An Efficient Cloud based Architecture for Integrating Content Management Systems.

Mr. Benny Thomas,Department of Computer Science, Christ University, Bangalore 560 029 India
Benny.thomas@christuniversity.in

Thrivani J , K R Venugopal,
University Visvesvaraya College of Engineering, Bangalore University, Bangalore.

Abstract—The use of digital content is increasing day after day and now it is an essential element of our day today life. The amount of stored information is so huge that it is highly difficult to manage the content especially in a distributed cloud environment. There are many open source software solutions available in cloud to handle huge amount of digital data. However none of these solutions addresses all the requirements needed to manage the content spread out in multiple systems effectively. The user has to relay on multiple content management systems to do the work. This turns into ever more unwieldy, time consuming and leads to loss of data. Using robust and integrated content management systems, these issues could be solved effectively. In this paper we have identified various challenges of using the content management system in the cloud after surveying many Content Management System related article and proposed an integrated solution named Cloud based Architecture integrating Content Management System which is capable of interfacing with various unique features available at different content management system installations in the cloud. This maximizes the functionality and performance of any Content management systems. The Representational State Transfer (REST) protocol is used to integrate the best features of various open source content management systems. REST provides higher level of security compared to existing systems as it does not store the user sessions. The users can interact with the system with the help of an interface which abstracts the complexities of multiple content management systems running in the cloud.

Keywords—CMS, Open Source, Cloud, REST, PhP, MYSQL,Web Services.

1. Introduction

Content management system, (CMS)[1] let people, with no technical background to manage websites, create, edit, organize and publish contents online. In other words [2] CMS is a tool that monitor the entire content on a website, without requiring any technical skills or knowledge of management. CMS are mostly used to launch e-commerce online shops, community portals, personal blogs, or organization websites. Thus, CMS play the role of collecting data and can become intelligent Internet systems. Most of the CMS uses its own unique methods and interfaces to handle the content. Therefore the features from one CMS to the other are different. The users often find it difficult to use the CMS without proper customization which is expensive and need lots

of technical competency. The [3] overabundance of CMS systems makes the decision-making process a challenging task since there is no available classification for CMS. Apart from this there are other challenges with respect to the use of multiple content management systems to full fill the user requirements. There is a potential need for efficient digital content management as the digital information is increasingly produced and used in all areas of day today life. At the same time, number of open source tools used to manage this information also increases every day, leaving behind the biggest challenge on how to select the cheap and best tool for managing large magnitude of digital content effectively.

Traditionally, the user chooses one content management system to store as well as manage all the contents in a way best suited with the requirements. Most of the time user finds it difficult when the required features are not available or not supported by the system. In such cases, the user has to depend on other CMS to do the required operation which is time consuming and cumbersome.

a. Content Management System (CMS)

The CMS is a software based on web technology which provides an environment where the creating, editing, managing of multimedia web content can be made on a reliable and secure way[5]. CMS also provide the best possible solution by organizing information, creating and managing an enterprise's knowledge[6]. The CMSs evolved as a result of large scale data storage and manipulation requirements in the Information systems especially in computer networks. These days there are huge amount of data coming into the information technology space through various heterogeneous sources such as internet, mobile phones and other wireless and digital devices. This leads to data explosion in the Information systems. These data needs to be efficiently managed. Content management system is software used to store and manage large amount of digital content in a computer networked environment such as cloud, intranet or internet. These systems are capable of managing the end-to-end flow of content in an organization, there by converting the raw data in to useful information and knowledge.
 In the olden days the contents are stored in a standalone system in the form of files. Later database management systems (DBMS) are used for storage and manipulation of information. DBMS stores and manages the raw data.

Therefore files and databases systems were sufficient to store and manipulate data with a single stand alone system.

With the emergence of networked system, the content management becomes more complex and network file systems and centralized file sharing systems were used to manage the content. This slowly directs to the development of content management systems. The user requirements on the content got gradually changed and end user started demanding for personalization, analytics and other similar features.

