COACTION PUBLISHING

Cloud computing for syndromic surveillance

Frank Hardisty*

The Pennsylvania State University, University Park, PA, USA

Objective

To describe how use of cloud computing resources can improve the timely provision of disease surveillance analyses.

Introduction

Two significant barriers to greater use of syndromic surveillance techniques are computational time and software complexity. Computational time refers to the time for many methods (for example, scan statistics and A Multidirectional Optimum Ecotope-Based Algorithm [AMOEBA] statistics) to create reliable results. Software complexity refers to the difficulty of setting up and configuring suites of software to collect data, analyze it and visualize the results. Both of these barriers can be partially surmounted by the use of cloud computing resources.

Methods

We used Amazon's EC2 (Elastic Compute Cloud) services to experiment with cloud computing for syndromic surveillance. We applied two cloud service models: Infrastructure as a Service (IaaS) and Software as a Service (SaaS).

Our first goal was to apply cloud computing technologies in order to reduce computational time needed for syndromic analyses. Scan statistics, due to their reliance on Monte Carlo simulation to find confidence levels, are particularly well suited to being improved by parallel computation. We used the R package DCluster to calculate scan statistics and combined that with the SNOW (Simple Network of Workstations) package. We also experimented with using cloud computing to parallelize the AMOEBA approach to cluster detection.

Our second goal was to determine the practicality of easing the barrier of software complexity. To that end, we created software packages that include data import, analysis and visual presentation of results and released them as freely available virtual machines, or images, for the public to use. The GeoViz Toolkit was one of the software packages delivered in this manner (Fig. 1).

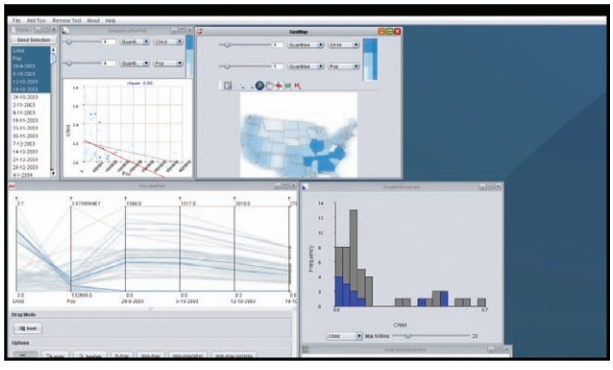

Fig. 1. The GeoViz Toolkit is shown with multiple views of rates of ILI. The spatiotemporal clustering algorithms in the GeoViz Toolkit, as well as the multiple view visualizations, are examples of syndromic analytic functions that can be improved using cloud computing.

Results

We found that both Infrastructure as a Service and Software as a Service cloud computing service models can help reduce barriers to effective use of sydromic surveillance methods. Easy provision of many computers allowed us to speed up the computational times by an order of magnitude. The creation of integrated software services to perform disease surveillance is the easiest way to deliver complex functionality.

Conclusions

We conclude that, in the future, cloud computing can and should play a more prominent role in disease surveillance.

Keywords

Cloud computing; scan statistics; informatics

Acknowledgments

We would like to thank the NGA University Research Initiative (NURI) program and Amazon Web Services, for their support. No endorsement is implied.

References

1. Hardisty F, Robinson A. The GeoViz Toolkit: using component-oriented coordination methods to aid geovisualization application construction. Int J Geogr Inform Sci. 2011;25:191–210.
2. Aldstadt J, Getis A. Using AMOEBA to create a spatial weights matrix and identify spatial clusters. Geogr Anal. 2006;38:327–43.
3. Kulldorff M. A spatial scan statistic. Commun Stat Theor Meth. 1997;26:1481–96.

***Frank Hardisty**
E-mail: hardisty@psu.edu

Citation: Emerging Health Threats Journal 2011, **4**: 11590 - DOI: 10.3402/ehtj.v4i0.11590