

Patrycja MARGOL¹
Paweł DYMORA²
Miroslaw MAZUREK³

STRATEGIE ARCHIWIZACJI I ODTWARZANIA BAZ DANYCH

Systemy bazodanowe są narażone na różnego rodzaju awarie, które mogą być spowodowane zarówno uszkodzeniem sprzętu, jak i błędem użytkownika czy systemu. Jednym ze skutecznych sposobów ochrony danych przed utratą jest tworzenie kopii zapasowych. Artykuł omawia kluczowe elementy dotyczące tworzenia strategii archiwizacji i odtwarzania danych. Ma na celu uświadomienie jak ważnym elementem pracy administratorów baz danych jest tworzenie odpowiednich strategii tworzenia kopii zapasowych i ich testowanie.

Słowa kluczowe: systemy bazodanowe, bazy danych, archiwizacja, odtwarzanie.

1. Awarie systemów bazodanowych

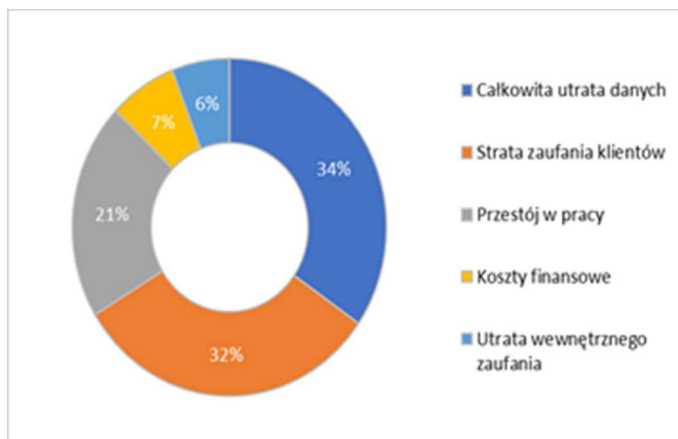
Dane, którymi firmy zarządzają są bardzo ważnym, o ile nie najważniejszym zasobem zapewniającym sprawne działanie wielu przedsiębiorstw [1]. Rys. 1 przedstawia możliwe konsekwencje utraty danych, których obawiają się osoby zarządzające danymi. Istnieje możliwość, że przypadku awarii utraczonych danych nie uda się odzyskać. Najczęstszą przyczyną utraty danych są awarie nośników na których dane są przechowywane. Nawet najdroższy i najnowocześniejszy dostępny sprzęt, nie jest w stanie zapewnić 100% bezawaryjności. Na wykresie, widocznym na rys. 2, można odczytać, że drugą z najczęstszych przyczyn jest błąd użytkownika. Błędy użytkownika skutkują chwilową oraz całkowitą utratą danych, w zależności od tego jakie strategię archiwizowania i odtwarzania danych zostały wprowadzone. Za błąd użytkownika uznaje się uszkodzenie struktury logicznej: usunięcie tabeli, zaktualizowanie niewłaściwego schematu oraz przypadkowe usunięcie plików fizycznych.

¹ Autor do korespondencji: Patrycja Margol, Politechnika Rzeszowska, adres e-mail: margol.pat@gmail.com

² Paweł Dymora, Politechnika Rzeszowska, Zakład Systemów Złożonych, pawel.dymora@prz.edu.pl

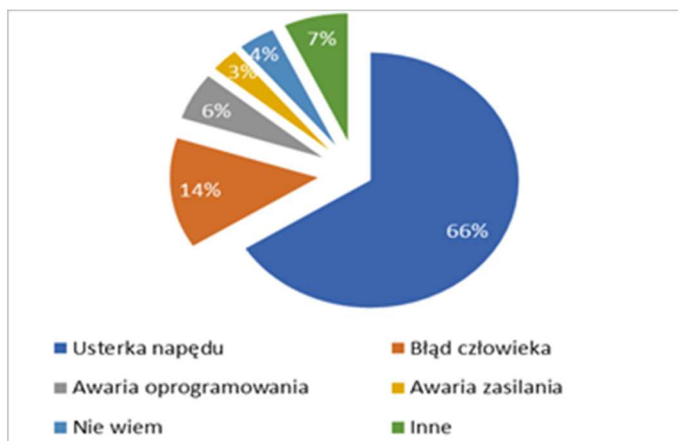
³ Miroslaw Mazurek, Politechnika Rzeszowska, Zakład Systemów Złożonych, mirosław.mazurek@prz.edu.pl

