

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ И ОТКАЗОУСТОЙЧИВОСТИ РЕЗЕРВНОГО КОПИРОВАНИЯ ЦИФРОВЫХ ДАННЫХ

Казаков В.Г., Сидоров Д.П., Федосин С.А.

В работе исследуются показатели, характеризующие эффективность и отказоустойчивость процессов резервного копирования, представлена методика их расчета, предложен способ избыточного копирования с повышенной эффективностью, произведен анализ результатов выполненных расчетов.

Ключевые слова: резервное копирование данных, восстановление информации, эффективность, отказоустойчивость.

Введение

Регулярное создание резервных копий является надежным способом защиты данных от утраты. Резервное копирование (далее РК) предполагает создание и хранение дополнительных, избыточных копий исходных данных на случай их частичной или полной утраты для возможности восстановления. При этом, как правило, ставится требование необходимости хранения истории изменений исходных данных для того, чтобы при необходимости было возможно вернуть версию данных по состоянию на определенный момент времени. Между тем, часть резервных копий также может быть утрачена, что сделает невозможным процесс восстановления. За счет введения дополнительной избыточности при резервном копировании можно добиться повышения вероятности восстановления при утрате некоторой части резервных копий, тем самым повысив отказоустойчивость процесса резервирования. В то же время избыточное хранение резервных копий повышает их объем. Встает необходимость исследования того, насколько эффективно избыточность резервных копий улучшает отказоустойчивость и другие характеристики восстановления.

В данной работе отказоустойчивость процесса резервного копирования рассматривается в аспекте возможности выполнения процедуры восстановления при утрате части сделанных копий. Под эффективностью резервного копирования понимается эффективность использования избыточного объема хранилища для увеличения скорости восстановления.

Постановка задачи

Под избыточностью при резервном копировании понимаются любые дополнительные ко-

пии исходных данных, кроме минимально необходимых для восстановления. Так, при полном резервном копировании [3] на каждый момент резервирования копируется весь объем исходных данных. Данные, остающиеся неизменными между моментами резервирования, дублируются, тем самым создавая избыточность. При восстановлении к состоянию на некоторый момент резервирования требуется только лишь соответствующая полная копия, а утрата других сделанных копий не нарушит процесс восстановления. Для сравнения: при инкрементном резервном копировании [3] изначально создается полная копия данных, затем на каждый момент резервирования создаются копии, содержащие только изменения состояния исходных данных по отношению к предыдущему моменту резервного копирования. Данный способ, или, другими словами, алгоритм резервного копирования не предполагает хранения избыточного объема данных, поэтому при восстановлении требуется использовать все копии, содержащие накопленные изменения, а утрата любой их части приведет к невозможности восстановления. Таким образом, за счет введения избыточности можно добиться повышения вероятности восстановления при утрате некоторой части копий, то есть повысить отказоустойчивость системы. При этом чем большая избыточность создается, тем больше увеличиваются объем копий и время их создания.

Способ резервного копирования определяет, какой набор копий создается на каждый момент резервирования, что в свою очередь имеет существенное влияние на базовые характеристики процессов резервирования и восстановления: скорость создания, скорость восстановления, объем требуемого хранилища. При выборе наиболее подходящего сочетания характеристик встает необходимость оценки того, насколько эффективно избыточность, порожденная алгоритмом РК, улучшает характеристики восстановления. Требуется количественно измерить влияние избыточности, создаваемой процессом резервного копирования, на количественные характеристики скорости процесса восстановления и отказоустойчивости, выявить наиболее эффективные способы резервирования и возможные пути совершенствования.

Характеристики эффективности и отказоустойчивости

Отказоустойчивость работы алгоритмов РК характеризуется возможностью восстановления данных в случае порчи или утраты некоторой части резервных копий. Восстановимостью далее будем называть показатель вероятности восстановления в случае утраты одной копии.

Если возможность порчи созданных резервных копий считать равновероятной, то на некоторый момент времени t_j для восстановления существует N_j созданных элементарных копий, а для восстановления требуется M_j недублированных копий. Тогда при утрате одной копии вероятность того, что восстановление на момент t_j окажется возможным, составит $\frac{N_j - M_j}{N_j} \cdot 100\%$, средняя восстановимость по моментам $t_1 \dots t_k$ будет равна:

$$RPE_{scheme}(t_k) = \frac{\sum_{i=1}^k \frac{N_i - M_i}{N_i}}{k} \cdot 100\%.$$

Эффективность процесса резервного копирования связана с эффективностью использования объема хранилища. Дополнительно созданная избыточность может снижать объем резервных данных, требующихся для восстановления, и тем самым существенно улучшать скорость восстановления [4]. Для того чтобы численно оценивать влияние дополнительно вводимой избыточности, создаваемой алгоритмом, на скорость восстановления, был введен показатель средней эффективности.

