

System Rozproszonej Bazy Danych dla Prywatnej Szkoły Podstawowej

Część 1: Wprowadzenie i Architektura

Autorzy: Piotr Ledwoch, Kacper Woźnica

Przedmiot: Rozproszone Bazy Danych

1. Wprowadzenie

1.1 Opis projektu

System Rozproszonej Bazy Danych (RBD) został zaprojektowany dla kompleksowego zarządzania prywatną szkołą podstawową. Projekt demonstruje praktyczne zastosowanie kluczowych konceptów rozproszonych baz danych w środowisku heterogenicznym, integrując trzy różne systemy bazodanowe.

1.2 Cele realizacji

- **Zarządzanie danymi edukacyjnymi** - przechowywanie informacji o uczniach, nauczycielach, ocenach i frekwencji
- **Obsługa finansowa** - kompleksowe zarządzanie kontraktami i płatnościami
- **System uwag pedagogicznych** - rejestrowanie obserwacji i komentarzy nauczycieli
- **Integracja systemów** - zapewnienie spójnego dostępu do danych rozproszonych

1.3 Znaczenie edukacyjne

System ilustruje najważniejsze aspekty rozproszonych baz danych:

- Integrację środowisk heterogenicznych
- Mechanizmy transakcji rozproszonych
- Strategie replikacji danych
- Optymalizację zapytań międzysystemowych

2. Architektura Systemu

2.1 Komponenty systemu

Microsoft SQL Server (Serwer główny)

- **Rola:** Centralny system zarządzania danymi szkolnymi
- **Zawartość:** Dane uczniów, nauczycieli, klas, przedmiotów, ocen, frekwencji
- **Uzasadnienie:** Doskonałe wsparcie dla linked servers i koordynacji transakcji

Oracle Database (Serwer finansowy)

- **Rola:** Obsługa aspektów finansowych szkoły
- **Zawartość:** Kontrakty finansowe, płatności, rozliczenia
- **Uzasadnienie:** Wysoką wydajność dla operacji finansowych i silne mechanizmy transakcyjne

PostgreSQL (Serwer uwag)

- **Rola:** System uwag i komentarzy pedagogicznych
- **Zawartość:** Uwagi nauczycieli, pochwały, obserwacje postępów
- **Uzasadnienie:** Elastyczność w przechowywaniu danych tekstowych i symulacja trzeciego środowiska

2.2 Strategia podziału danych

Podział funkcjonalny - każdy serwer obsługuje określoną domenę biznesową:

- **Separacja danych finansowych** - krytyczne dane finansowe odizolowane w dedykowanym systemie Oracle
- **Specjalizacja serwerów** - każdy system zoptymalizowany pod konkretne operacje
- **Niezależne skalowanie** - możliwość rozwoju każdego komponentu osobno

2.3 Schemat komunikacji

SQL Server jako centrum integracji:

- Łączy się ze wszystkimi systemami poprzez Linked Servers
- Udostępnia zunifikowane widoki danych rozproszonych
- Koordynuje transakcje rozproszone

Oracle z database links:

- Symuluje środowisko rozproszone poprzez połączenia między schematami
- Implementuje zaawansowane procedury finansowe
- Zarządza operacjami międzysystemowymi

PostgreSQL jako system zewnętrzny:

- Działa niezależnie z dostępem przez ODBC
- Przechowuje dane w izolowanym środowisku
- Symuluje integrację z zewnętrznym systemem pedagogicznym

3. Kluczowe Technologie

3.1 Linked Servers

- **Połączenia heterogeniczne** - integracja SQL Server z Oracle, PostgreSQL i Excel
- **Transparentny dostęp** - jednolity interfejs do różnych systemów
- **Optymalizacja zapytań** - inteligentne przekazywanie przetwarzania do systemów źródłowych

3.2 Database Links Oracle

- **Symulacja rozproszonego środowiska** - połączenia między schematami Oracle
- **Synonimy** - transparentny dostęp do zdalnych tabel
- **Wzajemne uprawnienia** - bezpieczna synchronizacja danych

3.3 Transakcje Rozproszone

- **MS DTC** - Microsoft Distributed Transaction Coordinator
- **Two-Phase Commit** - protokół zapewniający atomowość operacji
- **Error handling** - kompletny rollback przy błędach

3.4 Replikacja Transakcyjna

- **Publisher-Subscriber** - model replikacji między instancjami SQL Server
- **Continuous monitoring** - automatyczne śledzenie zmian
- **Disaster recovery** - mechanizmy odzyskiwania po awarii

4. Korzyści Rozwiązania

4.1 Funkcjonalne

- **Kompleksowe zarządzanie szkołą** - wszystkie aspekty w jednym systemie
- **Specjalizacja komponentów** - każdy system wykonuje swoje zadania optymalnie
- **Skalowalność** - możliwość niezależnego rozwijania modułów

