Inca: A Framework for Monitoring Grid Functionality and Performance

Shava Smallen, ssmallen@sdsc.edu
Kate Ericson, kericson@sdsc.edu

Supercomputing 2005
November 15 & 16
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?
  - …
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

- JUnit, PyUnit, Tinderbox
- Software Package (unit, integrated)
- Software Stack (interoperability)
- NMI
- 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

- Big Brother™
- Clumon
- Ganglia
- GridICE
- Hawkeye
- MDS
- MonALISA
- Nagios

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 (v2)
  • Archives full reports (v2)
  • Reporters can be run outside framework
**Unique features of Inca (cont.)**

- Compares results to a specification (v2)
- Easily and securely configured (v2)
  - Data collection
  - Installation
- Profiles and logs reporter resource use (v2)
Outline

• Inca in use

• Architecture overview

• Project plans
Inca today

- **Version 1**
  - aka 0.10.3
  - available from website and NMI distribution

- **Version 2**
  - scheduled for release early 2006

- *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 (v2)
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
1) Inca in use: TeraGrid software stack V&V

- **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

This page offers a summary of results for critical grid, development, and cluster tests (view list of tests). Details about a resource's test results are available by clicking on the resource name in the "Site-Resource" column of the table.

<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>50% passed</td>
<td>96% passed</td>
</tr>
<tr>
<td></td>
<td>75% passed</td>
<td>100% passed</td>
<td>50% 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: 28</td>
</tr>
<tr>
<td></td>
<td>5% passed</td>
<td>16% passed</td>
<td></td>
<td>9% passed</td>
</tr>
</tbody>
</table>

Expanded View of Errors

site1-resource1

Grid

1. globus-2.4.3-intel-r3: failed: duroc_mpi_helloworld_to_jobmanager-pbs test

Key

- All tests passed: 100%
- One or more tests failed: < 100%
- 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
1) Inca in use: Detailed Status View

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

Find Status of:
- atlas
- condor-g
- db2-client
- gcc
- globus
- gpt-3.0.1-ccc-r3
- hx-map
- hdf4
- hdf5
- mpich-g2-gcc
- mpich-p4-qcc
- myproxy
- openssh
- openssl
- python
- softenv
- srb-client
- tcl
- tusage
- urllib

<table>
<thead>
<tr>
<th>Resource</th>
<th>resource1</th>
<th>resource2</th>
<th>resource3</th>
<th>resource4</th>
<th>resource5</th>
<th>resource6</th>
<th>resource7</th>
</tr>
</thead>
<tbody>
<tr>
<td>atlas</td>
<td>3.4.1</td>
<td>3.4.1</td>
<td>3.4.1</td>
<td>3.4.1</td>
<td>3.4.1</td>
<td>3.4.1</td>
<td>3.4.1</td>
</tr>
<tr>
<td>condor-g</td>
<td>&gt;=6.5.3</td>
<td>&gt;=6.5.3</td>
<td>&gt;=6.5.3</td>
<td>&gt;=6.5.3</td>
<td>&gt;=6.5.3</td>
<td>&gt;=6.6</td>
<td>&gt;=6.6</td>
</tr>
<tr>
<td>db2-client</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>version</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unit tests</td>
<td>passed</td>
<td>error</td>
<td>passed</td>
<td>passed</td>
<td>passed</td>
<td>passed</td>
<td>passed</td>
</tr>
<tr>
<td>condor_g</td>
<td>8.1 (2 subpackages)</td>
<td>8.1</td>
<td>8.1</td>
<td>8.1</td>
<td>8.1</td>
<td>8.1</td>
<td>2 errors</td>
</tr>
<tr>
<td>unit tests</td>
<td>passed</td>
<td>error</td>
<td>passed</td>
<td>passed</td>
<td>passed</td>
<td>passed</td>
<td>passed</td>
</tr>
<tr>
<td>db2.connect</td>
<td>error</td>
<td>error</td>
<td>error</td>
<td>error</td>
<td>error</td>
<td>error</td>
<td>error</td>
</tr>
</tbody>
</table>