Суммарный объем резервных копий некоторого алгоритма РК «*scheme*» на некоторый момент t обозначим как $BVol_{scheme}(t)$, а время восстановления – $RTime_{scheme}(t)$. Средняя эффективность AER_{scheme} показывает, насколько создаваемая алгоритмом «*scheme*» избыточность снижает время, требуемое для восстановления. Значение данного показателя равно величине $(1/(RTime_{scheme})) / BVol_{scheme}$, нормированной на значение эффективности в идеальном случае $(1/(RTime_{min})) / BVol_{min}$. Таким образом, выражение для средней эффективности имеет вид:

$$AER_{scheme} = \frac{\sum_{i=1}^k \left(\frac{RTime_{min}(t)}{RTime_{scheme}(t)} \right) \left(\frac{BVol_{scheme}(t)}{BVol_{min}(t)} \right)}{k} \cdot 100\%.$$

Так как инкрементный алгоритм не резервирует дополнительных избыточных данных, то ему соответствует минимальный показатель $BVol_{min}$. Минимальное время восстановления требуется для алгоритма полного копирования. При дифференциальном резервировании [3] для восстановления требуется полная копия и одна копия с изменениями, что также характеризуется низким временем восстановления [5].

Теоретическая основа и методика расчета

Для вычисления показателей эффективности и восстановимости был применен разработанный способ моделирования работы процессов резервного копирования и восстановления [1]. Так, в модели рассматривается система резервного копирования, которая создает копии данных в последовательные моменты времени $\{t_k\}_{k=1, 2, \dots, T}$. Состояние данных для резервного копирования на момент времени t_j обозначается D_j , начальное состояние данных D_0 предполагается пустым. Элементарные копии, содержащие изменения между состояниями данных с момента t_s до момента t_{s+n} , обозначаются R_s^n или $R(s, n)$, где s и n – целые, причем $s \geq 0, n > 0$. Множество всех элементарных копий – Rep .

Так, в процессе инкрементного копирования, в моменты времени $\{t_k\}_{k=1, 2, \dots, T}$ создаются резервные копии $R_{incr}(t_k) = \{R_{k-1}^1\}$ (см. рис. 1), при этом на момент t_T будет иметься набор элементов: $R_{incr}(t_T) = \bigcup_{i=1 \dots T} R_{incr}(t_i) = \{R_0^1; R_1^1; \dots; R_{T-1}^1\}$.

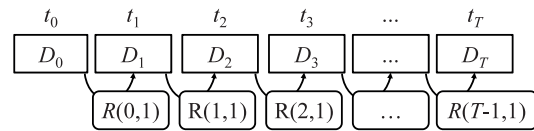


Рис. 1. Инкрементное копирование

Функция $Vol: Rep \rightarrow \mathbf{R}^+$ ставит в соответствие каждой элементарной копии положительное действительное число, обозначающее объем требуемой памяти для хранения:

$$Vol(R(t, m)) = Add(t, m) + Diff(t, m) + Serv(t, m),$$

где функция $Add(t, m)$ показывает объем всех добавленных данных между некоторыми состояниями данных D_t и D_{t+m} , $Diff(t, m)$ – объем всех измененных данных, $Del(t, m)$ – объем удаленных данных, $Serv(t, m)$ – объем служебных данных, хранимых в системе резервного копирования (см. рис. 2).

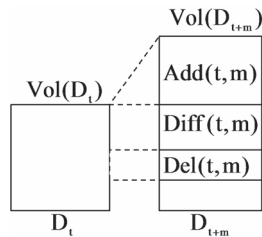


Рис. 2. Различие состояний данных

Объем хранилища резервных копий $Volume_{scheme}(t_j)$, требуемый для работы некоторого алгоритма РК («scheme») на некоторый момент времени t_j , будет равен совокупному объему всех созданных элементарных копий:

