BAKI UNİVERSİTETİNİN XƏBƏRLƏRİ

№4 ***Fizika-riyaziyyat elmləri seriyası*** ***2013***

İNFORMATİKA

УДК 681.3

АЛГОРИТМ СОЗДАНИЯ ПОЛНОГО РЕЗЕРВНОГО КОПИРОВАНИЯ В ОБЛАЧНЫХ ВЫЧИСЛЕНИЯХ

А.А.АЛИЕВ, Р.Б.САМЕДОВ
Бакинский Государственный Университет
aaliyev@mail.ru, samedov.ramin@gmail.com

В статье представлены современные технологии в системах хранения и резервного копирования. Рассмотрены традиционные методы шифрования. Дано определение Облачного резервного копирования. Предложен алгоритм полного резервного копирования в облачных вычислениях. Приведены результаты эксперимента.

Ключевые слова: облачные вычисления, резервное копирование, шифрование, архивирование, база данных

Информации, хранящейся в компьютерных системах, угрожает множество опасностей. Данные могут быть утеряны по причинам ошибок программного обеспечения, неумелой работы пользователей, сбоев физических носителей и средств связи, злонамеренной порчи данных. Абсолютной защиты от всех этих угроз не существует, риск утраты данных существует всегда [1].

Как показывает общемировая статистика [2], основными причинами потерь данных являются неисправная работа аппаратных средств (44%), человеческие ошибки (32%), в основном тех, кто имеет максимальный уровень доступа к системам хранения данных компании, 14% всех случаев потерь данных происходят вследствие ошибок программного обеспечения, другие 7% происходят из-за компьютерных вирусов, а вследствие стихийных бедствий – только 3%.

Сбои приводят к приостановлению бизнес-процессов и потере данных, тем самым ставят под вопрос существование бизнеса в целом. Пожалуй, единственный способ надежно сохранить нужную информацию - периодически создавать резервные копии [3].

Внедряя системы хранения данных и резервного копирования, любая компания сталкивается со сложными задачами оценки ее текущих потребностей, планировании будущих объемов данных, выбора технологий

и архитектур, которые должны максимально соответствовать требованиям безопасности, возможности последующего масштабирования, удовлетворять техническим требованиям скорости записи, чтения, восстановления данных и многим другим условиям. Выявить оптимальное решение очень непросто, особенно учитывая широкое многообразие существующих путей реализации систем хранения и резервного копирования, а также довольно высокую динамику изменения цен и появления новых технологий на IT рынке [2].

Основные определения

Резервная копия (англ. backup copy) – данные, хранимые на энергонезависимых носителях, обычно удаленно, предназначенные для восстановления, в случае если оригинал копии данных утерян или недоступен. Это опрделение по версии SNIA (Storage Networking Industry Association) [3].

Резервное копирование (далее будем называть бэкап) – процесс создания резервных копий.

Архивирование – это один из видов резервного копирования. Архивирование нужно не всем и не всегда. Чаще всего этот вопрос встает, когда компания начинает увеличиваться, и это касается обычно только файлов, финансовых баз, почтовых баз, т.е. тех документов, которые и неэлектронной жизни принято хранить в архивах (письма, приказы, бухгалтерская документация и т.п.).

Архивы обычно делаются раз в год и хранятся на внешних носителях где-нибудь в сейфе, в банковской ячейке. Процесс архивирование является очень простым, главное не забывать хотя бы раз в год проверять состояние архива, например, одновременно с записыванием нового восстановить часть данных из старого, следить за носителями, чтобы всегда можно было бы прочитать архив любой глубины. Например, может быть ситуация, когда потребуется архив, который был сделан 5 лет назад, и хранился на ленте, стример для которой не только был сломан и списан, но и уже давно не выпускался.

