CLOUD БАЗИРАНА СИСТЕМА ЗА СЪБИРАНЕ, СЪХРАНЯВАНЕ И АНАЛИЗ НА ЛОГОВИ СЪОБЩЕНИЯ

Стоян Попов, Елена Сомова

ПУ “Паисий Хилендарски”
stoyan.popov@gmail.com, eledel@uni-plovdiv.bg

***Резюме**: Работата представя софтуерна система CloudLog, чиято цел е да разреши основните проблеми при работата с логове от множество източници. CloudLog е cloud базирана система за събиране, централизирано съхраняване и анализ на логове. CloudLog използва cloud уеб услугите, предоставени от Amazon.com (AWS). Данни се съхраняват в cloud базираната база данни SimpleDB, а кодът на приложението може да се изпълнява на 2 или повече EC2 инстанции, разположени зад Elastic Load Balancer с цел постигане на по-голяма надеждност и хоризонтална мащабируемост. Цялата комуникация от и към уеб приложението е изцяло криптирана.*

***Ключови думи**: cloud computing, cloud базирани системи, логови съобщения*

1. Въведение

Всяка компютърна система, независимо от своя хардуер и операционна система, има механизъм за записване на събитията, които се случват при работа с нея, т. нар. ***логове***. Те не представляват интерес за обикновения потребител, но имат ключова роля в работата на системните и мрежови администратори. Първоначално логовете са използвани само като средство за откриване на проблеми, но в наши дни те служат и за много други цели като: оптимизация на системи и мрежи, записване на извършвани от потребителите дейности и откриване на опити за неправомерни операции.

Добрите практики, фирмените политики и дори законодателството в някои страни изискват от ИТ отделите да поддържат точен регистър на всички събития, случили се с управляваните от тях системи, с цел да се даде възможност за осъществяване при необходимост на допълнителен анализ. Правенето на такъв анализ за всяка една система поотделно е не само изключително времеемко, но и сравнително несигурно, защото например е възможно да се компрометира някоя от системите в дадена мрежа и да се прикрият следите като се заличат части от логовете на тази машина.

В практиката се използват 4 ***модела за съхраняване*** на логове в зависимост от броя на системите, които генерират данни и избора на място за съхраняване на тези данни: съхраняване във файлове при една система (обикновено на твърдия диск), централизирано съхраняване във файлове при

множество системи (обикновено на твърдия диск на сървър), централизирано съхраняване в база данни (на сървър) при множество системи и централизирано съхраняване чрез използване на междинни възли (които събират информация от няколко източника и я предават на един централен сървър). Последният модел е най-широко използван, поради прогресиращото разрастване на системите и мрежите.

Съществуват множество ***системи, обработващи логове***: *syslog-ng* [10] за *UNIX/Linux* с безплатна и платена версия (*syslog-ng Premium Edition*), работещи в режими – клиент, междинен възел или сървър; *syslog-ng Windows Agent* [10] за *Windows* за режим сървър; *Rsyslog* [11] с отворен код за *UNIX/Linux*, *BSD* и *Solaris*; *WinSyslog* [9] за *Windows*, работещо в режим сървър; *Kiwi Syslog Server* [5] за *Windows* с безплатна и платена версия, работещо в режим сървър; уеб базирано безплатно приложение *LogAnalyzer* [2], използвано с *Rsyslog* за преглед и анализ на логове; уеб базирано приложение *LogZilla* [6] за преглед и анализ на съобщения, работещо с *syslog-ng* и др.

Най-основните недостатъци на съществуващите системи са: работа само с определени операционни системи, ограничени възможности за графичен анализ, липса на възможност за анализ на данните, неудобно и сложно задаване на правилата за филтриране на съобщенията и неефективна класификация на съобщенията.

Използването на ***облачни услуги*** (cloud computing) [3, 4, 7, 8] е начин на организация и работа на компютърните системи, при който компютърни ресурси като процесорно време, оперативна памет и дисково място могат да бъдат динамично променяни чрез набор от предлагани на потребителя уеб услуги.