$$Volume_{scheme}(t_j) = \sum_{R \in \bar{R}_{scheme}(t_j)} Vol(R).$$

В зависимости от области применения резервного копирования вводятся выражения, определяющие характер роста, изменения состояния данных. Рассмотрим случай, когда модель характеризуется следующим рядом условий: в каждый последующий момент времени t_{k+1} по отношению к предыдущему t_k данные изменяются так, что $Diff(k, 1) = Vol(D_k) \cdot d$, и растут так, что $Add(k, 1) = Vol(D_k) \cdot p$, при этом часть данных удаляется $Del(k, 1) = Vol(D_k) \cdot m$. Объем изменений на каждый момент совпадает с изменениями предыдущего момента с уровнем D . При этом первичное состояние данных равно некоторому начальному объему V , $Vol(D_1) = V$. Параметр объема V измеряется в байтах, а параметры p , d и m – в процентах. Функции, описывающие изменение и рост данных, будут иметь вид

$$Add(s, n) = V \cdot \Psi^{s-1} \cdot (\Psi^n - (1-m)^n),$$

$$Diff(s, n) = V \cdot \Psi^{s-1} \cdot (d(1-m)^{n-1} + (1-d-m) \cdot A),$$

$$Vol(D_{s+n}) = V \cdot \Psi^{s+n-1},$$

$$\text{где } A = \left((1-m)^{n-1} - \left(1 - \frac{\Psi \cdot d \cdot (1-D)}{(1-d-m)} - m \right)^{n-1} \right)$$

и $\Psi = (1 + p - m)$.

Проверка адекватности данных, получаемых с использованием предложенной модели, была произведена экспериментально в несколько этапов с использованием специально разработанной программной системы резервного копирования на реальных данных.

Характеристики каждого способа резервного копирования зависят от характера и уровней роста, изменения, повторяемости изменений, объема данных и свойств используемых носите-

лей. Поэтому расчет единого показателя эффективности для алгоритмов РК необходимо вести в среднем по различным значениям выделенных влияющих факторов. В частности, зависимость показателя эффективности от параметров уровней роста и изменения данных для инкрементного алгоритма изображена на рис. 3. Иллюстрация разницы величин показателя эффективности в зависимости от фактора уровня повторения изменений ($D = 90\%$, $D = 10\%$) для инкрементного способа резервного копирования представлена на рис. 4.

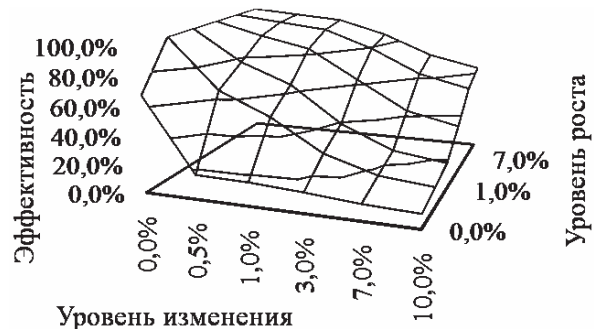


Рис. 3. Зависимость эффективности от уровней роста и изменения данных

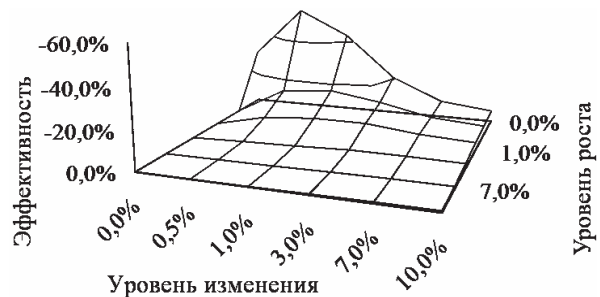


Рис. 4. Влияние совпадения изменений на показатель эффективности

Способ избыточного копирования с повышенной эффективностью

На основе описанной методики расчета предложенных показателей был исследован широкий спектр алгоритмов, включающий: полное, инкрементное, дифференциальное резервирование; мультиуровневый способ [5], алгоритм [2], «Z scheme» [5]. В результате проведенного исследования было выявлено, что существует возможность построения нового алгоритма, с регулируемым балансом базовых характеристик процесса РК, с повышенной степенью эффективности вводимой избыточности.

