**УДК 004.75**

***Самедов Р.Б.***
Бакинский Государственный Университет, Баку, Азербайджан
ramin.samedov@gmail.com

# АЛГОРИТМ СОЗДАНИЯ РЕЗЕРВНОЙ КОПИИ РАСПРЕДЕЛЕННОЙ БАЗЫ ДАННЫХ В ОБЛАЧНЫХ ХРАНИЛИЩАХ

*В статье представлены современные технологии в системах хранения и резервного копирования. Рассмотрены традиционные методы шифрования. Показаны недостатки традиционных методов резервного копирования. Разработан алгоритм создания резервной копии распределенной базы данных в облачных хранилищах. Приведены результаты эксперимента.*

***Ключевые слова:*** *облачные вычисления, распределенные базы данных, резервное копирование, шифрование.*

## Введение

Информация, хранящаяся в ИТ системах, каждодневно находится под риском взлома и различных атак. Информация может быть утеряна по разным причинам, таким, как ошибки программного обеспечения (ПО), ошибки пользователей при работе с системой, сбои аппаратных носителей и коммуникационных средств, а также при злонамеренной атаке на информацию. Защиты от всех угроз одним решением не существует, она требует постоянного контроля и постоянной работы, при этом риск потерять ценную информацию есть всегда [1].

Как демонстрирует общемировая статистика, ведущими причинами потери данных являются поломка физических и аппаратных средств (44%), ошибки пользователей системы (32%), более критичные ошибки администраторов, имеющих наибольший уровень доступа к системам хранения данных в организациях, 14% случаев потери данных происходят вследствие сбоев программного обеспечения, 7% случаев являются следствием атак вирусов на компьютерные системы, а наименьший процент (3%) приходится на долю стихийных бедствий.

Сбои становятся причинами прекращения работы функциональности и потери ценных данных, тем самым ставя под вопрос существование и работу всей системы в целом. Единственный метод надежного сбережения ценной информации − это время от времени делать резервные копии.

В процессе интеграции системы хранения данных с резервным копированием многие организации встречаются с комплексом сложных задач по оценке их текущих потребностей и нужд, необходимых в будущем объемов данных, выбору системы, которая будет удовлетворять всем условиям безопасности, скорости и надежности, а также возможности последующего масштабирования и многих других условий. Определить оптимальное решение становится сложной задачей, особенно учитывая большое разнообразие существующих схем реализации интеграции систем хранения и резервного копирования, а также довольно большую динамику изменения цен и появления новых технологий на IT рынке.

## О резервировании данных

Резервная копия – это запасная копия существующей электронной информации, хранящаяся в резервном месте. При необходимости или сбое основной системы необходимо воспользоваться резервной копией для полного восстановления системы [2].

Резервное копирование (далее резервирование) – процесс создания резервных копий.

Архивирование является подмножеством множества видов резервного копирования. Потребность в архивировании не всегда имеется. Чаще всего необходимость в этом появляется в организации при ее росте, причем это касается не только обычных файлов, но и финансовых

баз и данных, почтовых баз, т.е. тех документов, которые и в обычной бумажной версии принято хранить в архивах (письма, приказы, бухгалтерская документация и т.п.).

Частота заполнения архивирования бывает разной. Организации могут выполнять его раз в год и хранить данные на внешних носителях где-нибудь в помещении или в банковской ячейке. Процесс выполнения архивирования является очень простым, но главное не просто выполнять архивирование, но и периодически проверять состояние архива после его резервирования. Так, к примеру, необходимо не только записывать новые данные, но также восстанавливать хотя бы их часть из старых архивов, при этом не забывая следить за носителями, чтобы всегда можно было прочитать архив, выполненный ранее. Возможно возникновение ситуации, когда появится необходимость использования архивной копии данных, выполненной более пяти лет назад, и при этом оборудование, на котором хранились данные, сломалось и уже давно не выпускается. Естественно, такой ситуации необходимо избегать, храня архивы постоянно в актуальном и рабочем состоянии [3].

