

Hurtownie danych i przetwarzanie analityczne

Tematy projektów

1. Ocena funkcjonalności partycjonowania tabel i indeksów w SQL Server 2008 (3 osoby)

Wymagania

- opisać rodzaje partycjonowania tabel uwzględniając polecenia SQL, zilustrować własnymi przykładami opartymi o schemat benchmarku TPC-DS
- opisać rodzaje partycjonowania tabel uwzględniając polecenia SQL, zilustrować własnymi przykładami, odnieść do Oracle
- opisać DML na tabelach partycjonowanych
- dokonać analizy zapytań analitycznych z benchmarku TPC-DS na tabelach partycjonowanych i niepartycjonowanych, przedstawić i porównać plany wykonania zapytań
- odnieść ww. funkcjonalność do Oracle

2. Ocena funkcjonalności partycjonowania tabel i indeksów w IBM DB2 v.9.7 (3 osoby)

Wymagania

- opisać rodzaje partycjonowania tabel uwzględniając polecenia SQL, zilustrować własnymi przykładami opartymi o schemat benchmarku TPC-DS
- opisać rodzaje partycjonowania tabel uwzględniając polecenia SQL, zilustrować własnymi przykładami, odnieść do Oracle
- opisać DML na tabelach partycjonowanych
- dokonać analizy zapytań analitycznych z benchmarku TPC-DS na tabelach partycjonowanych i niepartycjonowanych, przedstawić i porównać plany wykonania zapytań
- odnieść ww. funkcjonalność do Oracle

3. Ocena funkcjonalności "indexed views" w SQL Server 2008 (3 osoby)

Wymagania

- przetestować funkcjonalność indexed views uwzględniając polecenia SQL, zilustrować własnymi przykładami opartymi o schemat benchmarku TPC-DS
- przetestować mechanizmy uaktualniania indexed views
- dokonać analizy zapytań analitycznych z benchmarku TPC-DS na tabelach i indexed views w kontekście przepisywania zapytań, przedstawić i porównać plany wykonania zapytań
- odnieść ww. funkcjonalność do Oracle

4. Ocena funkcjonalności "materialized query tables" w IBM DB2 v.9.7 (3 osoby)

Wymagania

- przetestować funkcjonalność MQT uwzględniając polecenia SQL, zilustrować własnymi przykładami opartymi o schemat benchmarku TPC-DS
- przetestować mechanizmy uaktualniania MQT
- dokonać analizy zapytań analitycznych z benchmarku TPC-DS na tabelach i MQT w kontekście przepisywania zapytań, przedstawić i porównać plany wykonania zapytań
- odnieść ww. funkcjonalność do Oracle

5. Ocena jakości kompresji danych i indeksów w SQL Server 2008 (3 osoby)

Wymagania

- dokonać eksperymentalnej oceny efektywności kompresji danych i indeksów uwzględniając polecenia SQL, zilustrować własnymi przykładami opartymi o schemat benchmarku TPC-DS
- porównać efektywność wykonywania zapytań z benchmarku TPC-DS na danych skompresowanych i nieskompresowanych
- odnieść ww. funkcjonalność do Oracle

6. Ocena jakości kompresji danych i indeksów w IBM DB2 v.9.7 (3 osoby)

Wymagania

- dokonać eksperymentalnej oceny efektywności kompresji danych i indeksów uwzględniając polecenia SQL, zilustrować własnymi przykładami opartymi o schemat benchmarku TPC-DS
- porównać efektywność wykonywania zapytań z benchmarku TPC-DS na danych skompresowanych i nieskompresowanych
- odnieść ww. funkcjonalność do Oracle

7. Analiza danych w implementacji MOLAP - Hyperion/Oracle (3 osoby)

Wymagania

- źródła danych: pliki csv lub xml
- technologia implementacyjna: Hyperion/Oracle
- sprawozdanie w formie tutorialu
- zaimplementować "kostkę" (definicje miar, wymiarów) umożliwiającą wykonanie analiz funduszy inwestycyjnych akcyjnych
- wczytać dane z plików
- zaimplementować analizy: ranking funduszy stabilnego wzrostu za rok 20011, analiza trendu funduszy stabilnego wzrostu w każdym z miesięcy roku 2011, predykcja trendu na rok 2012

8. Analiza danych w implementacji MOLAP - SQL Server 2008 (3 osoby)

Wymagania

- źródła danych: pliki csv lub xml
- technologia implementacyjna: SQL Server 2008 Analysis Services
- sprawozdanie w formie tutorialu
- zaimplementować "kostkę" (definicje miar, wymiarów) umożliwiającą wykonanie analiz kursów akcji
- wczytać dane z plików

- zaimplementować analizy: średnie miesięczne kursy akcji 5 wybranych spółek co najmniej 3 różnych sektorów za rok 2011, sumaryczna wartość akcji spółek w podziale na sektory w wybranym dniu, predykcja trendu wybranej spółki na czerwiec 2012

9. Porównanie efektywności struktur fizycznych w IBM DB2 v.9.7 (3 osoby)

Wymagania

- porównać w oparciu o benchmark TPC-DS efektywność wykonywania zapytań dla następujących scenariuszy testowych
 - utworzono standardowe indeksy na atrybutach połączeniowych i występujących w klauzuli WHERE
 - zdefiniowano multidimensional clusters o strukturach odpowiednich do zapytań (dobrać te struktury)
 - zdefiniowano clustered indexes o strukturach odpowiednich do zapytań (dobrać te struktury)
- przedstawić plany wykonania zapytań wraz z kosztami dla ww. scenariuszy testowych

10. Predykcja trendów w Oracle w oparciu o metody "nieliniowe" (3 osoby)

Wymagania

- dokonać analizy możliwości predykcji trendów w oparciu o różne metody "nieliniowe" w oparciu o dane dot. funduszy stabilnego wzrostu na rok 2012
- przedstawić implementacje tych analiz