Disease Surveillance Big Data Platform for Large Scale Event Processing

Silvino Neto[1] **and Felipe Silva Ferraz**[1]
[1]Recife Center for Advanced Studies and Systems, Recife, Brazil
silvino.neto@gmail.com, fsf@cesar.org.br

Abstract - *In a globalized world, disease outbreaks are likely to spread rapidly across the countries and territories. Therefore, early reports are extremely important in order to predict, identify, confirm, and respond to these occurrences, reducing the risks and consequences of large epidemics. This paper discusses the issues related to this topic and presents a Predictive Analytical Decision Support System (PADSS) integrated into a cloud-based Message-oriented Middleware (MOM) platform, developed to connect healthcare organizations in order to share electronic health records and statistical reports. This platform uses a customized version of the Health Level Seven International (HL7) Fast Healthcare Interoperability Resources (FHIR) specification to support system-level data exchange, enabling the prediction of disease outbreaks using a combination of techniques to perform real-time data analysis.*

Keywords: Disease Surveillance; Big Data; Message-oriented Middleware; FHIR; Predictive Analysis.

1 Introduction

In our previous work [1], we have presented the Platform for Real-Time Verification of Epidemic Notification (PREVENT), a cloud-based message-oriented middleware (MOM) platform for real-time disease surveillance. PREVENT was developed in order to process information streams originated from numerous sources allowing for the early identification of threats and prompt response. Modern communication infrastructure improves our capacity to report disease outbreaks worldwide in a timely manner. Institutions such as the World Health Organization (WHO) and the Centers for Disease Control (CDC) have been engaged in the development of an intelligence network in order to gather and process information from a wide range of sources, promoting a systematic event detection mechanism to support alert and response operations.

In order to support system-level exchange of clinical data in a large scale, PREVENT uses Health Level Seven International (HL7) Fast Healthcare Interoperability Resources (FHIR) specification [2] [3]. FHIR is HL7's new specification that comprises a set of international standards to exchange clinical and administrative data between healthcare applications. It vastly improves previous standards and technologies used in terms of simplicity, scalability, and extensibility. When compared with its predecessors, HL7 FHIR offers a whole new set of features, such as: support for multiple data formats like Extensible Markup Language (XML) and JavaScript Object Notation (JSON), an extensible data model, and a RESTful API.

In this paper, we introduce our new Predictive Analytical Decision Support System (PADSS) built on top of the PREVENT middleware platform. Aiming to quantify reported case numbers, PREVENT was designed to use a customized instance of the FHIR specification to carry statistical reports related to disease occurrences in order to monitor and notify disease outbreaks in real-time fashion. PADSS loads and analyzes data extracted from messages received from healthcare organizations, in order to identify patterns found in historical and transactional data enabling the anticipation of an outbreak occurrence.

This paper is further structured as follows: In section 2, we examine the background for this paper. In section 3, we present an overview of this platform system architecture and introduce PADSS design. In Section 4, we discuss our evaluation approach and present the results obtained. At last, section 5 closes the paper by presenting our conclusions.

2 Background

In this section, we introduce the concepts that served as basis for the development of this research.

2.1 Disease Surveillance

Disease surveillance is the ongoing systematic collection, analysis, and interpretation of outcome-specific data for use in planning, implementing and evaluating public health policies and practices [4].

In order to monitor potential threats, several initiatives have been coordinated by the WHO in collaboration with a wide network of institutions, such as the CDC, national public health institutes, and international healthcare agencies. These initiatives resulted in the development of a global surveillance network established to control and prevent disease outbreaks. This network shares information provided by a broad range of formal and informal sources, therefore every piece of information gathered needs to be verified, in order to assess its worth. In a joint effort, Health Canada and the WHO developed the Global Public Health Intelligence Network (GPHIN) which is a secure Internet-based early-warning tool that continuously searches global media sources to identify information about disease outbreaks and other incidents of potential international public health concern [5].

As a significant portion of initial outbreak reports comes from unofficial non-electronic sources, a large effort is required for the validation and verification of the information received. As a consequence, many researchers have been working on the development of disease surveillance platforms that use a structured approach, based on electronic information provided from reliable sources. The CDC in the United States developed the National Electronic Disease Surveillance System (NEDSS), which is a standards-based approach to facilitate electronic transfer of public health surveillance data from healthcare systems to public health departments [6]. NEDSS uses a system-level message exchange strategy, based on HL7 messaging standards to ensure interoperability between healthcare institutions.

