

2015

Integrating NoSQL in the Classroom

Brad Fowler

Georgia College and State University, brad.fowler@gcsu.edu

Heather Walker

Georgia College and State University, heather.massengill@bobcats.gcsu.edu

Jerica Williams

Georgia College and State University, jerica.williams@bobcats.gcsu.edu

Daniel McDonald

Georgia College and State University, daniel.mcdonald@gcsu.edu

Follow this and additional works at: <http://aisel.aisnet.org/sais2015>

Recommended Citation

Fowler, Brad; Walker, Heather; Williams, Jerica; and McDonald, Daniel, "Integrating NoSQL in the Classroom" (2015). *SAIS 2015 Proceedings*. 25.

<http://aisel.aisnet.org/sais2015/25>

This material is brought to you by the Southern (SAIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in SAIS 2015 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

INTEGRATING NOSQL IN THE CLASSROOM

Brad Fowler

Georgia College & State University
brad.fowler@gcsu.edu

Heather Walker

Georgia College & State University
heather.massengill@bobcats.gcsu.edu

Jerica Williams

Georgia College & State University
jerica.williams@bobcats.gcsu.edu

Daniel McDonald

Georgia College & State University
daniel.mcdonald@gcsu.edu

ABSTRACT

With the increasing popularity of big data, more and more organizations are turning to NoSQL databases as their preferred system for handling the unique demands of capturing and storing massive amounts of data. The likelihood that employees in all sizes of organizations will encounter NoSQL databases is growing every year. College students need to be exposed to this technology and begin to have a functional understanding of how it works and how to use it. This paper offers a teaching case for college instructors to integrate NoSQL into their existing database courses.

Keywords

Database, NoSQL, Teaching Case, CouchDB

INTRODUCTION

Big data is quickly becoming an integral part of modern information systems. With the relatively low price of data storage and the ever increasing speed of public Internets, data can be collected in massive amounts. This can be seen in modern web applications. Big data is prevalent in many of today's social media sites. Facebook logs 135 billion messages per month, and Twitter stores seven terabytes of data every day. This low cost of data storage coupled with cloud based infrastructures allows not only the major, multinational organizations but also small and medium enterprises to utilize big data (Purcell, 2014).

Some of the challenges in the use of big data are how to handle the velocity in which it is collected and how to store the data most effectively. In many cases, traditional Structured Query Language (SQL) databases cannot process data fast enough. In other cases, traditional relational database structures are not flexible enough to handle the varying data types and the inconsistent manner in which the data are collected. Major Internet companies, like Google, Facebook, Twitter, and Amazon have dealt with these challenges by moving to NoSQL databases. The term "NoSQL" was coined in the late 90's and means "Not Only SQL." It is used to describe database systems that are not relation based (Berg, Seymour, Goel, 2013). NoSQL databases are the best option for circumstances in which data inputs are expected to be extremely large at any given time or where data inputs are not entered in a consistent manner. Common uses for NoSQL databases are session stores, user profiles stores, mobile apps, e-commerce, and third-party data aggregation. NoSQL databases allow for large amounts of variable and unstructured data to be stored and used. Given the popularity of web applications in today's corporate world, Information Systems (IS) students need to develop a basic understanding of how NoSQL databases work in modern information environments. With the adoption of cloud-based infrastructures, it is very likely that they will use a NoSQL database during their careers. An understanding of what a NoSQL database looks like, how they differ from traditional SQL databases, and how to use them is going to be important for college graduates in the near future. In this paper we will build a teaching case for how to add a NoSQL lesson to an existing database course.

In the context of a typical database management course, students learn concepts and skills geared around traditional SQL relational databases. They learn about entities, relationships, normalization, and data structuring. SQL is typically used to create tables and store data into a database. Students will typically learn to query data in order to build knowledge from it. In this NoSQL lesson, the students will be exposed to a NoSQL database management system along with a fictitious data set to build a NoSQL database and then query information from it.

A BRIEF REVIEW OF THE LITERATURE

Paulo B. Goes, Editor-in-Chief of MIS Quarterly, identified an industry gap in the understanding of big data, its challenges, and its potentials. In the past several years, big data has emerged as a new trend in the field of information technology due to the increasing amounts of data being generated, collected, stored, and analyzed by organizations in all sectors. IS groups are taking advantage of this industry gap as an opportunity to offer academic programs that specialize in data and business analytics to develop students' skills in data science (Goes, 2014).