The present scenario is that the Content management systems are getting integrated with many other systems such as Customer relationship management, Analytics management and other heterogeneous tools and platforms in the cloud framework.

b. Cloud Computing

Cloud computing allows the organization to get access to computing based on demand, especially those that do not have the technical as well as financial capability to support their own infrastructure. To the end user, the cloud is not visible therefore the technology that supports the applications does not matter to them. For many organization, cloud computing offers a profitable solution to the problem of how to provide services to a large number of Internet users without spending capital in physical machines that needs to be maintained and upgraded on regular basis by its support staff.

c. Motivation

Most of the Open source CMS do not have all the necessary features needed for the creation of a standard web site. Therefore the additional features are integrated into the CMS with the help of plug-ins. Adding Plug-ins into the CMS needs technical expertise and it is a time consuming process. On top of that most of the time it is difficult to get the right plug-ins with the required features without affecting security and robustness of the system. Otherwise the plug-in with the right features need to be developed which is an expensive process. In order to overcome all these difficulties a cloud based integrated content management system (CicMS) is proposed, which addresses various challenges faced by the users of content management system (CMS) such as lack of essential features in CMS, Security, Performance issues and lack of data integrity. Here we identify the best and unique features of various open source CMS and integrate the best features exist in different CMS in to one common interface. More over [4] CMS cannot offer a standard representation of any content or metadata, instead it should be open to the integration of multiple systems and should provide a generic interface for the user interaction.

The proposed architecture is designed to work in the cloud environment where different installations of CMS from various venders' co exist. The framework is implemented with the help of RESTful apis' and works in private cloud. It is capable of connecting with the public cloud such as Amazon ec2 for backup operations. Cloud virtualization is done with the help of Virtual box in Ubuntu 16.04 environment

The remaining part of the paper consist of existing literature review, Methodology, Various technologies as well as the architecture of the system ,Conclusion and Reference.

III Literature Review

Accroding to S. Angelopoulos et al., [1] The existance of large amout of Content mangement systems makes the decissions making very challanging as there are no classification avaliable for CMS. This study try to identifiy the unique characteristics of CMS and classify them accroding to their performance. Here the CMSs are examined and their characteristcs are recorded in to a binary format. If the characteristcs matchs with the identified general list of characteristics, then it gets a value 1 for that feature.The examined CMS can maximum obtain a score of 51. These are the 51 unique featues of CMS. These are grouped in to six catagories such as Security, Management, Performance, Interoperability, Flexibility and Usability. The CMS with these features can be consided as the best one to select.

A Study on Content Management System using[2] SQL-to-NoSQL schema denormalization and migration, propose an autonomous SQL-to-NoSQL schema denormalization and migration. At the same time, it evaluates the proposed mechanism on the realistic CMS software. Based on their experimental results, this mechanism helps the current CMS software migrate into the cloud platform without re-designing their database schemas at the same time improves at least 45% access performance after schema migration. This research focus on the column-oriented NoSQL database and propose an autonomous SQL-to-NoSQL schema denormalization and migration. The proposed system can read the schemas from MySQL, and then automatically examine and produce the schemas for HBase.

Aakanksha Mirdha et al., [3] compares 7 CMSs, such as Wordpress, Typo3, Joomla, Dotnetnuke, Drupal, Alfresco and Plone on the scope namely, built-in features, system requirements, security, performance,   management and support . These CMSs are chosen based on their market share for e-commerce, page ranking in google, strength of the brand and adoption rate. This paper provides an insight into the open source CMSs available to choose to those who are looking for the best CMS to do online business through websites. As per this work, out of all CMs which provides good system performace and ease, CMS Alfresco proves to provide the maximum ease of use to the users

The research objective of Intuitive content management system [6] is to develop a CMS for non technical people using data-driven technologies. The existing CMSs solutions are not the best choice for non technical users as it is heavy loaded with technologies. Here the users only need to select the content. The rest of the thing will be taken care by the Intuitive CMS, which develop a beautiful website with the

content selected by the user.Here what you see is what you get style is used. So, users are able to recover the data fast and efficiently.It offeres superior performance by fast retrivel of all the required data. It provides a password protection to secure the system offer backup, content recycling and full editor activity logging service.