2.2 Healthcare Interoperability

Over the last three decades, there have been several attempts to improve interoperability between healthcare systems. Application vendors and scholars have collaborated on the development of a set of international standards that establish a framework for clinical and administrative data exchange between healthcare systems. As a result of this collaboration, in 1987 the HL7 was created. HL7 is a nonprofit international organization that supports and promotes the development of international interoperability standards and specifications for healthcare software applications [7].

In 1989, HL7 introduced the HL7 v2 messaging standard. HL7 v2 is an ad hoc messaging approach used to transfer several sorts of health-related information. HL7 v2 has become a widely used standard, being adopted and supported by most healthcare software application vendors in North America [2]. Despite HL7 v2 significant acceptance, the limitations imposed by its non-XML encoding syntax based on segments and delimiters have not allowed significant high scale use in larger multiplatform environments. The lack of a formal data model is considered a major drawback in HL7 v2 messaging approach. Several limitations were observed such as no common data dictionary or message transmission interfaces available.

HL7 v3 messaging standard emerged as a response to all the limitations and issues observed on the previous version. Despite its XML-based syntax and object-oriented approach, HL7 v3 was heavily criticized by healthcare software vendors for being inconsistent, overly complex and infeasible to implement or migrate to in production environments [2]. Given that HL7 had spent a considerable amount of time working on the development of the now unpopular HL7 v3 messaging standard, it seemed as if interoperability efforts for healthcare were stalled.

Hence, FHIR was created with the objective of being a simple, extensible, and scalable healthcare messaging standard. There have been several discussions towards a new messaging approach for data exchange in healthcare systems. As result, FHIR provides a Representational State Transfer (REST) interface, which is a simple, efficient and lightweight interoperable strategy for system integration.

FHIR provides a simple and modular object-oriented data model for exchanging electronic health records, but it also supports the data models introduced by HL7 previous specifications, in order to facilitate interoperation with legacy platforms. FHIR data model is extensible, allowing applications to define and use a set of customized resources and data structures. However, in order to declare and use new custom resource types, a set of requirements must be met, in order to guarantee the security and consistency of the model. FHIR supports both XML and JSON based syntax, it simplifies system-level communication by using a common set of interfaces. It also presents a resource interoperable design that allows information to be promptly distributed, providing an alternative to document-centric approaches by directly exposing data elements as services.

2.3 Big Data

Big Data refers to data sets so large, complex and dynamic that conventional data processing tools are insufficient to capture, store, manage and analyze. These data sets hold large volumes of many kinds of information that may be useful for several purposes, ranging from modeling customer behavior to disease outbreak tracking. Predictive analytics is a collection of methods used to extract value from stored data, in form of predictions about future events. The data analysis process can lead to improved decision making, which reflects in greater operational efficiency and risk reduction. Predictive analytics allows researchers to map correlations and identify trends that can help prevent disease outbreaks.

In the healthcare domain, Big Data is a relatively recent research topic. The work of Andreu-Perez et al. has presented an accurate assessment of recent developments in big data in the context of health informatics [15]. As observed in [15], there are several big data related topics concerning medical and healthcare computer-based solutions. Considering the scope of this paper, we focus mostly on the issues related to social, environmental and public health research fields.

Moreover, according to the work of Hay et al., one major limitation observed on most disease surveillance systems is related to their ability to combine static spatially continuous maps of infectious disease risk and continually updated reports of infectious disease occurrences [16]. Companies such as Epidemico [17] have been dedicated to working on commercial products that provide continuous monitoring of disease outbreaks, aiming to address this limitation.

Promising projects have emerged from researches related to the use of Big Data processing platforms. In China, Anying et al. have introduced a disease trend analysis model developed to predict regional outbreaks applying big data processing and data mining techniques to process data collected from drug sales reports, electronic medical records and geospatial data [18]. Recently, social media has surfaced as an important source of information to support real-time monitoring of disease occurrences, using geographical, temporal and text analysis. For instance, Lee et al. developed a real-time disease surveillance system that uses data captured from twitter messages to automatically track flu and cancer related activities [19].

2.4 Complex Event Processing (CEP)

CEP is the leading paradigm for the development of reactive monitoring systems, using an event-driven approach. CEP gathers and analyzes data from multiple distinct sources, allowing users of a system to receive chunks of information on the occurrence of certain circumstances. CEP processes streams of data in order to identify their significance within a cloud of information [8]. In order to accomplish that, CEP applies a set of rules to aggregate, filter and match low-level events, coupled with actions to generate new, higher-level ones derived from those events [9]. If combined with the appropriate technology, CEP enables the development of event-based information systems that are capable of performing real-time data analysis.