Big data is the term used to describe datasets so large and complex that they are difficult to process using traditional processing applications. Big data consists of both structured and unstructured data and may come from traditional or new data sources. Big data size is not a set value, but remains an ever-increasing, moving target. Over the years, the scope of the term big data has expanded significantly. Big data not only refers to the data itself, but also to a set of data management technologies that were first used by social media companies to process massive volumes of data. These technologies are used to capture, store, manage, and analyze large collections of data to solve problems (Shaw, 2014).

The concept of introducing NoSQL in a database curriculum was explored by Clare Stainer (2012). Stainer introduced the basic concepts of NoSQL databases, how they differ from relational databases, and what issues there may be in introducing NoSQL into a traditional database course. Stainer found that instructors are already challenged in covering the breath of information pertinent to modern databases and that the prevalence of using databases for business analytics increases this challenge. She also recommended that instructors teaching database courses follow the pace of the database industry and incorporate NoSQL into the curriculum (Stainer, 2012).

As mentioned above, traditional SQL databases have difficulties processing large, unstructured datasets, and therefore, NoSQL databases have become an alternative solution. From the review of the literature, no published teaching cases of NoSQL were found. Therefore, we have identified a gap in the literature that further justifies the need for big data research on the topic of NoSQL.

Conventionally, students only learn about SQL relational databases, but we have developed a NoSQL module using CouchDB that instructors can integrate into a database management course to give students experience with this emerging technology. CouchDB is open source NoSQL DBMS software that students can either install on their personal computer or use as a cloud-based service. Below is the teaching module we developed for MMIS graduate students, as well as MIS undergraduate students, enrolled in a database management course.

CONTEXT FOR LEARNING MODULE

Finding the best data set for instruction is important. NoSQL is usually implemented over relational databases for particular reasons. These reasons can vary from the volume of data collected in a given time frame to the unique quality of the data. In this learning module, we will use a NoSQL database for the purpose on storing semi-structured data. One of the best examples of semi-unstructured data is social media pages. Social media pages require data to be collected based on the user's preferences, as opposed to the developer's preferences. This means that amount and type of data entered into a social media page depend completely on the user. Users may choose to share a lot of information or very little. The database needs to be able to store this semi-structured data in a way that allows the social media page to use the data to enhance the page's value. Since many students use social media, this is a good example to use in the classroom. This lesson plan has the dual benefit of teaching students how their social media data is handled while preparing them for using NoSQL databases in the future.

COUCHDB – NOSQL DBMS

For this lesson, students will use CouchDB as their Database Management System. CouchDB is an open-source DBMS provided by the Apache Software Foundation. It uses JavaScript Object Notation (JSON) to store data and JavaScript as its language for queries. CouchDB is a document-oriented, or document-store database. In this style of NoSQL database, all data are stored in a "document" and identified by a unique key. Fields within a NoSQL document can vary from key to key, as opposed to a relational database in which fields must be consistent among records or keys. Other NoSQL database systems that would be suitable for use with the data set are MongoDB or Couchbase.

NOSQL DATA SET

Instructors will give students a textual case of information that represents possible data for users wanting to populate their social media page. The students will need to read the case to identify the data needed to be entered into the database. The textual case also will determine what fields and values students must create in their database.

In the textual case, the students will receive information on several users. The information will include data that is consistent with social media data. This information includes name, address, email, gender, interests, and friends. Some data will be consistent from person to person. Other data will vary from no data to many data points. Instructors can use Table 1 to illustrate what this type of data set would look like in a relational table to illustrate its flaws.