Бэкап систем — это, когда необходимо резервно копировать не отдельные файлы, а целиком всю систему, которая может состоять из нескольких компонентов, например, специального программного обеспечения, базы данных, а также файловых данных. Восстановливать системы лучше целиком, нежели по частям. Для этого сперва выбирается определенный бэкап-софт, где конечно вопрос цена, функционал, удобство играет свою не последнюю роль. Резервно копировать системы необходимо так, чтобы можно было гарантированно ее восстановить, даже в случае полной поломки сервера.

В целом, задача бэкапа — это хранить свежие резервные копии для быстрого восстановления «случайно\специально удаленного» или «сго-

ревшего», или «неправильно сконфигурированного». Если архивы обычно хранятся столько, сколько живет организация, а иногда и дольше, то для бэкапов уже можно ввести понятие глубина хранения, т.е. время, по истечении которого бэкап будет устаревать, и его можно перезаписать более свежими данными. Бэкап в целом, можно разделить на две основные части: бэкап данных и бэкап систем, и отдельно стоящий бэкап содержимого систем. Последнее очень обширная и конкретная тема, сильно зависящая от того, содержимое каких именно систем вы хотите резервно копировать, например, почтовый сервер, База данных, система CRM или настройки определенного программного обеспечения.

Репизиторий – это место, где хранятся и поддерживаются резервные копии данных.

Шаги необходимые для создания резервного копирования.

- Продумать, что именно необходимо резервно скопировать. Это значимо для универсальных серверов, когда, например, на одном сервере работает база данных, внутри корпоративный сайт и дополнительно смонтированы файловые ресурсы. В такой ситуации стоит естественно отдельно резервно копировать базу данных и отдельно необходимые файлы операционной системы.
- Программное обеспечение, выполнявшее бекап должно уметь восстанавливать бэкап на другую аппаратную архитектуру, отличающееся от текущего.
- «Системные» бэкапы, особенно систем постоянно используемых, не стоит хранить глубиной более чем неделя. Естественно в случае критической ситуации понадобится только самый свежий бэкап. А все остальные бекапы обычно необходимо бывают для подстраховки, на случай если «самый свежий» по каким-то причинам не отработал.
- Есть системы, которые сложно взаимосвязаны со всей инфраструктурой, и восстанавливать их, даже имея под рукой самый свежий «системный бэкап» конкретного сервера, бывает нелегко. Например: Active Directory, Exchange и т.д.
- Естественно необходимо выделить время, чтобы изучить документацию, и в тестовой среде, хотя бы один раз попробовать восстановить полностью систему.

Каждый бекап необходимо *шифровать*. Для шифрования обычно используется PGP. PGP (англ. *Pretty Good Privacy*) — компьютерная программа, также библиотека функций, позволяющая выполнять операции шифрования и цифровой подписи сообщений, файлов и другой информации, представленной в электронном виде, в том числе про-

зрачное шифрование данных на запоминающих устройствах, например, на жёстком диске [5].

Пользователь PGP создаёт ключевую пару: открытый и закрытый ключ. При генерации ключей задаются их владелец (имя и адрес электронной почты), тип ключа, длина ключа и срок его действия. Открытый ключ используется для шифрования и проверки цифровой подписи. Закрытый ключ - для декодирования и создания цифровой подписи.

Шифрование PGP осуществляется последовательно хешированием, сжатием данных, шифрованием с симметричным ключом, и, наконец, шифрованием с открытым ключом, причём каждый этап может осуществляться одним из нескольких поддерживаемых алгоритмов. Симметричное шифрование производится с использованием одного из семи симметричных алгоритмов на сеансовом ключе (AES, CAST5, 3DES, IDEA, Twofish, Blowfish, Camellia). Сеансовый ключ генерируется с использованием криптографического стойкого генератора псевдослучайных чисел. Сеансовый ключ зашифровывается открытым ключом получателя с использованием алгоритмов RSA или Elgamal (в зависимости от типа ключа получателя). Каждый открытый ключ соответствует имени пользователя или адресу электронной почты [5].