There are many forms of implementation of CEP based technologies both in the academy and industry. Most solutions are classified amongst two main categories: Aggregation-oriented CEP or Detection-oriented CEP. The first approach uses real-time processing of data captured from inbound events. Thus, on-line algorithms are executed in response to each event data unit entering the system. The second approach focuses on the examination of event data searching for patterns or recurring behaviors. Several aplicattions rely on a combination of both approaches [12].

Once data extraction has completed, the information obtained usually goes through a second-step analysis that aims to identify whether the events triggered at that time indicate a threat or opportunity. For instance, a significant amount of reports of an infectious disease coming from a geographically limited area could imply on an alert situation.

Therefore, if any threats have been identified, CEP solutions will immediately act on them, in order to initiate contingency measures. An event-driven processing strategy is an optimal choice for applications concerned with regular delivery of situational awareness and response [14]. For sensitive information that requires a high level of confidentiality and integrity, it is extremely important that CEP solutions provide security mechanisms that guarantee information safety. Migliavacca’s work introduces a robust middleware platform that enforces security policies for distributed event-driven applications [13].

3 Middleware

PREVENT is a MOM platform, designed to gather and process data received from healthcare organizations, ranging from hospitals to public health institutions. Data providers must go through an electronic subscription process in order to send and receive notifications. Data is exchanged in the form of electronic medical records that conform to the HL7 FHIR messaging standard. The data received goes through a real-time analysis process, in order to identify possible occurrences of disease outbreaks, using pre-calculated risk profiles. A CEP unit is used to identify recurring patterns and to infer trends over the analyzed information. Given that an alert situation has been detected, PREVENT will asynchronously notify subscribed healthcare organizations systems using an HTTPS push request that holds an extended HL7 FHIR message, improved to support statistical reports. The deployment diagram exhibited in Figure 1 illustrates the middleware system architecture.

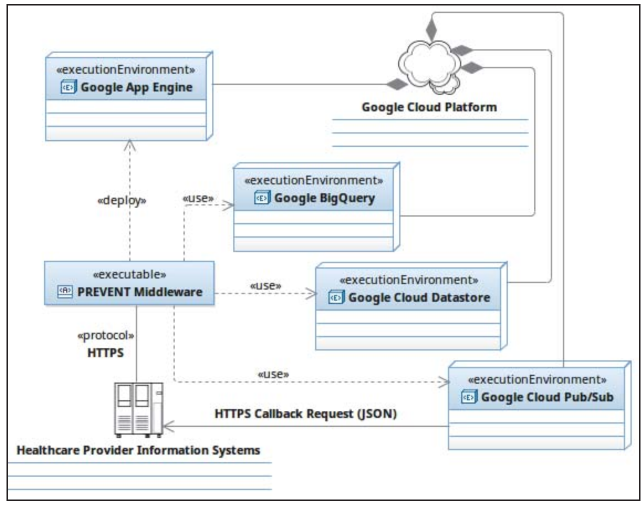

Fig. 1. Middleware System Architecture

In our initial work [1], we have established a small set of features and improvements that would further this research. In the remainder of this section, we introduce our most recent developments.

3.1 Decision Support System

The PREVENT platform was designed to acquire and process streaming data in real-time fashion, using stream analysis within a CEP engine. Given that PREVENT handles large data sets containing collections of data types, we concluded that the results obtained would be vastly improved by using Big Data analytics in a joint approach with CEP. As a consequence, PREVENT platform has been upgraded with a new decision support system, capable of performing large scale data analysis using a design strategy that resembles an Online Analytical Processing (OLAP) system. Prior to the big data hype, similar approaches have been attempted using data warehousing techniques, as the one introduced by Santos [11].

Considering that PADSS must be capable of processing large analytical datasets really quickly, PREVENT platform architecture has evolved to include the Google BigQuery framework, which is a cloud-based analytical database that is able to query massive datasets in few seconds [10]. Hence, it is capable of processing huge amounts of data in close to real-time. Google BigQuery implements a read-only dialect based on the Structured Query Language (SQL) standard, which is a simple and well-known domain-specific language used to manipulate data held in relational databases.

