Tool for Navigating Provenance Information

Vikas Deora, Arnaud Contes and Omer Rana

Cardiff University

PROVENANCE

Provenance

Contents

- Grid Provenance Project
- Tool Overview
- Tool for Provenance Visualization
- Screen shots
- Demo
- Further information

Grid Provenance Project

- **IBM United Kingdom Limited**
  - Alexis Biller
  - John Ibbotson
  - Neil Hardman
- **University of Southampton**
  - Luc Moreau
  - Paul Groth
  - Simon Miles
  - Victor Tan
  - Sheng
- **University of Wales, Cardiff**
  - Arnaud Contes
  - Omer Rana
  - Vikas Deora

- **German Aerospace Center (DLR)**
  - Andreas Schreiber
  - Guy K. Kloss
  - Tahjis
- **Universitat Politecnica de Catalunya**
  - Javier Vazquez
  - Sergio Alvarez
  - Steven Willmott
- **MTA SZTAKI Computer and Automation Research Institute, Hungarian Academy of Sciences**
  - Laszlo Varga
  - Tamás Kifor

Tool Overview

- The tool provide support to navigate through and analyze provenance information based on the use of a portal framework (eXo in our case).

Tools Overview
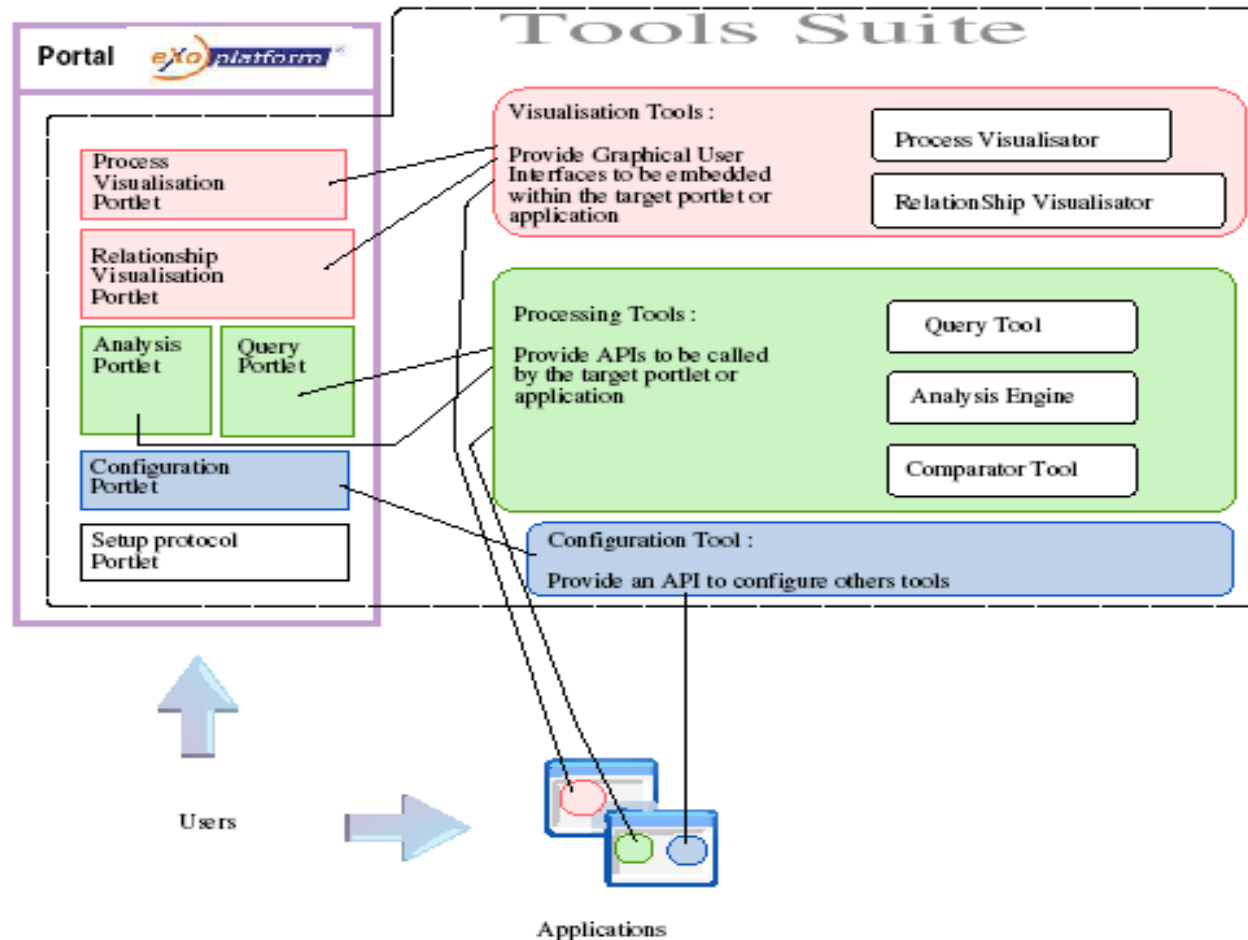
Concepts

- A set of provenance-aware actors involved in a process generate data about the execution. The data produced is composed of a set of *p-assertions.*

- Such set of p-assertions provide the description of the physical process.

- A p-assertion can be used to record one of the following events: an *interaction* between two actors, a *relation* between two events, or the state of an actor at a particular moment. In our system, interaction and relationship p-assertion are presently used.

Tools for Provenance Visualization

- The portal and portlets provide users with a set of tools to navigate through and analyze a set of p-assertions that represent an executed process.

- Interaction with a portal is made available using a Web client (browser).

- On receiving a user request to re-construct a process, the portlet interacts with local and external Provenance Store to retrieve all the p-assertions related to a particular process execution.

Visualization Portlet

- The visualization portlet displays two graphs:
    - process graph and
    - relationship graph

- The above are based on interaction and relationship p-assertions respectively.

Process Graph

- By capturing all the interactions that take place between actors involved in the computation of some data, one can replay an execution, analyze it, verify its validity or compare it with another execution.

- A crucial element of an interaction p-assertion is information to identify a message uniquely. Such information allows us to establish a flow of data between actors.

Relationship Graph

- While matching interaction p-assertions denote a flow of data between actors, relationships explain how data flows inside actors. Relationship p-assertions are directional since they explain how some data was computed from other data.

Screen shot – Process Visualization

Screen shot – Relationship Visualization

Screen shot – Provenance Challenge

Provenance Challenge Demo

- Demo

Further information

- http://sprocket-comsc.grid.cf.ac.uk:8080/portal

- Publication in Web Intelligence 2006 Conference