ISSN: 2167-1435
v29 n1956

CloudBees: A Resource Guide for Teaching Clouding Computing on a Java Platform

Michael Juntao Yuan
Michael@michaelyuan.com
Ringful Health, LLC
Austin, TX 78736

Ju Long
julong@txstate.edu
Dept of Computer Information Systems and Quantitative Methods
Texas State University-San Marcos
San Marcos, TX 78666

Abstract

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction. The projected popularity of the Cloud has motivated Information Systems educators to incorporate it in the information technology classroom (Frydenberg, 2011). In this research, we focused on a core service model of cloud computing, Platform as a Service (PaaS). In particular, we examined PaaS on a Java Platform, which is more relevant to the current IS education curriculum. We stated that finding the right resource to teach PaaS on a Java platform could be challenging and we set to identify a valuable resource –CloudBees for the IS educators. We argued that as a leader in the Java PaaS sector, CloudBees provides Information System educators a standard-based, integrated, and scalable platform for students who want to develop and deploy web applications in a cloud environment without administration headaches. We provided a detailed examination of CloudBees' core features. We hope our study will be beneficial to IS educators who wish to teach PaaS on a Java Platform. It could potentially be useful to corporate practitioners who need to select an appropriate platform provider for PaaS.

Keywords: Cloud Computing, Platform as a Service (PaaS), CloudBees, Resource Guide, Teaching Cloud Computing

1. INTRODUCTION

Cloud Computing and Platform as a Service (PaaS)

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction (Mell and Grance, 2011). Deployments of cloud computing are expected to be a $160 billion market in 2012 (Crossman, 2009) in a growth of 25% of all incremental investment in technology in 2012 – a growth indicated to be the largest since the Internet (Hamm, 2009).

Platform as a Service (PaaS) is one of the service model of cloud computing. The capability provided through PaaS to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud

infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment (Mell and Grance, 2011).

Cloud Computing and PaaS in Information Systems education

Cloud computing is clearly considered to be developing as an enabling model for improving the processes of firms. Schools of computer science and information systems may benefit by having students cognizant of the cloud computing technologies and strategies (Lawler, 2010 and 2011). In effect, several studies have pioneered in this front.

Among these studies, one describes an introductory lesson that builds on student familiarity with Google Docs to illustrate the concepts of Infrastructure, Software, Data, and Platform as a Service (Frydenberg, 2011). The study found that while many first-year college students have heard of the term Cloud Computing and used Google Docs, a popular web-based office suite of applications for collaboration, their knowledge of how the Cloud is used in a business environment is limited (Frydenberg, 2011).

Hollingsworth and Powell (2010) introduce Cloud Computing topics in a Web programming course where students write and configure applications deployed to Google's Cloud (App Engine). Malan (2010) created a network of virtual machines hosted on Amazon's Elastic Cloud platform for students to explore cloud concepts and build applications in an introductory computer science course. Rehman and Sakr (2010) chose Microsoft's Azure platform to design an undergraduate-level course in Cloud Computing where students learned about public and private clouds, virtualization, MapReduce and Hadoop parallel processing technologies, and other cloud technology concepts. Each of these courses requires programming knowledge in order to complete the projects (Frydenberg, 2011)

For PaaS, it is probably the most abstract cloud concept for students without programming background to understand. It is also the least accessible concept because most consumer cloud applications do not rely on PaaS (Frydenberg, 2011). One study proposed teaching PaaS in a capstone course. It compares two different options, a commercial platform or developing a private platform with technical support locally (Roggio, 2011).

2. RESEARCH PURPOSE

The cloud has become a widely used term in academia and the industry. Education has not remained unaware of this trend, and several educational solutions based on cloud technologies are already in place, especially for software as a service cloud.
However, despite some of the early research efforts, an evaluation of the educational potential of infrastructure and platform clouds has not been explored yet (Vaquero, 2011).

