Design and implementation of a social network infrastructure for designers of Multi-Cloud applications

Christos Papoulas

Thesis submitted in partial fulfillment of the requirements for the

Masters' of Science degree in Computer Science

University of Crete
School of Sciences and Engineering
Computer Science Department
Knossou Av., P.O. Box 2208, Heraklion, GR-71409, Greece

Thesis Advisors: Prof. *Kostas*, Dr. *Magoutis*

Your Title

Thesis submitted by
Author Name
in partial fulfillment of the requirements for the
Masters' of Science degree in Computer Science

THESIS APPROVAL

Author: _____
Author Name

Committee approvals: _____
Name of first member
Assistant Professor, Thesis Supervisor

_____
Name of second member
Associate Professor, Committee Member

_____
Name of third member
Professor, Committee Member

Departmental approval: _____
Name of Director of Graduate Studies
Professor, Director of Graduate Studies

Heraklion, July 2015

Abstract

In this work we propose a Social Network Platform, both front-end and back-end technology, for model-driven software design and deployment of applications in Multi-cloud environments. Nowadays, DevOps users wander around the web looking for automated tools like Chef supermarket, IBM bluemix and other deployment tools where they almost manually configure and deploy their applications without any model-driven concept. The users can create their own Cloud Application Modelling and Execution Language models or benefit from automatically generated models. We provide a Platform with an integrated community helping the DevOps users deploy their applications. We incentivize user to stay inside the Platform instead of roaming around the web looking for other Q&A sites like StackOverflow to find their answers. The runtime executions of applications can be uploaded and analysed by the Platform presenting cost effectiveness analysis.

Περίληψη

Στην εργασία αυτή . . .

Acknowledgements

Contents

II

List of Figures

IV

List of Tables

Chapter 1

Introduction

In this work, the design and implementation of a model-driven social networking platform for designers of Multi-Cloud applications is presented. The model-driven concept is not included in the Social Network Implementation but in the design of the users' application. In this targeted social networking platform, DevOps can benefit from other users' experience and answer design questions such as which is the most cost-effective deployment and which configuration fits their needs. This social networking platform binds all social networking concepts such as personal messaging, groups, new feeds with modelling-driven concepts of application composition and deployment, integrating a repository of cloud applications and infrastructure description based on Cloud Application Modelling and Execution Language (CAMEL).

This repository means to several benefits. An integrated environment can enrich user interactions with structured references to applications and their components, execution data, and mined knowledge from real deployments. Mined knowledge can be combined with user activity and profiles to provide personalized suggestions and hints. An improved mode of user interaction is expected to result to stronger incentives for DevOps users to contribute information to the underlying repositories. More content should lead to better quality of mined knowledge, benefiting the DevOps community and providing further incentive for contributions. The social networking platform designed to be closely integrated with a set of information repositories satisfying the following requirements: (R1) handle entire applications rather than just software components; (R2) abstract application structure through software modeling; (R3) capture and analyze application runtime performance. Raising the level of abstraction from components to applications and from code-centric to model-centric is expected to facilitate interaction between DevOps professionals. The analysis of application execution data can provide answers to many interesting questions of the community and support discussions and arguments with hard data. These requirements can provide software developers with strong incentives to contribute, leading to the sustainability and growth of information and derived knowledge in the repository.

1.1 Background

Application models inside social network platform are described in CAMEL. CAMEL integrates various domain-specific languages (DSLs). These DSLs cover a wealth of aspects of specification and execution of multi-cloud applications like CloudML, Scalability rules, WS Agreement, Saloon and Historical Execution Data. CAMEL is using the Eclipse Modelling Framework (EMF) on top of the Connected Data Objects (CDO). Applications is persisted on CDO repository. CloudML [1] is a recent approach that focuses on the provisioning and deployment of multi-cloud applications, is built upon MDE techniques and methods, and provides a models@run-time environment for enacting the provisioning. WS Agreement [2] is a Web Services protocol for establishing agreement between two parties, such as between a service provider and customer. Saloon [3] is an approach that uses models to represent clouds variability, as well as ontologies to describe the heterogeneous aspect of the cloud ecosystem. The CAMEL model couple together aAll those DSLs as shown in figure 1.1.

