Backup bazy do S3: agent backint i jego ograniczenia

Poziom: podstawy **Dotyczy: SAP HANA, agent Backint dla AWS S3** **Czas: ~15 min**

Skróty użyte na tej stronie (5)

Skrót	Co znaczy
AWS	Amazon Web Services: chmura Amazonu
HANA	SAP HANA: baza danych w pamięci firmy SAP, używana do dużych systemów księgowych i magazynowych
HTTP	Hypertext Transfer Protocol: protokół stron WWW
S3	Simple Storage Service: magazyn plików w chmurze AWS
SAP	Systeme, Anwendungen und Produkte in der Datenverarbeitung: niemiecki producent oprogramowania klasy ERP, właściciel silnika HANA i przejętego wcześniej Sybase ASE

Z życia wzięte

Z ŻYCIA WZIĘTE

O 4:10 nocne zadanie backupu bazy HANA kończy się błędem `Error reading backup from BACKINT`, a rano trzeba wyjaśnić, czy dane sprzed nocy są w ogóle bezpieczne. Tomek sprawdza w S3, czy oczekiwany klucz backupu w ogóle istnieje, i widzi, że plik jest, ale ma połowę spodziewanego rozmiaru, bo transfer do S3 urwał się w połowie przez chwilową awarię sieci. Uruchamia backup ad hoc jeszcze raz w ciągu dnia i dopiero po potwierdzeniu pełnego rozmiaru pliku w S3 uznaje kopię za odtwarzalną.

Czym jest backint

> **Pojęcie: backint.** Interfejs programowy SAP HANA (Backup Interface for SAP HANA), przez który silnik bazy danych oddaje strumień kopii zapasowej zewnętrznemu narzędziu zamiast zapisywać go bezpośrednio w systemie plików. HANA nie wie i nie musi wiedzieć, gdzie backint faktycznie zapisuje dane — może to być deduplikator, biblioteka backupowa albo, jak w tym przypadku, agent kierujący strumień wprost do magazynu obiektowego (S3). Z perspektywy HANA backup przez backint wygląda identycznie jak backup do pliku — różni się tylko tym, co dzieje się po drugiej stronie interfejsu.



HANA oddaje strumień backupu przez interfejs backint agentowi, który zapisuje albo odczytuje dane z magazynu S3 (kliknięcie otwiera pełny rozmiar)

Backup pełny, różnicowy i przyrostowy działają przez backint tak samo jak przez system plików — różnicę między tymi rodzajami backupu, niezależną od tego, gdzie trafia strumień, opisuje strona Backup pełny, różnicowy, przyrostowy, logi. Backint dokłada do tego wyłącznie warstwę transportu do zewnętrznego magazynu.

Zanim backup w ogóle trafi do S3, HANA musi wiedzieć, że ma go oddać przez backint, a nie zapisać na lokalny system plików — wybór ten ustawiamy w parametrze `catalog_backup_using_backint` (dla katalogu backupu) i analogicznych parametrach dla backupu danych i logów w `global.ini`, w sekcji `[backup]`. Sam agent backint dla S3 jest osobnym programem zewnętrznym, wskazanym w konfiguracji HANA ścieżką do binarki i do jego własnego pliku parametrów — to właśnie w tym pliku parametrów agenta leżą `MaximumConcurrentFilesForRestore` i `DownloadConcurrency` opisane niżej, nie w konfiguracji samej HANA.

Limity agenta backint dla S3

Agent backint kierujący dane do S3 ma własne, konfigurowalne limity równoległości, niezależne od limitów samego S3 czy przepustowości łącza — te limity decydują o tym, ile plików backupu agent przesyła (albo odzyskuje) naraz.

> **Pojęcie: `MaximumConcurrentFilesForRestore`.** Parametr konfiguracji agenta określający, ile plików backupu może być jednocześnie odczytywanych z S3 podczas operacji odtwarzania. Wyższa wartość przyspiesza odzyskiwanie przy wielu małych plikach kosztem większego obciążenia sieci i procesora hosta HANA w tym samym momencie; zbyt wysoka wartość na słabszym łączu wydłuża odtwarzanie zamiast je przyspieszyć, bo wszystkie równoległe

strumienie zaczynają się wzajemnie dławić.