Wiele przedsiębiorstw nie decyduje się na utworzenie strategii archiwizacji danych, tworzenia kopii zapasowych, gdyż uznaje ten proces za zbyt kosztowny oraz niepotrzebny. Dotychczasowe bezawaryjne działanie systemu nie oznacza, że dane nigdy nie zostaną utracone. Uszkodzenie jednego z plików fizycznych bazy danych może spowodować stratę danych niezbędną do spójnego działania przedsiębiorstwa. W przypadku awarii napędu, w końcowym rozliczeniu może się okazać, że stworzenie, wdrożenie i przestrzeganie obranych strategii archiwizowania danych jest o wiele tańsze niż próba odtworzenia danych. Przy braku kopii zapasowych nie ma pewności czy dane uda się odzyskać oraz jaki procent danych zostanie bezpowrotnie utracony.



Rys. 1. Konsekwencje utraty danych [2]

Fig. 1. Consequences of data loss [2]



Rys. 2. Przyczyny utraty danych [3]

Fig. 2. Causes of data loss [3]

2. Strategie archiwizacji danych

Kopie zapasowe są niezbędnym elementem do odtworzenia danych, jednak nie gwarantują, że odzyskane zostaną wszystkie utracone dane. Błędnie lub nieregularnie wykonywane backupy, mogą spowodować utratę części danych lub znacznie wydłużyć czas odtwarzania bazy danych. Kopie zapasowe nie służą jedynie do odtwarzania systemu bazodanowego po wystąpieniu awarii, są również używane do migracji danych, w przypadku przenoszenia bazy danych na inny nośnik lub serwer [4].

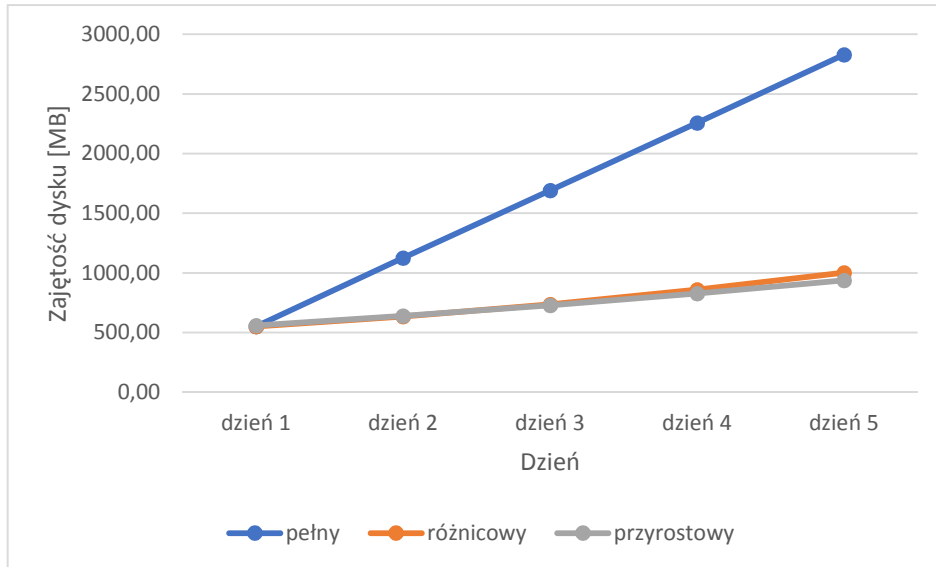
Przed stworzeniem strategii archiwizacji danych należy zastanowić się nad kilkoma bardzo ważnymi aspektami [4]:

- Czy możliwe jest wyłączenie bazy danych na czas wykonywania kopii zapasowej?
- Jaki rozmiar dodatkowej przestrzeni dyskowej można przeznaczyć na przechowywanie utworzonych kopii?
- Jak często będą wykonywane kopie i po jakim czasie zostaną usunięte?
- Jaką ilość danych firma jest skłonna utracić?
- Jak ważna jest szybkość odtworzenia danych po awarii?

