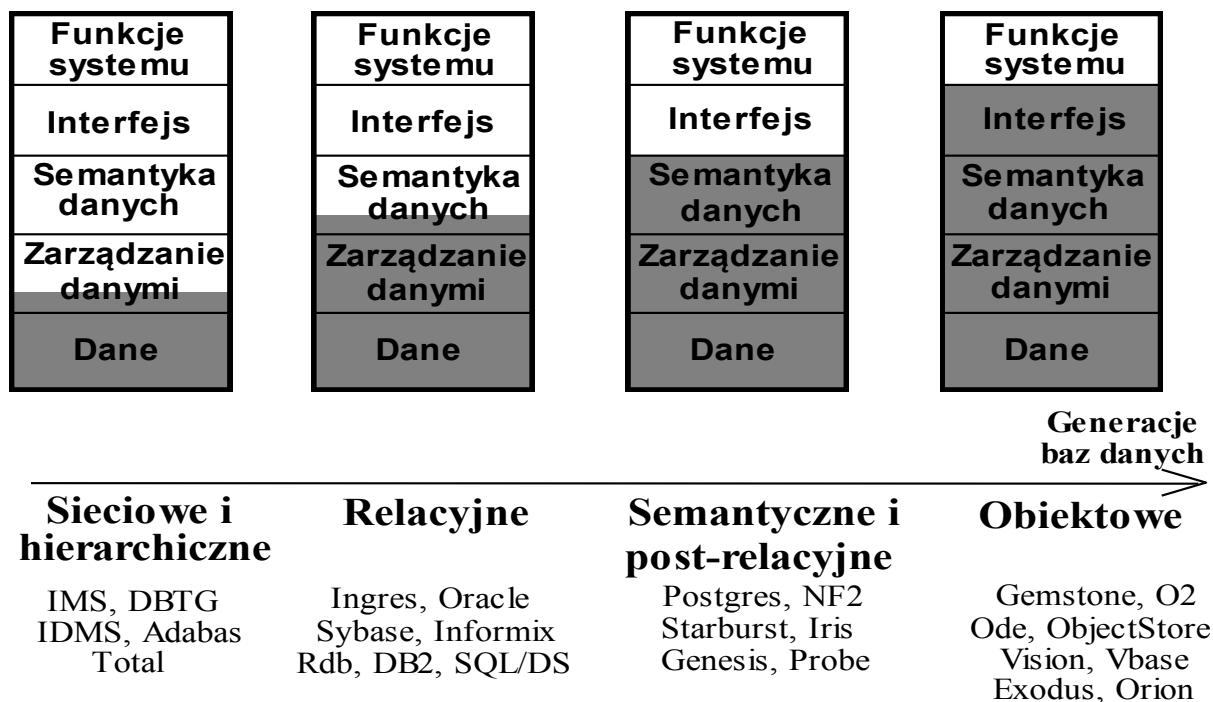


Omów ewolucję systemów opartych na bazach danych



Czym się różni modelowanie od strukturalizacji danych?

Modelowanie danych to tworzenie czysto teoretycznych koncepcji budowy baz danych na podstawie kontaktów z klientami natomiast strukturalizacja danych to proces przelewania idei i koncepcji pochodzących z procesu modelowania w konkretne struktury programistyczne.

Jakie aspekty modelowanych danych są istotne w późniejszym procesie ich strukturalizacji?

Ważne jest uchwycenie całości systemu który chcemy opisać za pomocą bazy danych wraz ze wszystkimi kluczowymi zjawiskami w nim zachodzącymi. Należy wyodrębnić wszelkie struktury (zadbać o opis całej struktury), nadać im odpowiednie atrybuty i wszystkie elementy połączyć odpowiednimi relacjami. Tylko poprawnie przeprowadzone modelowanie danych może zagwarantować sukces strukturalizacji.

Omów znaczenia czasu w modelowaniu i strukturalizacji danych

Ilość czasu w modelowaniu i strukturalizacji danych jest wprost proporcjonalna do jakości procesu modelowania i strukturalizacji.

Jakie są podstawowe składniki systemu przetwarzania danych i ich wzajemna relacja?

Są to maszyny (elementy wykonujące zadane operacje), programy (dane interpretowane przez maszyny jako algorytmy strukturalne lub obiektowe), dane (jako wejściowe dane "surowe" oraz wyjściowe, przetworzone przez system) oraz ludzie (manipulacja działaniem systemu i danymi, kontrola, administracja).

Scharakteryzuj każdy (lub wybrany) ze składników przetwarzania danych w kontekście jego roli w systemie i wykorzystywanych metod.