Keonsoo Lee et al., [7] proposed a Scenario based e-learing CMS, where a method of retaining learner's attention using the scenario based dynamic content management system is proposed. Here the content is personalized or customized to suite the requirement of particular learner using a Scenarios. Here a personlaized program of study is given to the learner using dynamic scnarios. To make the topic more interesting, the learning contents are composed with stories, stories are represented as scenarios, which are dynamically prepared according to the learner's behavior. Therefore, learners become the character of a story and make their own involvements. During these process, lessons are learned and tested as well . This method has improved the efficeny of the e-learing.

Cosmin A et al., [8] proposed a paper describing the main security issues of content management systems and suggested a solution, based on users experience. The issue reported by the users are sent to the developers in order to minimize the possible vulnerabilities over multiple platforms.The main security issues are Data manipulation, Accessing confidential data, Phising and Code execution. Here a web application with a graphical user interface is used with two tabs. One can be used to report the issue found by anyone while using the CMS. The other once consist of list of all the solved issues which are reported. That means the proposed system consist of a tool that allows the user to report any found issues in order to minimize possible security breaches for any kind of content management system.

A new architectural model for a content management system[9] is proposed which includes functionalities like multi-language, copyright, multichannel publishing, natural language processing, garbage collection management which are not included in the traditional ones.Here the issues related to the old CMSs are discussed and a new architechure is proposed to ovecome these issues.Natural language processing is able to recognize voice when it is being spoken.The developmet of NLP is one of the importance of this architecture.

Teduh Dirgahayu et al.,[10] propose an approach to information requirements engineering for specific CMSs. It includes information requirements elicitation and prioritization mehods in the website. This allows content publishers to provide information that meets readers' expectation. The elicitation lists all the information items that shall be published. The prioritization technique priorities information items to be publised. The approach is to facilitate content publishers in defining information to publish and its organization.

The choice of the right CMS solution is a important issue for organizations. Basar Oztaysi[11] proposed a multi critiria dicision model in the selection of CMS among available CMSs as choosing the right CMS is regarded as a critical decision making process in an organization. Here Analytic Hierarchy Process(AHP) is integrated (Grey-TOPSIS) with Technique for Order Preference by Similarity to Ideal Solution method. The proposel with decision model consists of 7 parameters as the criteria and built with four alternative. In the System, the weights of the criteria are determined by Analytic Hierarchy Process method and the alternatives are evaluated by Grey-TOPSIS. Due to the uncertainties, grey numbers are used for evaluations of the alternatives. One at a time sensitivity analysis is also provided in order to monitor the robustness of the method. Here seven criterias were determined for the selection. The most important one are Budget, Vendor reputation, capabilities ,usabilites, technological infrastructure, support criteria and duration.

Amjad Farooq et al., [12] provides a thorough comparative analysis of existing CMSs and suggested clear way to choose best Open Source Context Management System in multiple scenarios. The comparison is done based on the documentation, Administration, Plugins, Setups, Themes Flexibility, Security, Scalability and many other related factors. The paper helps to build powerful website with minimum knowledge on information technology by using cost effective content management systems.

iLaye, [13] an application model developed by Gorrell Cheek et al., provide access control mechanism to third party tools used to add additional functionality to the Content Management Systems such as search, site navigation and location services. It prevents the website from the security vulnerabilities and risks. It protests the Content management system from third party developed applications by using policy recommendations. The policies can be set and enforced by the CMS administrators. The system is made using the principle of least privileged model.

Semantic searches for extracting similarities in Content Management Systems published by Amitrh Ismail et al [14] studies the existing approaches for extracting semantic relationship and explain how Meta data can be applied to enhance the semantic information stored in a Content Management System. Here Meta data is used as a useful tool for addressing the similarities of contents stored to gather specific information. This model supports the formed ontology that is linked to a set of controlled expressions by metadata ontology.