As discussed in our previous work [1], PREVENT stores the notifications received from healthcare organizations in a NoSQL database. Each notification holds a HL7 FHIR message in JSON format, containing relevant information

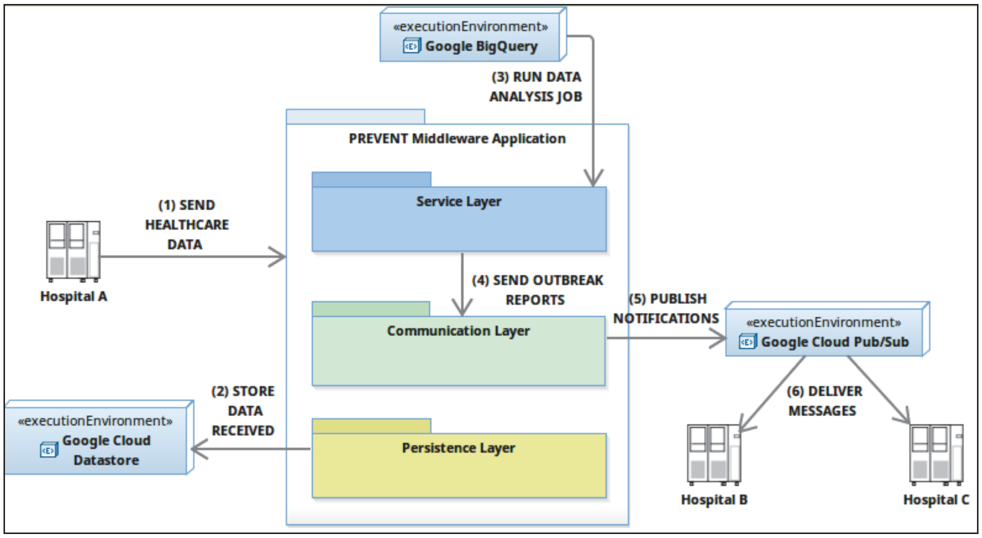

Fig. 2. Middleware Data Flow

about patient diagnostics, location, incidence rates, etc. The information provided is useful to identify recurring patterns that could be matched to historical data, favoring the prediction or early identification of outbreak occurrences.

In the following subsections, we explain how PADSS loads data from PREVENT NoSQL database into BigQuery and how it queries analytical data to be further processed by its CEP unit.

3.2 Data Analysis Process

The PADSS data analysis is performed as an asynchronous task, similarly to a batch process. Data selection operations are carried out in Query Jobs that are periodically executed. These jobs are designed to run a set of predefined queries that use aggregate and window functions in order to extract valuable information related to a specific time window.

The information collected is analyzed using statistical methods such as standard deviation and Z-scores. A threshold value is used to set the boundaries for an outbreak classification. For this particular case study, an arbitrary epidemic threshold of 1.96 times the standard deviation of the mean, for a two-week time window, is used for the characterization and detection of an outbreak [23]. A logistic regression model is used to estimate the probabilty of an outbreak occurrence. For However, given the adaptive nature of our rule-based processing engine, algorithms and rules used for outbreak classification may be easily modified to better adjust to seasonal and geographical variations. The collected data is used as input to PREVENT CEP engine, in order to decide whether an alert message should be dispatched to the subscribed healthcare applications. Traditional approaches used for mining historical data, in order to perform pattern recognition and trend prediction, relied mostly on statistical analysis. However, CEP-based strategies have proven to be much faster and scalable for high-frequency data, producing results of equivalent accuracy in time-constrained scenarios [14].

The strategy we used to integrate our CEP engine with an analytical database is partially inspired by the patterns presented on Maier's reference architecture proposal for big data technologies [20]. In his work, Maier performs an assessment of current state of art tools and technologies for big data, and introduces some patterns used by both industry and academy to implement predictive data analysis. One of the patterns discussed, presents the strategies implemented by commercial products to streamline analytical data into a CEP engine. As result, PREVENT platform is now capable of processing streaming data (real-time) in combination with historical datasets (batch), leveraging its confidence levels and accuracy. In order to systematically monitor the data received, PREVENT uses a set of SQL queries that retrieves historical information to be compared against real-time data. The objective of this process is to identify abnormal patterns or data spikes amongst the statistics collected from the messages received. The middleware data processing chain is exhibited in Figure 2.

BigQuery enhanced SQL dialect accelerates statistical analysis by providing a large collection of built-in aggregate and window functions. Additionally, it provides several utility functions that vastly simplify operations such as: Retrieving scalar values persisted on JSON data structures, using JSONPath expressions, regular expressions, date time operations, mathematical and trigonometric operations.