Resources

SW packages

- Select an Inca Status Page -
  - Select an Inca Status Page -
  Summary
  CTSS-Comute:
    Software/Services
    SoftEnv Keys
    User Environment
  CTSS-Viz:
    Software/Services
    SoftEnv Keys
    User Environment
  Other Status Monitors
# 1) Inca in use: Detailed view

<table>
<thead>
<tr>
<th>Reporter details:</th>
</tr>
</thead>
<tbody>
<tr>
<td>reporter name</td>
</tr>
<tr>
<td>description</td>
</tr>
<tr>
<td>version</td>
</tr>
<tr>
<td>status</td>
</tr>
<tr>
<td>url</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Execution information:</th>
</tr>
</thead>
<tbody>
<tr>
<td>inputs</td>
</tr>
<tr>
<td>verbose</td>
</tr>
<tr>
<td>help</td>
</tr>
<tr>
<td>log</td>
</tr>
<tr>
<td>host</td>
</tr>
<tr>
<td>ran at (GMT)</td>
</tr>
<tr>
<td>age</td>
</tr>
<tr>
<td>runs every</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporter system command log:</th>
</tr>
</thead>
<tbody>
<tr>
<td>The following are the <em>system</em> 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.</td>
</tr>
<tr>
<td>% globusrun -a -r test_hostname 2&gt;&amp;1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Host information:</th>
</tr>
</thead>
<tbody>
<tr>
<td>hostname</td>
</tr>
<tr>
<td>ipaddr</td>
</tr>
<tr>
<td>uname</td>
</tr>
</tbody>
</table>
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

![Grasp Gather Probe Execution Times on TeraGrid](image)

- **Time, in seconds, for probe to finish**
- **Cleanup**
- **Actual Queue**
- **Actual Computation**
- **Transfer Data from tg-login.purdue**
- **Transfer Data from tg-grid1.uc**
- **Transfer Data from lonestar.tacc**
- **Transfer Data from tg-login.1iu**
- **Transfer Data from tg-login.ornl**
- **Build Executable**
- **Stage Data**
- **Initialization**

**Time:**
- 9/20/2005 6:02
- 10/1/2005 5:03
- 10/16/2005 13:04
3) Inca in use: Monitoring Grid middleware reliability

Errors on Teragrid

Error Type

Error Count

0: No valid proxy found.
1: error code 12
2: Transport endpoint is not connected
3: error code 25
4: error code 70
5: error code 69
6: SoftEnv
7: error code 10
8: error code 5
9: Assertion GLOBUS FALSE
10: Network is unreachable
11: No such file or directory
12: Error with GSI credential
13: 425 425
14: error code 43
15: error code 7
16: Stale NFS file handle
17: Unexpected Gatekeeper or Service Name
18: error code 75
19: an end-of-file was reached
20: proxy is not valid long enough
21: 550 550
3) Inca in use: Error tracking over time
Outline

• Inca in use

• Architecture overview

• Project plans
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
Architecture overview: Scalable design
Outline

• Inca in use

• Architecture overview

• Project plans
Inca 2.0

• Initial version of Inca focused on basic functionality

• New features:
  • Improved archiving capabilities
  • Scalability - control and data storage
  • Usability - improved installation and configuration control
  • Monitor system impact (self-monitoring)
  • Security - SSL, proxy delegation
  • Condor integration

• Release early 2006
Improvements for TeraGrid

- **Improved GUIs**
  - User can select only information interested in

- **Historical Non-numerical data**
  - Error messages
  - Resource usage (CPU/memory used by reporters)

- **Additional reporters** - file transfer, batch queue and scheduler, compilers, HPSS, …
Summary

• Periodic, user-level functionality testing needed to monitor Grid reliability

• Inca provides a unique, automated framework for functionality testing and performance measurement

• Inca is successfully deployed on several Grids today
Future work

- Integration with knowledge base
- Standardized suites
- Automatic tuning of reporter execution frequencies to ensure low impact on resources
More information

Inca 2.0 to be released January 2006
Inca workshop tentatively scheduled for February 2006

• Email: inca@sdsc.edu

• Website: http://inca.sdsc.edu

• Supported by:
  - SDSC
  - TeraGrid™
  - NMI
  - PMaC

SAN DIEGO SUPERCOMPUTER CENTER