Maszyny za pomocą struktur mechanicznych i logicznych wykorzystujących prawa fizyki wykonują i interpretują algorytmy. Algorytmy to oprogramowanie, które służy do obróbki i modelowania danych wejściowych w wyjściowe. Ludzie sprawują kontrolę nad systemem i mają możliwość modyfikowania danych oraz administracji procesem. Dodatkowo w literaturze czasami pojawia się dodatkowy składnik "procedury". Procedury precyzują zasady projektowania i użytkowania bazy danych.

Co to są procesy ETL?

Procesy ETL (Extract, Transform and Load) to narzędzia wspomagające proces pozyskania danych dla baz danych, szczególnie dla hurtowni danych (baz danych, które są zorganizowane i zoptymalizowane pod kątem pewnego wycinka rzeczywistości którego opisują).

Zadaniem narzędzi ETL jest:

- pozyskanie danych ze źródeł zewnętrznych,
- przekształcenie danych w formę odpowiadającą strukturze bazy danych,
- załadowanie danych do bazy danych (zazwyczaj będącej hurtownią danych).

Scharakteryzuj rodzaje architektur baz danych.

Architektura systemów baz danych rozwijała się od tzw. architektury jednowarstwowej, w kierunku architektury wielowarstwowej.

Architektura jednowarstwowa (tzw. architektura CS, Client-Server) to architektura najprostsza. Termin architektura typu klient-serwer wywodzi się od sposobu interakcji komponentów softwarowych z systemem, mianowicie klient jest procesem który potrzebuje pewnych zasobów a proces serwera tych zasobów dostarcza. W dzisiejszych czasach praktycznie nieużywana w zastosowaniach profesjonalnych.

Architektura dwuwarstwowa wyróżnia dwie warstwy oprogramowania - warstwę serwera (proces serwera) oraz warstwę klienta (proces klienta). Lokalizacje obu procesów mogą znajdować się na jednym komputerze, choć zazwyczaj serwer umieszczany jest na innym komputerze niż procesy klienta komunikując się poprzez sieć. Na serwerze znajdują się dane i oprogramowanie zapewniające dostęp do danych (np. interpretator wywołań SQL) a po stronie klienta realizowana jest logika aplikacji. Klient odpowiada także za przetwarzanie otrzymanych danych i formę przedstawienia wyników.

Zalety: bezpieczeństwo serwera, mały ruch sieciowy, przetwarzanie danych na serwerze.

Wady: Utrudnienia związane z administracją wielu komputerów, duże koszty eksploatacji klientów, brak kontroli nad działaniami użytkowników.

Architektura trójwarstwowa dzieli aplikację bazy danych na trzy części: warstwę dolną (realizuje dostęp do bazy danych), warstwę środkową (reguły dziedziczenia danych) oraz warstwę górną (interfejs użytkownika). Warstwa dolna to prawie zawsze programy wykonywane na serwerze, obsługujące zlecenia warstwy środkowej. Warstwa środkowa może być klientem lub serwerem, natomiast górna - zawsze klientem. W arch. trójwarstwowej mamy doczynienia z tzw. "chudym (cienkim) klientem" (interfejs WWW przez przeglądarkę internetową) lub "grubym klientem" (klient realizuje dodatkowo bardziej złożone interfejsy użytkownika jak i część logiki aplikacji).

Zalety: proste projektowanie, szybka implementacja, podział na komponenty

Wady: Potrzebny silny sprzęt serwerowy, trudniejsze technologie, zwiększony ruch sieciowy.

Jakie są główne problemy przetwarzania numerycznego?

Główne problemy przetwarzania numerycznego to potrzeba tworzenia stosunkowo skomplikowanych i wielkoobjętościowych algorytmów obliczeniowych w stosunku do małej ilości danych obliczanych tą metodą.

Wyjaśnij pojęcie OLTP.

Systemy informatyczne możemy podzielić na transakcyjne(OLTP) i analityczne (OLAP). Generalnie można przyjąć, że systemy OLTP dostarczają danych źródłowych do hurtowni danych, natomiast systemy OLAP pomagają w ich analizie.

OLTP (On-line Transaction Processing) charakteryzuje się dużą ilością prostych transakcji zapisu i odczytu. Główny nacisk kładziony jest na zachowanie integralności danych w środowisku wielodostępowym oraz na efektywność mierzona liczbą transakcji w danej jednostce czasu.

OLAP (On-line Analytical Processing) charakteryzuje się natomiast stosunkowo nielicznymi, ale za to złożonymi transakcjami odczytu. Miara efektywności jest czas odpowiedzi. Powszechnie wykorzystuje się go w technikach związanych z Data Miningiem.

Scharakteryzuj bazy operacyjne od strony technicznej.