Метод полного резервирования системы в облачные вычисления

Вместе с быстрым ростом объемов хранимых данных возрастает сложность их защиты, используя стандартные традиционные алгоритмы резервного копирования. Каждый алгоритм резервного копирования делает компромисс между основными характеристиками процессов создания копий и операций восстановления данных. Важнейшими из них являются скорость репликации, требуемый объем памяти для хранения резервных копий, скорость восстановления [6].

Обычно каждая сделанная копия данных сохраняется в некотором хранилище данных, который физически находится рядом с источником, которого и выполняли бекап. Основные недостатки полного резевного копирования без использования облака: копирование всех файлов является медленным, а хранение полных резервных копий на каждый момент времени требует много места. Но, несмотря на это у этого алгоритма есть большой минус, то, что бекап производится не в облачные вычисления, а в обычное хранилище данных (далее будем называть сторадж). И потеря этого стораджа приведет потери всех бекапов. Для решения данной проблемы предлагается новый метод полного резервного копирования в облака, таким образом, что каждый набор бекапа записывался бы в облако. Но при этом облака это отдельная система, которая физически располагается где-то в Интернете. Естественно при этом риск конфиденциальности сто-

ит особенно остро и в наш метод, естественно, необходимо добавить шифрование данных, прежде чем отправить эти данные в облака.

Алгоритм полного резервного копирования

Шаг 1: Единицу времени обозначим t, которое принимает значение от 0 до F.

Шаг 2: Систему обозначим буквой S и в каждый момент времени t состояние системы будет постоянно меняться.

Шаг 3: Систему S будем резервировать. Для этого после бекапа она будет попадать в репозиторий. Репозиторий обозначим буквой R и в единицу времени t от 0 до F в репозиторий будут попадать бекапы.

Шаг 4: В системе шифрования PGP создадим закрытый ключ OpenPGP key pair, который создается локально и, который можно передать на другие компьютеры для расшифровки данных зашифрованных этим закрытым ключем.

Шаг 5: Полученные в шаге 3 файлы будем шифровать ключом, созданным в шаге 4. После шифрования PGP запросит ключевую фразу. Этой ключевой фразой будут зашифрованы файлы.

Шаг 6: Шифрованные файлы, полученные после шага 5 обозначим статусом 1. А не зашифрованные файлы обозначим статусом 0.

Шаг 7: Все файлы из репозитория R, имеющие статус 1, будем переносить в облака. Облака обозначим C и в единицу времени от 0 до F в облаке будет находится все зашифрованные файлы созданные в эту единицу времени t.

На рис.1 приведена схема алгоритма полного резервного копирования.

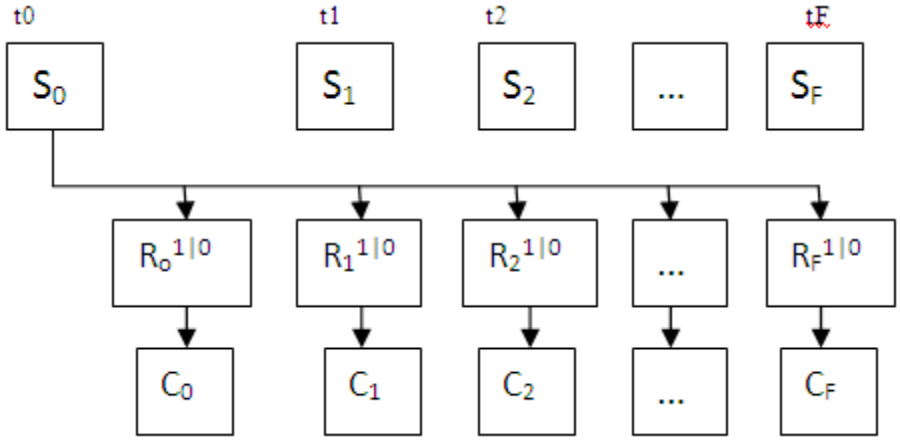

Рис.1. Схема алгоритма полного резервного копирования

Предложенный метод является более безопасным в отличие от традиционного. Так как здесь нет зависимости от единственного места хранения данных. А есть новое еще одно место облако, где риск потери данных намного меньше, чем риск потери хранилища в традиционным алго-

ритме полного резервного копирования. Также данные перед отправкой в облака шифруются, дабы избежать риска потери данных и использование их другими лицами.