Chef . . .

Figure 1.1: CAMEL DSLs.
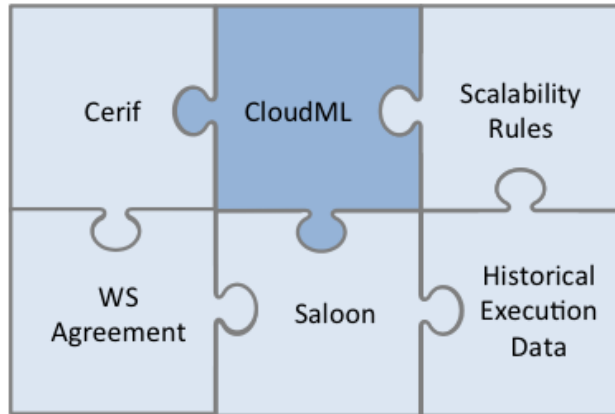
1.2 Related Work

This section describes related work for other professional networks and their caching architecture.

1.2.1 Caching data

Facebook, the largest social network, serves billions of requests per second using memcached [4]. In this magnitude of scale, Facebook has several pools of memcached servers (regional pools) around the globe. A request for a single page can produce hundred of requests to the back-end system. Memcached is used to store

not only key-value from MySQL queries but also pre-computed results from sophisticated algorithms. In order to achieve a near real time communication experience to the end user, memcached servers have to be efficient, reducing latency.

The research question in such systems is when a particular key will be invalidated. This problem occurs according to [4] in two cases: (1) *stale sets* and (2) *thundering herds*. A stale set occurs when a web server sets a value to the memcached that does not reflect the real value of the database. Thundering herds occur when a specific key has a heavy read and write activity at the same time. Stale sets are resolved by an N-bit token, that is bound to a specific key and sent from the memcached to the web server that wants to update the key when a cache miss occurs. If a delete request is received, the request for updating this value from the client is rejected. The thundering herbs are solved by configuring the memcached servers to return an N-bit token only once every ten seconds per key.

Linkedin, the largest professional network, stores hundreds of terabytes of data to Project Voldemort [5], a key-value store, inspired by Amazon Dynamo [6]. Linkedin stores to Voldemort pre-computed offline data. For example, the results of data mining applications, such as features like "People You May Know", that are running on hundreds of terabytes to make an estimation and are using Hadoop as the computational component of those estimations. Voldemort and Dynamo have the same following requirements: (1) a simple *get/put* application interface (2) A *replication* factor, the number of replicas for each key-value tuble, implemented using vector clock, (3) a *required read* factor to succeed a get request and (4) a *required write* factor to succeed a put request.

1.2.2 Professional Networks

IT professionals use a variety of online sources as aids in their daily tasks. Developers typically prefer community-moderated forums over vendor-moderated sites. Social networks focusing on software technology in particular provide developers with the opportunity to leverage the knowledge and expertise of their peers.

One of the most popular such platforms is GitHub [7], a collaborative revision control platform for developers launched in April 2008, and arguably the largest code-hosting site in the world. GitHub provides social networking functionality such as feeds, followers, wikis and a social network graph that captures how developers work on versions of their repositories, which version is newest, etc. Gitter [8] is a related service that facilitates discussions between members of GitHub communities by providing a long-term chat integrated with code and issues. Sourceforge [9] was the first code-hosting platform offered to open-source projects. It was launched in 1999 and offered IT professionals the ability to develop, download, review, and publish open-source software. Sourceforge is similar to GitHub in its support for social features. Other similar code-hosting platforms are Google Code [10] and Microsoft CodePlex [11]. None of those platforms collect, analyze, or use information from executions of application deployments to improve the level of technical discussion between users or abstract code structure through modelling or enhance

user interactions through the use of analytics over application execution histories.