4 Evaluation

Our evaluation approach aims to portray real-life usage scenarios, employing a set of simulations that will be further described. This strategy attempts to illustrate effective use of this platform in order to anticipate and respond to disease outbreaks. The criteria established for this validation is based on the timely delivery of outbreak reports according to the results obtained by analytical data processing.

In our experiments, we evaluate the middleware by using a set of simulation tools. The test environment prepared for this evaluation is composed by a cloud-based instance of PREVENT/PADSS middleware application deployed into the Google Cloud Platform. A set of 50 HTTPS endpoints have been deployed in order to simulate individual health care units that were previously registered on the PREVENT middleware platform, to send and receive notifications. Furthermore, to emulate enlisted healthcare systems, each HTTPS endpoint is either a Java Servlet class or a PHP file that logs the messages received and returns an HTTP status code 200 (OK) to acknowledge successful reception of notifications delivered. This evaluation process is comprised by a single test scenario that affects all the components that are connected to perform real-time predictive analysis, using a combined approach that includes big data analytics and large scale event processing.

4.1 Simulation Description

The simulation experiment performed in the scope of this evaluation is based on the ongoing dengue fever outbreak reported in the Recife metropolitan area. Dengue fever is a mosquito-borne tropical disease, which is commonly reported in developing countries. Given the unavailability or restricted access to actual electronic medical records, in order to perform this experiment, we had to prepare a simulated dataset, comprised of hypothetical and fictional information, but representative of the occurrences we expect to observe. Thus, we created a dataset composed by a total of 100,000 HL7 FHIR messages. Each message holds information related to: healthcare unit location (latitude and longitude) and patient diagnostics. The message samples are comprised of several distinct ICD-10 codes, representing numerous diseases or health conditions. A FHIR message sample is illustrated in Figure 3.


```json
{
 "resourceType": "DiagnosticReport",
 "id": "f001",
 "contained": [
  {
   "resourceType": "DiagnosticOrder",
   "id": "req",
   "subject": {
    "reference": "Patient/f001",
    "display": "J. Ferreira da Silva"
   },
   "orderer": {
    "reference": "Practitioner/f001",
    "display": "P. Vieira Neto"
   }
  },
  {
   "resourceType": "DiagnosticStatiscalReport",
   "id": "r001",
   "ICD10": "A90",
   "occurrences": "73",
   "casualties": "2",
   "periodInDays": "15",
   "latitude": "-8.0475458",
   "longitude": "-34.8769621"
  }
 ],
 "extension": [{
  "url": "http://crucial-quarter-94700.appspot.com/PREVENT/DiagnosticStatiscalReport",
  "valueReference": {
   "reference": "DiagnosticStatiscalReport/r001"
  }
 }],
 "effectiveDateTime": "2013-04-02",
 "issued": "2013-05-15T19:32:52+01:00",
 "performer": {
  "reference": "Organization/f001",
  "display": "Hospital das Clínicas da Universidade Federal de Pernambuco"
 }
}
```


Fig. 3. FHIR Message Sample

A total of 10,000 messages have been customized to carry statistical data reports as an extension of the HL7 FHIR messaging specification. The FHIR message samples used in this experiment are represented in JSON format, including the extensions of the FHIR model introduced by this research, which are highlighted in red as depicted in Figure 3. The customization of the FHIR standard was necessary given the current limitations observed in the FHIR messaging data model. Up to now, there is no available support for aggregated data reporting, including public health information. A complete list of pending improvements and additions that are expected to be included in the FHIR specification can be found at [21].

In addition to the information previously described, each message contained in the dataset is associated with one specific sender. The sender is one of the 50 subscribed HTTPS endpoints that have been set up in order to simulate health care unit applications. Each health care unit represents a facility that could be a hospital, health care agency, or any other health-related government organization. Finally, each sender is linked with a single location, using traditional geographical coordinates. Chosen locations are restricted to the cities located in the State of Pernambuco, Brazil. If the data analysis outcome indicates that an outbreak occurrence has been identified, the PREVENT platform will start dispatching alert notifications to the registered HTTPS endpoints that are located within a limited risk zone. The alert messages dispatched are customized HL7 FHIR messages improved to report alert and emergency events, given that there is currently no available support for outbreak reporting in the FHIR specification. To confirm that the HTTPS endpoints, which are eligible to receive notifications, have been properly notified, we developed a shell script that is able to extract information from log files using a set of regular expressions, allowing us to identify that an outbreak report has been received.