Практическое применение метода

Для применения этого метода возьмем базу данных Oracle 11g, Windows 7, Kleopatra PGP и Google Drive.

Сперва создадим полную резервную копию базы данных в архивном репозитории базы данных при помощи Oracle утилиты RMAN Recovery Manager: Release 10.2.0.4.0

```
C:\Users\rsamadov>rman target /
Recovery Manager: Release 10.2.0.4.0 - Production on Thu Nov 21 12:41:54 2013
Copyright (c) 1982, 2007, Oracle.  All rights reserved.
connected to target database: TESTDB(DBID=2040826276)
RMAN> backup database plus archivelog;
Starting backup at 21-NOV-13
current log archived
using target database control file instead of recovery catalog
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:02
Finished backup at 21-NOV-13
```

Анализируем полученный бекап в архивном репозитории базы данных.

```
C:\Users\rsamadov>cd C:\Backup\2013_11_21\
C:\Backup\2013_11_21>dir
 Volume in drive C has no label.
 Volume Serial Number is 706B-F53F
 Directory of C:\Backup\2013_11_21
o1_mf_annnn_TAG20131121T124252_98vkwhg5_.bkp
o1_mf_annnn_TAG20131121T124646_98vl3qkq_.bkp
o1_mf_nnndf_TAG20131121T124407_98vkyr75_.bkp
o1_mf_annnn_TAG20131121T124252_98vkxnbc_.bkp
o1_mf_ncnnf_TAG20131121T124407_98vl3m7t_.bkp
```

Все эти файлы являются полным бекапом базы данных Oracle. Приступаем шифрованию полученных бекапов при помощи открытого ключа PGP. Для этого в программе Kleopatra выбираем соответствующие файлы, вводим ключевую фразу. Далее выбираем метод шифрации и шифруем файлы. В результате получаем информацию успешного шифрования данных (рис.2.)

o1_mf_annnn_TAG20131121T124252_98vkwhg5_.bkp → o1_mf_annnn_TAG20131121T124252_98vkwhg5_.bkp.gpg: Signing and encryption succeeded.
01_mf_annnn_TAG20131121T124646_98vl3qkq_.bkp → 01_mf_annnn_TAG20131121T124646_98vl3qkq_.bkp.gpg: Signing and encryption succeeded.
o1_mf_nnndf_TAG20131121T124407_98vkyr75_.bkp → o1_mf_nnndf_TAG20131121T124407_98vkyr75_.bkp.gpg: Signing and encryption succeeded.
o1_mf_annnn_TAG20131121T124252_98vkxnbc_.bkp → o1_mf_annnn_TAG20131121T124252_98vkxnbc_.bkp.gpg: Signing and encryption succeeded.
o1_mf_ncnnf_TAG20131121T124407_98vl3m7t_.bkp → o1_mf_ncnnf_TAG20131121T124407_98vl3m7t_.bkp.gpg: Signing and encryption succeeded.

Рис. 2. Процесс успешного шифрования файлов.

Полученные шифрованные файлы отправляем в облака для хранения их там.

Эксперимент анализа метода полного резервирования системы в облачных вычислениях

Для эксперимента возьмем базу данных Oracle объемом 12G, объемом 29G, объемом 41G и будем их резервировать стандартным алгоритмом и алгоритмом полного резервирования в облака.

