ABSTRACT
An Educational Data Portal (EDP) play important role in teaching and learning as it contains useful resources. Every big educational institutions such as university shall build an EDP soon or later. The aim of this study is to utilize Big Data solutions in building a Dashboard for an Education Data Portal. The proposed EDP is envisioned to be a core tool for all students and learning agencies, providing support for many types of views and content/instructional resources to allow effective data-driven decision-making for students, teacher and the public, based on recent standards. It supports many features such as accessibility of data and content anywhere, scalability, extensibility of functionality, and extensibility of the technology architecture to support integration with the Shared Learning Infrastructure (SLI).

The Data Dashboard is highly scalable and extensible architecture that will grow, if necessary, to meet the needs of students, and educators

General Terms
Big Data, MapReduce, Hadoop

Keywords
Big Data; MapReduce; Hadoop, Educational Data Portal

1. INTRODUCTION
An Educational Data Portal (EDP) play important role in teaching and learning as it contains useful resources[1]. Many successful implementation of portals are proposed such as StarBRITE, Vanderbilt University Biomedical Research Integration, Translation and Education portal [2]. Every big educational institutions such as university shall build an EDP soon or later. The aim of this study is to utilize Big Data solutions in building a Dashboard for an Education Data Portal. Section 2 reviews main terms and tools for Big Data solutions. Section 3 is the proposed framework. Section 4 invetisates the proposed framework.

2. PREVIOUS WORK
It is often assumed that Big Data resources are too large and complex for human comprehension. The analysis of Big Data is best left to software programs. Not so.

When data analysts go straight to the complex calculations, before they perform a simple estimation, they will find themselves accepting wildly ridiculous calculations. [3] Rapid increases in high performance computing sets the stage for so-called “big data” analysis challenges. However, conventional climate analysis techniques are inadequate in dealing with the complexities of today’s data. [4]. Big data are of special Volume, Variety, and Velocity([5], [6]). A data warehouse stores a substantial amount of historical data. Users of this system are able to continuously ask or query it to retrieve data for analysis. [7]. A data warehouse is a database containing data from multiple operational systems that has been consolidated, integrated, aggregated, and structured, so that it can be used to support the analysis and decision-making process of a business. [8] The data warehouse model is constructed from two relational data model schemas covering demographics and inventory-accounting. [9]

Building a data warehouse requires focusing closely on understanding three main areas: the source area, the destination area, and the mapping area (Extraction–transformation–loading, ETL processes). [10]. MapReduce has become an important distributed processing model for large-scale data-intensive applications like data mining and web indexing. There is a predictive schedule and prefetching (PSP) mechanism, that reduces the execution time, increases the overall throughput and improves the I/O utilization. [11]. MapReduce is a parallel programming model to process large datasets, and it was inspired by the Map and Reduce primitives from functional languages. Its first implementation was designed to run on large clusters of homogeneous machines ([12],[13]) MapReduce’s execution model includes an all-map-to-all-reduce communication, called the shuffle, across the network bisection. [14]. Hadoop—an open-source implementation of MapReduce is widely used for short jobs requiring low response time. [15]. Hadoop Technology Stack is shown in figure 1.

Figure 1. Hadoop Technology Stack [16]
Hive is a batch-oriented, data-warehousing layer built on the core elements of Hadoop (HDFS and MapReduce). It provides users who know SQL with a simple SQL-lite implementation called HiveQL without sacrificing access via mappers and reducers.

As a result, Hive is best used for data mining and deeper analytics that do not require real-time behaviors.[12] Hive uses three mechanisms for data organization: Tables, Partitions and Buckets. [17]. Hive, allows SQL developers to write Hive Query Language (HQL) statements that are similar to standard SQL statements, HQL statements are broken down by the Hive service into MapReduce jobs and executed across a Hadoop cluster. [18]

Pig was initially developed at Yahoo! To allow people using Hadoop to focus more on analyzing large data sets and spend less time having to write mapper and reducer programs. Like actual Pigs, who eat almost anything, the Pig programming language is designed to handle any kind of data. [18]. Pig was designed to make Hadoop more approachable and usable by nondevelopers. [17]

3. PROPOSED FRAMEWORK
The proposed EDP is envisioned to be a core tool for all students and learning agencies, providing support for many types of views and content/instructional resources to allow effective data-driven decision-making for students, teacher and the public, based on recent standards [19].

To support this vision, the solution architecture must have the following characteristics:

- Accessibility of data and content anywhere and anytime by students, public, teachers, and other educators at home, school, and via mobile devices.
- Scalability to accommodate students, public, and teachers/principals/other educators.
- Extensibility of functionality in the system environment.
- Extensibility of the technology architecture to support integration with the Shared Learning Infrastructure (SLI).

The following diagram provides a graphical view of the proposed architecture for EDP.

4. IMPLEMENTATION
We implement the proposed framework in third author website http://eldosuky.com. Each student, either under graduate or post-graduate, has a profile in the web site, as shown in the next figure.
To access or modify one's data, student is asked to enter credentials, as shown in figure 4.

Figure 4. Logging to the web site

Figure 5. XP desktop for each student

The website has access to the student affaire data warehouse, this enables courses to tracks students scoring, as shown in the following figures.

Figure 6. Administration

Figure 7. XP look and feel
Courses taught in 2013-1014 for Undergraduate:

- Discrete Mathematics
- Algorithm Analysis and Design
- Assembly language

Students of Algorithm Analysis and Design

1. Section 1
2. Section 2
3. Section 3
4. Section 4
5. Section 5
6. Section 6
7. Section 7
8. Section 8

The full list is found here

Figure 8. some courses

Figure 9. students list

5. CONCLUSION AND FUTURE WORK

The proposed EDP is envisioned to be a core tool for all students and learning agencies, providing support for many types of views and content/instructional resources to allow effective data-driven decision-making for students, teacher and the public, based on recent standards. It supports:

- Accessibility of data and content anywhere and anytime by students, public, teachers, and other educators at home, school, and via mobile devices.
- Scalability to accommodate students, public, and teachers/principals/other educators.
- Extensibility of functionality in the system environment.
- Extensibility of the technology architecture to support integration with the Shared Learning Infrastructure (SLI).

The Data Dashboard is highly scalable and extensible architecture that will grow, if necessary, to meet the needs of students, and educators. In the future we plan to extend the underlying infrastructure, as well as linking it with many other databases. Also, we would like to incorporate other services such as recommending subjects for students [20] and opinion mining [21].

6. REFERENCES