Концепцията *cloud computing* промени начина, по който бизнесът оперира, съхранява и защитава своите информационни ресурси. Все повече приложения преминават към използване на cloud базирани решения поради тяхната огромна гъвкавост по отношение на оперативни разходи. Основните характеристики на cloud базираните решения са:

- динамично увеличаване и намаляване на използваните ресурси в зависимост от текущите нужди на приложението;
- заплащане само за използваните ресурси;
- липса на предварителна инвестиция в хардуер.

CloudLog е софтуерна система, чиято цел е да разреши основните проблеми при работата с логове от множество източници. Приложението дава възможност на системни администратори и разработчици да следят какво се случва с дадена ИТ инфраструктура в почти реално време. Системата предоставя набор от уеб услуги за запис на логови данни и уеб интерфейс за преглед и анализ на събраната информация. *CloudLog* използва cloud уеб

услугите, предоставени от *Amazon.com (Amazon Web Services* (*AWS*)). Данни се съхраняват в SimpleDB, а кодът на приложението може да се изпълнява на 2 или повече EC2 инстанции, разположени зад Elastic Load Balancer, с цел постигане на по-голяма надеждност и хоризонтална мащабируемост. Цялата комуникация от и към уеб приложението е изцяло криптирана. Процесът на анализ на събраната информация е направен интуитивен и приятен чрез използването на графики и интерактивни таблици.

CloudLog е изградена като разширение на системата *S-Syslog* [1], като използва напълно функционалността на нейната част *S-Syslog Web*.

2. Система *S-Syslog*

S-Syslog [1] е система за събиране, класификация, централизирано съхраняване и анализ на събития, регистрирани чрез syslog протокола от устройства като персонални компютри, принтери, рутери и други. Системата реализира модела за централизирано съхраняване на логове чрез използване на междинни възли. Тя осигурява едновременно максимална защита на събраната информация и пълен достъп до нея. *S-Syslog* улеснява процеса на анализ чрез използването на графики, които могат да се променят от потребителя при необходимост.

Системата е съставена от два компонента – *S-Syslog Collector* и *S-Syslog Web*.

S-Syslog Collector е *JAVA* приложение, което работи като syslog сървър и изпълнява ролята на междинен възел. То приема съобщения, класифицира ги и изпраща обработените данни към *S-Syslog Web*. Там те се съхраняват в база данни, готови за анализ. Комуникацията между двете приложения се изгражда чрез *RESTful API*.

При стартирането си *S-Syslog Collector* се свързва със *S-Syslog Web* и използвайки API функции прави заявка за актуалните правила за класификация. След като ги получи се създава тяхно локално копие, което се използва при обработката на данните. При получаването на съобщение то се класифицира и крайният резултат се изпраща към *S-Syslog Web*. *S-Syslog Web* дава възможност на системните и мрежовите администратори да следят какво се случва с цялата ИТ инфраструктура на организацията в почти реално време от всяко място и по всяко време.

S-Syslog Collector приема syslog съобщения по 3 различни канала: стандартната за syslog протокола *UDP* връзка, надеждната *TCP* връзка и *TCP* връзка, защитена с *SSL*.

За класификация на съобщенията се използва дървовидна йерархия от регулярни изрази, т. е. правилата за класификация се представят в дърво, чиито възли съдържат регулярни изрази. С цел постигане на по-голяма

ефективност възлите с общ родител, които са на едно ниво в йерархията, получават различни приоритети. След натрупване на достатъчно данни приложението може автоматично да определи приоритета на отделните правила въз основа на броя съобщения, които те са класифицирали.

Комуникацията между всички участници и *S-Syslog Web* протича изцяло чрез *HTTPS* протокола. Това прави невъзможно пасивното подслушване на трафика и осигурява необходимото ниво на сигурност за системата.

S-Syslog Web приложението е изградено в обекто-ориентиран стил на програмиране от няколко модула за: работа със съобщения, графичен анализ, управление на потребителите, системни статистики и работа с правила за класификация. Приложението използва предварително проектиран и реализиран модел за многоезична поддръжка с възможност за лесно добавяне на нови езици.

Модулът за работа с правила за класификация има следните дейности: дървовидно представяне на наличните правила; добавяне на ново правило; редактиране на съществуващо правило; изтриване на съществуващо правило; преместване в дървото на съществуващо правило и автоматично определяне на приоритетите на правилата на дадено ниво въз основа на натрупаните до момента данни.