Резервирование систем – это необходимость выполнения резервного копирования не отдельных файлов, а целиком всей системы, которая может состоять из нескольких компонентов, например, распределенной системы, состоящей из специального программного обеспечения, базы данных, а также файловых данных. Восстанавливать системы лучше целиком, а не по частям. Для этого сперва выбирается определенное ПО для резервирования, конечно, с учетом цены, функционала, удобства, играющих не последнюю роль. Резервное копирование системы необходимо также для того, чтобы можно было гарантированно ее восстановить даже в случае полной поломки сервера.

Основной целью резервирования является хранение последних резервных копий для быстрого восстановления в случае той или иной необходимости. Если архивы хранятся столько лет, сколько существует организация (иногда и дольше), то для резервных копий необходимо ввести понятие “глубина хранения”, т.е. время, после которого резервная копия устаревает, и данные на них необходимо перезаписать, то есть сделать свежую копию данных. Резервные копии делят на две части: резервные копии информационной системы со всей структурой и непосредственно только данные системы. Естественно, резервные копии информационной системы со всей структурой являются очень важной частью политики безопасности. К примеру, в организациях необходимо резервно копировать почтовый сервер, базы данных Oracle, системы CRM.

Репозиторий – это место, где хранятся и поддерживаются последние и свежие резервные копии данных.

**Распределенные базы данных**

Системы распределенной обработки, или распределенные системы (РС), функционирующие в компьютерных сетях, являются одной из наиболее перспективных и быстро развивающихся областей информатики. Такое место они заняли благодаря их существенным преимуществам по сравнению с изолированными системами, функционирующими на базе отдельных компьютеров. Наиболее успешным и часто применяемым на практике типом РС является распределенная база данных (РБД), представляющая собой интеграцию автономных локальных баз данных, географически распределенных и связанных посредством компьютерной сети. Все локальные БД предполагаются изначально целостными и непротиворечивыми. Узлы взаимодействуют между собой путем обмена сообщениями. Средством взаимодействия пользователя с РБД являются транзакции. Транзакцией называется последовательность операций (подтранзакций) на РБД, переводящая ее из одного непротиворечивого состояния в другое непротиворечивое состояние. Каждая подтранзакция перед началом своей работы должна захватить в каждом узле ресурс. Две транзакции вступают в конфликт тогда и только тогда, когда они работают с одним и тем же общим ресурсом и, по крайней мере, одна из

реализуемых ими операций является записью. Порядок выполнения действий двух транзакций существен только в том случае, если они конфликтуют. Выполнение каждой отдельной транзакции сохраняет целостное состояние РБД. Следовательно, несколько последовательно выполняемых транзакций также сохраняют целостное состояние РБД.

**Создание резервной копии распределенной базы данных**

Для создания резервной копии распределенной базы данных необходимо продумать, в каком состоянии требуется выполнить резервирование. Существуют два метода для создания резервной копии распределенной базы данных [4]:

первый метод – это создание резервной копии распределенной базы данных в тот момент, когда база данных потушена. Но большим минусом данного метода является то, что базу данных необходимо полностью выключить. Этот метод не подходит для систем, работающих в режиме 24x7;

второй метод – это создание резервной копии распределенной базы данных в рабочем ее состоянии. Создание резервной копии базы данных занимает немалое время и при этом сами данные постоянно изменяются. Поэтому процесс создания согласованной резервной копии базы данных усложняется. Для этих целей обычно используется утилита RMAN для баз данных Oracle.

