

Project 1706 NetKarma:

Description of Prototype

Sep 30, 2010

→ netKarma persistent service is available at

<http://bitternut.cs.indiana.edu:42816/axis2/services/KarmaService>

→ NetKarma Provenance Repository resides on a server in the GENI Meta-Operations Center (GMOC), located at Indiana University. Further information about the NetKarma and GENI Adaptor software is available at http://pti.iu.edu/d2i/provenance_netkarma.

Prototype Description

The netKarma persistent service is Axis2 web service is available at <http://bitternut.cs.indiana.edu:42816/axis2/services/KarmaService>. It is an Axis2 web service and has a WSDL access API so provenance can be stored and retrieved programmatically. NetKarma Provenance Repository resides on a server in the GENI Meta-Operations Center (GMOC), located at Indiana University. Further information about the NetKarma and GENI Adaptor software is available at http://pti.iu.edu/d2i/provenance_netkarma.

NetKarma is a tool and repository for capturing and accessing the workflow of GENI experiments at multiple layers in the PlanetLab stack, including slice creation, topology of the slice, operational status, and links to measurement data. It does so from a provenance perspective, which is an orientation around the contextual information of an experiment. The purpose of an integrated provenance view is not so much reproducibility of the experiment, but in exposing a view of the experiment as collected in real time that can aid in its understanding and can facilitate sharing of results because provenance underlies perceptions of quality. NetKarma consists of Karma provenance collection tool and persistent web service as an underlying infrastructure, and the Advanced Message Queuing Protocol standards compliant RabbitMQ messaging service used to ingest notifications to Karma database. As part of this prototype, we developed a GENI version of the Adaptor instrumentation type. The GENI adaptor provides an interface that uses the GENI experiment log files and a set of rules to derive provenance information and maps them into the Karma repository. The adaptor is simply a generic log processing unit for GENI log files, which comprise of two sub-units: Log Parser, Notification Generator. The Log Parser module is used to process log files to extract provenance information, while the Notification Generator is used for generating and sending provenance notifications to Karma repository. The GENI Adaptor has also capability to create visual graph from Open Provenance Model v.1.1 formatted workflow graphs.

We apply provenance capture to an experiment running on GENI network and capture provenance information. We focused on provenance collection from Gush (<http://gush.cs.williams.edu/trac/gush>) to capture provenance information available at the experiment level about a particular experiment and its execution. Gush is an extensible execution management system for GENI. To facilitate testing of the prototype, we used Twister (<http://www.iterativemapreduce.org>), a parallel, iterative version of the MapReduce to execute a crawling application using breath-first search through a large-scale random graph. By utilizing MapReduce programming framework, the application explores the nodes of the same level of the graph in parallel, and then goes to the nodes in next levels iteratively. This way, it is able to process breadth-first graph search in parallel. We utilized PlanetLab, where we run Gush to deploy and execute the Twister experiment. Gush requires that users describe

their experiments or computation in an XML document. It uses this document to locate and access the remote resources in PlanetLab. In its execution flow, Gush contacts a host to deploy a twister server, which then reads a configuration file and internally connects to other hosts where the application needs to run. In this experimental study, provenance of both successful and failed executions (because of improper setup of the application) captured.