4.2 Results

Our previous experiments [1] have successfully demonstrated the effectiveness of the PREVENT platform concerning QoS requirements. In this paper, our evaluation approach performs a case study based on a Dengue fever outbreak, using a simulated dataset. The results obtained by this evaluation show that a mixed approach, that uses both CEP and Big Data technologies, enables the timely delivery of outbreak reports through the implementation of pattern recognition and statistical methods, allowing a more agile decision making.

The use of large analytical datasets improves our ability to identify emerging patterns, by matching historical data with current information. As depicted on Figure 5, in the following SQL query, we match historical aggregated and geospatial data, in order to identify the locations with higher incidence rates of Dengue fever.

The information obtained by the query exhibited in Figure 5 is used to map the historical data retrieved with newly reported cases. This query obtains the total and average case numbers for each quarter of the year, followed by the standard deviation. The results are restricted to Dengue fever cases

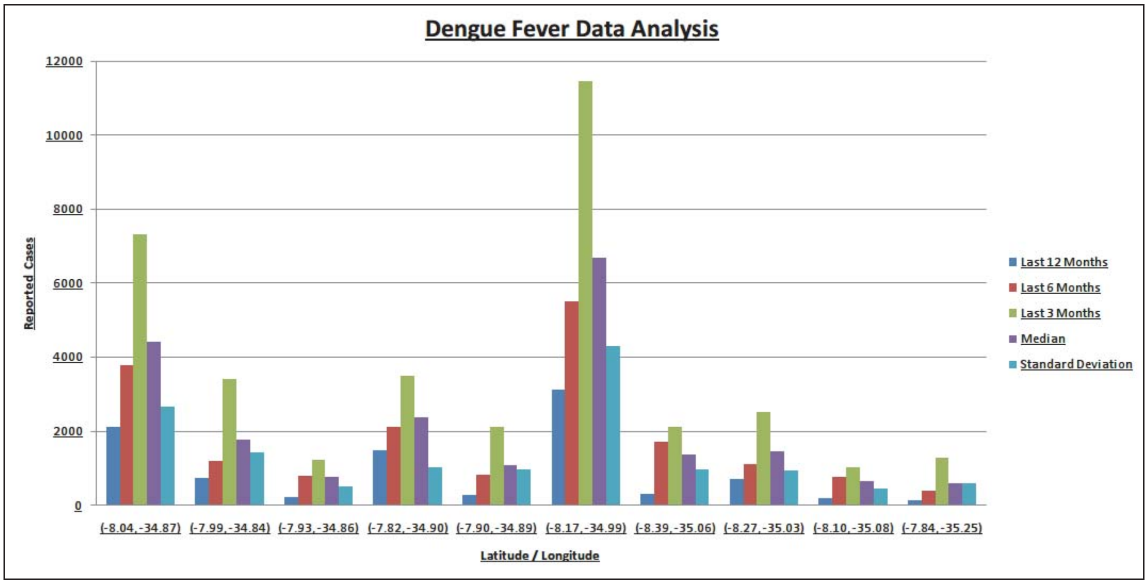

Fig. 4. Case Study Data Analysis

only, occurred in the past 12 months, and grouped by latitude and longitude. Considering the results obtained, we are able to build a column chart that compares each location statistics, and analyzes deviation patterns occurred in the processed datasets, as demonstrated in Figure 4, located on the next page.

```
SELECT
    COUNT(COALESCE(FHIR_MESSAGES.occurrences, 1)) AS totalNumber,
    AVG(FHIR_MESSAGES.occurrences) AS mean,
    STDDEV(FHIR_MESSAGES.occurrences) AS standardDeviation,
    FHIR_MESSAGES.latitude,
    FHIR_MESSAGES.longitude,
    (
      // Recife - Latitude: -8,0475458
      // Recife - Longitude: -34,8769621
      (ACOS(SIN(-8.0475458 * PI() / 180) * SIN((latitude/1000) * PI() / 180) +
      COS(-8.0475458 * PI() / 180) * COS((latitude/1000) * PI() / 180) *
      COS((-34.8769621 - (longitude/1000)) * PI() / 180)) * 180 / PI()) * 60 * 1.1515
    ) AS distance,
    QUARTER(dateTime) as quarter
  FROM
    (SELECT
      JSON_EXTRACT(message, '$.occurrences') AS occurrences,
      JSON_EXTRACT(message, '$.ICD10') AS ICD10,
      JSON_EXTRACT(message, '$.Location.latitude') AS latitude,
      JSON_EXTRACT(message, '$.Location.longitude') AS longitude,
      dateTime
     FROM
      PREVENT:notifications
     WHERE
  dateTime >= DATE_ADD(CURRENT_TIMESTAMP(), -12, "MONTH")) AS FHIR_MESSAGES
  WHERE
    // ICD-10 Code for "Dengue Fever (Classical Dengue)"
    FHIR_MESSAGES.ICD10 = 'A90'
  GROUP BY
    FHIR_MESSAGES.quarter,
    FHIR_MESSAGES.latitude,
    FHIR_MESSAGES.longitude,
    FHIR_MESSAGES.distance;
```