**Проблема взлома данных, шифрования резервной копии**

При работе с конфиденциальными данными всегда есть риск потери этих данных. В случае потери критичных данных само существование организации может оказаться под вопросом. Во избежание несанкционированного доступа к данным рекомендуется шифровать данные. Также предлагается шифровать каждую резервную копию системы. Для шифрования можно использовать утилиту PGP. PGP – это ПО с библиотекой процедур и функций, при помощи которых возможно выполнять операции шифрования сообщений, файлов, а также цифровых подписей. При этом имеется возможность шифровать также любую информацию в электронном виде, в том числе выполнить прозрачное шифрование данных на разных устройствах хранения, например, на жестких дисках. Клиент PGP создает ключевую пару: открытый и закрытый ключи. При создании пары ключей указываются их владелец с указанием имени и электронного адреса почты, тип ключа, его длина и срок завершения работы ключа. При этом с помощью открытого ключа шифруются и проверяются цифровые подписи. Закрытый ключ необходим для расшифровки и создания цифровой подписи. При помощи PGP шифрования возможно использовать функции хеширования, сжатия данных, шифрования с симметричным ключом и, наконец, шифрования с открытым ключом, причем каждый этап возможно реализовать при помощи любого из нескольких поддерживаемых алгоритмов. Симметричное шифрование возможно осуществлять при помощи одного из семи симметричных алгоритмов на сеансовом ключе (AES, CAST5, 3DES, IDEA, Twofish, Blowfish, Camellia). Для генерации сеансового ключа необходимо использовать криптографические стойкие генераторы псевдослучайных чисел. Сеансовый ключ в итоге зашифровывается открытым ключом получателя с использованием алгоритмов RSA или Elgamal, в зависимости от типа ключа, который будет использован получателем. Все открытые ключи должны соответствовать имени пользователя или адресу электронной почты, указанной при создании ключа [5].

**Проблемы создания резервной копии РБД**

Каждый день можно наблюдать быстрый рост объемов хранимых данных, вместе с этим возрастает сложность их защиты. При помощи стандартных средств резервирования распределенной системы все данные хранятся в обычном хранилище данных. Обычное хранилище данных обладает риском потери данных путем сбоев физических накопительных дисков или из-за ошибок администраторов, обслуживающих систему хранения данных. При

этом в случае потери диска с резервной копией данных имеется риск восстановления этих данных со стороны злоумышленника. В организациях часто отсутствует полное копирование системы хранения данных. В случае ее утраты все данные будут утрачены.

Для решения вышеописанной проблемы разработан алгоритм, при помощи которого данные резервируются в облачное хранилище данных. В качестве облачного хранилища данных используется Google Drive. Согласно описаниям функциональности, облачное хранилище данных Google Drive обладает большим штатом профессиональных администраторов, следящих за работой системы хранения данных. Коэффициент отказоустойчивости работы всей системы облачного хранилища намного выше локальной системы хранения данных.

При использовании облачного хранилища данных увеличивается отказоустойчивость всей системы, но при этом стоит вопрос конфиденциальности информации. Для решения этого вопроса необходимо использовать шифрование данных перед отправкой резервной копии в облачное хранилище данных. Таким образом, в случае потери данных в облачном хранилище данных злоумышленники не смогут ими воспользоваться ввиду нечитабельности зашифрованных данных.

**Алгоритм создания резервной копии распределенной базы данных**

Шаг 1. Обозначим через символ *“B”* единицу измерения времени. *“B”* будет принимать значения от 0 до *“П”*.

Шаг 2. Обозначим через символ *“C”* распределенную базу данных. В каждый момент времени *“B”* состояние РБД будет постоянно меняться.

Шаг 3. Необходимо создать резервную копию *“C”*. С этой целью каждую новую резервную копию *“C”* будем отправлять в репозиторий хранения резервных копий *“C”*. Репозиторий обозначим буквой *“P”* и в единицу времени *“B”* от 0 до *“П”* в репозиторий будут попадать резервные копии РБД.

Шаг 4. При помощи библиотек шифрования создадим открытый и закрытый ключи шифрования. Закрытый ключ шифрования необходимо передать другим системам, которым необходимо расшифровать данные, зашифрованные этим закрытым ключом.

