by resource providers
- System administrators
  - Notified of problems
  - Enough information to understand context of problem
- End users
  - View results and compare to problems they are having
  - Debug user account/environment issues
  - Advanced users: feedback to Grid/VO

Inca

- Inca is a framework for the automated testing, benchmarking and monitoring of Grid resources
- Inca provides:
  - Scheduled execution of information gathering scripts (reporters)
  - Data management
    - collection
    - archiving
    - publishing
Related Grid monitoring tools

Inca’s primary objective: user-level Grid functionality testing and performance measurement

Unique features of Inca

- Debugging
  - Runs under a regular user account
  - Flexibly expresses results
  - Captures reporter execution context
  - Securely re-runs reporters
  - Archives full reports
  - Reporters can be run outside framework
Unique features of Inca (cont.)

• Compares results to a specification
• Easily and securely configured
  • Data collection
  • Installation
• Profiles and logs reporter resource use

Outline

• Inca in use
• Architecture overview
• Software Status
Inca today

• Version 1
  • aka 0.10.3
  • available from website and NMI distribution

• Version 2 pre-release
  • Available as of 02/06
  • Production version available in 1-3 months

• Both versions of Inca are currently being used in production environments

Inca in use

1) Software stack validation and verification (v1)

2) Network bandwidth measurements (v1)

3) Grid benchmarking
1) Inca in use: TeraGrid software stack V&V

- TeraGrid - an “enabling cyberinfrastructure” for scientific research
  - ANL, Indiana Univ., NCSA, ORNL, PSC, Purdue Univ., SDSC, TACC
  - 40+ TF, 1+ PB, 40Gb/s net

- Common TeraGrid Software & Services
  - Common user environment across heterogeneous resources
  - TeraGrid VO service agreement

- Common software stack:
  - 20 core packages: Globus, SRB, Condor-G, MPICH-G2, OpenSSH, SoftEnv, etc.
  - 9 viz package/builds: Chromium, ImageMagick, Mesa, VTK, NetPBM, etc.
  - 21 IA-64/Intel/Linux packages: glibc, GPFS, PVFS, OpenPBS, intel compilers, etc.

- 50 version reporters: compatible versions of SW
- 123 tests/resource: package functionality
  - Services: Globus GRAM, GridFTP, MDS, SRB, DB2, MyProxy, OpenSSH
  - Cross-site: Globus GRAM, GridFTP, OpenSSH
1) Inca in use: TeraGrid deployment

- 8 sites/17 resources
- Run under user account inca

1) Inca in use: Summary status page

<table>
<thead>
<tr>
<th>Site-Resource</th>
<th>Grid</th>
<th>Development</th>
<th>Cluster</th>
<th>Total Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>site1-resource1</td>
<td>Pass: 32 Fail: 1</td>
<td>Pass: 23 Fail: 0</td>
<td>Pass: 1 Fail: 1</td>
<td>Pass: 56 Fail: 2</td>
</tr>
<tr>
<td></td>
<td>96% passed</td>
<td>100% passed</td>
<td>93% passed</td>
<td>96% passed</td>
</tr>
<tr>
<td></td>
<td>75% passed</td>
<td>100% passed</td>
<td>93% passed</td>
<td>85% passed</td>
</tr>
<tr>
<td>site2-resource1</td>
<td>Pass: 1 Fail: 18</td>
<td>Pass: 2 Fail: 10</td>
<td>n/a</td>
<td>Pass: 3 Fail: 9</td>
</tr>
<tr>
<td></td>
<td>1% passed</td>
<td>16% passed</td>
<td>n/a</td>
<td>9% passed</td>
</tr>
</tbody>
</table>

Key:
- Green: All tests passed: 100%
- Red: One or more tests failed: < 100%
- Gray: Tests not applicable to machine or have not yet been ported

History of percentage of tests passed in "Grid" category for a 6 month period

Expanded View of Errors

site1-resource1

Grid
- globus-2.4.3-nilx-3 failed: ducr_mpi_helloworld_to_jobmanager-pbs
1) Inca in use: Detailed Status View