StackOverflow [12] advances on earlier Q&A sites in which users ask and answer questions. Users can vote up or down questions and answers and earn *reputation points* and *badges* in return for their active participation. Although StackOverflow and GitHub address different aspects of software development (StackOverflow is not a code-hosting platform) there is a synergy and correlation between the two [13]. The proposed social network platform extends StackOverflow through the use of social networking features that enable users interested in reasoning about application deployments to use and share knowledge drawn from analyses of information repositories.

IBM's BlueMix [14] is a development and support platform for communities of DevOps users wishing to compose distributed applications out of components drawn from libraries and deploy them at IBM-provided and supported cloud infrastructure. BlueMix is a key component of IBM's DevOps best practices [15] for achieving rapid prototyping, automated deployment, and continuous testing of software. BlueMix encourages its users to ask their questions to StackOverflow but also includes a community forum [14] with rating of answers contributes to eventually building a basic knowledge base, similar to traditional approaches such as StackOverflow. The proposed social network platform system differs from BlueMix in its support for expressing applications as models (CloudML, CAMEL) and its use of two information repositories, the PaaSage repository of models and execution histories and Chef supermarket, and the use of analytics over past executions to enable users to reason about application deployments. A common feature between the proposed social network platform and BlueMix is support for deployment of distributed applications.

LinkedIn widely adopted across a range of professional communities due to its robust set of social features (and to some extent due to its use of extensive analytics over collected information [16]), LinkedIn provides no specific support for software engineering activities and thus more closely resembles traditional social networking platforms such as Facebook.

The lack of Social Networking features of github came to fill the Geeklist platform [17], where developers and IT companies can discover and share the work they have done, connect with other companies in a social network manner or join development communities. Another code hosting platform is Snipplr [18], where developers can upload short code snippets but not full programs, in order to keep all of their frequently used code in one place that is accessible from any computer and any user. Masterbranch [19] is a new under development platform that allows collating and sharing of projects within a user's profile. This profile works similarly to LinkedIn and has an incentivisation scheme called DevScore, coupled with unlockable achievements that add a gamification element. Dzone [20] is essentially a link repository for developers allowing link sharing and incentivisation based on voting for the popular links. The Code project [21] website and forum allow code-specific discussion and share relevant articles and news, contains blogs, newsletter and a questions and answers section.

The above systems can be further classified based on whether they use a repository to store software-related information (code, models, configuration, or execution histories) and whether this information is shared and raised through crowd sourcing [22]. GitHub, GoogleCode, CodePlex, SourceForge, BlueMix, Chef Supermarket, and our platform store at least one type of software-related information and all systems but BlueMix are raising shared content in their software-related repositories via crowd sourcing. Our professional network is the only solution that analyzes information in its software-related repositories to assist users with suggestions and hints.

Table 1.1: Feature comparison.

	User Interaction				Repository					Repo assisted hints [b]	Application deployment
	Social features [a]	Groups	Q & A	Personal messaging	Software code	Software models	Software config	Execution histories	Crowd sourced		
GitHub	✓	✗	✗	✗	✓	✗	✗	✗	✓	✗	✗
Sourceforge	✓	✗	✓	✓	✓	✗	✗	✗	✓	✗	✗
GoogleCode	✗	✗	✓	✗	✓	✗	✗	✗	✓	✗	✗
CodePlex	✓	✗	✓	✓	✓	✗	✗	✗	✓	✗	✗
StackOverflow	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗
BlueMix	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✓
Chef Supermarket	✗	✗	✗	✗	✗	✗	✓	✗	✓	✗	✗
LinkedIn	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗
geeklist	✓	✓	✗	✗	✗	✗	✗	✗		✗	✗
Snipplr	✗	✗	✗	✗	✓	✗	✗	✗		✗	✗
Masterbranch	✗	✗	✗	✗	✓	✗	✗	✗		✗	✗
Dzone	✓	✗	✗	✗	✓	✗	✗	✗		✗	✗
codeproject	✓	✗	✓	✗	✓	✗	✗	✗		✗	✗
PaaSage SN	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓

[a] Features: follow and news feed
[b] User assistance based on data analysis of the repository

Chapter 2

Social Network User Interface

In this chapter described the User Interface of Social Network implementated based on 104 mock-ups created by HCI expert team. The key design objective of the social network platform is to create a strong bond between (i) software engineering services for managing and deploying cloud-targeted application models; and (ii) community-oriented facilities for communication and collaboration between users. The interconnections between the two in the design of the user interface are depicted in Figure 2.1. The prototype implementation is publicly accessible on-line at http://socialnetwork.paasage.eu.