USERS										
USER_ID	FIRST_NAME	LAST_NAME	ADDRESS	CITY	STATE	ZIPCODE	EMAIL	GENDER	INTERESTS	FRIENDS
41001	John	Doe	1234 Hancock St.	Milledgeville	GA	31061	john.doe@gmail.com	male	running	Jane Doe
									games	Jeff Doe
									bikes	Jerry Doe
41002	Jane	Doe	456 Jefferson St.	Milledgeville	GA	31061	jane.doe@gmail.com	female	bikes	John Doe
									shopping	Jerry Doe
									nascar	
41003	Jerry	Doe	789 Briarcliff Rd.	Atlanta	GA	31061	jerry.doe@gmail.com	male	nascar	John Doe
									football	Jane Doe
									games	
41004	Jeff	Doe	2345 Gray St.	Gray	GA	31061	jeff.doe@gmail.com	male	nascar	John Doe
									football	Jane Doe
									hockey	Jerry Doe
									sports	Jason Doe
										Janet Doe
										Jordan Doe
41005	Jason	Doe	678 Montgomery St.	Milledgeville	GA	31061		male	TV	Janet Doe
									movies	Jane Doe
									sports	Jordan Doe
									girls	Jessica Doe
41006	Janet	Doe	987 Clarke St.	Milledgeville	GA	31061	janet.doe@gmail.com	female	shopping	Jessica Doe
									being social	John Doe
									pets	Jane Doe
41007	Jordan	Doe	4321 Wayne St.	Milledgeville	GA	31061	jordan.doe@gmail.com		hunting	Jane Doe
									shopping	John Doe
									pets	Jerry Doe
									running	Jessica Doe
41008	Jessica	Doe	3219 N. Columbus St.	Milledgeville	GA	31061	jessie.doe@gmail.com	female	shopping	John Doe
									running	Jane Doe
									sports	Jerry Doe
									pets	Jeff Doe
									family	Jason Doe
									friends	Janet Doe
										Jordan Doe

Table 1: Example Relational Database Dataset

SOCIAL MEDIA PAGE

The students will create a database through CouchDB based on the information obtained from the textual case as well. Once the database is created, they will enter each person’s data into a new document. Each document will store the data with a field and value. An example of a field and value would be “name” as the field and “John Doe” as the value or “interests” as the field and “running, games, bikes” as the value. The instructor can elect to allow the students to create other fields that they think will be useful for a social media page.

QUERIES

CouchDB uses JavaScript as its querying language. Querying is done in the Views section of the CouchDB interface. Querying with JavaScript can be very complex. We will present a few simple queries for the student to get a basic understanding of how to get data out of a NoSQL database.

To make a query in CouchDB, select "Temporary view" from the "View" dropdown window. There will be a default function in the "Map Function" window.

This function is:

```
function(doc) {
    emit(null, doc);
}
```

This function will return a null response for the "Key" and the complete documents for the "Value" for all documents in the SocialMediaPage database.

The first function used will be:

```
function(doc) {
  emit(doc.name, doc.address);
}
```

This function will return the "name" fields of the documents for the Key and the corresponding "address" fields for the "Value" for all documents in the SocialMediaPage database.

The second function used will be:

```
function(doc) {
  if(doc.gender == "male") {
    emit(doc.name, doc.address);
  }
}
```

This function will return the "name" fields for the Key and "address" fields for the Value for all documents that have a male for their gender field.

CONCLUSION

We have developed a full NoSQL learning module that can be integrated into the end of a database management course to augment concepts and skills geared around traditional SQL relational databases. The assignment that will be given to the students is provided in Appendix A and the CouchDB tutorial showing the steps of how to complete the assignment are provided in Appendix B. Implementing this NoSQL lesson will help students understand why NoSQL is an alternative solution for processing large datasets. Students will use CouchDB as their database management system to create a database from a textual case for populating social media pages. As we believe that most students use a social media page, this is a good example to use in the classroom. This allows them to apply the lesson information to understand how their data are handled in their social media pages and prepares them to use NoSQL databases in the future. Developing students' understanding of the reach of Big Data, as well as the almost limitless potential for its analysis and application, is a growing necessity for educators concerned with students' information literacy and future earning potential. This NoSQL module offers instructors the opportunity to integrate a lesson on information literacy at the same time they're introducing timely concepts about the application of databases in the modern technology environment.

REFERENCES

1. Berg, K. L., Seymour, T., & Goel, R. (2013). History of Databases. *International Journal of Management & Information Systems*, Vol. 17(Issue 1), pp. 29.
2. CouchDB [Software]. Available from Apache CouchDB: <http://couchdb.apache.org>
3. Goes, P. B. (2014). Big Data and IS Research. *MIS Quarterly*, 38(3), iii-viii.
4. Purcell, B. (2014). Big Data using Cloud Computing. *Journal of Technology Research*, Vol. 5, pp. 1–8.
5. Shaw, J. (2014, March-April). Why big data is a big deal. *Harvard Magazine*. Retrieved from <http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal>
6. Stanier, C. (2012). Introducing NoSQL into the Database Curriculum. In *TLAD 2012* (pp. pp. 61–71). University of Hertfordshire: Higher Education Academy.