4.2 Techniczne

- **Wydaźność** - równoległe przetwarzanie na różnych serwerach
- **Bezpieczeństwo** - separacja krytycznych danych finansowych
- **Elastyczność** - wykorzystanie najlepszych cech każdego systemu

4.3 Edukacyjne

- **Praktyczne zastosowanie** - wszystkie kluczowe koncepty RBD w jednym projekcie
- **Różnorodność technik** - od podstawowych linked servers po zaawansowane transakcje rozproszone
- **Realistyczne scenariusze** - problemy i rozwiązania typowe dla środowisk produkcyjnych

5. Implementacja Rozwiązania

5.1 Konfiguracja środowiska

Infrastruktura techniczna:

- **SQL Server 2019+** - instancja główna (port 1433) i replika (port 1434)
- **Oracle Database 19c+** - trzy schematy (FINANCE_DB, REMOTE_DB1, REMOTE_DB2)
- **PostgreSQL 12+** - baza remarks_system ze schematem remarks_main
- **Dodatkowe komponenty** - ODBC drivers, OLE DB providers, MS DTC

Mechanizmy połączeń:

- **Linked Servers** - ORACLE_FINANCE, POSTGRES_REMARKS, MSSQL_REPLICA, EXCEL_DATA
- **Database Links** - połączenia między schematami Oracle
- **ODBC DSN** - konfiguracja PostgreSQL30 dla dostępu z SQL Server

5.2 Struktura danych

Rozkład tabel w systemie:

- **SQL Server:** 15 tabel (uczniowie, nauczyciele, oceny, frekwencja, słowniki)
- **Oracle:** 4 tabele główne (contracts, payments, contracts_remote, payment_summary)
- **PostgreSQL:** 1 tabela (remarks z metadanymi)

Strategia kluczy:

- Wspólne identyfikatory (studentId, teacherId, parentId) jako klucze logiczne
- Sekwencje Oracle z automatycznymi triggerami
- Identity columns w SQL Server
- Serial type w PostgreSQL

5.3 Mechanizmy integracji

Zapytania wielosystemowe:

- **OPENQUERY** - przetwarzanie zdalne w Oracle i PostgreSQL
- **OPENROWSET** - zapytania ad-hoc i eksport do Excel
- **Four-part naming** - standardowy dostęp przez linked servers

Widoki rozproszone:

- **vw_DistributedStudentData** - dane z wszystkich trzech systemów
- **vw_StudentFinancialInfo** - połączenie SQL Server z Oracle
- **vw_StudentCompleteInfo** - kompletny profil ucznia

6. Funkcjonalności Systemowe

6.1 Operacje CRUD

Zarządzanie uczniami:

- Pełny cykl życia ucznia (tworzenie, aktualizacja, usuwanie)
- Walidacja danych i integralności referencyjnej
- Obsługa relacji z rodzicami i klasami

Zarządzanie nauczycielami:

- Rejestracja nauczycieli z danymi kontaktowymi
- Przypisywanie do klas jako wychowawcy
- Zarządzanie planem lekcji

System ocen:

- Wystawianie ocen z wagami i komentarzami
- Kalkulacja średnich (zwykłych i ważonych)
- Statystyki per uczeń i per przedmiot

6.2 Operacje rozproszone

Integracja z PostgreSQL:

- Dodawanie i usuwanie uwag pedagogicznych
- Dynamiczne budowanie zapytań OPENQUERY
- Zabezpieczenie przed SQL injection

Integracja z Oracle:

- Zarządzanie kontraktami finansowymi
- Przetwarzanie płatności (w tym częściowych)
- Synchronizacja między schematami zdalnymi

6.3 Transakcje rozproszone

Operacje atomowe:

- Dodawanie ucznia z jednoczesnym utworzeniem kontraktu finansowego
- Przetwarzanie płatności z aktualizacją statusów w wielu systemach
- Obsługa błędów z automatycznym rollback

Mechanizmy spójności:

- Two-Phase Commit protocol
- MS DTC dla koordynacji transakcji
- Walidacja krzyżowa między systemami

6.4 Replikacja danych

Konfiguracja replikacji:

- Publikacja tabeli students z SQL Server głównego
- Subskrypcja push na instancji repliki
- Automatyczne agenty replikacji (Log Reader, Distribution Agent)

Monitoring i zarządzanie:

- Sprawdzanie statusu replikacji
- Obsługa konfliktów (Publisher wins strategy)
- Procedury recovery w przypadku awarii

7. Zaawansowane Funkcjonalności

7.1 Pakiet Oracle pkg_DistributedFinance

Komponenty pakietu:

- **Funkcje PIPELINED** - wydajne zwracanie zbiorów danych finansowych
- **Procedury autonomiczne** - tworzenie kontraktów z automatycznymi płatnościami
- **Raporty finansowe** - różne typy analiz (SUMMARY, DETAILED, OVERDUE)