In particular, Platform-as-a-Service (PaaS) has gained great popularity among developers in recent years. It has the potential to significantly reduce IT administration and management complexities associated with developing and deploying web applications.

To incorporate PaaS in the IS curriculum, it is important to understand the strengths, weaknesses, and constraints of the platform, especially security, compliance, and perhaps legal constraints (Roggio, 2011).

As suggested by extant research, it is not recommended to develop the platform in house. Establishing a set of development tools (solution stack) in a platform can be very time consuming, and keeping these plug-ins up to date is daunting administrative work. Administrators of such a platform require a significant competence in a variety of skills (Roggio, 2011).

Therefore, our research focuses on examining the commercial PaaS providers and their potential to be utilized in IS education. Traditionally, PaaS providers have focused on Ruby, Python and PHP platforms. The Java platform has so far been under-served by PaaS providers due to the complexity of managing and scaling Java application servers that can be very resource hungry. On the other hand, the Java platform is widely used in most Information System Education curricula. For IS educators, finding the right resource to teach clouding computing on a Java platform could be challenging.

Since 2010, several companies have tried to tackle the Java PaaS problem, and among them, CloudBees has emerged as an early leader of this front, as noted in several independent

comparison studies such as InfoQ's "A Java Developer's Guide to PaaS" (Yuan, 2011).
In this research, we will discuss key concepts behind Java PaaS and the key features of CloudBees. We argue that as a leader in the Java Platform-as-a-Service (PaaS) sector, CloudBees provides Information System educators a standard-based, integrated, and scalable platform for students who want to develop and deploy web applications in a cloud environment without administration headaches.

3. ADVANTAGES OF CLOUDBEES

A standard-based Java platform

The Java platform has a strong tradition of standardization – both in the programming language itself as well as in the libraries for the platform (aka "write once, run anywhere"). In the world of enterprise Java, much of the innovation in the platform happens in application frameworks and application servers. The ability to use any framework on a variety of different application servers is a core requirement for a Java PaaS.

In the past, we have seen Java PaaS efforts that are based on non-standard technology. A very good example is Google App Engine for Java (GAE). GAE provides a Java runtime that limits access to file system, networking, threading and other Application Programming Interface (API) methods. It provides a non-relational database, and supports a limited set of outdated Java Data Object (JDO) API on top of it. While Google provided great technical rationale on why they did this, the non-standard approach created many problems and gotchas for Java developers. Fast forward to 2011, when Google decided to move GAE out of beta and dramatically raised the price, many GAE developers found themselves stuck in that PaaS environment not being able to move their application and data elsewhere because their applications were written specifically to GAE's non-standard Java runtime platform.

By comparison, CloudBees strongly believes in the benefit of standardization and portability of the Java platform. As a result, the CloudBees PaaS environment supports plain old Tomcat, Java Platform Enterprise Edition (Java EE) certified JBoss 6 Application Server (JBoss 6 AS), and MySQL database. You can deploy just any standard-based Java web application to CloudBees, and easily take your CloudBees application to any other PaaS or in-house providers of Tomcat / Java EE application servers.

With the CloudBees Software Development Kit (SDK), you can easily test and validate your application on their local computer inside an Eclipse IDE. But, because of the support for standard-based technology, you do not need the SDK or Eclipse to be a CloudBees developer. You can use your favorite tools – any IDE or even text editors such as EMACS or VIM.

Since CloudBees deploys both the application server and the database for you, it provides a tightly integrated experience just as an in-house IT team might provide for you. The database is simply available as a Java Naming and Directory Interface (JNDI) resource as soon as it is deployed – there is no need to chase down Java Database Connectivity (JDBC) drivers, connection strings, and configure JNDI. The same goes for clustering. The CloudBees PaaS deploys application servers running your application behind a load balancer. Of course, your application does need to be "cluster aware" by saving HTTP sessions in a database or a shared in-memory cache, but all the hard work of managing the cluster and load balancer has been take care of for you.