2.1 User Interface

The discrete entities, which bound together the Social Networking with model aspects of Platform are:

- *Application Models.* An example is shown if figure 2.2, consisting of a human friendly description (label 1 in fig.2.2), the Camel Description of the model (label 2 in fig.2.2), reviews about the model (label 3 in fig.2.2). An overview of engineering aspects such as version and runs (label 4 in fig.2.2) and an overview of social aspects such as share and watches (label 5 in fig.2.2)

- *Components.* We have integrated the Chef supermarket components into Social Network Platform. The components help the DevOps users to generate their application models.

- *Groups.*

- *Users.*

2.2 Gamification

Figure 2.1: The engineering & social activities are seamlessly within the Platform.

Figure 2.2: The application model home page
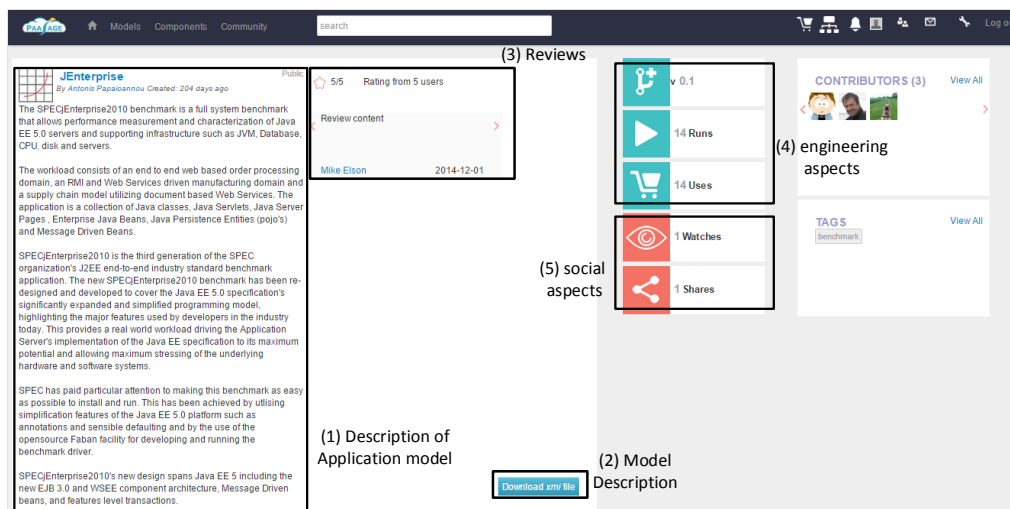
Chapter 3

Implementation

This section describes the implementation of social network site and how the system scales.

The system is composed by the following components, as shown in figure 3.1: At the first layer lives (1) the Social Networking engine, which runs all PHP scripts and described in section 3.1. At the second layer lives (2) the Memcached caching system, which described in section 3.3. At the third layer lives (3) the Social Network MySQL database, and (4) the CDO server - client components and the CDO repository.

Achieving the scalability of the system, two system architectures are examined at two layers of the system: (1) We added more than one Social Network engine at the first layer of the system. In this implementation, in order to keep the file system in consistent mode we integraded Apache Zookeeper[23]. (2) We added more than one memcached machines at the second layer in order to add more cpu capacity and improve the system response time.

3.1 Implementation of Social Network

The social networking platform is implemented over the extensible Elgg social network framework[24]. Elgg is open source software written in PHP, uses MySQL for data persistence and supports jQuery [25] for client-side scripting. The architecture of Elgg Social Network shown in figure 3.3. The Model of the framework is structured around the following key concepts as shown in figure 3.2

- *Entities*, classes capturing social networking concepts: users, communities, application models. Elgg Core comes with four basic objects: ElggObject, ElggUser, ElggGroup, ElggSite, ElggSession, ElggCache and a lot of other classes necessary for the proper engine operation.

