RPO i RTO: dwie liczby, od których zaczyna się każdy plan backupu

Zanim ktokolwiek zapyta, jakim narzędziem robimy backup, ktoś powinien odpowiedzieć na dwa pytania, które w ogóle nie są techniczne: ile danych wolno nam stracić i jak długo wolno nam nie działać.

Skróty użyte na tej stronie (3)

Skrót	Co znaczy
RPO	Recovery Point Objective: ile danych wolno stracić przy awarii
RTO	Recovery Time Objective: jak długo wolno nie działać przy awarii
SLA	Service Level Agreement: umowa o poziomie usług, np. 99,9% dostępności

Po co to istnieje

Pojęcie: RPO (Recovery Point Objective, docelowy punkt odtworzenia). Maksymalna ilość danych, wyrażona jako odcinek czasu, jaką organizacja godzi się bezpowrotnie stracić w razie awarii. RPO równe piętnastu minutom oznacza, że po najgorszej możliwej awarii dopuszczalne jest odtworzenie bazy do stanu sprzed najwyżej piętnastu minut — wszystko, co wydarzyło się w tym oknie, może przepaść i to jest zaakceptowane z góry, nie efekt niedopatrzenia.

Pojęcie: RTO (Recovery Time Objective, docelowy czas odtworzenia). Maksymalny czas, jaki może upłynąć od momentu awarii do momentu, w którym baza znowu obsługuje ruch produkcyjny. RTO równe dwóm godzinom oznacza, że cały proces — wykrycie awarii, decyzja, uruchomienie procedury odtworzenia, weryfikacja, przełączenie aplikacji — musi

> zmieścić się w tym oknie, niezależnie od tego, o której godzinie w nocy awaria się zdarzy.

Obie liczby są decyzją biznesową, nie techniczną, z jednego prostego powodu: to nie administrator bazy wie, ile godzin utraconych zamówień sklep internetowy jest w stanie przełknąć bez katastrofy finansowej, ani ile godzin przestoju systemu księgowego firma zniesie bez naruszenia terminów ustawowych. Administrator dostaje RPO i RTO jako dane wejściowe od właściciela systemu czy biznesu i dopiero na ich podstawie dobiera architekturę backupu i odtwarzania — nie odwrotnie. Plan backupu ustawiony bez uprzedniej rozmowy o RPO i RTO jest zgadywanką, która zwykle okazuje się zbyt optymistyczna dopiero w dniu prawdziwej awarii.

Analogia

RPO i RTO tłumaczą się dobrze przez pryzmat ubezpieczenia samochodu. RPO to odpowiedź na pytanie „ile ostatnich kilometrów jazdy jestem gotów odtworzyć z pamięci, jeśli licznik przepadnie w wypadku” — im częściej robimy sobie zdjęcie licznika (im częściej backup), tym mniej kilometrów trzeba by rekonstruować z domysłu. RTO to z kolei odpowiedź na pytanie „ile dni mogę być bez samochodu, zanim to naprawdę zaszkodzi” — od tej liczby zależy, czy warczymy się na naprawę w zwykłym warsztacie (długo, tanio) czy na wypożyczenie zastępczego auta tego samego dnia (szybko, drogo). Dokładnie tak samo RPO decyduje o częstotliwości zabezpieczania danych, a RTO o tym, jak kosztowny i szybki mechanizm przywracania trzeba mieć w gotowości — i obie decyzje należą do właściciela samochodu (biznesu), nie do mechanika (administratora bazy), który tylko realizuje to, co zamówiono.

Jak to działa

□□ Rozwiń: jak to działa pod spodem

Od RPO do częstotliwości backupu. RPO wyznacza wprost, jak często musi powstawać nowy punkt odtworzenia. RPO liczone w godzinach da się pokryć samym backupem pełnym albo różnicowym w odpowiednim odstępie. RPO liczone w minutach wymaga już backupu logów, opisanego na stronie Backup pełny, różnicowy, przyrostowy, logi — bo tylko ciągły zapis logu transakcyjnego pozwala odtworzyć stan sprzed kilku minut, a nie sprzed ostatniego backupu głównego zestawu danych. RPO równe zeru — brak akceptowalnej utraty jakichkolwiek zatwierdzonych danych — nie da się zrealizować samym backupem w żadnej postaci, bo backup z definicji jest zawsze o krok w tyle za bieżącym stanem; wymaga repliki synchronicznej, tematu strony Replikacja: synchroniczna, asynchroniczna, logiczna, fizyczna, w której każda transakcja jest potwierdzona klientowi dopiero, gdy zapisze się także na drugim serwerze.