Был разработан алгоритм резервного копирования, обозначаемый $Braid^k$ (от англ. braiding – плетение) и в общем виде задаваемый функцией выхода следующего вида:

$$\begin{aligned}\bar{R}_{Braid^k}(t_T) &= \{\bar{R}_{Init^k}(t_1, t_T); \bar{R}_{Braid^k Element}(t_{k+1}, t_T)\}, \\ \bar{R}_{Init^k}(t_{start}, t_{end}) &= \{\{R_0^{start}\}_{end-start \geq 0}; \{R_1^{start}\}_{end-start \geq 1}; \dots; \\ &\dots; \{R_{start}^{k-1}\}_{end-start \geq k-1}\}_{end \geq start}, \\ \bar{R}_{Braid^k Element}(t_{start}, t_{end}) &= \{\{R_{start-k+i}^k\}_{i=\lfloor i/k \rfloor, k=0}; \\ &\{R_{start-k+i}^k\}_{i=\lfloor i/k \rfloor, k=1}; \dots; \{R_{start-k+i}^k\}_{i=\lfloor i/k \rfloor, k=k-1}\}_{end \geq start, i=0, \dots, end-start}.\end{aligned}$$

Параметр k называется степенью и обозначает количество уровней создаваемых копий. За счет изменения количества уровней можно добиться различного соотношения степени избыточности и скорости восстановления, при этом возможно динамическое изменение количества уровней в процессе работы системы резервного копирования.

Был разработан ряд усовершенствований, улучшающих показатели работы предложенного алгоритма. Так как наиболее приоритетным, как правило, является состояние данных последнего резервирования, то возможно временное создание дополнительных элементарных копий вида $R(k-1, 1)$ на конечный момент t_k , которое позволит повысить вероятность восстановления к данному состоянию до 100%. Создание и хранение периодических дополнительных копий вида $R(x, 1)$ служат «мостами»-переходами между уровнями алгоритма, которые позволяют значительно увеличить вероятность восстановления при выходе из строя носителей резервных копий.

Анализ результатов расчетов

Рассмотрим результаты расчетов со следующими параметрами: значения роста и изменения в экспоненциальной модели в пределах от 0,001% до 1%, совпадение изменений 10-90%, данные за 365 моментов резервирования. Средние значения показателей эффективности (AERTime) и восстановимости (RProbE), рассчитанные для алгоритмов, представлены на рис. 5.

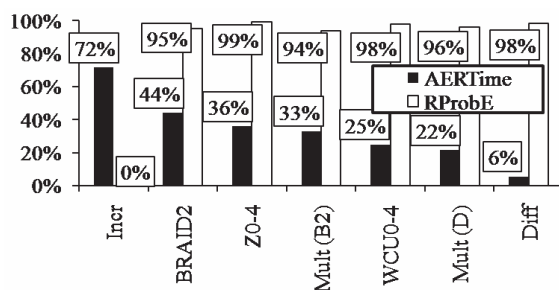


Рис. 5. Эффективность и восстановимость алгоритмов РК

Как можно видеть из представленных данных, наибольшей эффективностью обладает ин-

крементный алгоритм («Incr»), но отсутствие избыточности обуславливает нулевое значение показателя восстановимости. Вторым по показателю средней эффективности является алгоритм плетения второй степени и с создаваемыми мостами переходов каждые десять моментов РК («BRAID2»). Его эффективность на 22% превосходит эффективность «Z scheme» с четырьмя уровнями («Z0-4»). Мультиуровневый алгоритм на основе схемы плетения с двумя уровнями («Mult(B2)») показывает на 50% большую эффективность, чем мультиуровневый алгоритм («Mult(D)»).

Соотношения величин объемов, требуемых для хранения копий различными способами РК, представлены на рис. 6. Важной особенностью является то, что для работы схемы плетения требуется сравнительно малый объем памяти для хранения копий. Согласно рис. 6, для работы схемы «BRAID2» потребовался в 2 раза больший объем репозитория, чем для инкрементного подхода (оптимального в данном смысле), что, в свою очередь, является более чем в два раза меньшим требованием, чем для следующего по показателю эффективности алгоритма «Z scheme», обозначенного «Z0-4». Для «Mult(D)» потребовался в два раза больший объем, чем для схемы «Mult(B2)». Алгоритм Костелло с соавт. («WCU0-4») показал средние результаты.

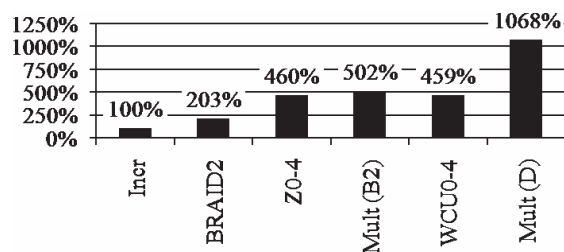


Рис. 6. Соотношение объемов копий различных алгоритмов РК

Заключение

Предложенные показатели позволяют характеризовать отказоустойчивость и эффективность процессов резервного копирования, производимого разными способами. Представленная методика расчета, основанная на разработанном способе моделирования работы процессов резервного копирования, позволяет получить значения данных показателей аналитически, с учетом различных уровней влияния факторов изменения исходных данных.