VisualRest:[15] A proposal for the cloud based content management system studies the challenges of managing the digital content in a distributed environment. Here the technologies that can be used to implement the distributed CMS is studied and introduced the VisualRest which provides a uniform way to manage the content stored in heterogeneous devices. It has removed the physical structure and complicated process from the user. The system is made for the distributed and heterogeneous environments. Here users can register their different devises in to a single entry point called the VisualRest server. The servers provide the required information to the connected devices.

IV. Methodology

Here three most popular CMS namely Joomla, Drupal and wordpress are taken for the case study. The features, functionalities and the use of these systems are studied thoroughly. The limitations of these systems are identified. First a feature comparison of the three systems is done. Architecture is developed to increase the functionality, performance and security by integrating the unique features of these three systems. A prototype model for software development is used for implementing the system.

a.Comparative Study of Existing CMS

CMS allows us to develop a website faster. There are many CMS available these days with different functionalities. Among the many collections of CMSs the most popular are Word press, Joomla and Drupal. These are open source systems built on PHP+MySQL technology, however they vary significantly in their features and capabilities. Comparisons of these CMSs' on the bases of its varying features are included. This helps the individual to understand the three CMSs features in-depth as well as it help to choose the best one which suites the user requirements.

JOOMLA

One can develop web sites and web applications with Joomla CMS. It is written using PHP and is free open source software. It provides many built in features that help us to create powerful and dynamic website with many formatting features.

DRUPAL

Drupal is another open source CMS which uses the same technology .It can be setup in Linux, Windows and in Macintosh OS. It uses the presentation abstraction layer architecture. The PAC layer works independently and connects with each other to produce the final result. Unlike Joomla, Drupal does not use the Model View Controller Design pattern.

WORDPRESS

Wordpress is free software used for creating Blogging website dynamically. Blog is used for reviews and opinion on regular basis about any pertinent topic. The word press can be tailored in to a content management system. It is easy to install as well as to maintain and it provides good search engine optimization features.

Features Comparison Table

Features	Drupal	Joomla	Word press
Templates	Yes	Yes	Yes
Modules	Yes	Yes	Yes
Multiple-user -	Yes	Yes	Yes
Multiple-level menu system	Yes	Yes	No
Multiple-site support -	Yes	No	No
Pages	Yes	No	Yes
Blog	No	No	Yes
Contact Form	Yes	No	No
Forum	Yes	No	No
Polls	Yes	Yes	No
Search	Yes	Yes	Yes
Upload	Yes	No	No
Statistics	Yes	No	No
Taxonomy	Yes	No	No
News Feed	Yes	Yes	Yes
Comment and Tracker	Yes	No	No
Book	Yes	No	No
Content translation	Yes	No	No
Ping	Yes	No	No
Profile	Yes	No	Yes
Throttle	Yes	No	No
Trigger	Yes	No	No
Article	No	Yes	No
Banner	No	Yes	No
Weblinks	No	Yes	No

In the feature comparison table, some of the most prominent features of CMS are included. The presence or absence of these features in three CMS is compared in the table. Out of the Twenty five features mentioned on the table, Word press has only eight features, Joomla has ten, and where as drupal has twenty one features. It is also observed that all these features are present in one or the other CMS.
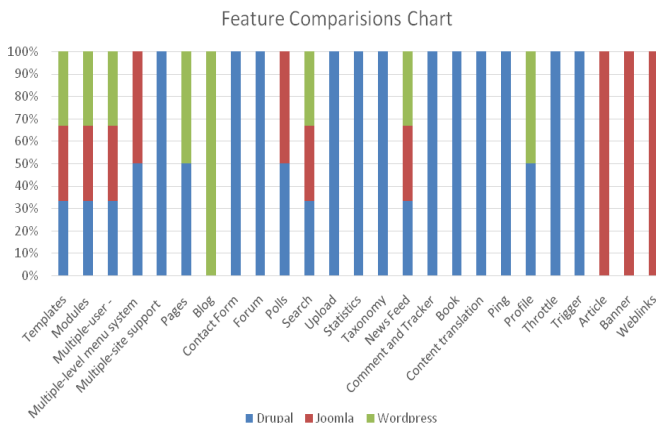
The Graph here shows a clearer picture of the presence of features in the CMS. The line with one color shows that the feature is present only in one CMS. The line with two colors shows that the feature is present in two CMS. The line with three colors shows that the feature is present in all the three CMS