Останалите модули за работа със съобщения, графичен анализ, управление на потребителите и системни статистики са детайлно описани в т. 5, т. к. те се използват само с малки подобрения в представяната cloud базирана система.

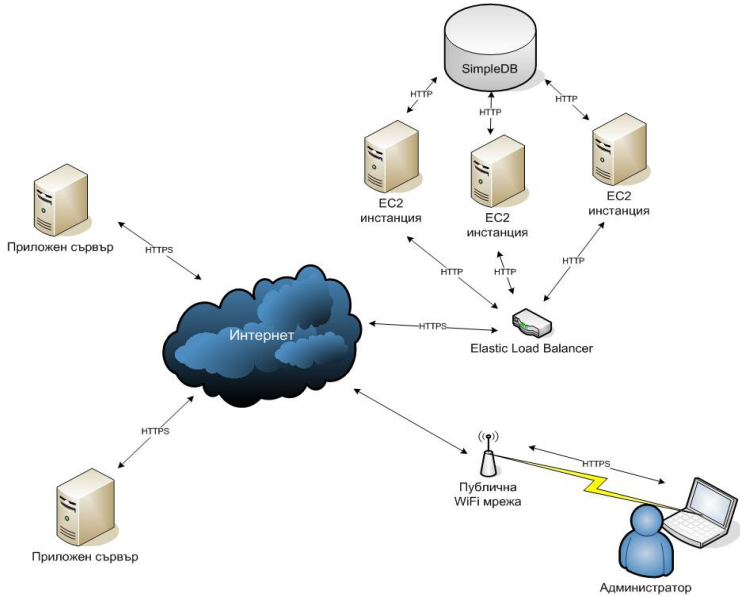
3. Обща схема на *CloudLog*

CloudLog е cloud базирана система за събиране, централизирано съхраняване и анализ на логове. На Фигура 1 е представена схема на работата на *CloudLog*.

Приложение, работещо на даден приложен сървър изпраща логови данни към *CloudLog*, извършвайки *HTTPS POST* заявка към адрес, принадлежащ на Elastic Load Balancer (ELB). ELB автоматично разпределя входящия трафик към няколко различни Amazon EC2 инстанции. Така се постига по-голяма надеждност и мащабируемост на услугата, тъй като за нея отговарят няколко независими една от друга машини. Във всеки момент от време могат да бъдат добавени нови EC2 инстанции, ако текущото натоварване на системата го изисква. ELB следи състоянието на всяка отделна инстанция, свързана с него и при проблем с някоя от тях, трафикът се пренасочва към друга, докато той не бъде отстранен.

ELB поддържа прекъсване на *SSL* комуникацията на ниво load balancer. По този начин се осигурява централизирано управление на *SSL* сертификатите и

изнасяне на задачата по декриптиране на трафика от крайните сървъри към load balancer-а. Комуникацията между ELB и EC2 инстанциите се осъществява чрез *HTTP* протокола.



Фигура 1. Обща схема на работата на CloudLog

CloudLog дава възможност на системните и мрежовите администратори да следят какво се случва с цялата ИТ инфраструктура на организацията в почти реално време. Те могат да използват приложението от всяко място и по всяко време. Единственото, което им е необходимо е Интернет връзка и компютър с инсталиран уеб браузър.

CloudLog предоставя възможност за запис на нови съобщения чрез уеб услуги (web services). Това прави лесна интеграцията ѝ с вече съществуващи софтуерни и хардуерни системи.

Уеб услугите са ключово средство при интеграцията на приложения, написани на различни езици и работещи под различни платформи. Съществуват два основни подхода за реализирането им: *SOAP* и *REST*. Въпреки голямата популярност на *SOAP* протокола в миналото, днес се наблюдава постепенното му изместване от *REST* при проектирането и изграждането на уеб услуги. *REST* е избран като протокол за уеб услугите, предлагани от *CloudLog*. Основната причина за това е леснотата на създаване и използване на *REST* базирани услуги. Всеки разработчик познаващ *HTTP* и

XML може лесно и бързо да създаде *REST* базирана уеб услуга без да има нужда от каквито и да е допълнителни средства и инструменти.

