

УДК 519.688

МЕТОДИКА ОПТИМИЗАЦИИ РАБОТЫ СЛУЖБЫ КАТАЛОГОВ

А. В. Андреев^{а, 1}, системный администратор, аспирант

^аСанкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, РФ

Постановка проблемы: при обращении к службе каталогов множества сервисов (прокси-сервер, файловый сервер и т. д.) сервер каталога должен обрабатывать тысячи записей. При этом наблюдается низкая масштабируемость и, как следствие, увеличение времени доступа к данным. Целью работы является уменьшение аппаратных и временных затрат при чтении, записи, удалении данных каталога. **Методы:** проведены сравнительные анализы аппаратных и временных затрат при стандартных настройках каталога и при работе с СУБД PostgreSQL, методов индексирования в СУБД PostgreSQL, аппаратных и временных затрат при неиндексированных и индексированных данных в СУБД PostgreSQL. **Результаты:** определены сервисы, работающие со службой каталогов, и их операции (добавление пользователя, получение почты и т. д.). Получены числовые данные аппаратных и временных затрат для каждой конфигурации каталога при разных типах запросов. Определено подмножество атрибутов сервисов из общего числа атрибутов каталога. Определен метод индексирования данных в СУБД PostgreSQL с учетом типов данных каталога. Разработана и описана последовательность шагов оптимизации службы каталогов при работе в высоконагруженных вычислительных сетях с заданным набором сервисов. **Практическая значимость:** данный метод позволил добиться значительного уменьшения аппаратных и временных затрат для стандартных операций службы каталогов. Результаты исследования использованы при разработке серверных решений LinuxWizard.

Ключевые слова — служба каталогов, базы данных, методика, индексирование, анализ.

Введение

Служба каталогов (Directory Service) — средство иерархического представления различных ресурсов и хранения информации об этих ресурсах. В качестве ресурсов выступают материальные ресурсы, персонал, сетевые ресурсы и т. д.

Самой простой службой каталогов является служба имен (a naming service), отображающая имена сетевых ресурсов и соответствующие им сетевые адреса. При организации такой службы каталогов пользователь не должен помнить физический адрес сетевого ресурса — служба имен будет определять местоположение ресурса. Каждый ресурс при этом будет считаться объектом в службе каталогов.

Информация об определенном объекте (ресурсе) хранится как значения атрибутов этого объекта.

Служба каталогов в контексте компьютерных сетей — программный комплекс, позволяющий администратору работать с упорядоченным рядом признаков массивом информации о сетевых ресурсах (общие папки, серверы печати, принтеры, пользователи и т. д.), хранящимся в едином месте, что обеспечивает централизованное управление как самими ресурсами, так и информацией о них, а также позволяющий контролировать использование их третьими лицами [1].

В течение 1980-х гг. международной организацией International Telegraph and Telephone

Consultative Committee разрабатывался общий стандарт для службы каталогов, в дальнейшем названный X.500, частью которого является протокол доступа к каталогу DAP (Directory Access Protocol), используемый в современных службах каталогов в облегченном варианте LDAP (Lightweight Directory Access Protocol) по причине первоначальной всеобъемлющей функциональности, оказавшейся неуместной для персональных компьютеров.

OpenLDAP — открытая реализация протокола LDAP, разработанная проектом OpenLDAP Project. Распространяется под собственной лицензией, называемой OpenLDAP Public License. Для хранения записей OpenLDAP использует базу данных Berkeley DB, но доступны различные модули для хранения данных в других базах данных [2].

Berkeley DB (BDB) — база данных, реализованная в виде библиотеки. BDB является нереляционной базой данных — она хранит пары ключ/значение как массивы байтов и поддерживает множество значений для одного ключа [3].

Когда появляется необходимость обращения к LDAP-серверу с помощью различных приложений, LDAP-сервер должен обрабатывать тысячи записей LDAP. При этом наблюдается низкая масштабируемость и, как следствие, увеличение времени обращения к данным.