Kalkulacje finansowe:

- Wyliczanie należności na podstawie czasu trwania kontraktu
- Obsługa płatności częściowych
- Synchronizacja podsumowań między schematami

7.2 System raportowy

Raporty wielosystemowe:

- Kompletny profil ucznia z danymi ze wszystkich systemów
- Statystyki frekwencji z analizą obecności
- Raporty finansowe z zagregowanymi danymi płatności

Eksport danych:

- Wieloarkuszowy eksport do Excel
- Automatyczne formatowanie nagłówków
- Obsługa błędów połączenia z poszczególnymi systemami

7.3 Optymalizacja wydajności

Strategie przetwarzania:

- Maksymalizacja obliczeń na serwerach źródłowych
- Minimalizacja transferu danych przez sieć
- Inteligentne planowanie zapytań (remote vs local)

Mechanizmy cachowania:

- Connection pooling dla linked servers

- Ponowne wykorzystanie połączeń ODBC
- Optymalne limity czasowe dla zapytań

8. Testowanie i Weryfikacja

8.1 Zakres testów

Testy integracji:

- Operacje CRUD na wszystkich systemach
- Integralność danych między systemami
- Funkcjonalność transakcji rozproszonych

Testy wydajnościowe:

- Obciążenie linked servers
- Latencja replikacji
- Throughput zapytań wielosystemowych

Testy odporności:

- Symulacja awarii poszczególnych systemów
- Test recovery po przerwach w połączeniu
- Weryfikacja mechanizmów rollback

8.2 Rezultaty testów

Funkcjonalność:

- Wszystkie operacje CRUD działają poprawnie
- Transakcje rozproszone zapewniają atomowość
- Replikacja synchronizuje dane w czasie rzeczywistym

Wydajność:

- Optymalne czasy odpowiedzi dla zapytań lokalnych
- Akceptowalna latencja dla operacji rozproszonych
- Efektywne wykorzystanie zasobów wszystkich systemów

Niezawodność:

- Graceful degradation przy awarii pojedynczych systemów
- Skuteczne mechanizmy recovery
- Spójność danych po operacjach rollback

9. Analiza Wyników

9.1 Realizacja założeń projektowych

Wypełnienie wszystkich wymagań:

- **Środowisko heterogeniczne** - SQL Server, Oracle, PostgreSQL
- **Linked Servers** - pełna integracja między systemami
- **Database Links** - symulacja rozproszonego środowiska Oracle
- **Zapytania ad-hoc** - OPENQUERY i OPENROWSET
- **Transakcje rozproszone** - MS DTC i Two-Phase Commit
- **Replikacja** - automatyczna synchronizacja danych

Dodatkowe funkcjonalności:

- Zaawansowane procedury Oracle w pakiecie pkg_DistributedFinance
- Kompletny system raportowy z eksportem do Excel
- Widoki rozproszone z inteligentną optymalizacją zapytań
- Mechanizmy monitoringu i diagnostyki

9.2 Architektura jako rozwiązanie wzorcowe

Podział funkcjonalny systemów:

- **SQL Server** - doskonały hub centralny z możliwościami integracji
- **Oracle** - efektywny system finansowy z zaawansowanymi procedurami

- **PostgreSQL** - elastyczne rozwiązanie dla danych tekstowych

Strategia komunikacji:

- **Synchroniczna integracja** - linked servers dla dostępu w czasie rzeczywistym
- **Asynchroniczna integracja** - replikacja dla wysokiej dostępności
- **Optymalizacja wydajności** - inteligentne przekazywanie przetwarzania

9.3 Wyzwania i rozwiązania

Wyzwania techniczne:

- **Heterogeniczność systemów** - różne typy danych i składnie SQL
- **Latencja sieci** - wpływ na wydajność zapytań rozproszonych
- **Koordinacja transakcji** - zapewnienie spójności ACID

Zastosowane rozwiązania:

- **Standaryzacja przez casting** - unifikacja typów danych
- **Przetwarzanie zdalne** - minimalizacja transferu danych
- **MS DTC** - profesjonalna koordynacja transakcji rozproszonych

Mechanizmy niezawodności:

- **Graceful degradation** - system kontynuuje pracę przy częściowych awariach
- **Szczegółowe logowanie** - kompletna informacja o błędach
- **Procedury recovery** - jasne kroki przywracania systemu

10. Wartość Edukacyjna

10.1 Demonstracja konceptów RBD

Strategie dystrybucji:

- **Podział funkcjonalny** - różne typy danych w wyspecjalizowanych systemach
- **Replikacja master-slave** - zapewnienie wysokiej dostępności
- **Federacja logiczna** - integracja bez konsolidacji fizycznej