W przypadku systemów bazodanowych dostępnych 24/7 backupy wykonuje się w trakcie działania bazy danych. Należy wtedy wybrać moment, w którym baza danych jest najmniej obciążona innymi operacjami. Kopie wykonywane w trakcie działania systemu bazodanowego to tzw. „gorące” kopie zapasowe. Kopie te są kopiami niespójnymi, dlatego zalecane jest wykonywanie w regularnych odstępach czasu „zimnych” kopii zapasowych. Można wykonywać kopie plików fizycznych, takich jak: pliki kontrolne, dzienniki powtórzeń, zarchiwizowane dzienniki powtórzeń, pliki danych, przestrzenie tabel oraz logicznej struktury bazy danych i pojedynczych tabel [5].

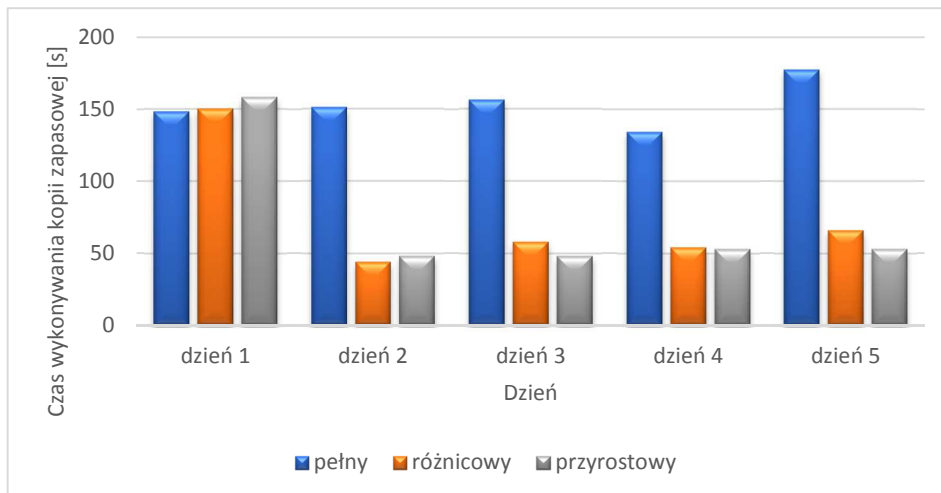
Kopie zapasowe dzieli się na kopie pełne, przyrostowe i różnicowe. Kopie pełne zawierają całą fizyczną strukturę systemu, przyrostowe zaś zawierają zmiany wykonane po utworzeniu ostatniej kopii pełnej lub przyrostowej. Natomiast w kopiach różnicowych zostają umieszczone zmiany powstałe od czasu wykonania ostatniej pełnej kopii zapasowej [4, 5]. Od rodzaju kopii zapasowej zależy jak dużą przestrzeń dyskową należy przeznaczyć na przechowywanie backupów. Rys. 3 przedstawia ilość miejsca zajmowanego przez kopie zapasowe po tygodniu codziennego, regularnego tworzenia w testowym środowisku bazodanowym. Widoczne jest, że największej przestrzeni pamięciowej wymagają pełne kopie zapasowe. Tworzone są one najdłużej. Czas tworzenia kopii zapasowych przedstawia rys. 4. Przyrostowe kopie zapasowe zajmują najmniej miejsca. Pozwala to na zastosowanie hybrydowych rozwiązań archiwizowania danych.

W testowanych przypadkach, raz w tygodniu tworzono kopię pełną, w pozostałe dni tworzono przyrostowe kopie zapasowe, a co godzinę kopie zarchiwizowanych dzienników powtórzeń. W przypadku utraty danych, takie rozwiązanie pozwala na odtworzenie ich z godziną dokładnością.