- *Metadata* describing and extending entities (e.g., a response to a question, a review of an application model, etc.).

Figure 3.1: The overall architecture of Social Network.
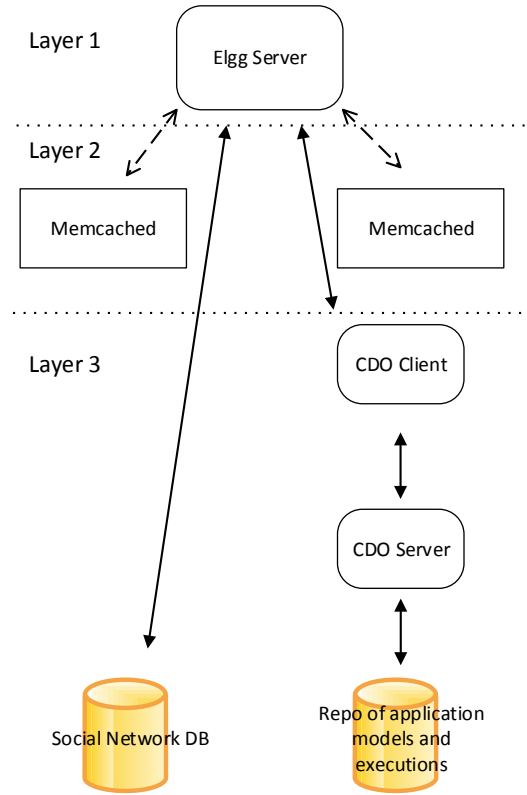
- *Relationships* connect two entities (e.g., user A is a friend of user B, user C is a contributor to an application model, etc.) and are persisted in the Social Network DB.

- *Annotations* are pieces of simple data attached to an entity that allow users to leave ratings, or other relevant feedback.

All Elgg objects inherit from ElggEntity, which provides the general attributes of an object. Elgg core comes with the following basic entities: ElggObject, ElggUser, ElggGroup, ElggSite, ElggSession, ElggCache, as well as other classes necessary for the operation of the engine.

Elgg comprises a core system that can be extended through plugins (examples are the Cart system or the handling of Application Models). Plugins add new functionality, can customize aspects of the Elgg engine, or change the representation of pages. A plugin can create new objects (e.g., ApplicationObject) characterized (through inheritance of ElggEntity) by a numeric globally unique identifier (GUID), owner GUID, Access ID. Access ID encodes permissions ensuring that when a page

Figure 3.2: The Elgg Engine Data model.
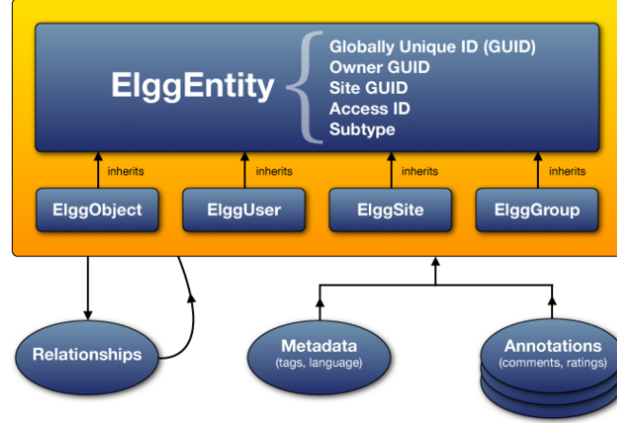
requests data it does not touch data the current user does not have permissions on.

Figure3.3 shows the model, view, and control parts of Elgg's architecture. In a typical scenario, a web client requests an HTML page (e.g., the description of an application model). The request arrives at the *Controller*, which confirms that the application exists and instructs *Model* to increase the view counter on the application model object. The controller dispatches the request to the appropriate handler (e.g., application model, component handler, community handler) which then turns the request to the view system. View pulls the information about the application model and creates the HTML page returned to the web client.

