

Monitoring Grid Functionality and Performance with Inca

ABSTRACT

Determining whether a Grid is “up” in the face of complex Grid software deployments can be difficult and depends on the types of applications and users that utilize it. By detailing a set of software, services, and features that should be available on a Grid in a machine-readable format, a Grid can be tested periodically by an automated system to verify its health and usability to users. To this end, we’ve developed Inca as a flexible framework for the automated testing, benchmarking and monitoring of Grid systems. Inca includes mechanisms to schedule the execution of information gathering scripts, and to collect, archive, publish, and display data. While initially developed for the TeraGrid project, Inca has been successfully deployed to other Grids worldwide. This tutorial will cover the Inca 2.0 software and engage attendees in hands-on sessions to customize their own Inca installations.

TUTORIAL DESCRIPTION

Grids users execute their applications at a large scale by accessing a number of compute, data, and visualization resources. Running applications on a Grid, however, can be difficult due to the distributed, heterogeneous nature of Grid resources and their susceptibility to failures. From a user's perspective, a Grid is "up" if they can run their particular application on a set of resources, access their data and the software/services their application requires, and get reasonably good performance. A Grid may appear to be "down" to a user due to Grid software stack, hardware, and network failures. In order to provide a unified, usable system for running scientific applications, Grid architects and system administrators define a set of Grid functionality requirements and test that resource providers are meeting those requirements. Due to the dynamic nature of Grids, periodic, automated testing is necessary in order to have confidence that users are able to successfully run their applications.

This tutorial presents Inca, a unique tool for verifying the reliability, functionality, and performance of today's production Grids. Inca offers user-level, automated, continuous checking and benchmarking of a Grid's services, software and environment. Inca includes mechanisms to schedule the execution of information gathering scripts, and to collect, archive, publish, and display data. The Inca project began three years ago in order to validate and verify software stack and services for the TeraGrid, a scientific discovery infrastructure with resources from eight U.S. partner sites. Inca continues to be the only tool whose primary objective is to provide *user-level* Grid functionality testing and performance measurement. A production release of Inca 2.0 software is planned for the summer of 2006 and will be ready to extend to the wider Grid community at SCI06.

Grid managers, site system administrators, and others responsible for delivering a usable Grid are the target audience for this tutorial. Grid managers can use Inca to verify that their resource provider requirements are being met and system administrators can use it to debug user-level problems detected by Inca. Attendees will learn about the architecture and features of the Inca 2.0 software and participate in hands-on sessions to write Grid tests and deploy them within an Inca installation. Self-guided sessions throughout the tutorial will be provided for participants to work at their own pace and interests. Attendees are expected to be familiar with Grid systems, software and services and to have programming or scripting experience.

Material is based on a successful 2-day workshop provided at SDSC in February 2006 and will be updated for SCI06 with new content based on upgrades to the Inca software in 2006.

Tutorial Content:

Inca 2.0 overview:

We will describe the challenges and requirements of Grid testing and performance monitoring, provide an overview of Inca and its approach to Grid testing and performance monitoring and discuss how Inca relates to other monitoring tools. We will then show examples of Inca in use today including the Inca deployment on TeraGrid. Finally, we will discuss the status of the Inca 2.0 software and roadmap.

Working with Inca Reporters:

A reporter is an executable program that tests or measures some aspect of a Grid system or installed software. Attendees will learn about the input/output requirements for an Inca reporter and the process for publishing them into a repository where an Inca deployment can discover them. This session will also cover the Perl API libraries that facilitate the writing of reporters and will review the standard set of reporters available to users.

Inca Control Infrastructure:

The control infrastructure stages, schedules, and executes reporters on Grid resources. Attendees will learn about the control infrastructure architecture, design goals and configuration for executing sets of reporters on Grid resources. We will also describe features designed to ease installation and maintenance of an Inca deployment.

Administering Inca with incat:

Incat (Inca Administration Tool) is the GUI interface to Inca's control infrastructure. This talk parallels much of the control infrastructure talk and will illustrate how a user would configure Inca to test and/or measure performance of their Grid resources using incat.

Inside the Inca Depot:

The depot is the component of Inca responsible for storing data produced by reporters and deployment configuration. Attendees will learn about the architecture and design goals of the depot and the types of queries (including historical queries) and triggers (e.g., email) that it supports.

Inca data display:

Attendees will learn how to use JSP, tag libraries, stylesheets, and Jetty to display views of the testing and performance data collected by Inca in web pages. Specifically, we will describe an example to display the status of a common Grid software stack and environment.

DESCRIPTION OF HANDS-ON EXERCISES

1. **Reporters:** attendees will write sample reporters that exercise the Perl APIs described in Session II.
2. **Sample Inca Deployment:** attendees will learn how to specify resources, select and configure reporters, and set scheduling options. Users will install incat on their laptop and use it to connect to an Inca deployment on a machine located at SDSC in a self-guiding 2-part hands-on session. Attendees will then use incat to change the default configuration (add/delete/change reporters and add an additional resource located at SDSC).
3. **Data Display:** attendees will make small changes to the display of the Grid software stack web page from the default installation created in the hands-on exercise #2 above.

Requirements

Attendees will need laptops with SSH capability and JDK 1.4.2_09 or later. We will make the sample exercises available remotely.

TUTORIAL OUTLINE

Course Outline (morning):

- I. Inca 2.0 overview (45 mins)**
 - a. Grid testing and performance monitoring challenges and requirements
 - b. Inca goals and objectives
 - c. Inca in use
 - i. Software stack validation and verification
 - ii. Network bandwidth measurements
 - iii. Grid benchmarking
 - d. Inca 2 Architecture/Components
 - e. Inca Software Status
- II. Working with Inca Reporters (40 mins)**
 - a. Goals and objectives
 - b. Input/Output requirements
 - c. Perl Reporter APIs and examples
- III. Hands-on: Reporter API and Repository (35 mins)**
- IV. Inca Control Infrastructure (30 mins)**
 - a. Goals and objectives
 - b. Description of control components
 - c. Configuration options
 - d. Installation and maintenance
- V. Administering Inca with incat (30 mins)**
 - a. Installation and execution
 - b. Specifying reporter repositories
 - c. Specifying resources
 - d. Managing suites and report series

Course Outline (afternoon):

- VI. Hands-on: Inca deployment - part 1 (30 mins)**
- VII. Inside the Inca Depot (30 mins)**
 - a. Goals and objectives
 - b. Architecture
 - c. Storage capabilities
 - d. Queries and triggers
- VIII. Inca data display (25 mins)**
 - a. Overview
 - b. Querying the depot
 - c. Rendering depot query results in a web page
- IX. Hands-on: Inca data display (25 mins)**
- X. Hands-on: Inca deployment - part 2 (60 mins)**