

NetKarma Quarterly Report 10/1/2009-12/31/2009

Beth Plale – Principal Investigator

Chris Small – Co-Principal Investigator

Summary

- *Web Site and Internal Wiki deployed*
- *Selection of GUSH and Raven projects as initial integration partners*
- *Karma scalability improvements*
- *Karma 3.0 release expected 2010 Q1*
- *Development area for integration efforts*

Major Accomplishments

Milestones Achieved

Project Web Page: <http://www.dataandsearch.org/provenance/?q=node/33> documents activities in the NetKarma project. This page will link to future documentation and code developed for NetKarma. An internal wiki has also been created to allow for internal documentation of our test infrastructure and development.

Integration partners: We established communication with John Hartmann, University of Arizona, on Raven, and Jean Albrecht, Williams College, on GUSH. This initial stage has been one of relationship building and gaining a better understanding of the monitoring and logging support these tools provide.

Deliverables Made

As per our milestones, we delivered a whitepaper on Karma at the termination of GEC6. The whitepaper can be found at http://groups.geni.net/geni/attachment/wiki/netKarma/NetKarma_GEC7_deliverable_Plae-Small.pdf

Description of Work Performed During Last Quarter

Activities and Findings

Karma scalability improvements: We developed a workload generator that can generate large amounts of provenance information. The workload we are using is made up of a collection of scientific workflows that were gathered from a survey we conducted. The workload generator generates pseudo-realistic provenance information using process execution times and data product sizes derived from real workflows. We are using this workload generator to build a multi-Gigabyte database of provenance information and using it to test the provenance server's ingest speeds, query performance, and the overall scalability of the data storage server.

Development Area: Created a development area with installations of the GUSH and Raven software. Using an existing Planetlab slice to create client side logs for possible ingestion into the NetKarma framework. Restored Indiana University Planetlab site to working condition.

Project Participants

During this time, key participants in the NetKarma project included:

Beth Plale, PI
Chris Small, Co-PI
Girish Subramanian, PhD student
Sharanya Chinnusamy, MS student
David Ripley, technical staff
Robert Ping, Project and Information Management

Publications & Documents

NetKarma: GENI Provenance Registry *Poster presented at the 6th GENI Engineering Conference*
http://groups.geni.net/geni/attachment/wiki/netKarma/NetKarma_Poster.pdf

GENI Documents:

Plale, B., Small, C., NetKarma: a tool for obtaining a provenance-based record of experimentation

http://groups.geni.net/geni/attachment/wiki/netKarma/NetKarma_GEC7_deliverable_Plaie-Small.pdf

Outreach Activities

This is the first quarter of the project; no outreach efforts to report.

Collaborations

Meeting with Prasad Calyam of the OnTimeMeasure Spiral 2 Project. Discussed future integration of measurement data captured by the OnTimeMeasure infrastructure into the NetKarma provenance collection.

Planned Activities for Q2

The upcoming quarter will be spent developing instrumentation extensions to Karma to ingest provenance from GUSH. This will likely be through building an adapter that takes in log files, audit streams, and/or performance monitoring streams, and sorts through the information, converting provenance-useful information into messages using the Karma provenance tracking library.

We also need to set up a meeting with Christopher Small to discuss and identify partners at the control plane layer.

We intend to release Karma 3.0 the first quarter of 2010. Karma v3.0 will support instrumentation through Axis 2 handlers in addition to Java applications. It will also include asynchronous communication using WS-Messenger, a publish-subscribe system that is an implementation of latest WS-Notification and WS-Eventing.

A demo of interoperability with GUSH is planned for GEC7 in March 2010.

.