Pojęcie: `DownloadConcurrency`. Liczba równoległych połączeń HTTP, jakie agent otwiera do S3 przy pobieraniu pojedynczego pliku backupu — niezależna od `MaximumConcurrentFilesForRestore`, które mnoży liczbę plików, nie liczbę połączeń na plik. Dla dużych plików backupu (pełny backup wielogigabajtowej bazy) wyższa wartość `DownloadConcurrency` skraca czas pobrania jednego pliku, dzieląc go na równoległe fragmenty.

Przykładowa konfiguracja tych dwóch parametrów w pliku konfiguracyjnym agenta:

```
MaximumConcurrentFilesForRestore: 5
DownloadConcurrency: 10
```

Uważaj: oba parametry razem określają łączną liczbę jednoczesnych połączeń do S3 — przy `MaximumConcurrentFilesForRestore: 5` i `DownloadConcurrency: 10` agent może w szczycie otwierać do 50 połączeń naraz. Na hoście z ograniczoną przepustowością sieci warto obniżyć jeden z tych parametrów, zanim obniży się drugi, żeby zachować kontrolę nad tym, który wymiar równoległości faktycznie ogranicza.

Czas trwania kopii logów i „ostatnia strona" pliku logu

Pojęcie: kopia logów transakcyjnych (log backup). Ciągły, przyrostowy zrzut logu transakcyjnego HANA, uruchamiany automatycznie po zapełnieniu bieżącego segmentu logu — w odróżnieniu od backupu pełnego i różnicowego, log backup nie jest jednorazową operacją na żądanie, tylko procesem działającym w tle przez cały czas pracy bazy.

Przy kopiowaniu backupu logów do innej lokalizacji (np. przy replikacji katalogu backupu do ośrodka zapasowego) istotne jest to, że plik ostatniego, aktualnie otwartego segmentu logu (`log_backup_0_0_0_0` i podobne) rośnie i zmienia zawartość, dopóki segment nie zostanie zamknięty. Skopiowanie takiego pliku w trakcie jego zapisu daje kopię niekompletną — jeśli klucz obiektu w S3 został skopiowany wcześniej, a segment logu urósł już po tej kopii, trzeba skopiować nowszą wersję tego samego klucza ponownie, inaczej odtworzenie bazy z tej kopii zatrzyma się dokładnie na granicy, w której ostatnia skopiowana wersja się urwała.

Uważaj: przy każdym odświeżeniu kopii katalogu backupu do innej lokalizacji sprawdzamy, czy ostatni plik logu w źródle ma nowszą datę modyfikacji (albo inny rozmiar) niż jego kopia w celu — jeśli tak, kopiujemy go ponownie, zanim uznamy synchronizację za zakończoną.

Typowe komunikaty i co oznaczają

Komunikat	Znaczenie
`RECOVER DATA finished with error: recovery could not be completed`	proces odtwarzania zatrzymał się przed ukończeniem — właściwa przyczyna jest zwykle w komunikacie poniżej tej linii, nie w niej samej
`Error reading backup from BACKINT`	agent backint nie zwrócił oczekiwanych danych dla żądanego pliku backupu — najczęściej plik jest niekompletny albo nie istnieje pod oczekiwanym kluczem w S3
`Not all data could be written: Expected 4096 but transferred 0`	agent otworzył strumień do pliku backupu, ale nie odczytał z niego ani bajtu — typowy obraz sytuacji, w której kopiowany plik logu był w trakcie dopisywania (patrz sekcja wyżej) albo klucz w S3 wskazuje na starą, niepełną wersję obiektu
`Backint cannot find file`	agent nie znalazł w katalogu backupu (w S3, pod wskazanym prefiksem) pliku, którego żąda HANA na podstawie katalogu backupu bazy — zwykle plik został przeniesiony, usunięty przez regułę cyklu życia opisaną na stronie S3: kubełki, obiekty, uprawnienia, albo nigdy nie dotarł do docelowej lokalizacji

Diagnozowanie tych komunikatów zaczynamy zawsze od sprawdzenia, czy plik backupu, o który pyta HANA, faktycznie istnieje pod oczekiwanym kluczem w S3 i czy jego rozmiar odpowiada oczekiwanemu — dopiero potem sprawdzamy uprawnienia i limity równoległości agenta.