PostgreSQL — свободная объектно-реляционная система управления базами данных (СУБД). PostgreSQL базируется на языке SQL и поддерживает многие из возможностей стандарта SQL:2003 (ISO/IEC 9075) [4].

Поскольку в службе каталогов OpenLDAP доступны модули sql, можно реализовать хранение

¹ Научный руководитель — кандидат технических наук, доцент кафедры вычислительных систем и сетей Санкт-Петербургского государственного университета аэрокосмического приборостроения В. В. Балберин.

данных в СУБД PostgreSQL и проанализировать масштабируемость и изменения времени обращения к данным [5].

Для анализа текущей задачи можно выполнить ряд шагов:

1) провести анализ временных и аппаратных затрат при работе службы каталогов с BDB;

2) определить минимальный необходимый и достаточный набор индексов для работы заданного набора сервисов;

3) реализовать совместную работу службы каталогов OpenLDAP и PostgreSQL;

4) провести сравнительный анализ временных и аппаратных затрат при работе служб каталогов с СУБД postgresql и bdb;

5) проиндексировать таблицы postgresql для различного набора сервисов. Провести сравнитель-

ный анализ временных и аппаратных затрат при работе службы каталогов с заданным набором индексов и исходным вариантом.

Средства и результаты анализа

Для проведения сравнительного анализа достаточно стандартных средств операционной системы GNU/Linux, таких как syslog, time, top.

Для индексирования таблиц СУБД PostgreSQL предпочтителен только один метод btree как единственный, удовлетворяющий всем типам данных, используемых в службе каталогов [6].

Результаты анализа при работе службы каталогов при стандартных настройках и после индексирования таблиц СУБД PostgreSQL представлены в табл. 1 и 2.

■ Таблица 1. Временные затраты, с

Сервис	Аутентификация		Добавление пользователя		Удаление пользователя		Отправка почты		Прием почты	
	I	II	I	II	I	II	I	II	I	II
Прокси-сервер	0,230	0,08	–	–	–	–	–	–	–	–
Файл-сервер	0,285	0,132	0,440	0,230	0,41	0,15	–	–	–	–
Почтовый сервер 1	0,235	0,084	–	–	–	–	0,228	0,830	0,330	0,125
Сервер аутентификации	0,282	0,097	–	–	–	–	–	–	–	–
Почтовый сервер 2	0,280	0,101	–	–	–	–	0,290	0,110	0,330	0,123

I — стандартные настройки; II — после индексирования.

■ Таблица 2. Аппаратные затраты, %

Сервис	Аутентификация		Добавление пользователя		Удаление пользователя		Отправка почты		Прием почты	
	I	II	I	II	I	II	I	II	I	II
Прокси-сервер	ЦПУ-70 ФС-70	ЦПУ-33 ФС-14	–	–	–	–	–	–	–	–
Файл-сервер	ЦПУ-84 ФС-92	ЦПУ-54 ФС-52	ЦПУ-86 ФС-100	ЦПУ-59 ФС-69	ЦПУ-79 ФС-100	ЦПУ-59 ФС-62	–	–	–	–
Почтовый сервер 1	ЦПУ-73 ФС-69	ЦПУ-40 ФС-22	–	–	–	–	ЦПУ-73 ФС-72	ЦПУ-47 ФС-37	ЦПУ-73 ФС-75	ЦПУ-43 ФС-36
Сервер аутентификации	ЦПУ-71 ФС-71	ЦПУ-35 ФС-16	–	–	–	–	–	–	–	–
Почтовый сервер 2	ЦПУ-75 ФС-70	ЦПУ-43 ФС-20	–	–	–	–	ЦПУ-77 ФС-75	ЦПУ-50 ФС-33	ЦПУ-74 ФС-78	ЦПУ-42 ФС-34

I — стандартные настройки; II — после индексирования.

ЦПУ — загруженность центрального процессорного устройства.

ФС — загруженность файловой системы.

Последовательность шагов для оптимизации службы каталогов

Для решения такой задачи была использована следующая последовательность действий.