За идентификация на API потребителите се използва *HTTP Basic* методът. Той е проектиран, за да даде възможност на браузъра или друга клиентска програма да изпрати потребителско име и парола към уеб сървъра заедно със заявка за конкретен ресурс. В *CloudLog* тези данни са напълно защитени, тъй като се използва *HTTPS* и съдържанието на цялата комуникация е криптирано.

4. Данни на *CloudLog*

Данните, с които работи системата се съхраняват в cloud базираната база данни *SimpleDB*. Тази база данни е избрана, защото не е необходимо да се закупува, инсталира и поддържа никакъв хардуер, нито пък да се конфигурира някакъв софтуер. Освен това времето за дизайн на базата данни е изключително малко, тъй като промените в структурата са лесни за изпълнение и не пречат на нормалната работа на приложенията, които я използват. *SimpleDB* автоматично осигурява достатъчно дисково пространство и процесорно време, като разпределя натоварването между множество дискове и отделни сървъри, когато това е необходимо. Тя е изключително надеждна, тъй като данните се съхраняват на няколко места едновременно. Това изключва нуждата от наблюдение, създаване на резервни копия, намеса от страна на системен администратор и осигурява необходимото ниво на надеждност за системата *CloudLog*. Достъпът до *SimpleDB* е реализиран чрез *SOAP* базирани уеб услуги.

SimpleDB се различава от традиционните релационни бази данни по това, че при нея липсва концепцията за таблица. Вместо в таблици данните се групират в домейни (domains). Всеки домейн съществува в рамките на конкретен *AWS* акаунт и има уникално име.

SimpleDB се използва за съхраняване на частично структурирани данни като отделните записи са подобни, но не е задължително да са еднакви по структура. Всички данни, съхранени в *SimpleDB*, се индексират автоматично.

Всеки запис в един *SimpleDB* домейн има атрибути и име, което трябва да е уникално в рамките на домейна. Атрибутите са двойки от вида име-стойност. Всички стойности на атрибути се третират като низове.

Добавянето, изтриването и промяната на записи става чрез *SimpleDB API*. За извършване на заявки се използва *SQL* подобен език. В *SimpleDB* не се поддържат съединения (joins) на данни от различни домейни. Вместо това данните се съхраняват в денормализиран вид с цел по-бързия достъп до тях.

За реализирането на системата са създадени 4 домейна, в които се пазят данните за потребителите на приложението, API потребителите, потребителските сесии на уеб интерфейса и регистрираните съобщения.

Паролите на потребителите на приложението и API потребителите се съхраняват като *SHA-256* хеш с цел осигуряване на сигурност.

5. Възможности на *CloudLog*

Приложението е изградено в обектно ориентиран стил на програмиране. Използвана е *MVC* (Model-View-Controller) архитектура, която позволява разделението на бизнес логиката от графичния интерфейс и данните.

Системата използва предварително проектиран и реализиран модел за многоезична поддръжка. Този подход прави изключително лесно добавянето на нови езици.

Приложението съдържа няколко модула за: работа със съобщения, графичен анализ, системни статистики, управление на потребителите, автоматично известяване и импорт на данни.

Модулът за работа със съобщения предлага следните възможности:

- Списък с всички регистрирани събития с възможност за сортиране по показаните характеристики;
- Филтриране на съобщенията по дата, източник, хост и съдържание на съобщението;
- Детайлен преглед на съобщение.

Модулът за графичен анализ дава възможност за:

- Задаване на филтри за съобщенията, участващи в анализа – дата, източник, хост и съдържание на съобщението;
- Графики на съобщенията, групирани по час, ден от седмицата, седмица, месец или година.

Модулът за автоматично известяване предоставя на потребителите възможност да дефинират правила за изпращане на e-mail съобщения. Когато системата регистрира логово съобщение, отговарящо на критериите на дадено правило, изпраща e-mail на посочен от потребителя адрес. По този начин се намалява времето за реакция при регистриране на критични събития. Получателите, съдържанието и темата на известието се задават като част от дефиницията на правилата за известяване.