Odtwarzanie bazy z backupu w S3

Proces odtwarzania (recovery) w studiu administracyjnym HANA pyta kolejno o lokalizację backupu danych, backupu logów i katalogu backupu — przy backincie wskazującym S3 każda z tych lokalizacji jest ścieżką wewnątrz kubełka (prefiksem), a nie ścieżką w lokalnym systemie plików. Zanim proces odtwarzania cokolwiek zmieni, HANA opcjonalnie sprawdza dostępność wszystkich wymaganych backupów delta i logów — jeśli którychś brakuje, odtwarzanie zatrzymuje się przed wprowadzeniem jakiejkolwiek zmiany, zamiast ryzykować częściowe odtworzenie. Warto zostawiać tę weryfikację włączoną mimo wydłużenia czasu przygotowania — powtórzenie całego odtwarzania od zera po odkryciu braku w jego trakcie kosztuje zdecydowanie więcej czasu niż samo sprawdzenie na starcie.

Częste błędy

Kopiowanie katalogu backupu do ośrodka zapasowego bez ponownego sprawdzenia ostatniego segmentu logu. Jednorazowe skopiowanie całego katalogu backupu wygląda na kompletne, ale jeśli ostatni segment logu wciąż rósł w trakcie kopiowania, odtworzenie z tej kopii urwie się dokładnie na tym segmencie — konieczne jest odświeżenie przynajmniej ostatniego pliku przed uznaniem kopii za gotową do użycia.

Podniesienie `MaximumConcurrentFilesForRestore` bez sprawdzenia przepustowości łącza. Wyższa wartość nie przyspiesza odtwarzania automatycznie — na wolniejszym łączu równoległe strumienie konkurują o to samo pasmo i całkowity czas odtworzenia może się wydłużyć zamiast skrócić.

Traktowanie komunikatu `Backint cannot find file` jako problemu z samą HANA. Przyczyna leży niemal zawsze po stronie magazynu backupu (usunięty obiekt, zła nazwa prefiksu, wygasła reguła cyklu życia) — HANA tylko przekazuje błąd zwrócony przez agenta backint.

□□ Ćwiczenia

> Zadania sprawdzające, bez dostępu do systemu SAP HANA — wystarczy zrozumienie mechanizmu.

1. Wyjaśniamy, dlaczego HANA „nie wie", że jej backup trafia do S3, mimo że faktycznie tam ląduje.
2. Opisujemy różnicę między `MaximumConcurrentFilesForRestore` a `DownloadConcurrency` i podajemy, w jakiej sytuacji obniżenie jednego z nich pomaga bardziej niż obniżenie drugiego.
3. Odtwarzanie bazy zatrzymuje się z komunikatem `Not all data could be written: Expected 4096 but transferred 0`. Wymieniamy dwie najbardziej prawdopodobne przyczyny i sposób ich rozróżnienia.
4. Uzasadniamy, dlaczego warto zostawić włączoną weryfikację dostępności backupów delta i logów przed rozpoczęciem odtwarzania, mimo że wydłuża to czas przygotowania procesu.

Powiązane: Backup pełny, różnicowy, przyrostowy, logi · S3: kubełki, obiekty, uprawnienia · Architektura SAP HANA: instancja, numer, nazwa SID, procesy

Źródła: dokumentacja SAP HANA, interfejs Backint · dokumentacja agenta backint dla Amazon S3

Revision #37
Created 2026-09-25 08:26:24 UTC by Tomasz Konieczny
Updated 2026-10-26 14:16:28 UTC by Tomasz Konieczny