b. RESTful Web Services

Initially RESTful Web Services architecture was developed to use in distributed hyper media platforms. Currently it is majorly used in the industry as the guiding principle for designing the web services. It uses the W3C/IETF standards and infrastructure. REST provides a simpler method to access the net

Representational State Transfer (ReST) is an extensively resource oriented architecture used between client application and web servers with ReST API . ReST is a set of ideas about how data can be consumed elegantly to different kind of application, and although it's not a part of HTTP, it is generally discussed in the context of HTTP because resources are stored by servers and used by client using HTTP method, where resources are contacted via its unique URI and transfers its state using its state representation. The connection between client and the server in REST is stateless; hence a server need not require maintaining state with each individual client. A client authenticates itself by a client ID or token during each transaction.

RESTful Web Services have the following characteristics
a. *Addressability:* Resources are identified with Uniform Resource Identifier. They do not require a separate resource discovery and location mechanism such as UDDI. They can be accessed once exposed to the web
b. *Statelessness:* Each transaction is independent of the other and unrelated to previous ones. Servers do not keep states between requests.
d. *Uniform Interface: M*anipulation of the resource is done using a fix set of HTTP methods (GET, PUT, DELETE and POST) without special code to deal with each one.
RESTful Web Services are easy to install, distribute, find and invoke. Resources are exposed dynamically computed on the

Web. The elasticity and loose coupling these services pay for are highly useful to both clients and servers

V, Architecture of CiCMS

The implementation is done with the help RESTful services which are used in the contest of web application. The REST is architecture where the resources are stored in the servers and used by the clients' application using HTTP methods. A resource is connected via the Uniform resource locator and the state is transmitted using its representation state.
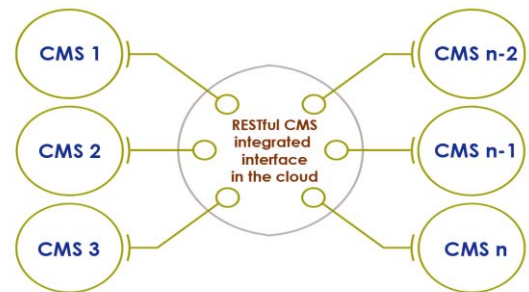

Figure 2: Architecture of CicMS.

The purpose of CicMS is to build anarchitecture to integrate the best features of various open source Content Management systems in the cloud. The various CMS in cloud has their own unique features that make them to stand out. At the same time these unique features may not be available with the other open source CMS. Therefore most of the time the end user has to depend on more than one content management system to do their intended work to get the result. In such situations the data portability from one CMS to another CMS becomes cumbersome and arduous. This can also lead to data lose and it in turn affect the data integrity. The architecture implements an interface, which connects various content management systems present in the cloud. The best features of these CMSs will be integrated in to the interface. For the end user this interface act as a CMS with all the unique combination of best features from other CMSs. Using this functionality of CMS can be maximized. The system will be faster and secure as it uses RESTful webservices.

In order to implement the CicMS interface in the cloud first identify the open source CMS exist in the cloud and study the best unique features and common features of each of these CMSs. The common features or overlapping features can be extracted into the CicCMS interface from one of these most user friendly CMSs. The interface should have the capability to interact with databases of these CMSs. The data storage can be done in one of the selected master databases. However the interface should have the capability to store the data in any CMS databases based on the requirement.