1. Определение динамического набора сервисов и запросов к службе каталогов [7]; например, для таких сервисов, как прокси, файловый сервер, почтовый сервер, сервер аутентификации, будут актуальны следующие запросы:

- аутентификация;
- добавление записи;
- удаление записи;
- отправка почты;
- прием почты.

2. Настройка сервера с соответствующими настройками для работы с СУБД PostgreSQL [8].

3. Определение исходного множества используемых атрибутов A .

4. Выделение необходимого подмножества атрибутов B_n , принадлежащих множеству A ($A \subset B_n$) для каждого сервиса.

5. Выбор уникального множества индексов C из подмножеств атрибутов для текущего набора сервисов согласно формуле

$$C = ((B_i \setminus B_{i+1}) \cup \dots \cup (B_{n-1} \setminus B_n)) \cup (B_n \setminus ((B_i \setminus B_{i+1}) \cup \dots \cup (B_{n-1} \setminus B_n))),$$

где C — уникальное множество индексов; B_i — множество атрибутов для определенного сервиса, $i = 1..n$, n — количество сервисов.

Множество индексов C состоит из объединенных подмножеств B_n всех атрибутов сервисов с удалением дублирующихся атрибутов для каждого из них.

6. Индексирование согласно уникальному множеству индексов C [9].

Заключение

В данной статье представлен способ оптимизации работы службы каталогов в высоконагруженных локальных вычислительных сетях. Все действия, описанные в статье, следует применять только на Linux-сервере с установленными пакетами `openldap` и `postgresql-server`. В таком решении данные службы каталогов хранятся в СУБД PostgreSQL, что является уже редким случаем. Индексирование позволяет добиться более масштабируемого решения по сравнению с неиндексированным и тем более стандартным случаем с BerkeleyDB. Целесообразность индексирования обуславливается типами запросов к каталогу, так как они являются однотипными (поиск записи, удаление записи и добавление записи).

Сравнительный анализ проводился по двум основным характеристикам: временные затраты и аппаратные. Результаты после процедуры индексирования представлены для обеих характеристик в таблицах.

Согласно анализу, временные затраты для индексированного решения меньше на 50 % по сравнению с решением на BerkeleyDB, а преимущество по аппаратным затратам более 30 %.

Следует отметить, что данное улучшение наблюдается только для служб каталогов с количеством записей более 600, что было определено экспериментально.

Выражение для поиска уникального множества индексов службы каталогов универсально для различных конфигураций и может быть использовано для всех типов каталогов (Active Directory, OpenLDAP, 389 DS).

Предлагаемые шаги оптимизации службы каталогов универсальны и эффективны для динамического набора сервисов.

Результаты данного исследования используются в разработках компании LinuxWizard.

Литература

1. Свободная энциклопедия — Служба каталогов. http://ru.wikipedia.org/wiki/Служба_каталогов (дата обращения: 03.09.2012).
2. Свободная энциклопедия — OpenLDAP. <http://ru.wikipedia.org/wiki/Openldap> (дата обращения: 03.09.2012).
3. Свободная энциклопедия — BerkeleyDB. http://ru.wikipedia.org/wiki/Berkeley_DB (дата обращения: 03.09.2012).
4. Свободная энциклопедия — PostgreSQL. <http://ru.wikipedia.org/wiki/PostgreSQL> (дата обращения: 03.09.2012).

5. Руководство администратора. <http://www.openldap.org/doc/admin24/> (дата обращения: 15.08.2012).
6. Douglas K., Douglas S. PostgreSQL. — Sams, 2006. — 1032 p.
7. Документация об открытой реализации протокола LDAP. <http://www.openldap.org> (дата обращения: 07.09.2012).
8. Алгоритм настройки службы каталогов для работы с СУБД PostgreSQL. http://open.linuxwizard.ru/wiki/index.php/OpenLDAP_PostgreSQL (дата обращения: 20.09.2012).
9. Smith G. PostgreSQL 9.0 High Performance. — Packt Publishing, 2010. — 468 p.

