ISDS 2014 Conference Abstracts

Evaluation of Praedico™, A Next Generation Big Data Biosurveillance Application

Mark Holodniy*[1], Carla Winston[1], Cynthia A. Lucero-Obusan[1], Gina Oda[1], Anosh Mostaghimi[1], Julie A. Pavlin[3], Payam Etminani[2], Chris Lee[2] and Farshid Sedghi[2]

[1]Office of Public Health, Department of Veterans Affairs, Palo Alto, CA, USA; [2]Bitscopic Inc., Redwood City, CA, USA; [3]Armed Forces Health Surveillance Center, Silver Spring, MD, USA

Objective

The purpose of our study was to conduct an initial assessment of the biosurveillance capabilities of a new software application called Praedico™ and compare results obtained from previous queries with the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE).

Introduction

The National Strategy for Biosurveillance promotes a national effort to improve early detection and enable ongoing situational awareness of all-hazards threats. Implicit in the Strategy's implementation plan is the need to upgrade capabilities and integrate multiple disparate data sources, including more complete electronic health record (EHR) data into future biosurveillance capabilities. Thus, new biosurveillance applications are clearly needed. Praedico™ is a next generation biosurveillance application that incorporates cloud computing technology, a Big Data platform utilizing MongoDB as a data management system, machine-learning algorithms, geospatial and advanced graphical tools, multiple EHR domains, and customizable social media streaming from public health-related sources, all within a user friendly interface.

Methods

Initially, 1 million VA patient records were analyzed as a training set for validation of the Extract-Transform-Load (ETL) layer into a designated biomart within the VA Healthcare-Associated Infection and Influenza Surveillance System (HAIISS) and analyzed for utilization, processivity, searchability, display functions, timing and accuracy. In addition, a validation set of combined VA and Department of Defense (DoD) biosurveillance data comprising 17 million DoD and 25 million VA records respectively, was used to assess the performance of Praedico™ using ICD-9 encounter codes from known influenza-like illness (ILI) outpatient visits previously analyzed using ESSENCE alone[1].

Results

Utilizing a big data NoSQL database backend, Praedico™ required significantly less data storage and utilization, while providing the same analytical capabilities. Praedico™ was flexible in terms of hardware and software requirements using a modular design that could run on commodity hardware and any major operating system including Linux, Windows and OS X. Using a single operation in Praedico™, the number of ILI encounters in VA and DOD as a proportion of the total number of cases during that same 2007-2008 time period (Figure 1) could be queried and displayed in seconds. Praedico™ dashboard also geographically displayed the results for both VA and DoD on a single map using a geospatial tool embedded directly in the application. Both applications displayed the same number of VA ILI cases. Additionally, new ILI queries could be run within the Praedico™ dashboard and updated results displayed in real time.

Conclusions

Our analysis demonstrates improved resource utilization and response time for Praedico™, and results were comparable to VA ESSENCE.

Figure 1. Praedico dashboard displaying VA and DoD ILI Data

Keywords

biosurveillance; Veterans; Big Data; system architecture; ESSENCE

Acknowledgments

References

1. Pavlin JA, et al. Combining Surveillance Systems: Effective Merging of U.S. Veteran and Military Health Data. PLoS ONE 2013 epub Dec 26.

*Mark Holodniy
E-mail: mark.holodniy@va.gov