Od ostatniej kopii do awarii uplywa RPO, od awarii do przywrocenia uslugi uplywa RTO (kliknięcie otwiera pełny rozmiar)

Od RTO do metody odtworzenia. RTO wyznacza, jaka metoda odtwarzania w ogóle mieści się w dostępnym czasie. Odtworzenie wielkiej bazy z backupu logicznego, wymagające odbudowania każdego indeksu od zera, może trwać godziny niezależnie od tego, jak szybki jest sam odczyt pliku eksportu — nie mieści się w RTO liczonym w minutach. Backup fizyczny odtwarza się szybciej, bo przywraca gotowe pliki, ale wciąż wymaga czasu na skopiowanie ich z nośnika backupu i na odegranie logów do właściwego momentu. RTO liczone w sekundach albo pojedynczych minutach realizuje się już nie odtwarzaniem z backupu, tylko przełączeniem na gotową, na bieżąco aktualizowaną replikę — proces opisany jako failover na stronie o replikacji.

RPO / RTO	Co to oznacza w praktyce	Co technicznie jest do tego potrzebne
RPO 24h / RTO 8h	akceptujemy utratę do doby danych, dzień na przywrócenie	nocny backup pełny albo różnicowy, odtwarzanie ręczne na życzenie
RPO 1h / RTO 4h	strata do godziny danych, kilka godzin na przywrócenie	backup pełny + przyrostowy co godzinę, procedura odtworzenia przetestowana i udokumentowana
RPO 15min / RTO 1h	strata kilkunastu minut, godzina na przywrócenie	backup pełny/różnicowy + backup logów co kilkanaście minut, PITR
RPO ~0 / RTO kilka minut	praktycznie zero utraconych danych, przełączenie niemal natychmiastowe	replika synchroniczna z automatycznym albo półautomatycznym failoverem

Okno backupowe. Pojęcie: okno backupowe. Przedział czasu, w którym backup może się wykonywać bez nadmiernego wpływu na obciążenie produkcyjne — zwykle noc albo inny okres niskiego ruchu. Okno backupowe jest twardym ograniczeniem, z którym RPO musi dać się pogodzić: żądanie RPO równego piętnastu minutom przy backupie pełnym trwającym trzy godziny jest sprzecznością samą w sobie, bo w ciągu tych trzech godzin backup pełny fizycznie nie zdąży się powtórzyć wystarczająco często — rozwiązaniem jest tu przesunięcie ciężaru na backup logów, który jest lekki i szybki, przy zachowaniu backupu pełnego rzadziej.

Koszt każdego poziomu. Każde obniżenie RPO i RTO o rząd wielkości kosztuje więcej — nie tylko licencji czy sprzętu, ale też złożoności operacyjnej, którą trzeba utrzymywać i regularnie testować. Backup nocny jest tani: jeden proces, jeden harmonogram, minimalny narzut na produkcję. Backup logów co kilka minut wymaga już monitorowania ciągłości łańcucha logów i

większej przestrzeni na ich przechowywanie. Replika synchroniczna wymaga drugiego serwera o porównywalnej mocy, stałego łącza o niskim opóźnieniu między nimi i akceptacji tego, że każdy `COMMIT` czeka na potwierdzenie z drugiej strony — koszt widoczny wprost jako wydłużony czas odpowiedzi każdej transakcji zapisującej. Rozmowa o RPO i RTO powinna więc od razu obejmować rozmowę o budżecie, bo liczba „zero minut utraty danych” wpisana bez zastanowienia w umowę SLA pociąga za sobą architekturę, której koszt bywa nieproporcjonalny do rzeczywistej wagi systemu.

Częste błędy

RPO zero zapisane w umowie przy nocnym backupie w praktyce. Klauzula SLA obiecująca zerową utratę danych, podczas gdy jedyną realnie wdrożoną metodą jest nocny backup pełny, jest zobowiązaniem niemożliwym do spełnienia od pierwszego dnia — nikt tego jednak nie zauważa, dopóki nie zdarzy się awaria w środku dnia roboczego, po której okazuje się, że można odtworzyć bazę wyłącznie do stanu sprzed kilkunastu godzin, a nie sprzed kilku sekund. Rozbieżność między RPO zapisanym na papierze a RPO faktycznie osiąganym przez wdrożoną architekturę jest jednym z najbardziej kosztownych błędów planowania backupu, bo ujawnia się dokładnie w najgorszym możliwym momencie i zwykle obciąża administratora, nie autora umowy.