Zarządzanie transakcjami:

- **Distributed ACID** - spójność między systemami
- **Protokół 2PC** - atomowość operacji rozproszonych
- **Poziomy izolacji** - równowaga między wydajnością a spójnością

Przetwarzanie zapytań:

- **Optymalizacja kosztowa** - wybór między przetwarzaniem lokalnym a zdalnym
- **Strategie JOIN** - planowanie zapytań z uwzględnieniem sieci
- **Materializacja wyników** - efektywny transfer danych

10.2 Aplikowalność w środowiskach produkcyjnych

Scenariusze enterprise:

- **Integracja systemów legacy** - praca z istniejącymi rozwiązaniami
- **Migracja stopniowa** - podejście krok po kroku do modernizacji
- **Środowisko wieloendorowe** - zarządzanie różnymi technologiami

Compliance i governance:

- **Suwerenność danych** - różne systemy dla różnych typów danych
- **Wymagania regulacyjne** - wyspecjalizowane systemy dla wrażliwych danych
- **Ślady audytu** - rozproszone logowanie i monitoring

11. Możliwości Rozwoju

Dodatkowe moduły:

- **System biblioteki** - katalog książek w PostgreSQL
- **Zarządzanie laboratoriami** - sprzęt w dedykowanym schemacie Oracle

- **Kalendarz wydarzeń** - integracja z systemami NoSQL

Zaawansowana analityka:

- **Machine learning** - modele predykcyjne dla wyników uczniów
- **Automatyczne raporty** - integracja z SQL Server Reporting Services

12. Wnioski Końcowe

12.1 Sukces implementacji

Projekt System Rozproszonej Bazy Danych dla prywatnej szkoły podstawowej stanowi **kompleksową demonstrację** wszystkich kluczowych aspektów rozproszonych baz danych. System łączy **teoretyczne koncepty** z **praktyczną implementacją**, tworząc rozwiązanie o jakości produkcyjnej.

Główne osiągnięcia:

- **Pełna realizacja wymagań** - wszystkie założenia projektu zostały zrealizowane
- **Architektura klasy enterprise** - zastosowanie wzorców używanych w środowiskach produkcyjnych
- **Kompleksowa strategia testowa** - weryfikacja wszystkich aspektów systemu
- **Szczegółowa dokumentacja** - wsparcie dla dalszego rozwoju i utrzymania

12.2 Innowacje projektu

Elementy wykraczające poza podstawowe wymagania:

- **Pakiet PL/SQL** - zaawansowane procedury finansowe z funkcjami PIPELINED
- **Sophisticated error handling** - zaawansowana obsługa błędów w środowisku rozproszonym
- **Optymalizacje wydajności** - techniki minimalizujące latencję sieci
- **Realistyczne scenariusze** - uwzględnienie problemów środowisk produkcyjnych

12.3 Rekomendacje dla dalszego rozwoju

Środowiska produkcyjne:

- **Wzmocnienie bezpieczeństwa** - szyfrowanie i zaawansowana autoryzacja
- **Narzędzia monitoringu** - kompleksowe śledzenie zdrowia systemu
- **Strategie backup** - procedury spójnych kopii zapasowych między systemami
- **Dokumentacja operacyjna** - żywa dokumentacja dla zespołów eksploatacyjnych

Rozwój edukacyjny:

- **Integracja NoSQL** - dodanie MongoDB lub Cassandra jako kolejnego węzła
- **Wdrożenie chmurowe** - dystrybucja w środowisku Azure/AWS
- **Architektura mikroservisów** - dekompozycja zorientowana na usługi
- **Wzorce event-driven** - kolejki wiadomości i streaming zdarzeń

12.4 Wartość dla nauki i praktyki

System stanowi **solidną podstawę** dla zrozumienia konceptów rozproszonych baz danych oraz **praktyczne odniesienie** dla implementacji podobnych rozwiązań w środowiskach enterprise. Projekt demonstruje, że profesjonalne systemy RBD mogą być zarówno **funkcyjnie bogate**, jak i **edukacyjnie przejrzyste**.

Długoterminowa wartość:

- **Wzorzec architektoniczny** - referencyjne rozwiązanie dla podobnych projektów
- **Baza wiedzy** - zbiór sprawdzonych technik i rozwiązań
- **Platforma rozwoju** - fundament dla dalszych innowacji
- **Narzędzie edukacyjne** - praktyczny przykład teorii RBD w działaniu

Podsumowanie: Projekt System Rozproszonej Bazy Danych dla prywatnej szkoły podstawowej jest **pełnym sukcesem** realizacji założeń edukacyjnych i technicznych. Stanowi przykład **profesjonalnego podejścia** do projektowania systemów rozproszonych, łącząc **solidne podstawy teoretyczne** z **praktyczną implementacją** na poziomie enterprise.