Each CMS can maintain their own local data as well as it can connect its data with the other CMS. The interface follows service oriented architecture, selects one or more

CMS running in the background on the fly to do the user job. The interface should hide the complexity of the CMS selection from user. I.e. The existence of multiple autonomous Content management systems running in the background in cloud will be hidden from the end user.

The interface should support an access control list (ACL) and there can be users with different kinds of privileges ranging from administer to ordinary end user. The administrator can set the main database storage and provide access control list to users. Based on theses access permissions users can use various best features available in the CicMS.

VI. Conclusion

In this paper the architecture, design of cloud based integrated content management system interface is implemented which combine the best features of various content management systems. This addresses the problem of how to integrate the best features of various content management systems in the cloud. This architecture is good for any organization who wishes to implement a cloud based content management system. We are using only open source software to keep the cost as low as possible without compromising the quality of the implementation. All the tools used to develop the system can be downloaded from the respective website. The system has good flexibility and reliability. In future it can further be expanded to incorporate automatic process for resource allocation and more secure user authentication method.

References

[1] C. H. Lee and Y. L. Zheng, "SQL-To-NoSQL Schema Denormalization and Migration: A Study on Content Management Systems," *Proc. - 2015 IEEE Int. Conf.*
   Syst. Man, Cybern. SMC 2015, pp. 2022–2026, 2016.

[2] A. Mirdha, A. Jain, and K. Shah, "Comparative Analysis of Open Source Content Management Systems," *IEEE Int. Conf. Comput. Intell. Comput. Res.*, pp. 1–4, 2014.

[3] S. Angelopoulos and F. Kitsios, "Decision making for content management systems: Criteria identification and categorisation," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, vol. 2015–March, pp. 1233–1239, 2015.

[4] A. Ulrich Mauthe and P. Thomas, *Professional Content Management Systems. Handling Digital Media Assets*. 2004.

[5] A. Rivera, S. Candela, and C. R. García, "WikiDIS: A case of collaborative content management system for educative community," *2010 IEEE Educ. Eng. Conf. EDUCON 2010*, pp. 601–610, 2010.

[6] C. C. Hoong, "Intuitive Content Management System," no. I4ct, pp. 541–543, 2015.

[7] K. Lee, S. Rho, and E. Hwang, "Scenario Based Dynamic Content Management System for e-Learning Environment," *2016 30th Int. Conf. Adv. Inf. Netw. Appl. Work.*, pp. 183–186, 2016.

[8] E. C. Popovici, O. Fratu, and I. Technology, "Security Issues in Most Popular Content Management Systems," no. January, pp. 277–280, 2016.

[9] A. Saha and S. K. Setua, "Architectural model for next generation content management system," *8th Int. Conf. Electr. Comput. Eng. Adv. Technol. a Better Tomorrow, ICECE 2014*, pp. 437–440, 2015.

[10] T. Dirgahayu, N. Setiani, and Z. Zukhri, "Information requirements engineering for specific content management systems," pp. 54–59, 2014.

[11] B. Oztaysi, "A decision model for information technology selection using AHP integrated TOPSIS-Grey: The case of content management systems," *Knowledge-Based Syst.*, vol. 70, pp. 44–54, 2014.

[12] A. Farooq, F. Javed, M. Hussain, T. Abbas, A. Hussain, and C. Science, "Open Source Content Management Systems :," pp. 38–43, 2012.

[13] G. Cheek, M. Shehab, T. Ung, and E. Williams, "iLayer: Toward an Application Access Control Framework for Content Management Systems," *Policies Distrib. Syst. Networks (POLICY), 2011 IEEE Int. Symp.*, pp. 65–72, 2011.

[14] A. Ismail and M. Joy, "Semantic searches for extracting similarities in a content management system," *2011 Int. Conf. Semant. Technol. Inf. Retrieval, STAIR 2011*, no. June, pp. 113–118, 2011.

[15] N. Makitalo, H. Peltola, J. Salo, and T. Turto, "VisualREST: A Content Management System for Cloud Computing Environment," *37th EUROMICRO Conf. Softw. Eng. Adv. Appl.*, pp. 183–187, 2011.