

Rozproszona baza danych

Rozproszoną bazą danych nazywamy zbiór baz danych umieszczonych na wielu różnych węzłach sieci komputerowej lokalnej lub rozległej. Z rozproszonymi bazami (jak ze wszystkim) wiąże się kilka udogodnień i kilka problemów.

Zalety

- Rozproszona baza danych dobrze odzwierciedla strukturę organizacji,
- lokalna autonomia - lokalne oddziały (np. filia firmy) kontroluje tylko te dane, które jej dotyczą,
- bezpieczeństwo ważnych danych - jeśli spłonie serwerownia, to traci się tylko część danych,
- większa wydajność - ruch może być rozłożony na wiele serwerów,
- względy ekonomiczne - taniej jest złożyć sieć słabych komputerów niż kupić jeden mocny,
- modularność - można usuwać i dodawać moduły bez burzenia integralności,
- redundancja danych na serwerach ułatwia tworzenie transakcji.

Wady

- Złożoność - konieczne są dodatkowe systemy zarządzające, aby klient końcowy miał wrażenie, że korzysta z "normalnej" bazy,
- dochodzą ekstra koszty na infrastrukturę,
- bezpieczeństwo - należy dobrze zabezpieczyć wszystkie węzły,
- problemy z integralnością - zwłaszcza w przypadku sieci globalnej (przesyłanie zmian może zabić połączenie z Internetem),
- wymagani są pracownicy z większym doświadczeniem,
- brak standardów - nie ma opracowanych prostych metod konwersji bazy danych w bazę rozproszoną,
- trudniej zaprojektować taką bazę,
- wymagane jest dodatkowe (często płatne) oprogramowanie,
- problemy z współbieżnością, konieczność stosowania timestampingu i blokowania.

SZRBD (DDBMS)

Zbiór narzędzi umożliwiających połączenie baz w całość to System Zarządzania Rozproszoną Bazą Danych (SZRBD). Całość systemu składa się ze:

- stacji roboczych,
- oprogramowania i sprzętu sieciowego,
- mediów komunikacyjnych (kable ;),
- procesor transakcji (zwany też menedżerem transakcji lub procesorem aplikacji)
 - czyli kawałek softu obecny na każdej maszynie, który wysyła zapytania
- procesor danych lub menedżer danych
 - czyli kawałek softu obecny na każdej maszynie, który zarządza obecnymi na niej danymi.

W zależności od tego, ile jest uruchomionych procesów obsługujących dane i ile komputerów dane te trzyma, mamy 3 kombinacje:

	Dane na jednej maszynie Single side data (SD)	Dane na wielu maszynach Multiple side data (MD)
Jeden proces Single side process (SP)	SPSD Zwyczajna baza danych	-
Wiele procesów Multiple side process (MP)	MPSD Serwer plików + stacje robocze	MPMD Pełne rozproszenie

- **SPSD**
 - Typowa baza danych,
 - dane przechowywane są na serwerze,
 - serwer wykonuje operacje,
 - klient jedynie zamawia dane i je otrzymuje.
- **MPSD**
 - Wiele procesów odpalonych w sieci dzieli te same dane,
 - wymaga serwera plików (na przykład NFS).
- **MPMD**
 - Pełne rozproszenie z obsługą wielu menedżerów transakcji i menedżerów danych na wielu hostach,
 - może być homogeniczny
 - integruje systemy baz danych jednego typu (na przykład MySQL),
 - lub heterogeniczny
 - integruje systemy baz danych różnych rodzajów (MySQL + Oracle).
 - a nawet w pełni heterogeniczny
 - integruje systemy baz danych o różnych modelach danych (baza obiektowa + relacyjna) uruchomione na różnych systemach (na przykład mainframe + zwyczajny PC).

Przezroczystość

Rozproszona baza danych musi też spełniać warunek przezroczystości - klient musi widzieć całą rozproszoną bazę danych jako zwykłą bazę. Przezroczystość odbywa się na kilku płaszczyznach:

- Przezroczystość rozproszenia,
 - pozwala na zarządzanie rozproszoną bazą tak, jakby była to zwyczajna baza.
- Przezroczystość transakcji,
 - każda transakcja jest spójna i pozostawi każdą BD systemu spójną.
- przezroczystość błędów, wiążąca się z powyższą.
- Przezroczystość wydajności,

- chodzi o minimalizację kosztów wykonania zapytania. Na koszty składa się:
 - czas dostępu (I/O),
 - czas komunikacji,
 - czas pracy procesora.
- Przezroczystość heterogeniczności,
 - nie ma znaczenia, jakie to systemy się ze sobą łączą.

Projektowanie

Aby zaprojektować rozproszoną bazę danych trzeba sobie zadać pytania o 3 rzeczy. Są to:

- Fragmentacja danych,
 - jak podzielić dane:
 - fragmentacja pozioma - podzielić relację na podzbiory krotek (wierszy),
 - na przykład: studenci do jednej bazy, pracownicy do drugiej,
 - fragmentacja pionowa - podzielić relację na wzdłuż atrybutów (kolumn),
 - na przykład: imię i nazwisko do jednej bazy, PESEL do drugiej,
 - fragmentacja mieszana — wiadomo.
- Replikacja danych,
 - które dane mają się powtarzać,
 - replikacja zapewnia szybszy dostęp,
 - skraca czas komunikacji,
 - zapewnia większe bezpieczeństwo kluczowych danych,
 - jest parę modeli:
 - pełna replikacja (w praktyce niestosowana),
 - replikacja niektórych fragmentów bazy danych,
 - brak replikacji.
- Alokacja danych,
 - gdzie umieścić fragmenty danych i ich kopie,
 - wiadomo, 3 modele:
 - centralna alokacja (jedna baza na jednym hoście),
 - rozdzielona alokacja (baza dzielona na fragmenty i trzymana w różnych miejscach),
 - replikowana alokacja (baza na jednym hoście, a jej kopia na innym).

Źródła

- [Distributed Database Management System](#) - prezentacja po angielsku, dość kompletna,
- [Rozproszone bazy danych](#) - prezentacja do wykładu na PW
- [Rozproszone bazy danych](#) - prezentacja do wykładu na Politechnice Śląskiej, raczej mętna
- [Hasło](#) na angielskiej Wiki.