Rys. 3. Rozmiar miejsca zajmowanego przez kopie zapasowe

Fig. 3. The amount of space occupied by backups



Rys. 4. Czas tworzenia kopii zapasowych

Fig. 4. Backup time

Przyrostowe kopie zapasowe są tworzone najszybciej, jednak odtworzenie bazy danych na ich podstawie wymaga najwięcej czasu. Rozwiązaniem pozwalającym zmniejszyć rozmiar pamięci zajmowanej przez kopie zapasowe, z równoczesnym zmniejszeniem czasu odtwarzania danych, mogą być różnicowe kopie zapasowe. Łączą one zalety kopii pełnych i przyrostowych. Odtwarzanie bazy danych na ich podstawie jest nie wiele dłuższe niż w przypadku pełnej kopii zapasowej, przy czym zajmują one mniej przestrzeni dyskowej. Pozwala to na dłuższe przechowywanie starszych kopii zapasowych oraz używanie dodatkowych metod archiwizacji danych fizycznych i logicznych.

W przypadku awarii pamięci dyskowej, na której przechowywana jest baza danych, odtworzenie danych odbywa się jedynie na podstawie najnowszej, dostępnej i sprawnej kopii zapasowej. Oznacza to, że w przypadku, gdy codziennie o godz. 18 tworzone są pełne kopie zapasowe, w razie uszkodzenia dysku o godz. 16, jedyną dostępną kopią jest kopia z dnia poprzedniego. Zmiany wykonane po utworzeniu ostatniego backupu zostaną utracone. Z tego powodu dobrą opcją mogą okazać się różnicowe kopie zapasowe, połączone z codziennym tworzeniem kopii zarchiwizowanych dzienników powtórzeń. Dodatkowo kopie logicznej struktury danych pozwalają na szybkie odtworzenie danych po wystąpieniu błędu logicznego.

W ramach testów stworzone zostały trzy przykładowe strategie archiwizacji danych. Wybrany do testów systemem bazodanowym był RDBMS Oracle 11g Enterprise Edition. Archiwizowanie odbywało się za pomocą mechanizmów udostępnianych wraz z systemem bazodanowym. Pierwsza strategia opierała się na pełnych kopiach zapasowych tworzonych codziennie. Kopie te wymagały dużych zasobów pamięciowych, z tego powodu nie były stosowane żadne dodatkowe sposoby archiwizacji danych. Zaletą tej strategii było szybkie odtworzenie danych w przypadku awarii. Wadą tego rozwiązania była zajmowana pamięć oraz czas tworzenia backupu.

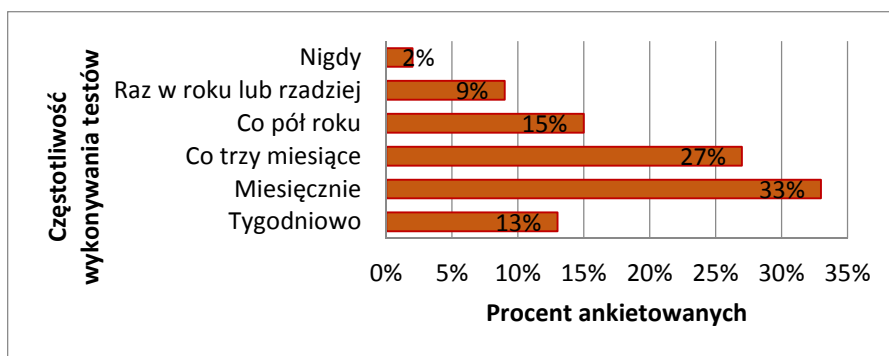
W drugiej strategii, raz w tygodniu tworzone pełną kopię zapasową, a w pozostałe dni natomiast kopie przyrostowe. Dodatkowo co 6 godzin była wykonywana kopia logicznej bazy danych, a co godzinę tworzone backupy zarchiwizowanych dzienników powtórzeń. Wadą tej strategii był czas odtwarzania danych. Dane musiały zostać odtworzone z kopii pełnej oraz wszystkich kopii przyrostowych, co znacznie wydłużyło proces odtworzenia bazy danych. Znaczącym plusem jest znacząco mniejsza ilość miejsca zajmowana przez kopie. W kopiach tych znajdowały się jedynie dane zmienione od czasu wykonania ostatniej kopii zapasowej. Również tworzone są szybko co pozwala na minimalne dodatkowe obciążenie systemu bazodanowego.