No vender lock-in

The CloudBees PaaS supports Java standards like servlets, Java EE and others, and is based on industry standard application servers such as Tomcat and JBoss. As such, the CloudBees PaaS does not require you to use proprietary APIs or other non-standard Java libraries, ensuring that your Java code remains portable. The CloudBees PaaS simply runs your Java code. If you are unhappy with CloudBees for any reason, you can take your code and run it elsewhere.

Easy Continuous Integration

While all other PaaS providers focus on providing a hosted runtime environment for applications, CloudBees take the "platform" concept further to support the entire development, testing, and deployment life cycle of Java applications. Specifically, CloudBees supports private source control repositories, private Maven repositories for team's internal use, and on-demand continuous build servers managed by Jenkins.

Maven is a widely used build tool for managing dependencies in a Java project. A typical enterprise Java project has hundreds of external dependencies for libraries and frameworks. The exact version of each of those dependency Java ARchive (JARs) must be carefully managed to make sure that they all work together in a particular version of the application server. Maven helps us automatically and declaratively manage those dependencies via configuration files.

While Maven is very useful, not all JARs are available in public Maven repositories. That is especially true for internal library JARs you might build. Without internal Maven repositories, you would fall back to manually copying JARs and putting those JARs into their individual computer's local Maven repositories – a process that is not only error prone, but forgoes much of the benefits Maven brings. CloudBees provides you with private Maven repositories for both development and production use. It is a great feature to take full advantage of Maven.

Perhaps even more significantly, CloudBees is the first company to provide "Jenkins in the cloud" services to support continuous integration processes required by many Agile development teams. The concept of continuous integration is that software products needs to be built and validated against an array of tests every time a developer checks code into the source control repository, to make sure that the software is always in the "ready to be released" state. If the build fails or any test fails, appropriate members of the team need to be notified. Jenkins is a very popular tool to manage the continuous build and testing process.

However, the challenge of using Jenkins in house is relatively high infrastructure and IT administration overheads. You have to run Jenkins on an in-house server or a cluster of servers, and have an IT administrator to manage it. That barrier often prevents continuous integration. CloudBees provides an on-demand Jenkins service that spawns a new Jenkins build server instance every time the code changes in the source control (or at fixed times during the day). The build server runs the Maven script against your private Maven repository to build the software application, and then automatically runs all the unit tests and end-user tests via a script. CloudBees then reports the results back to you via a dashboard and notifications.

If you have always wanted to apply Agile methods in your curriculum to build more reliable software faster, but did not have time / resources to set up the required technical infrastructure, CloudBees could be the ideal choice.

An ecosystem of add-on service providers

Since CloudBees' PaaS offering covers the entire cycle of application development and deployment, it provides an excellent platform for tool providers to directly integrate into applications. Indeed, CloudBees has built an eco-system of service providers that supports deep integration into the PaaS platform itself. That is in contrast against some of the other PaaS eco-systems where service providers / partners simply get a logo and link on the PaaS's web site.

Notable examples of services providers in the CloudBees eco-system include the following.

- New Relic is an application performance monitoring service. It embeds its software agent in to your application to collect detailed performance information about your application at real time. In CloudBees, the New Relic software agent is automatically instrumented into your application at the time of deployment without the developer needing to change any source code. Once New Relic is instrumented, you can access its performance dashboard from your CloudBees account.

- Paper Trail provides aggregation and analysis of log files. While CloudBees' management console provides access to the server's raw log file. Paper Trial adds value by combining logs from different sources, keeping them around for longer periods of time, and providing summary analysis based on the logs. Once enabled, Paper Trials automatically collects and reports the log data from your running CloudBees applications.

- Sauce Labs is the company behind the very popular Selenium open source project. Selenium is a tool for automated in-browser testing of web applications. You can develop a Selenium script that operates the web site step-by-step in a

browser you specify. That is a great tool for continuous integration. After a Jekins server builds the application, we can use Maven to run automated unit tests, and then run Selenium scripts to test the application from the user's perspective to make sure that the build is valid. The entire process can be automated within CloudBees as your application is continuously built and integrated.