Figure 3.3: Architecture of the Elgg Social Networking engine.
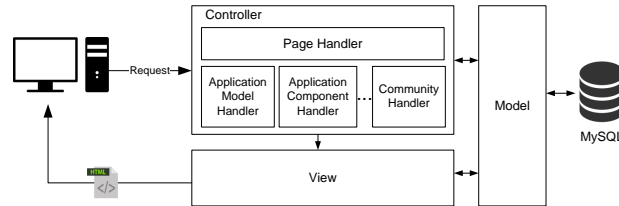

The extensibility of Elgg can be established not by modifying the core system but by introducing new plug-ins which follow the MVC model. A new plug-in can create a new entity. Thus, each entity is characterized by a numeric Globally Unique Identifier and Access ID. The Access ID determines the permissions that other users have. Thus, when a page requests data, it never touches those data that the current user does not have permission to see. All plug-ins share a common structure of folders and PHP files, following the MVC model of figure 3.3.

The hierarchy of a plug-in is shown in figure 3.4. Folder *actions* includes the actions applied on application models. Every active participation by the user is

performed via an action. Logging in, creating, updating or deleting content are all generic categories of actions. The *views* folder contains the *php* forms applied on application models, *river* events (Elgg terminology for live feeds). Viewss are responsible for creating the output for the client browser. Generally, this will be HTML, but it can be also JSON or other format. *Pages* overrides elements of core Elgg pages and can be from chunks of presentation output (like sidebars) down to individual html code. The *js* and *lib* folder provides javascript and *php* library functions. Finally, the *vendors* folders include third-party frameworks such as Twitter's bootstrap front-end [26]. The most important file of a plug-in is the *start.php* script, which contains the *page handler*. Page handler is a function manages the plug-in pages enabling custom url redirect to a specific page. The plug-in initialization is also defined in the start.php and registers actions, events and determines the views.

Figure 3.4: The structure of the application description plug-in.
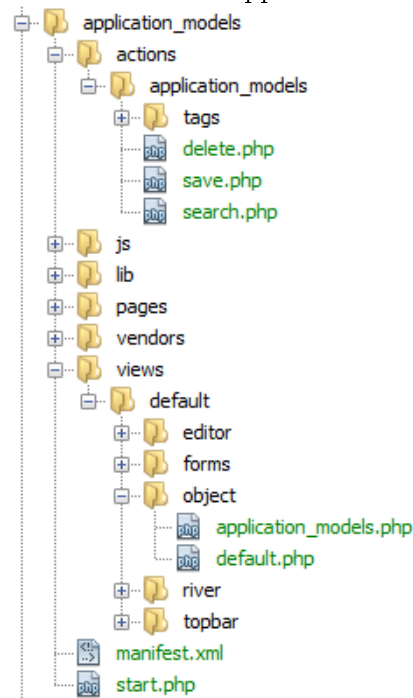
The execution history of deployments of application models and the description of those models is stored in the CAMEL information repository, which is implemented as an Eclipse CDO server. The exchange of information between Elgg and the CAMEL information repository is going through CDO Client who retrieves the information from CDO server and sent it to the Elgg over sockets.

3.2 User interface

The design of User interface of Social Network based on 102 mock-ups designed by HCI experts. In order to support those look & feel and the functionality of those mock-ups 25K lines of php, js and css code is written. . . .
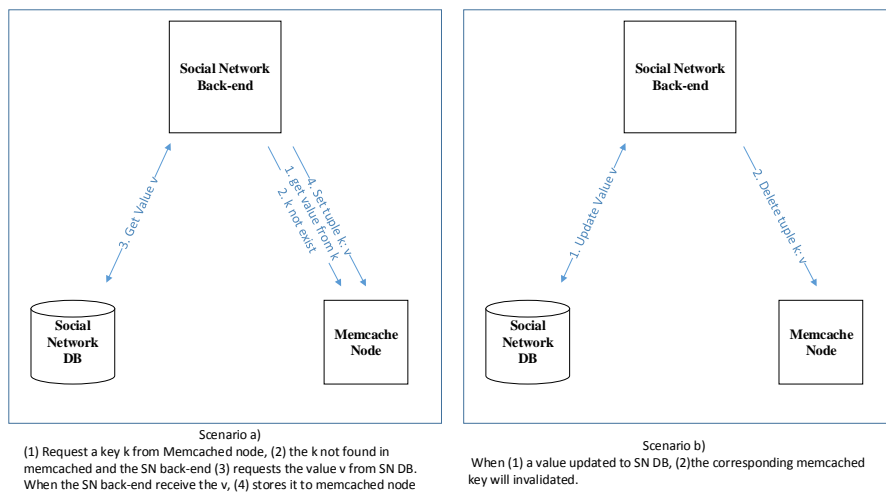
3.3 Memcache