Ostatnią stworzoną strategią była strategia oparta na różnicowych kopiach zapasowych. Umożliwiała również odtworzenie logicznych danych za pomocą uruchomionych wcześniej mechanizmów. Dodatkowo co godzinę tworzone by-

ły kopie zarchiwizowanych dzienników powtórzeń. Czas wykonania różnicowej kopii zapasowej był krótszy niż w przypadku kopii pełnej. Ponadto odtwarzanie za ich pomocą było znacznie szybsze niż przy kopiach przyrostowych (zakładając awarię po 5 dniach wykonywania backupów) [6].

3. Testy strategii odtwarzania kopii zapasowych

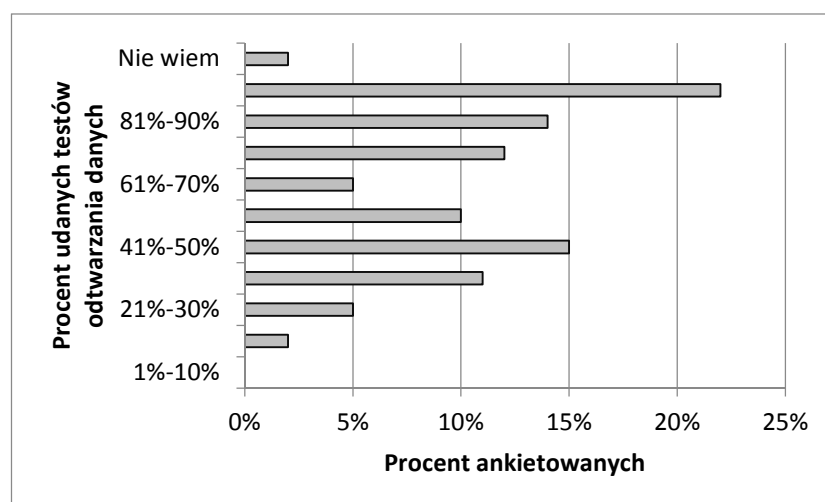
Regularne tworzenie kopii zapasowych, według wcześniej obranych strategii archiwizowania jest bardzo ważne, jednak w trakcie tworzenia backupu może pojawić się awaria prądu lub zasoby dysku mogą ulec wyczerpaniu. Uszkodzony backup łatwo jest przeoczyć w przypadku, gdy stosuje się kilka rodzajów kopii i są one tworzone z dużą częstotliwością. Odtworzenie z uszkodzonej kopii może okazać się niemożliwe. Odtworzona zostanie tylko część danych znajdujących się w kopii. Testowe odtwarzanie danych pozwala sprawdzić skuteczność tworzenia backupów oraz czy utworzone strategie spełniają założone wymagania. Rys. 5 przedstawia częstotliwość testowego odtwarzania danych przez osoby, których praca związana jest z zarządzaniem danymi.



Rys. 5. Częstotliwość testowego odtwarzania danych [7]

Fig. 5. Frequency of test data recovery [7]

Rys. 6 pokazuje procentową ilość udanych testów odtworzenia danych. 33% ankietowanych odpowiedziało, że tworzone kopie zapasowe są testowane co miesiąc. Istnieje możliwość, że kopie zapasowe wykonane po ostatnim testowym odtworzeniu danych ulegną uszkodzeniu, np. na dysku przechowującym kopie zapasowe zabraknie pamięci. W tym przypadku jedynym backupem umożliwiającym odtworzenie danych, jest ostatnia działająca kopia. Nie każdy test zakończy się wynikiem pozytywnym. Skuteczność mniejszą niż 50% wskazało około 30% ankietowanych.



Rys. 6. Liczba udanych testów odtworzenia bazy danych [7]

Fig. 6. Number of successful database recovery tests [7]

Testowanie kopii zapasowych może zająć od kilku minut do kilku godzin, w zależności od wielkości bazy danych i rodzaju wykonywanego testu. Testy powinny przeprowadzać się na specjalnie stworzonej do tego celu bazie danych. Uchroni to główną bazę danych przed przypadkowym jej uszkodzeniem.