- MongoHQ provides hosted MongoDB services. It is great for users who want to use a NoSQL solution for database instead of the default MySQL service provided by CloudBees itself. Similarly, Cloudant provides Apache CouchDB as a hosted service for CloudBees applications.

- SendGrid provides a hosted SMTP service to send emails either a mailing list or to individual users. It is setup so that email messages from SendGrid are unlikely to be categorized as SPAM by ISPs. The SendGrid service, once enabled, is available as a JNDI resource in your CloudBees application. Once you get a JavaMail session from JNDI, you can simply use the JavaMail API to send messages without worrying about connection URLs and credentials.

High performance at low cost

A common question about cloud computing is the cost. It is especially important for using PaaS in an educational setting. CloudBees has a free tier of 5 applications and 5 shared databases. It provides sufficient functionalities to teach PaaS on a Java platform. There are limitations for the free applications. For instance, they cannot be clustered, cannot support CNAME based URLs, and the database is limited to 10MB of data, which usually is not a barrier for the platform adoption in education setting. The free tier is great for running a "staging environment" to test applications built from the continuous integration process. For instance, you can run Selenium scripts from Sauce Lab to test the freshly built application in your staging environment.

In CloudBees terms, the capability provided by the single free web application is called an "AppCell". In a paid subscription plan, each AppCell costs $0.05 per hour. In addition, for $25 a month, you can get a dedicated database with 1GB of storage.

To evaluate the computing power provided in the free tier application, we conducted a simple load testing experiment. The test application is a single servlet that mimics a user login action. It queries a single database table to retrieve the username and password, and then decide whether to allow or reject the login. We used blitz.io service to ramp up 250 concurrent users to access the web application and database. As seen from Figure 1 (Appendix), the application scales up to handle the increased load in a near linear fashion with little increase of response time. That indicates that the system is still well within its capabilities at 250 concurrent users. The 250 requests per second "hit rate" translates to 22 million web requests per day, which is sufficient for many applications.

4. CONCLUSIONS

As IS educators, we need to recognize the rapid migration toward cloud computing as the framework for many modern application needs (Roggio, 2011). The IS 2010 model curriculum (Topi, et al., 2010) notes that "service-oriented architecture, Web services, software as a service, and Cloud Computing are all important elements in the new way of organizing the fundamental architecture for computer-based systems and solutions that is gradually becoming the dominant paradigm of organizational computing." (p. 6) Given the projected rise in usage and popularity of Cloud applications in the next decade for consumers and the enterprise, Cloud Computing is an appropriate topic to present in the information technology classroom (Frydenberg, 2011).

In this research, we focused on a core service model of cloud computing, PaaS. In particular, we examined PaaS on a Java Platform, which is more relevant to the current IS education curriculum. We stated that finding the right resource to teach PaaS on a Java platform could be challenging and we set to identify a valuable resource –CloudBees for the IS educators. We argued that as a leader in the Java Platform-as-a-Service (PaaS) sector, CloudBees provides Information System educators a standard-based, integrated, and scalable platform for students who want to develop and deploy web applications in a cloud environment without administration headaches. We provided an overview of CloudBees' core features. We hope

ISSN: 2167-1435
v29 n1956

our study will be beneficial to IS educators who wish to teach PaaS on a Java Platform. It could potentially be useful to corporate practitioners who need to select the most appropriate platform provider for PaaS.

5. REFERENCES

Crosman, P. (2009). Securing the clouds. Wall Street and Technology, December, 23.

Frydenberg, M. (2011). The Silver Lining: A Teaching Case Using Google Docs to Illustrate Cloud Computing Concepts. Information Systems Educators Conference 2011 ISECON Proceedings Wilmington North Carolina, USA v28 n1641

Hamm, S. (2009). Cloud computing: Big bang for business. Business Week, June 15, 044.

Hollingsworth, J., & Powell, D. (2010). Teaching Web Programming Using the Google Cloud. Proceedings of the 48th Annual Southeast Regional Conference (ACM SE '10) (p. 5). New York, NY: Association for Computing Machinery.