Fig. 5. PADSS Query Used For Evaluation Experiment

Also, using the geographical coordinates retrieved, the incidence rates calculated, and the spherical law of cosines [22], we have drawn a circle on the map representing the locations that are at increased risk of exposure. This is significantly valuable information, considering that we may be able to assess the spreading patterns of an outbreak, improving our response and control mechanisms. Figure 6 exhibits the outbreak incidence map for our Dengue fever study case, based on the data processed.

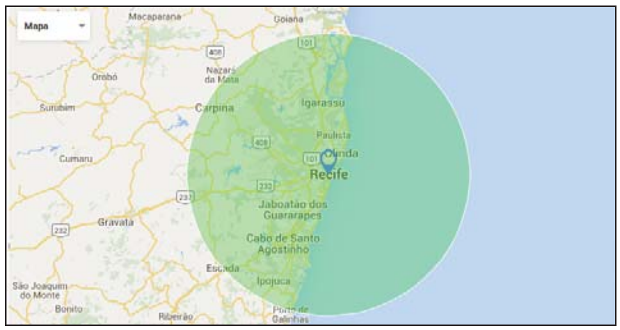

Fig. 6. PADSS Limited Risk Zone Circle

Finally, based on the information extracted from the log files generated by each HTTPS endpoint, the following metrics have been collected:

TABLE I. EVALUATION METRICS

Registered HTTPS endpoints	*Notified HTTPS endpoints*	*Processed Data Amount*	*Data Analysis Execution Time(s)*	*Message Delivery Execution Time(s)*	*Total Execution Time(s)*
50	37	300MB	6.321	6.862	13.183

A few observations must be made based on the results gathered. First, the SQL queries used in this experiment needs to be tuned in order to process real large datasets. Regardless of the capacity offered by Google Cloud infrastructure, when processing large datasets in such time constrained scenarios, it is advisable to improve your queries to optimal levels. Second, this present research is not addressing statistical errors due to the computation of false positive results, considering that a wrong diagnosis may have been given by a physician. Last, for accurate results, profile settings must be enabled in order to use adaptive rules for decision making.

5 Conclusions

This paper has introduced an analytical decision support system, designed to perform predictive analysis using big data technologies combined with large scale event processing. There is an increasing need for real-time processing of analytical data that improves pattern recognition and trends prediction using very large historical datasets. In the context of this work, merging big data predictive analysis with real-time event processing improves our ability to predict or identify outbreak occurrences at early stages. Therefore, healthcare organizations are better equipped to respond to these occurrences in a swift and appropriate manner.

This platform uses the FHIR specification in order to exchange system-level notifications, enabling the massive dissemination of real-time outbreak reports using a standards-based messaging model. Since the beginning of this research, the FHIR standard has evolved significantly. There is much work left to be done as demonstrated by [21], but we are optimistic about the outcome.

The results presented demonstrate the benefits of using an approach that integrates big data analysis and streaming data processing, achieving a higher level of accuracy by matching transactional and historical data. The possibility of performing real-time predictive analysis even for very large datasets is a major advantage when compared with traditional data mining strategies. Coupling the outcomes of analytical data processing with a rule-based CEP engine improves our ability to perform pattern recognition or trends prediction, in a much faster and scalable manner, especially in scenarios that involves high-frequency data such as electronic medical records. As future works, we intend to perform field experiments using real heterogeneous data, given that in the scope of this paper we were unable to perform experiments with larger and richer datasets.

6 References

[1] Silvino Neto, Márcia Valéria, Plínio Manoel, and Felipe Ferraz, "Publish/Subscribe Cloud Middleware for Real-Time Disease Surveillance," 10th International Conference on Software Engineering Advances, 2015, pp. 131-138.

[2] D. Bender and K. Sartipi, "HL7 FHIR: An Agile and RESTful approach to healthcare information Exchange," Computer-Based Medical Systems (CBMS), IEEE 26th International Symposium, 2013, pp. 326-331.