4. Scenariusze odtwarzania kopii zapasowych

Scenariusze odtwarzania pozwalają przetestować utworzone kopie zapasowe oraz stworzyć lub dopracować strategię odtwarzania danych. Po wystąpieniu awarii, nie ma czasu na zastanawianie się jakie procedury należy podjąć, w celu naprawiania szkód. Z tego powodu tworzone są strategię odtwarzania danych. Strategię te zawierają sposób postępowania oraz dopuszczalny czas przestoju pracy bazy danych. Testowe odtwarzanie danych powinno spełniać przyjęte wcześniej normy i założenia. W przypadku komplikacji istnieje możliwość zmiany przyjętych wcześniej procedur. Przewidzenie każdego typu awarii jaki może wystąpić i zabezpieczenie się przed nim jest niemożliwe. Możliwe jest jednak stworzenie strategii postępowania w przypadku najczęściej występujących typów awarii bazy danych. Przypadki losowe bardzo często będą wariacją opracowanych wcześniej zdarzeń. Opcją ratującą może być odtworzenie całej bazy danych, nie zawsze konieczne i wymagające najdłuższego przestoju w pracy przedsiębiorstwa [4, 5].

W rozdziale, przedstawiono i omówiono kilka przykładowych scenariuszy odtwarzania, które miały na celu przetestowanie stworzonych strategii archiwizacji

danych. Scenariusze te zakładają wystąpienie najczęstszych awarii, takich jak: utrata całej bazy danych, uszkodzenie plików koniecznych do otwarcia bazy danych, uszkodzenie plików użytkownika oraz przypadkowe usunięcie jednej z tabel. Natomiast wykonane testy miały na celu porównanie stworzonych strategii archiwizacji na podstawie szybkości odtworzenia bazy danych oraz strat danych.

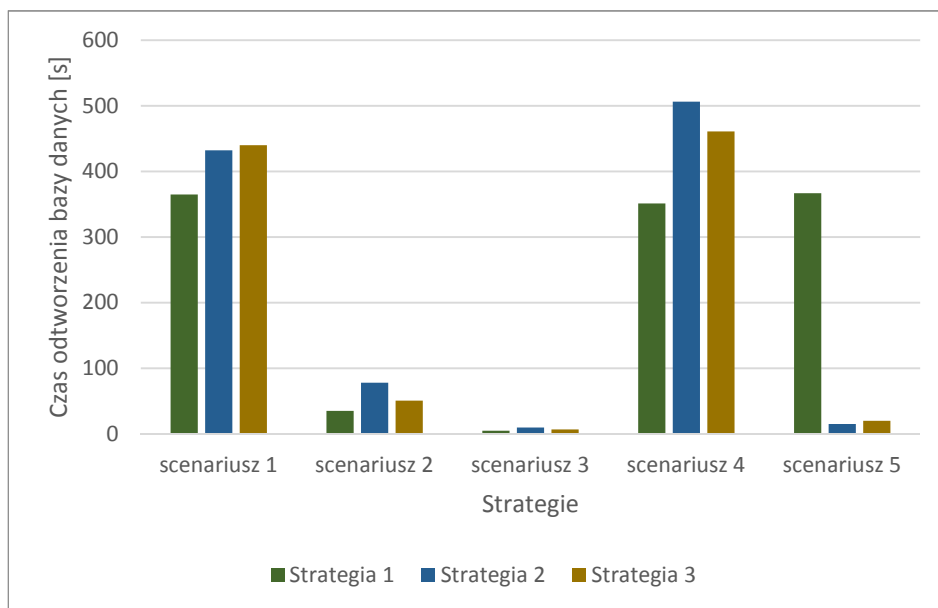
Pierwszy scenariusz przedstawia awarię całej bazy danych, usunięte zostały wszystkie pliki fizyczne. Odtwarzanie bazy danych odbyło się na podstawie kopii zapasowych, stworzonych zgodnie z założeniami wcześniej przedstawionych strategii archiwizacji danych. Najszybsze odtworzenie utraconych danych było możliwe przy użyciu pełnych kopii zapasowych (strategia pierwsza). Poza codziennym tworzeniem pełnych kopii zapasowych, strategia ta nie zakładała użycia żadnych dodatkowych form archiwizacji danych. Z tego powodu nie było możliwe odtworzenie zmian wykonanych po utworzeniu kopii zapasowej. W strategii drugiej oprócz codziennych przyrostowych kopii zapasowych były tworzone cogodzinne kopie zarchiwizowanych dzienników powtórzeń. Utracie mogły ulec jedynie zmiany wykonane w trakcie 1 godziny, jednak samo odtworzenie danych było czasochłonne, szczególnie po 5 dniach wykonywania kopii zapasowych. Trzecia strategia zakładała codzienne tworzenie różnicowych kopii zapasowych. Cogodzinne kopie zarchiwizowanych dzienników powtórzeń oraz zaimplementowane dodatkowe mechanizmy pozwalają na odtwarzanie danych logicznych. Strategia ta pozwalała na odtworzenie danych z dokładnością do 1 godziny, zaś przestój w pracy przedsiębiorstwa był krótszy niż w przypadku strategii 2.