Тест 1 производится с базой объемом 12G. На стандартное резервирование базы данных ушло 3 минуты 16 секунд.

На архивирование базы данных в облако вместе с шифрацией уходит дополнительное время: 42 секунды на шифрацию, 2 минуты и 15 секунд на загрузку данных в облачное хранилище.

Тест 2 производится с базой объемом 29G. На стандартное резервирование базы данных ушло 4 минуты 32 секунды.

На архивирование базы данных в облако вместе с шифрацией уходит дополнительное время: 1 минута 11 секунд на шифрацию и 3 минуты и 40 секунд на загрузку данных в облачное хранилище.

Тест 3 производится с базой объемом 41G. На стандартное резервирование базы данных ушло 6 минут 15 секунд.

На архивирование базы данных в облако вместе с шифрацией уходит дополнительное время: 2 минуты 3 секунды на шифрацию и 5 минут и 10 секунд на загрузку данных в облачное хранилище.

Таблица результата эксперимента:

Длительность (в секундах)	*Test №*	*Описание теста*
196	1	Полное резервирование
196	1_1	Полное резервирование
42	1_2	Шифрация данных
135	1_3	Загрузка данных в облако
272	2	Полное резервирование
272	2_1	Полное резервирование
71	2_2	Шифрация данных
220	2_3	Загрузка данных в облако
375	3	Полное резервирование
375	3_1	Полное резервирование
123	3_2	Шифрация данных
310	3_3	Загрузка данных в облако

Заключение

Таким образом, полное резервное копирование данных в облачное хранилище является наиболее безопасным способом резервного копирования. В работе рассмотрен алгоритм, а так же показано практическое применение алгоритма и приведены результаты эксперимента, проведенного над тремя разными базами данных.

ЛИТЕРАТУРА

1. Казаков В.Г., Федосин С.А. Технологии и алгоритмы резервного копирования / Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению “Информационно-телекоммуникационные системы”, 2008, 49 с.
2. Давлетханов М. Новое слово в корпоративном резервном копировании // Softkey.info. 2008. URL: http://www.softkey.info/reviews/ review 4797. php
3. Storage Networking Industry Association (SNIA). A Dictionary of Storage Networking Terminology. http://www.snia.org/education/dictionary/, 2008.
4. Википедия Свободная энциклопедия http://ru.wikipedia.org/wiki/PGP
5. Егоров А. Резервное копирование данных пока обойдется без инноваций // CNews. URL: http://www.cnews.ru/reviews/free/infrastructure2007/articles/reservecopying.shtml
6. Казаков В.Г. Избыточность в алгоритмах резервного копирования // Системы управления и информационные технологии. 2009, №2.2(36), с. 252-256.

HESABLAMA BULUDLARINDA TAM EHTİYAT NÜSXƏSİNİN YARADILMASI ALQORİTMİ

Ə.Ə.ƏLİYEV, R.B.SƏMƏDOV

XÜLASƏ

Məqalədə, sistemlərin ehtiyat nüsxələrin yaradılması və bekapı göstərilmişdir. Ənənəvi şifrələmə üsullarından istifadə edilməsi. “Hesablama buludlarına” tam ehtiyat nüsxəsinin yaradılması alqoritmi verilmişdir. Alqoritmin sınaq nəticələri göstərilmişdir.

Açar sözlər: Hesablama buludları, ehtiyyat nüsxə, şifrələnmə, arxivləşdirmə, verilənlər bazası

AN ALGORITHM OF A FULL BACKUP IN CLOUD COMPUTING

A.A.ALIYEV, R.B.SAMADOV

SUMMARY

This article presents the latest technology in storage systems and backup and describes the encryption methods, the definition of cloud backup, an algorithm of a full backup in cloud computing. The results of the experiment are shown in the graphic.

Key words: Cloud computing, making backup, cryptography, archiving, database

Поступила в редакцию: 25.12.2013 г.
Подписано к печати: 27.12.2013 г.