W hurtowniach danych bazy operacyjne to takie bazy, które zawierają często aktualizowane bieżące operacje organizacji. Bazy operacyjne często opierają się na modelu OLTP który jest zoptymalizowany do szybszego przeprowadzania transakcji: wstawiania, usuwania i aktualizowania danych. Operacyjne bazy danych często dokładane są do hurtowni danych dla wzrostu wydajności.

Co to jest transakcja w rozumieniu baz danych i jaką pełni rolę?

Transakcja to zbiór operacji na bazie danych, które stanowią w istocie pewną całość i jako takie powinny być wykonane wszystkie lub żadna z nich. Warunki jakie powinny spełniać transakcje bardziej szczegółowo opisują zasady ACID (Atomicity, Consistency, Isolation, Durability - Atomowość, Spójność, Izolacja, Trwałość).

Jakie są podstawowe architektury systemów OLTP?

Są to architektury SOA (zorientowane na usługi systemy interfejsowe) oraz architektury Web.

Na czym polega mechanizm archiwizacji danych?

Archiwizacja danych w rozumieniu informatyki, jest to czynność wykonywania kopii danych w pamięci masowej, w celu ich długotrwałego przechowywania. Istnieje ponad 40 mechanizmów archiwizacji danych z czego najprostszym jest kopiowanie plików na nośnik inny, niż oryginalnie dane się znajdują.

Przedstaw koncepcję Hurtowni Danych i scharakteryzuj ich rodzaje.

W praktyce hurtownie są bazami danych integrującymi dane z wszystkich pozostałych systemów bazodanowych w przedsiębiorstwie. Ta integracja polega na cyklicznym zasilaniu hurtowni danymi systemów produkcyjnych (może być tych baz lub systemów dużo i mogą być rozproszone).

Architektura bazy hurtowni jest zorientowana na optymalizację szybkości wyszukiwania i jak najefektywniejszą analizę zawartości. Stąd bywa, że hurtownie danych nie są realizowane za pomocą relacyjnych baz danych, gdyż takie bazy ustępują szybkością innym rozwiązaniom.

W praktyce w ramach architektury hurtowni wyróżniany jest poziom danych detalicznych oraz warstwa agregatów/kostek tematycznych.

Cele HD to przetwarzanie analityczne (OLAP), wspomaganie decyzji (DSS), archiwizacja danych, analiza efektywności i wsparcie dla systemów CRM (np. dobieranie strategii marketingowych na podstawie danych o klientach i sprzedaży).

Jaką rolę pełnią metadane w HD?

W przypadku HD, metadanymi są definicje tabel, widoków, kluczy itp. natomiast danymi – zawartość tych tabel, widoków – rekordy. Istnieją także dodatkowe metadane wspomagające i optymalizujące transakcje w HD.

Porównaj systemy OLTP i Hurtownie Danych (jako przykład OLAP).

Ogólnie można podzielić systemy IT na transakcyjne (OLTP) i analityczne (OLAP). Generalnie możemy założyć, że systemy OLTP są źródłami danych do hurtowni danych podczas gdy hurtownie danych (OLAP) pomagają w analizowaniu ich.

Omów pojęcia faktów i wymiarów w kontekście systemów OLAP.

Fakty i wymiary powiązane są z Kostką OLAP, która jest strukturą danych pozwalającą na szybką analizę danych. Przechowuje ona dane w sposób bardziej przypominający wielowymiarowe arkusze kalkulacyjne niż tradycyjną, relacyjną bazę danych. Można ją również zdefiniować jako zdolność manipulowania i analizowania danych z różnych punktów widzenia. Rozmieszczenie danych w kostkach pokonuje ograniczenia relacyjnych baz danych.

Wymień i omów schematy danych wykorzystywane w HD.

- Gwiazda – istnieje pojedyncza centralna tabela faktów, a wymiary są zdenormalizowane. Pojedyncza tabela miar/faktów jest połączona z tabelami wymiarów przez klucze główne i obce.
- Płatek śniegu – istnieje pojedyncza, centralna tabela faktów. Wymiary są znormalizowane.
- Konstelacja faktów – kombinacja wielu schematów gwiazd ze wspólnymi wymiarami. Różne tabele faktów mogą odwoływać się do różnych poziomów danego wymiaru.

Co to jest i jak przebiega eksploracja danych?

Eksploracja danych (spotyka się również określenie drążenie danych, pozyskiwanie wiedzy, wydobywanie danych, ekstrakcja danych) (ang. data mining) - jeden z etapów procesu odkrywania wiedzy z baz danych. Idea eksploracji danych polega na wykorzystaniu szybkości komputera do znajdowania ukrytych dla człowieka (właśnie z uwagi na ograniczone możliwości czasowe) prawidłowości w danych zgromadzonych w hurtowniach danych poprzez wyspecjalizowane algorytmy lub sieci neuronowe.