Kolejnym sposobem na przetestowanie skuteczności stworzonych strategii był scenariusz drugi, w którym usunięte zostały z bazy danych pliki przestrzeni tabel (zostały wybrane przestrzenie tabel, bez których otworzenie bazy danych jest niemożliwe). Możliwe było odtworzenie jedynie utraconych przestrzeni tabel, bez konieczności odtwarzania całej bazy danych. Po odzyskaniu plików baza danych była odtwarzana za pomocą zarchiwizowanych dzienników powtórzeń, dzięki temu żadne dane nie zostały utracone na stałe. Scenariusz trzeci w sposobie odtwarzania danych był bardzo podobny do scenariusza drugiego. W tym przypadku usunięte były pliki danych użytkownika. Możliwe było funkcjonowanie bazy danych bez tych plików (po zmianie ich trybu na offline). Po ich odtworzeniu ponownie zostały przeniesione w tryb online. Takie postępowanie pozwoliło na odtworzenie danych bez przestoju w pracy bazy danych.

Uszkodzeniu mogą ulec również pliki dziennika powtórzeń co obrazuje scenariusz czwarty. Tutaj sposób postępowania był podobny do odtworzenia całej bazy danych. Ilość utraconych danych zależała od tego czy utracony plik, był plikiem aktualnie otwartym (nie został zarchiwizowany). W przypadku aktualnie otwartego pliku utracone zostały wszystkie zmiany w nim zawarte. Ostatni scenariusz (piąty) miał na celu przetestowanie stworzonych strategii pod

względem odtworzenia danych po wystąpieniu błędu logicznego. Usunięta została jedna z tabel, zawierająca 4000 rekordów. Strategia pierwsza nie miała zaimplementowanych żadnych metod odtwarzania danych logicznych. Z tego powodu konieczne było odtworzenie całej fizycznej struktury bazy danych. Żadne dane nie zostały utracone na stałe, jednak cały proces wymuszał przestój w pracy bazy danych na czas jej odtworzenia. W strategii drugiej, co 6 godzin, tworzone były logiczne kopie bazy danych. Na ich podstawie możliwe było odtworzenie utraconej tabeli. O ile wystąpiły jakieś straty, dotyczyły one niewielkiej ilości danych. Strategią, w której były zaimplementowane dodatkowe mechanizmy odtwarzania logicznej struktury danych, była strategia trzecia. Uruchomione mechanizmy pozwalały na odtworzenie usuniętej tabeli w ciągu kilku sekund, bez konieczności przestoju w pracy bazy danych [8].

Porównanie stworzonych strategii archiwizacji przedstawia rys. 7. Strategia pierwsza umożliwiała najszybsze odtworzenie bazy danych, w przypadku utraty całej bazy danych. Większa ilość danych była tracona niż w przypadku pozostałych dwóch strategii. Najdłużej trwało odtwarzanie danych przy stosowaniu strategii drugiej. Najlepszą strategią pod względem czasu odtwarzania, ilości zajmowanego miejsca i ilości odtworzonych danych okazała się strategia trzecia, pozwalająca na odtworzenie plików fizycznych i danych logicznych.



Rys. 7. Czas odtworzenia bazy danych

Fig. 7. Data recovery time

5. Podsumowanie

Dane są bardzo ważnym elementem każdego przedsiębiorstwa, zaś utrata ich może przynieść wiele niekorzystnych konsekwencji, takich jak utrata klientów lub w końcowym efekcie upadek przedsiębiorstwa. Bazy danych coraz częściej pracują non-stop (24/7), stąd też bardzo niekorzystny jest każdy przestój w ich pracy np. z powodu wystąpienia awarii. Ponadto istotne jest minimalizowanie strat powstałych wskutek uszkodzeń tj. utraconych danych jak i czasu niedostępności (minimalizacja czasu odtworzenia). Ważne jest również odpowiednie dobranie częstotliwości wykonywania kopii zapasowych do ilości przetwarzanych danych.