Модулът за импорт на данни дава възможност за:

- Импорт на данни чрез директна връзка с *MySQL* сървър. Поддържа се syslog-ng формат на базата данни;
- Импорт на данни от *CSV* файлове. Потребителят задава връзка между данните от *CSV* файла и поддържаната в *CloudLog* информация.

Чрез модула потребителят създава заявка за извършване на импорт. Самият процес на импорт може да бъде много времеемък и затова се извършва от регулярно изпълняващи се скриптове извън уеб приложението.

Модулът за системни статистики съдържа:

- Списък с IP адрес, дата и час на последните пет влизания на потребители на уеб интерфейса;
- Списък с IP адрес, дата и час на последните пет влизания на API потребители;
- Брой регистрирани съобщения за последните 24 часа;
- Информация за използваните *SimpleDB* ресурси.

Модулът за управление на потребителите предлага следните функционалности:

- Показване на списък с всички потребители;
- Добавяне на нов потребител;
- Изтриване на потребител;
- Редактиране на данните на потребител;
- Блокиране на достъпа на потребител.

CloudLog обслужва три групи от потребители, които имат различни права за достъп:

- API потребители;
- Администратори;
- Анализатори.

Правата на ***API потребителите*** са ограничени само до извикването на API функции. За този тип потребители приложението предлага отделна входна точка.

Потребителите от другите групи нямат права за достъп до API функциите. След вход в приложението в зависимост от потребителската си група, те имат достъп до различни негови части.

Анализаторите могат да извършват действия, свързани с анализа на събраните данни. Те имат достъп до модула за съобщения и модула за графичен анализ. След вход в системата те се препращат към списък с регистрирани съобщения.

Администраторите имат пълен достъп до системата. Освен модулите за съобщения и графичен анализ, те могат да използват още модул за управление на потребителите и модул за системни статистики. След вход в системата те се пренасочват към модула за системни статистики.

Заключение

Представената софтуерна система *CloudLog* разрешава основните проблеми при работата с логове от множество източници. *CloudLog* е cloud базирана система за събиране, централизирано съхраняване и анализ на логове, използваща cloud уеб услугите, предоставени от *Amazon.com*.

Благодарности

Разработката е частично финансирана по проект ДО 02-308 към Националния научен фонд.

Литература

1. Попов С., Е. Сомова, Система за събиране, класифициране, съхраняване и анализ на syslog съобщения, IV Национална научна конференция 2011 за студенти, докторанти и млади учени, Пловдив, 30 април 2011 г.
2. Adiscon LogAnalyzer – syslog web viewer, analysis and reporting tool, http://loganalyzer.adiscon.com/.
3. Barr J., Host Your Website In The Cloud, SitePoint, 2010.
4. Habeeb M., A Developer's Guide To AmazonSimpleDB, Addison-Wesley, 2011.
5. Kiwi Syslog, http://www.kiwisyslog.com/kiwi-syslog-server-overview/.
6. Logzilla is Syslog | Proactive Network Management Using Syslog – LogZilla, http://www.logzilla.pro/.
7. Masse M., REST API Design Rulebook, O'Reilly Media, 2011.
8. Powers D., PHP Object-Oriented Solutions, Friends of ED, 2008.
9. Syslog for Windows – The Enhanced Syslog Daemon for the Windows Platform, http://www.winsyslog.com/.
10. syslog-ng – Multiplatform Syslog Server and Logging Daemon, http://www.balabit.com/network-security/syslog-ng.
11. The enhanced syslogd for Linux and Unix rsyslog, http://www.rsyslog.com/.

CLOUD-BASED SYSTEM FOR COLLECTING, STORING AND ANALYSIS OF LOG MESSAGES

Stoyan Popov, Elena Somova

***Abstract**: The work presents a software system CloudLog, whose goal is to solve main problems in working with many log sources. CloudLog is a cloud-based system for collecting, centralized storing and analysis of log messages. CloudLog uses cloud web services provided by Amazon.com (AWS). Data is stored in cloud-based database SimpleDB. The code of the application can be executed on 2 or more EC2 instances placed behind an Elastic Load Balancer in order to achieve high availability and horizontal scalability. The whole communication coming in and out the web application is completely encrypted.*