Mylenie RPO z RTO. Zdanie „przywrócimy dane sprzed najwyżej godziny” miesza w jedną liczbę dwa osobne pytania — ile danych stracimy (RPO) i jak długo to potrwa (RTO). System może mieć RPO równe piętnastu minutom (backup logów co kwadrans) i jednocześnie RTO równe sześciu godzinom (bo samo odtworzenie dużej bazy z backupu fizycznego i odegranie logów zajmuje tyle czasu) — to dwie zupełnie niezależne liczby i osobno się je ustala, osobno testuje.

Ustalanie RPO i RTO bez testu. Liczby zapisane w dokumencie, których nigdy nie zweryfikowano realnym odtworzeniem na czas, są tylko przypuszczeniem. Dopiero test odtworzenia z zegarkiem w ręku pokazuje, czy RTO deklarowane na dwie godziny da się faktycznie osiągnąć, czy w praktyce trwa sześć, bo nikt nie uwzględnił czasu na skopiowanie plików backupu z wolnego nośnika archiwalnego.

□□ Ćwiczenia

> Zadania sprawdzające.

Ćwiczenie 1 — dobór architektury do wymagań

System zamówień ma ustalone RPO na pięć minut i RTO na trzydzieści minut. Obecnie działa wyłącznie nocny backup pełny. Jakie dwie zmiany w architekturze backupu i odtwarzania są konieczne, żeby faktycznie spełnić te wymagania, i dlaczego sam częstszy backup pełny by nie wystarczył?

Odpowiedź

Po pierwsze, potrzebny jest ciągły backup logów (albo replikacja) — tylko log transaksyjny pozwala odtworzyć stan sprzed pięciu minut, backup pełny nawet wykonywany co godzinę wciąż zostawia lukę rzędu godziny. Po drugie, potrzebna jest przetestowana, zautomatyzowana procedura odtwarzania do punktu w czasie, bo ręczne odtwarzanie dużej bazy z backupu pełnego plus ręczne odegranie logów zwykle nie mieści się w trzydziestu minutach — trzydzieści minut RTO wymaga gotowego skryptu albo mechanizmu, nie improwizacji w trakcie awarii. Sam częstszy backup pełny nie rozwiązuje problemu RPO, bo backup pełny co pięć minut przy większej bazie fizycznie nie zdąży się wykonać ani nie zmieści w rozsądnym oknie zasobów.

Pytania kontrolne

1. Czym różni się RPO od RTO w jednym zdaniu?

Odpowiedź

RPO mówi, ile danych wolno stracić (mierzone czasem wstecz od momentu awarii), a RTO mówi, jak długo wolno nie mieć działającej bazy (mierzone czasem od awarii do przywrócenia usługi) — jedno dotyczy utraconych danych, drugie dotyczy czasu przestoju, i są od siebie niezależne.

2. Dlaczego RPO równe zeru nie da się zrealizować żadną formą backupu, niezależnie od tego, jak często ten backup się wykonuje?

Odpowiedź

Backup, w każdej postaci, jest zawsze kopią stanu sprzed pewnego momentu w przeszłości — nawet ciągły backup logów ma opóźnienie rzędu sekund czy minut między zatwierdzeniem

transakcji a bezpiecznym zapisaniem tego zdarzenia w archiwum logu. RPO równe zeru wymaga, żeby żadna zatwierdzona transakcja nie mogła zostać utracona, co jest możliwe wyłącznie wtedy, gdy zapis istnieje jednocześnie w dwóch niezależnych miejscach w momencie potwierdzenia `COMMIT` u klientowi — a to jest już definicja repliki synchronicznej, nie backupu.

Powiązane: Backup pełny, różnicowy, przyrostowy, logi · Odtwarzanie do punktu w czasie (PITR) · Replikacja: synchroniczna, asynchroniczna, logiczna, fizyczna · Co monitorujemy w bazie danych

Źródła: dokumentacja PostgreSQL, rozdział „Continuous Archiving and Point-in-Time Recovery” · Microsoft Docs, „Recovery models” i „Business continuity and database recovery” · Oracle Database Backup and Recovery User's Guide, rozdział „Backup and Recovery Concepts”

Revision #44
Created 2026-09-24 23:07:43 UTC by Tomasz Konieczny
Updated 2026-10-26 14:11:53 UTC by Tomasz Konieczny