UDC 519.688

Methods of Optimizing the Directory ServiceAndreyev A. V.^a, System Administrator, Post-Graduate Student, aathexf@gmail.com^aSaint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaja St., 190000, Saint-Petersburg, Russian Federation

Purpose: When there is a need to address a LDAP-server using different services (proxy server, file server etc.) the LDAP-server should handle thousands of records. This is characterized by low scalability and, as a result, it increases time of access to data. The goal of this paper is to decrease hardware and time expenditures while reading, recording and deleting directory data. **Methods:** There have been conducted comparative analyses of hardware and time expenditures at standard settings of a directory service and during work with PostgreSQL; there have been presented methods of indexing in PostgreSQL; hardware and time expenditures at indexed and non-indexed data in PostgreSQL. **Results:** There have been defined services working with a directory service and their operations (adding a user, receiving mail etc.). There have been obtained numerical data of hardware and time expenditures for each configuration of a directory for different types of queries. There has been sorted out a subset of service attributes from the total of directory services. There has been defined a method of indexing data in PostgreSQL taking into account types of directory data. There has been described a sequence of steps for optimizing hardware and time expenditures of a directory service when working at high-load LAN with the given set of services. **Practical relevance:** The method has allowed achieving significant decrease in hardware and time expenditures for standard operations of a directory service. The research results have been implemented in developing software for Linux Wizard servers.

Keywords — Directory Service, Databases, Methodology, Indexing, Analysis.

References

1. *Svobodnaia entsiklopediia — Sluzhba katalogov* [The Free Encyclopedia — Directory Service]. Available at: http://ru.wikipedia.org/wiki/Служба_каталогов (accessed 3 August 2012).
2. *Svobodnaia entsiklopediia — OpenLDAP* [The Free Encyclopedia — OpenLDAP]. Available at: <http://ru.wikipedia.org/wiki/Openldap> (accessed 3 September 2012).
3. *Svobodnaia entsiklopediia — BerkeleyDB* [The Free Encyclopedia — BerkeleyDB]. Available at: http://ru.wikipedia.org/wiki/Berkeley_DB (accessed 3 September 2012).
4. *Svobodnaia entsiklopediia — PostgreSQL* [The Free Encyclopedia — PostgreSQL]. Available at: <http://ru.wikipedia.org/wiki/PostgreSQL> (accessed 3 September 2012).
5. *Rukovodstvo administratora* [OpenLDAP Admin Guide]. Available at: <http://www.openldap.org/doc/admin24/> (accessed 15 August 2012).
6. Douglas K., Douglas S. *PostgreSQL*. Sams, 2006. 1032 p.
7. *Dokumentacija ob otkrytoj realizacii protokola LDAP* [OpenLDAP Documentations]. Available at: <http://www.openldap.org> (accessed 7 September 2012).
8. *Algoritm nastroiки sluzhby katalogov dlia raboty s SUBD PostgreSQL* [How to setup OpenLDAP with PostgreSQL]. Available at: http://open.linuxwizard.ru/wiki/index.php/OpenLDAP_PostgreSQL (accessed 20 September 2012).
9. Smith G. *PostgreSQL 9.0 High Performance*. Packt Publishing, 2010. 468 p.

ПАМЯТКА ДЛЯ АВТОРОВ

Поступающие в редакцию статьи проходят обязательное рецензирование.

При наличии положительной рецензии статья рассматривается редакционной коллегией. Принятая в печать статья направляется автору для согласования редакторских правок. После согласования автор представляет в редакцию окончательный вариант текста статьи.

Процедуры согласования текста статьи могут осуществляться как непосредственно в редакции, так и по e-mail (ius.spb@gmail.com).

При отклонении статьи редакция представляет автору мотивированное заключение и рецензию, при необходимости доработать статью — рецензию. Рукописи не возвращаются.

Редакция журнала напоминает, что ответственность за достоверность и точность рекламных материалов несут рекламодатели.