Horrigan, J. (2008, September 12). Use of Cloud Computing applications and services. Retrieved June 1, 2011, from Pew Internet and the American Life Project: http://www.pewinternet.org/Reports/2008/Use-of-Cloud-Computing-Applications-and-Services/Data-Memo.aspx

Lawler, J.P. (2011). Cloud Computing in the Curricula of Schools of Computer Science and Information Systems. Information Systems Education Journal, 9 (2), 34-54.

Lawler, J.P., Joseph A. (2011). Cloud Computing as a Core Discipline in a Technology Entrepreneurship Program. Information Systems Education Journal, 10 (3), 55-66.

Malan, D. J. (2010, June). Moving CS 50 into the Cloud. Journal of Computing Sciences in Colleges, 25(6), 111-120.

Mell, P., Grance, T. (2011). The National Institute of Science and Technology Definition of Cloud Computing. National Institute of Science and Technology Special Publication 800-145, U.S. Department of Commerce, September, 2011.

Rehman, M. S., & Sakr, M. F. (2010). Teaching in the Cloud. Proceedings of the 2011 IEEE Global Engineering Education Conference (pp. 875-879). Amman, Jordan: IEEE Society.

Roggio, R.F. (2011). Cloud Computing for Capstone Software Development Courses, Information Systems Educators Conference 2011 ISECON Proceedings, v28 n1692.

Sultan, N. (2010, April). Cloud Computing for education: A new dawn? International Journal of Information Management, 30(2), 109-116.

Topi, H., Valacich, J. S., Wright, R. T., Kaiser, K. M., Nunamaker, J. F., Sipior, J. C., et al. (2010). IS 2010 Curriculum guidelines for Undergraduate Degree Programs in Information Systems. Retrieved July 9, 2012, from Association for Computing Machinery Current Curricula: http://www.acm.org/education/curricula/IS%202010%20ACM%20final.pdf

Vaquero, L.M. (2011). EduCloud: PaaS versus IaaS Cloud Usage for an Advanced Computer Science Course, IEEE Transactions on Education, 54(4).

Yuan, M (2011). A Java Developer's Guide to PaaS. InfoQ, December 2011, retrieved on July 1st, from http://www.infoq.com/articles/paas_comparison

ISSN: 2167-1435
v29 n1956

Appendix

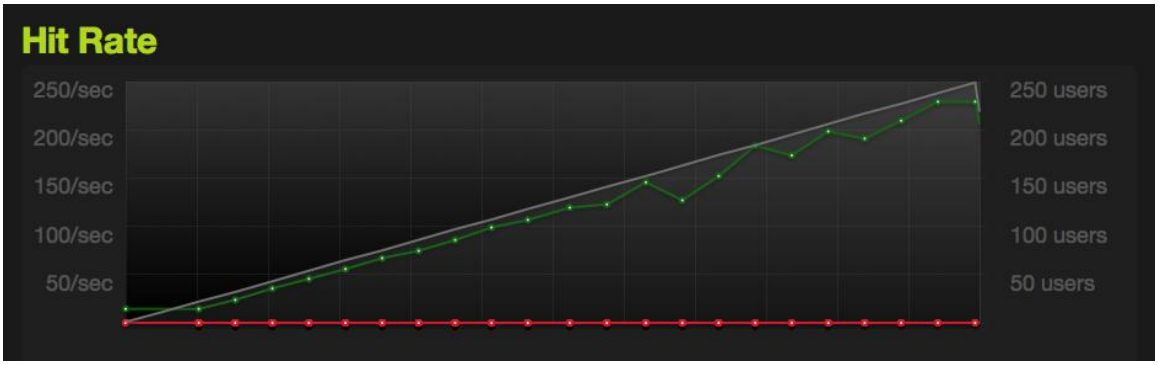

Figure 1. Linear scalability for 250 concurrent users with one AppCell