Шаг 5. Полученные в шаге 3 резервные копии *“C”* будем шифровать ключом, созданным в шаге 4. При шифрации библиотекой шифрования необходимо будет ввести ключевую фразу, которой будут зашифрованы файлы.

Шаг 6. Зашифрованные файлы, полученные после шага 5, обозначим статусом 1, а незашифрованные файлы обозначим статусом 0.

Шаг 7. Все файлы из репозитория *“P”,* имеющие статус 1, будем переносить в облака. Облака обозначим через *“O”,* и в единицу времени от 0 до *“П”* в облако будут отправлены все резервные копии со статусом 1.

На рис.1. графически отображена работа алгоритма создания резервной копии РБД в облачных технологиях.

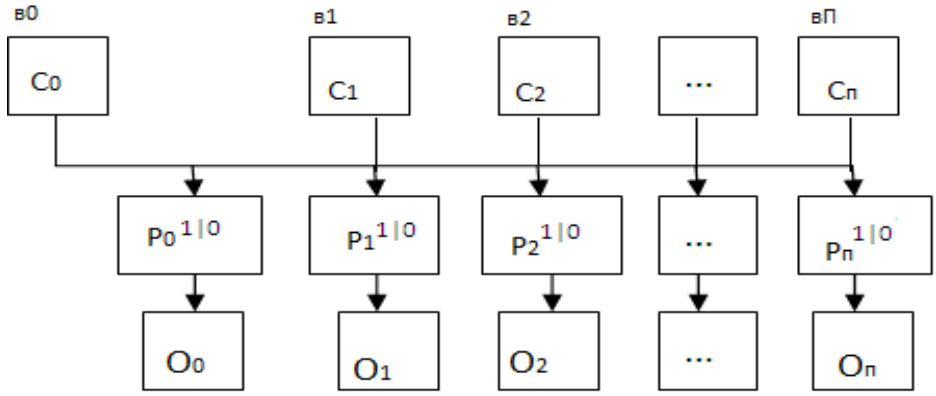


Рис.1. Схема работы алгоритма создания резервной копии РБД в облачных хранилищах

**Описание технологий и команд для проведения эксперимента работы алгоритма**

Для проведения эксперимента на основе алгоритма создания резервной копии РБД в облачных хранилищах будут использованы следующие программные средства: РБД Oracle версии 11g, операционная система Windows 7, библиотека шифрования Kleopatra PGP и облачные технологии Google Drive.

Для создания резервной копии базы данных и отправки созданной резервной копии в репозиторий необходимо использовать утилиту Oracle RMAN Recovery Manager, а именно команду:

RMAN> backup database plus archivelog.

По результатам работы вышеописанной команды будут созданы файлы, которые являются резервной копией работающей базы данных Oracle. Далее необходимо приступить к шифрованию полученных резервных копий РБД при помощи открытого ключа библиотеки шифрования. Для этого в программном обеспечении Kleopatra необходимо выбрать соответствующие файлы, ввести ключевую фразу. После выбора метода шифрования и запуска процесса файлы шифруются с использованием ключа и ключевой фразы. Полученные шифрованные файлы отправляются в облачные технологии Google Drive.

**Результаты экспериментов**

Для проведения экспериментов необходимо использовать несколько баз данных с разным объемом данных. Для проведения первого эксперимента требуется база данных с объемом данных 12G, для второго эксперимента – база данных с объемом данных 29G, для третьего эксперимента – база данных с объемом данных 41G. Для всех трех РБД двумя способами создаются резервные копии.

Первый способ – это способ создания резервной копии путем копирования резервных данных на локальные диски хранения данных, а второй способ основан на алгоритме, описанном в данной статье.

На основе проведенных трех экспериментов были получены результаты, отраженные в таблице 1.