This section describes the experience gained by using memcached[27]. Memcached is an open source, high-performance, distributed memory object caching system. We choose memcached, because is a generic simple in-memory key-value store. It has a powerful API available for PHP. After memcached integration the system increase the response time and performance.

Memcached added in layer 2 of system architecture and used for storing the following key-value tuples: (1) values from Social Network Database such as entities of Social Network, applications, components, users, group discussions, (2) evaluated javascript code results and (3) executions histories from repository of application models. Storing the executions of applications at Memcached the response time of the system increased because the PHP modules do not need to go through the heavy CDO client but get directly the executions of applications from Memcached.

The apache jmeter[28] was used to measure the responce time of the system and the sysstat tool[29] was used to measure the cpu usage. Section 4.1 shows the performance results of this implementation.

All tuples at Memcached are inserted with maximum key expiration time of thirty days. When a value in social network is updated, the memcached key will deleted 3.5.

Figure 3.5: The scenario a depicts a request from memcached when the key does not exist and scenario b depicts a updated operation of a value)

Scenario a)
(1) Request a key k from Memcached node, (2) the k not found in memcached and the SN back-end (3) requests the value v from SN DB. When the SN back-end receive the v, (4) stores it to memcached node

Scenario b)
When (1) a value updated to SN DB, (2)the corresponding memcached key will invalidated.
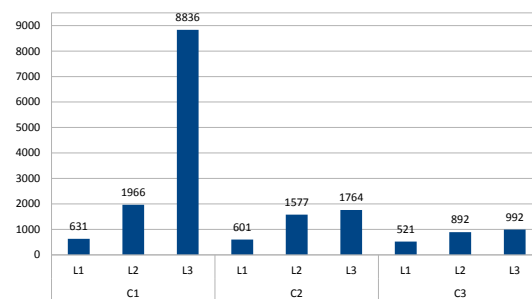
Chapter 4

Evaluation

This chapter describes the evaluation of the two different implementation of the system architecture. (1) By introducing more than one memcached instances at layer 2 as figure 3.1 shows and (2) by introducing more than one Social Network engines at layer 1.

4.1 Improving Performance with memcached

The test perform with the following loads: (L1) ten users requests two applications consecutively one hundred times, (L2) ten users requests four applications consecutively one hundred times and (L3) ten users requests eight applications consecutively one hundred times. In this experiment we kept constant the following components of system: the Elgg front-end apache server, the Social network database, and the CDO server - client communication. In the system, increased the number of memcached nodes. The figure **??** shows the average, minimum and maximum response time with the following system configuration: (C1) no memcached node, (C2) one memcache node added and (C3) two memcache nodes added.

4.2 Improving Performance with engine

Figure 4.1: The average response time for all configurations.

Chapter 5

Comparison

Compare your work . . .

5.1   AA

5.2   BB

Chapter 6

Conclusions and Future Work

Bibliography

[1] N. Ferry, A. Rossini, F. Chauvel, B. Morin, and A. Solberg, "Towards model-driven provisioning, deployment, monitoring, and adaptation of multi-cloud systems," in *Proceedings of CLOUD 2013: 6th IEEE International Conference on Cloud Computing*, L. O'Conner, Ed. IEEE Computer Society, 2013, pp. 887–894.

