

A Grid Based Approach to Share Public Health Surveillance Applications - The R Example

Kailah Davis*¹ and Julio Facelli^{1,2}

¹Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA; ²Center of High Performance Computing, University of Utah, Salt Lake City, UT, USA

Objective

This poster describes an approach which leverages grid technology for the epidemiological analysis of public health data. Through a virtual environment, users, particularly epidemiologists, and others unfamiliar with the application, can perform on-demand powerful statistical analyses.

Introduction

Currently, there's little effective communication and collaboration among public health departments. The lack of collaboration has resulted in more than 300 separate biosurveillance systems (1), which are disease specific, not integrated or interoperable, and may be duplicative (1). Grid architecture is a promising methodology to aid in building a decentralized health surveillance infrastructure because it encourages an ecosystem development culture (2), which has the potential to increase collaboration and decrease duplications.

Methods

This project had two major steps: creation and validation of the grid service. For the first step [creation of the service], we first determined the parameter set required to execute R from the command line. We then used the caGrid Introduce toolkit (3) and Grid Rapid Application Virtualization Interface (gRAVI) (4) to wrap the R command line interface into a grid service. The service was then deployed to the caGrid training grid. After deployment, the service was invoked using the R grid service client which was automatically created by Introduce and gRAVI.

Our second step was aimed at validating the service by using using the grid service client to illustrate the working principles of R in a grid environment. For this illustration, we selected the article by Hohle et al (5). In this article, the 'surveillance' package was developed to provide different algorithms for the detection of aberrations in routinely collected surveillance data. For validation purposes, only a subset of the analyses presented in the article, namely the Farrington and CUSUM algorithms, were reproduced. Using the grid web client, we uploaded the necessary data files for processing, as well as the Rscript which was used to replicate the results of (5). The application then ran the R script on the execution machine; this machine had all the necessary R packages needed for the specific scenario.

Results

The implementation of was validated by showing that the results of the original paper can be reproduced using grid based version of R. Figure 1 shows the plots related to the steps described above; the plots illustrating the Farrington and CUSUM algorithms are seen to be identical to that in (5).

Conclusions

We demonstrated that it is possible to easily deploy applications for public health surveillance uses. We conclude that the techniques we used could be generalized to any application that has a command line

interface. Future work will be aimed creating a workflow to access data services and grid-enabled text processing and analytic tools. We believe that by providing a set of examples to demonstrate the benefit of this technology to public health surveillance infrastructure may provide insight that may lead to a better, more collaborative system of tools that will become the future of public health surveillance.

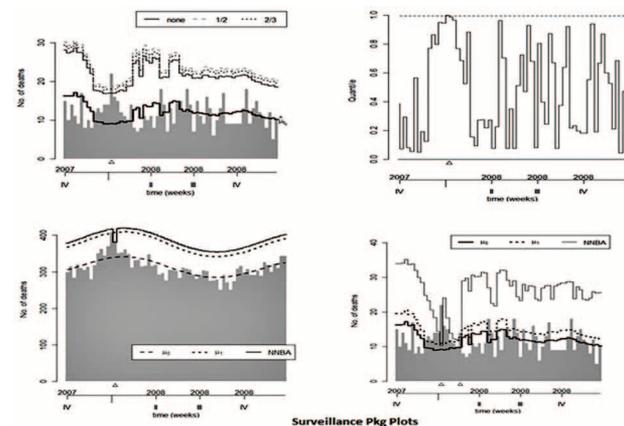


Fig 1. Recreated Plots

Keywords

Grid computing; Public health grid; analytical service

Acknowledgments

This work was supported by NLM training grant #T15LM007124 and CDC Center of Excellence for Public Health Informatics # 1P01HK000069-10.

References

1. Subcommittee NBA. Improving the Nation's Ability to Detect and Respond to 21st Century Urgent Health Threats: First Report of the National Biosurveillance Advisory Subcommittee. 2009.
2. Facelli JC. An agenda for ultra-large-scale system research for global health informatics. *ACM SIGHT Record*. 2012;2(1):12-.
3. Hastings S, Oster S, Langella S, Ervin D, Kurc T, Saltz J. Introduce: an open source toolkit for rapid development of strongly typed grid services. *Journal of Grid Computing*. 2007;5(4):407-27.
4. Chard K, Tan W, Boverhof J, Madduri R, Foster I, editors. Wrap scientific applications as WSRF grid services using gRAVI. 2009: IEEE.
5. Höhle M, Mazick A. Aberration detection in R illustrated by Danish mortality monitoring. *Biosurveillance: Methods and Case Studies*. 2010:215-37.

*Kailah Davis

E-mail: kailah.davis@utah.edu

