한국 보건의료 빅데이터 플랫폼에서 웹 기반 OLAP 서버 구현

An Implementation of Web-Enabled OLAP Server in Korean HealthCare BigData Platform

Ly Pichponreay, 김 진 혁 , 정 승 현, 이 경 희, 조 완 섭
충북대학교

Ly Pichponreay, Kim jin-hyuk, Jung seung-hyun, Lee kyung-hee Lee, Cho wan-sup
Chungbuk National University

ABSTRACT

In 2015, Ministry of Health and Welfare of Korea announced a research and development plan of using Korean healthcare data to support decision making, reduce cost and enhance a better treatment. This project relies on the adoption of BigData technology such as Apache Hadoop, Apache Spark to store and process HealthCare Data from various institution. Here we present an approach a design and implementation of OLAP server in Korean HealthCare BigData platform. This approach is used to establish a basis for promoting personalized healthcare research for decision making, forecasting disease and developing customized diagnosis and treatment.

Keywords - Big Data; Apache Hadoop; Apache Spark

Ⅰ. Introduction

Lately, the world is paying attention to healthcare. The efficient of using this large amount of healthcare data will lead to many potential benefit to the nation. Electronic Health Record adoption is a part of data collection which help Healthcare data analysis in reduce cost, improve medical research and improve population health [1]. In Korean HealthCare BigData platform there is a critical problem we are facing is the lack of relevant information to make timely decision. Even various institution data has been collected and integrated but statistical aggregation information need to be carried out by using OLAP. So, an implementation need to be done for different kinds of dataset despite architecture, data size, and analysis options.

This paper presents the use of tools, particularly Spark, Redis and Flask for processing large personalized healthcare dataset queue. This web-based OLAP system also incorporate the web service which provide access to the analysis functionalities over the network.

Ⅱ. Related Work

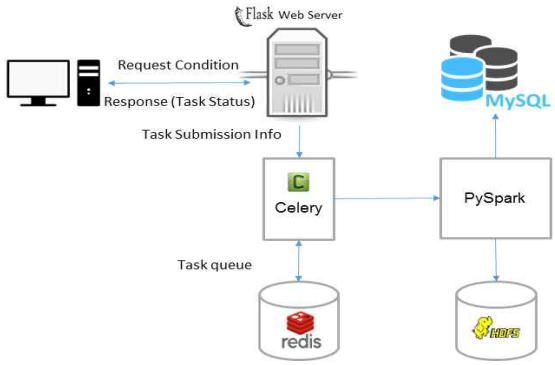
For efficient response and analyze data, a number of technologies have been proposed, including web service, OLAP, Queueing system. Reference [2] incorporated Java and Microsoft SQL Server reporting tools for multilevel viewing data in Flu disease data warehouse. Reference [3] provided Multidimensional Analysis for Traditional Chinese Medicine Diagnosis and Treatment on Hepatitis Diseases. It supported web-based OLAP, reporting and a notable report exporting as various file format.

However, most of this kind of systems don't support queuing system and web invoking service. Our work is unique primarily by focusing on providing a reliable and flexible service which can be used to invoke functionality of analysis service despite type of dataset compared to existing OLAP [2] [3]. Our approach aims to solve architecture constraint by providing web service and resource problem by queuing system.

Ⅲ. Web-Enabled OLAP System

A. System Architecture

Figure 1 shows an overview of our system. This OLAP server was built with four open-source platforms, Redis, Celery Flask and Spark. Dataset Data was generated based on previous user's dataset in Hadoop File system. There are three fundamental part involved in our implementation. The first step is loading data to in-memory database of Spark for OLAP analysis, the second part is creating web service to provides functionalities of analysis over the network and the last part is providing queuing system for each task submission.



▶▶ Figure 1. System Architecture

B. OLAP Analysis

OLAP is a set of operations that one can do on a data set, such as pivoting, slicing, dicing, drilling. For example, one can do OLAP operations with Excel PivotTables. Apache Spark [4] known as a fast distribute processing engine for large dataset through its in-memory cluster computing and because of dataset is too large to process on a single system, Spark is used to create OLAP cube of each dataset on request. By doing the multidimensional analysis we can get the different combination of each dataset based on user's preference.
E.g. Effect of hypertension control on complications in new hypertension patients (2008-2012)

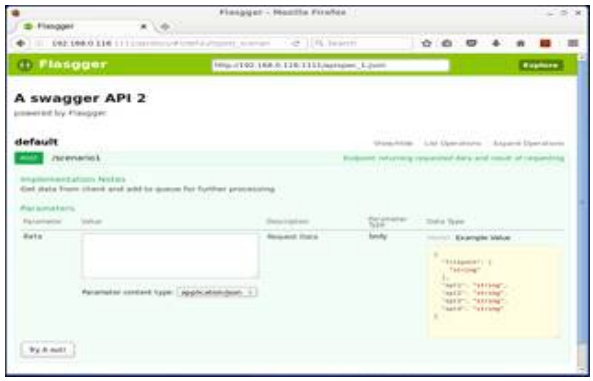
	특별시 및 광역시		기타 도시	
고혈압조절군	306	464	149	
고혈압비조절군	1364	1472	43	499
30-39세	80	112	156	943
40-49세	291	394	235	843
50-59세	593	585	158	402
60-69세	498	533	118	
70세 이상	208	312		

▶▶ Figure 2. OLAP Cube for hypertension control dataset_1

As shown in figure 2, by applying OLAP to dataset we can generate a combinative view of hypertension control dataset of specific user's dataset. OLAP generate an aggregation of count for each area, age group and control group.

C. Web Service

Web service is a software component that implements a dynamic system environment so that all applications used in different computing environments can communicate and execute directly. Web services provide data and services to other applications using standardized Web protocols and data formats such as JSON and XML. This ensures seamless data flow between all computers, regardless of whether they are connected or not. Here we used Flask to create web service so user can invoke the analysis through HTTP protocol. Figure 2 is the representation of our interactive documentation of our RESTful API.



▶▶ Figure 3. API Documentation

D. Queuing System

Because of dataset is too big and our clusters is not big enough, we have implemented this queuing system to allow each task to process in queue. When request arrive, all request information will be store in an in-memory database(Redis)[4]. Celery[5] will responsible for task manipulation. By doing this, we can avoid resource limitation problem. By using celery configuration, we can limit workers to process the task as well.

Ⅳ. Conclusion

This paper presents an architecture and implementation of OLAP server in Korean healthcare BigData platform. This web-based OLAP system enables programmer to invoke analysis service via HTTP protocols without architecture, location and time constraint. we strongly believe that this proposed system will provide a convenient, effective and reliable information for decision-making related to health care. However, several areas remain for further work and performance evaluation need to done as well.

▮ References ▮

[1] Menachemi, N. and T. H. Collum (2011). "Benefits and drawbacks of electronic health record systems." Risk Manag Healthc Policy 4: 47-55.
[2] Bi, L., et al. (2012). Multidimensional analysis for traditional Chinese medicine diagnosis and treatment on hepatitis diseases. e-Health Networking, Applications and Services (Healthcom), 2012 IEEE 14th International Conference on, IEEE.
[3] Dutta, R. (2013). "Health care data warehouse system architecture for influenza (flu) diseases." Computer Science & Information Technology 77.
[4] Redis. Retrieved April 19 2017, from http://redis.io
[5] Celery.Retrieved April 19 2017, from https://celeryproject.org