Разработан алгоритм резервирования данных с возможностью регулировки баланса базовых характеристик процессов РК. Согласно предложенной методике были произведены расчеты

показателей средней эффективности и восстановимости. Полученные результаты позволяют сделать вывод о том, что предложенный способ обладает повышенным средним уровнем эффективности по сравнению с другими алгоритмами с избыточностью.

Работа поддержана Аналитической ведомственной целевой программой «Развитие научного потенциала высшей школы 2009-10 г.г.» (№ 2.1.2/2485), а также Фондом содействия развитию малых форм предприятий в научно-технической сфере в рамках программы «УМНИК» 2008-10 г.г. (№ 08-2-8446).

Литература

1. Казаков В.Г., Федосин С.А. Метод моделирования алгоритмов резервного копирования для получения оценок объема репозитория // ИКТ. Т.6, № 3, 2008. – С. 126-132.
2. Costello A., Umans, C., Wu F. Online backup and restore / UC Berkeley, USA. – 1998. – URL: <http://www.nicemice.net/amc/research/backup/>
3. Description of Full, Incremental, and Differential Backups / Microsoft Corporation. 2006. – URL: <http://support.microsoft.com/kb/136621>
4. Kazakov V.G., Fedosin S.A. Selecting the Optimal Recovery Path in Backup Systems // Advanced Techniques in Computing Sciences and Software Engineering. S. Tarek. Springer, 2010. – P. 295-299.
5. Kurmas Z., Chervenak A. Evaluating backup algorithms // Proc. of the Eighth Goddard Conference on Mass Storage Systems and Technologies. 2000. URL: <http://www.cis.gvsu.edu/~kurmasz/papers/kurmas-MSS00.pdf>

A STUDY ON THE FAULT-TOLERANCE AND EFFICIENCY OF DIGITAL DATA BACKUP

Kazakov V. G., Fedosin S. A., Sidorov D. P.

In this paper we evaluate fault-tolerance and efficiency characteristics of backup and recovery processes, present an estimation method for proposed indicators. A new efficient backup method is described and its performance has been evaluated in compare with other algorithms.

Keywords: data backup, restore, fail-safe recovery, efficiency, fault-tolerance.

Казаков Виталий Гайясович, к.т.н., директор Центра дистанционного образования Мордовского государственного университета (МГУ) им. Н. П. Огарева. Тел. 8-917-999-28-11. E-mail: vitalykg@mail.ru

Сидоров Дмитрий Петрович, к.т.н., доцент Кафедры «Автоматизированные системы обработки информации и управления» (АСОИУ) МГУ им. Н. П. Огарева. Тел.: (8-917-996-85-91, (8-342) 47-86-91. E-mail: fedosinsa@mrsu.ru

Федосин Сергей Алексеевич, к.т.н., профессор, заведующий Кафедрой АСОИУ МГУ им. Н.П. Огарева. Тел.: 8-927-276-40-33, (8-342) 47-86-91. E-mail: fedosinsa@mrsu.ru

НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 681.518.2

ПРИМЕНЕНИЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ ДЛЯ ОПТИМИЗАЦИИ БИЗНЕС-ПРОЦЕССОВ

Димов Э.М., Луковкин С.В., Третьяков Р.В.

В статье рассматривается вопрос применения генетических алгоритмов в качестве средства повышения эффективности управления бизнес-процессами. Приведен алгоритм, позволяющий определить наиболее эффективные параметры управляющих воздействий.

Ключевые слова: бизнес-процесс, оптимизация, генетический алгоритм, целевая функция, управление, управляющее воздействие.

Введение

Повышение качества управления бизнес-процессами в сложных социально-экономических

системах уже долгое время остается актуальной задачей и имеет важное научное и практическое значение. Для уменьшения рисков при оценке эффективности управляющих воздействий применяются новые информационные технологии, в частности, имитационное моделирование. Однако зачастую даже при наличии имитационной модели бизнес-процесса определить, какие из возможных вариантов управляющих взаимодействий приведут к желаемой цели, бывает довольно сложно. В таких случаях вместо того, чтобы перебирать все возможные варианты вручную,