[3] "FHIR Specification v. 1.0.2", 2015. [Online].Available: http://www.hl7.org/FHIR/index.html. [Accessed: 02-Nov-2015].

[4] "Communicable Disease Surveillance and Response Systems", 2006. [Online]. Available: http://www.who.int/csr/resources/publications/surveillance/WHO_CDS_EPR_LYO_2006_2.pdf?ua=1 [Accessed: 04-Oct-2015]

[5] Eric Mykhalovskiy, Lorna Weir, "The Global Public Health Intelligence Network and Early Warning Outbreak Detection," Canadian Journal of Public Health, 2006, v. 97, n.1, pp. 42-44.

[6] National Electronic Disease Surveillance System Working Group, "National Electronic Disease Surveillance System (NEDSS): a standards-based approach to connect public health and clinical medicine," Journal of Public Health Management and Practice, 2001, v. 7, n. 6, pp. 43-50.

[7] "Health Level Seven International". [Online]. Available: http://www.hl7.org/. [Accessed: 02-Nov-2015]

[8] V. Vaidehi, R. Bhargavi, K. Ganapathy, and C.S. Hemalatha, "Multi-sensor based in-home health monitoring using Complex Event Processing," International Conference on Recent Trends In Information Technology (ICRTIT), 2012, pp. 570-575.

[9] D. Robins, "Complex event processing," Second International Workshop on Education Technology and Computer Science, Wuhan, 2010. [Online]. Available: http://goo.gl/jLROpc . [Accessed: 30-May-2015].

[10] "Google Big Query". [Online]. Available: https://cloud.google.com/bigquery/. [Accessed: 30-Oct-2015].

[11] Ricardo Jorge Santos and Jorge Bernardino, "Global Epidemiological Outbreak Surveillance System Architecture," 10th International Database Engineering and Applications Symposium, 2006, pp. 281-284.

[12] Yingmei Qi, Lei Cao, Medhabi Ray, and Elke A. Rundensteiner, "Complex event analytics: online aggregation of stream sequence patterns," ACM SIGMOD International Conference on Management of Data, 2014, pp. 229-240.

[13] Matteo Migliavacca et al., "Distributed Middleware Enforcement of Event Flow Security Policy," ACM/IFIP/USENIX 11th International Middleware Conference, 2010, pp. 334-354.

[14] Eva Zámečníková and Jitka Kreslíková, "Design of Adaptive Business Rules Model for High Frequency Data Processing," in Information Systems Architecture and Technology: System Analysis Approach to the Design, Control and Decision Support, Wrocław, Poland: Oficyna Wydawnicza Politechniki Wrocławskiej, 2014, pp. 75-84.

[15] Javier Andreu-Perez et al., "Big Data for Health," IEEE Journal of Biomedical and Health Informatics, 2015, v. 19, n. 4, pp. 1193-1208.

[16] S. I. Hay, D. B. George, C. L. Moyes, and J. S. Brownstein, "Big data opportunities for global infectious disease surveillance," PLoS Medicine, 2013, v. 10, n. 4, e1001413.

[17] "Epidemico – Website", 2015. [Online]. Available: http://www.epidemico.com/. [Accessed: 28-Oct-2015].

[18] Li Anying, Chen Ke, Song He, and Lei Yu, "The Industry Data Analysis Processing Model Design: The Regional Health Disease Trend Analysis Model," International Conference on Cloud Computing and Big Data, 2014, pp. 130-133, doi: 10.1109/CCBD.2014.11.

[19] Kathy Lee, Ankit Agrawal, and Alok Choudhary, "Real-Time Disease Surveillance Using Twitter Data: Demonstration on Flu and Cancer," 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013, pp. 1474-1477, doi: 10.1145/2487575.2487709.

[20] M. Maier, "Towards a Big Data Reference Architecture," Master's Thesis, Eindhoven University of Technoly, Eindhoven, The Netherlands, 2013.

[21] "FHIR v.1.0.2 – Outstanding Issues", 2015. [Online]. Available: http://hl7.org/fhir/todo.html. [Accessed: 02-Nov-2015].

[22] Weisstein, Eric W. "Spherical Trigonometry." From MathWorld – A Wolfram Web Resource. http://mathworld.wolfram.com/SphericalTrigonometry.html [Accessed: 02-Nov-2015]

[23] R. Snacken, J. Lion, W. Aelvoet, "Five years of sentinel surveillance of acute respiratory infections (1985–1990): The benefits of an influenza early warning system", European Journal of Epidemiology, 1992, v. 8, n. 4, pp. 485-490.