LaTeX Triana for Big Data

ABHISHEK NAIK[1]

[1] *School of Informatics and Computing, Bloomington, IN 47408, U.S.A.*
[*] *Corresponding authors: ahnaik@indiana.edu*

paper-1, February 25, 2017

Overleaf system or Sharelatex systems.

Keywords: Cloud, I524

https://github.com/absnaik810/sp17-i524/blob/master/paper1/S17-IR-2022/report.pdf

INTRODUCTION

Triana is an open source Problem Solving Environment (PSE) that is supported with a powerful data analysis tool. It is predominantly used for image, text and signal processing; and comes with a host of different tools for analysis. Besides, it also provides features for easy integration of custom analysis tools. It thus focuses on supporting services from various environments, like peer-to-peer (P2P) and Grid.

BACKGROUND

[1] Triana has a graphical interactive environment that can be used by the users to create their applications and specify their behavior. In this, it is similar to tools like [2] draw.io and [3] IBM's Rational Rose that provide the user with a graphical user interface (GUI) to draw the UML diagrams. The users can model the work flow using the various units onto the workspace and then depict the relationship between them with some interconnection.

[4] Triana was originally developed for the GEO600 gravitation project. In this project, it is still being used for analysing the gravitational wave signals that emanate from the laser interferometer based out of Germany. [**?**] Triana uses a pluggable architecture, that is it checks the inputs coming to, and the outputs coming from various units in real time to perform data-type checking and conformation. It can also be used to monitor the work flow and run the executables in standalone mode. [5] It has a custom work flow language, although it can be integrated with others like the Business Process Execution Language (BPEL). This helps it in analysing large data sets and makes it particularly important in big data analysis.

INFRASTRUCTURE

[**?**] The Triana infrastructure consists of the Triana Controlling Service (TCS) that has a Triana Engine, implemented as a Triana service. The Triana engines are free to carry out the execution either locally or on distributed servers, as per the implementation policies in force. Communication in the later case can be carried out using pipelined work flow distributions.

The Triana Distributed implementation makes use of Triana Group units. They have the same features of the normal units (like input/output, etc.) and so they can be connected to the other Triana Units using the standard connection mechanism. The tools need to be group so that they can be distributed; and Triana has two distribution policies:

- Parallel: a 'farming-out' mechanism involving no communication between the hosts.
- Pipeline: this involves distributing the group 'vertically'.

Groups can in turn contain groups and each group can in turn, have its own distribution policy to be followed. The Triana distribution mechanisms are based upon the concept of GAT (Grid Application Toolkit). The GAT aims to shield the applications from the implementation details using a standard Application Programming Interface (GAT-API). It also provides a set of Grid services for carrying out tasks such as resource and information management.

PROJECTS

[6] Triana has been involved in a lot of projects (including some Big Data projects) like the [7] Data-Mining Grid, [8] Scalable Robust Self-organizing Sensor Network Project (SRSS) and [9] SHaring Interoperable Work flows for large-scale scientific simulations on Available DCIs (SHIWA), etc. As a part of the Data-Mining Grid, Triana was predominantly used to model and then manage the planning, development and the execution of data-mining work flows in grid computing environments. [8] The SRSS project is carrying out a research about the various communication protocols that can be leveraged in distributed and self-organizing networks; and they are using Triana's P2P binding to simulate various P2P networks. The European project, SHIWA, mainly focuses on the interoperability of myriad European Work flow Systems and has been intrinsically integrated

with Triana to create [9] SHIWA bundles. Other than these, Triana is also being actively used in some other projects like EDGI, TRIACS, EDGES, WHIP (Work flows Hosted In Portals), DART (Distributed Audio Retrieval using Triana), GridOneD, GEO600, BiodiversityWorld, DIPSO, GEMMS, etc. many of which are related to Big Data.

CONCLUSION

Thus, to conclude, we presented Triana and the Triana PSE. We focused on the Triana infrastructure as well as its distributed implementation. We also presented some Big Data projects that used Triana for their development.

REFERENCES

[1] S. Majithia, I. Taylor, M. Shields, and I. Wang, "Triana as a graphical web services composition toolkit," in *Cardiff School of Computer Science, Cardiff University, Cardiff*, 2003. [Online]. Available: http://www.trianacode.org/papers/pdf/TrianaAHM03WS.pdf

[2] "Draw-io-website," Code Repository, JGraph, accessed: 2017-2-20. [Online]. Available: https://github.com/jgraph/draw.io

[3] "Ibm-rational-rose-website," Web Page, IBM, accessed: 2017-2-20. [Online]. Available: http://www-03.ibm.com/software/products/en/rosemod

[4] "Geo600-website," Web Page, Max Planck Gesellschaft, accessed: 2017-2-20. [Online]. Available: www.geo600.uni-hannover.de

[5] F. Pop, J. Kolodziej, and B. DiMartino, Eds., *Resource Management for Big Data Platforms*. Springer Nature, 2016, pp. 48–49.

[6] "Triana-project-list," Code Repository, Cardiff School of Computer Science & Informatics, accessed: 2017-2-20. [Online]. Available: https://github.com/CSCSI/Triana/wiki/Projects

[7] "Data-mining-grid," Web Page, University of Ulster, accessed: 2017-2-20. [Online]. Available: http://www.datamininggrid.org/

[8] "Srss," Web Page, US Naval Research Laboratory, accessed: 2017-2-20. [Online]. Available: http://srss.pf.itd.nrl.navy.mil

[9] "Shiwa-workflow," Web Page, Computer and Automation Research Institute, Hungarian Academy of Sciences - MTA SZTAKI, accessed: 2017-2-20. [Online]. Available: http://www.shiwa-workflow.eu/