- Select an Inca Status Page
- Select an Inca Status Page - Summary

**TeraGrid**

Common TeraGrid Software and Services 2.0: CTISS-P2
Page generated by Inca - 06/14/05 20:22 CDT

File Status of:
- atlas
- kilo2
- kilo-inv
- slshk
- slshk-inv
- bull
- mshl-q2-xc
- gmx
- mshl-q2-xc
- atlas
- kilo2
- kilo-inv
- slshk
- slshk-inv
- bull
- mshl-q2-xc
- gmx

**Resources**

- atlas
- kilo2
- kilo-inv
- slshk
- slshk-inv
- bull
- mshl-q2-xc
- gmx

**SW packages**

- atlas
- kilo2
- kilo-inv
- slshk
- slshk-inv
- bull
- mshl-q2-xc
- gmx

---

1) Inca in use: Detailed view

**Reporter details:**

- **reporter name**: grid<@>liferay@library.university keeper
- **description**: This test runs `globalkm -v [hostname]` to check that the gatekeeper at the host is accessible from the local machine.
- **version**: 1.4
- **status**: production
- **url**: [http://www.ncsa.uiuc.edu/People/libasnev/teragrid-setup-test.html](http://www.ncsa.uiuc.edu/People/libasnev/teragrid-setup-test.html)

**Execution information:***

- **inputs**: `verbose=1`, `help=no`, `log=3`, `host=test_hostname`
- **ran at (GMT)**: Wed Jun 15 00:10:02 2005
- **time**: 27 mins
- **run every**: 1 hour(s)

**Reporter system command log:**

The following are the *system* commands executed by the reporter. Note that the reporter may execute other actions in between system commands (e.g., change directories). Please click the on reporter name above for the full reporter code.

```
% globalkm -v test_hostname
```

**Host information:**

- **hostname**: ran_on_hostname
- **ipaddr**: 192.168.0.0.0
- **uname**: Linux **** (AMD) Fri Jun 3 11:44:48 EST 2005 (686) i686 i386-0M/i686
2) Inca in use: Comparison of end-to-end bandwidth measurement tools

- Joint work with Margaret Murray (TACC) and Martin Swany (UDel)
- Compare bandwidth measurement tools:
  - Pathload [Dovrolis]
  - Pathchirp [Ribeiro]
  - NWS ping [Wolski]
- Deployed to TeraGrid, GEON
- Poster presented at Grid 2005

3) Inca in use: Grid benchmarks

- GrASP: Grid Assessment Probes
  - Set of probes designed to emulate Grid applications
  - Deployed to GEON and TeraGrid
- C. Olschanowsky, O. Khalili, J. He, H. Casanova, A. Snavely. Acquiring and Using Benchmark Data from Computational Grids, submitted for publication
3) Inca in use: Measuring Grid middleware performance

3) Inca in use: Monitoring Grid middleware reliability
3) Inca in use: Error tracking over time

Outline

• Inca in use
• Architecture overview
• Software Status
1. Create a "suite"
2. Submit suite to Reporter Agent
3. Reporter Agent invokes Reporter Managers and distributes suite and reporters
4. Reporter Managers send data to Depot
5. GUIs can display collected data by querying Depot

Architecture overview: Scalable design
Outline

• Inca in use

• Architecture overview

• Software Status

New features of v2

• Full report archiving
• Flexible querying interface
• Improved installation and configuration control
  • GUI tool for centralized administration
  • Proxy management via MyProxy
  • Reporter sharing via repositories
  • Binary distribution
• Profile reporter system usage
• Inca components communicate using SSL
Software Status

• 2.0 Pre-release
  • Available as of February 6, 2006
  • More integration/stability testing
  • Not recommended for production deployments
  • Binary distribution
• 2.0 Production release in 1-3 months
  • Source and binary distributions

http://inca.sdsc.edu/prerelease.html