Таблица 1

Результаты выполнения экспериментов алгоритма создания резервной копии РБД в облачных хранилищах

| Продолжительность (сек.) | Номер эксперимента | Тип операции |
|---|---|---|
| 196 | 1 | Создание резервной копии на локальных дисках |
| 196 | 1.а | Создание резервной копии в репозитории |
| 42 | 1.б | Шифрование резервной копии |
| 135 | 1.в | Отправка шифрованных данных в облако |
| 272 | 2 | Создание резервной копии на локальных дисках |
| 272 | 2.а | Создание резервной копии в репозитории |
| 71 | 2.б | Шифрование резервной копии |
| 220 | 2.в | Отправка шифрованных данных в облако |
| 375 | 3 | Создание резервной копии на локальных дисках |
| 375 | 3.а | Создание резервной копии в репозитории |
| 123 | 3.б | Шифрование резервной копии |
| 310 | 3.в | Отправка шифрованных данных в облако |

Как видно из таблицы 1, времени для создания резервной копии РБД в облачных хранилищах требуется больше. Но за счет использования облачных хранилищ

увеличивается отказоустойчивость системы хранения резервной копии базы данных. В случае сбоя или пожара всей серверной инфраструктуры не будет потери резервных копий РБД. Благодаря этому будет возможным восстановить работоспособность всей системы.

**Заключение**

Таким образом, создание резервной копии РБД в облачных хранилищах увеличивает отказоустойчивость всей системы за счет более безопасного хранения резервных копий. В работе разработан алгоритм, показана схема его работы, а также описано практическое применение алгоритма.

**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. www.ict.edu.ru/ft/005653/62330e1-st17.pdf
2. Jansen W. A. Cloud hooks: Security and privacy issues in cloud computing / Proceedings of the 44th Hawaii International Conference on System Sciences, 44th Hawaii International Conference System Sciences (HICSS), January 04–07, 2011, pp.1–10. doi>10.1109/HICSS.2011.103
3. Rahumed A. A secure cloud backup system with assured deletion and version control / Proceedings of the 40th International Conference Parallel Processing Workshops (ICPPW), 13-16 Sept. 2011, Taipei City, Taiwan, pp.160–167. DOI: 10.1109/ICPPW.2011.17
4. Li Q., Xu H. Research on the backup mechanism of oracle database / Proceedings of the International Conference Environmental Science and Information Application Technology (ESIAT), 4–5 July 2009, pp.423–426. DOI: 10,1109 / ESIAT.2009.294
5. Zimmermann P. R. The official PGP user's guide, MIT press, 1995, 21 p.

**UOT 004.75**

**Səmədov Ramin B.**
Bakı Dövlət Universiteti, Bakı, Azərbaycan
ramin.samedov@gmail.com

**Hesablama buludlarında paylanmış verilənlər bazalarının ehtiyat nüsxələrinin yaradılması alqoritmi**

Məqalədə paylanmış verilənlər bazalarının ehtiyat nüsxələrinin yaradılması göstərilmişdi. Ənənəvi şifrləmə üsullarında istifadə edilmişdi. Paylanmış verilənlər bazalarının ehtiyat nüsxələrinin hesablama buludları mühitində yaradılması alqoritmi verilmişdi. Alqoritmin sınaq nəticələri göstərilmişdi.

***Açar sözlər**: hesablama buludları, ehtiyat nüsxə, şifrlənmə, paylanmış verilənlər bazaları.*

**Ramin B. Samadov**
Baku State University, Baku, Azerbaijan
ramin.samedov@gmail.com

**The algorithm creating backup copies of database distributed in cloud storage**

The paper presents the modern technologies in storage and backup systems. Traditional methods of encryption are examined. The shortcomings of traditional backup methods are shown. A backup copy creating algorithm of a distributed database in cloud storage is developed. The results of the experiment are presented.

***Keywords**: cloud computing, distributed databases, backup, encryption.*