[2] A. Andrieux, K. Czajkowski, A. Dan, K. Keahey, H. Ludwig, T. Nakata, J. Pruyne, J. Rofrano, S. Tuecke, and M. Xu, "Web services agreement specification (ws-agreement)," in *Open Grid Forum*, vol. 128, 2007, p. 216.

[3] C. Quinton, N. Haderer, R. Rouvoy, and L. Duchien, "Towards multi-cloud configurations using feature models and ontologies," in *Proceedings of the 2013 international workshop on Multi-cloud applications and federated clouds*. ACM, 2013, pp. 21–26.

[4] R. Nishtala, H. Fugal, S. Grimm, M. Kwiatkowski, H. Lee, H. C. Li, R. McElroy, M. Paleczny, D. Peek, P. Saab *et al.*, "Scaling memcache at facebook." in *nsdi*, vol. 13, 2013, pp. 385–398.

[5] R. Sumbaly, J. Kreps, L. Gao, A. Feinberg, C. Soman, and S. Shah, "Serving large-scale batch computed data with project voldemort," in *Proceedings of the 10th USENIX conference on File and Storage Technologies*. USENIX Association, 2012, pp. 18–18.

[6] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall, and W. Vogels, "Dynamo: amazon's highly available key-value store," in *ACM SIGOPS Operating Systems Review*, vol. 41, no. 6. ACM, 2007, pp. 205–220.

[7] "Github," http://github.com/, [Online; accessed 24-May-2015].

[8] "Gitter: The chat for github," http://gitter.im/, [Online; accessed 24-May-2015].

[9] "Sourceforge," http://sourceforge.net, [Online; accessed 24-May-2015].

[10] "Google code," https://code.google.com, [Online; accessed 24-May-2015].

[11] "Codeplex: Project hosting for open source software," https://www.codeplex.com, [Online; accessed 24-May-2015].

[12] "Stackoverflow," http://stackoverflow.com/, [Online; accessed 24-May-2015].

[13] B. Vasilescu, V. Filkov, and A. Serebrenik, "Stackoverflow and github: Associations between software development and crowdsourced knowledge," in *Social Computing (SocialCom), 2013 International Conference on*, Sept 2013, pp. 188–195.

[14] "Ibm blue mix for developers," http://https://developer.ibm.com/bluemix/, [Online; accessed 24-May-2015].

[15] "Ibm devops best practices," http://www.ibm.com/developerworks/devops/practices.html, [Online; accessed 24-May-2015].

[16] R. Sumbaly, J. Kreps, and S. Shah, "The big data ecosystem at linkedin," in *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. ACM, 2013, pp. 1125–1134.

[17] "Geeklist," https://geekli.st/, [Online; accessed 10-June-2015].

[18] "Snipplr," http://snipplr.com/, [Online; accessed 10-June-2015].

[19] "Masterbranch," https://masterbranch.com/, [Online; accessed 10-June-2015].

[20] "Dzone," http://www.dzone.com/, [Online; accessed 10-June-2015].

[21] "The code project," http://www.codeproject.com/, [Online; accessed 10-June-2015].

[22] J. Howe, "The rise of crowdsourcing," *Wired magazine*, vol. 14, no. 6, pp. 1–4, 2006.

[23] A. Zookeeper, "Apache zookeeper," https://zookeeper.apache.org/, 2015, [Online; accessed 20-May-2015].

[24] E. S. N. Engine, "Elgg social networking engine," http://elgg.org/, 2015, [Online; accessed 19-May-2015].

[25] jQuery, "jquery," https://jquery.com/, 2015, [Online; accessed 19-May-2015].

[26] "Bootstrap front-end framework," http://getbootstrap.com/2.3.2/, [Online; accessed 24-May-2015].

[27] Memcache, "Memcache," http://memcached.org/, 2015, [Online; accessed 18-May-2015].

[28] A. jMeter, "Apache jmeter," http://jmeter.apache.org/, 2015, [Online; accessed 19-May-2015].

[29] sysstat, "sysstat," https://github.com/sysstat/sysstat, 2015, [Online; accessed 19-May-2015].