Niniejszy artykuł przybliży administratorom baz danych zagadnienie zarządzania kopiami bezpieczeństwa oraz przedstawia rezultaty testów efektywności (przeprowadzonych dla różnych scenariuszy) zaproponowanych strategii archiwizacji i odtwarzania danych, które mogą okazać się cennymi wskazówkami w ich pracy. Przed stworzeniem strategii archiwizacji danych należy dokładnie zastanowić się jakie dane zostaną skopiowane, jak je zabezpieczyć przed niepożądaną działalnością osób trzecich, jakie rodzaje kopii zapasowych będą najkorzystniejsze. Samo wykonywanie kopii zapasowych może być niewystarczające, szczególnie jeżeli będą one wykonywane w nieregularnych odstępach czasu, część danych nie zostanie skopiowana lub w wykonywanych kopiach pojawią się błędy. Po wystąpieniu awarii, nie ma czasu na poszukiwanie najlepszego sposobu odtworzenia danych. Utworzona kopia zapasowa może okazać się uszkodzona, a jedyny dostępny działający backup przestarzały, dlatego w efekcie utracone zostaną zmiany wykonane po jego stworzeniu. Testy kopii zapasowych pozwalają wykluczyć uszkodzone kopie oraz dopracować strategię archiwizacji i odtwarzania bazy danych. Scenariusze odtwarzania umożliwiają przetestowanie tworzonych kopii zapasowych. Niemożliwe jest stworzenie jednego, uniwersalnego skryptu, pozwalającego na odtworzenie bazy danych. Cała procedura zależy od tego jakie dane uległy uszkodzeniu oraz dostępności zarchiwizowanych dzienników powtórzeń.

Tworzenie i testowanie strategii może zostać uznane za czasochłonne, wymagające dodatkowych kosztów. Niemniej łatwiej i taniej jest stworzyć zabezpieczenia pozwalające na odtworzenie danych, niż podejmować próby odzyskania danych, które w żaden sposób nie były archiwizowane.

Literatura

- [1] <http://searchdatacenter.techtarget.com/feature/New-options-to-evolve-your-data-backup-and-recovery-plan>, [Dostęp dnia: 15.04.2017].
- [2] <http://www.cioandleader.com/zerto/pdf/Research-Report-The-Evolving-Business-Continuity-and-Disaster-Recovery-Landscape.pdf>, [Dostęp dnia: 03.04.2017].

- [3] <http://media.krollontrack.pl/pr/281382/uzytkownicy-wszystkich-typow-nosnikow-narazeni-na-ustrate-danych>, [Dostęp dnia 05.04.2017].
- [4] DiSario D.: *Backup Fanatic. 90-Minute Books*. Winter Haven, 2015.
- [5] Farooq T., Ault M., Portugal P., Hourri M., Hussain S. J., Czuprynski J., Harrison G.: *Oracle Database Problem Solving and Troubleshooting Handbook*. Addison-Wesley. USA, 2016.
- [6] Kuhn D.: *Oracle RMAN Database Duplication*. Apress. New York, 2015.
- [7] <http://www.cioandleader.com/zerto/pdf/Research-Report-The-Evolving-Business-Continuity-and-Disaster-Recovery-Landscape.pdf>, [Dostęp dnia: 20.04.2017].
- [8] Raghav R.: *Oracle Recovery Appliance Handbook. An Insider's Insight*. iUniverse. Bloomington, 2016.

DATABASE BACKUP AND RECOVERY STRATEGIES

Summary

Database systems are prone to different types of failure. Cause of them could be both the hard disk failure or user error. One way to protect data against loss is by making backups. This article discusses key elements in development of database backup and recovery strategies. It aims to make people aware of how important is to do and test backups.

Keywords: backup, recovery, database, database systems

DOI: 10.7862/re.2017.15

Tekst złożono w redakcji: wrzesień 2017

Przyjęto do druku: październik 2017