

SZKOLENIE ŚREDNIO ZAAWANSOWANE

Projektowanie hurtowni danych z wykorzystaniem pakietu Pentaho Business Intelligence

PENTAHO/DW

Czas trwania: 5 dni

Cele szkolenia

- Zdobyć przekrojową wiedzę dotyczącą projektowania, implementowania, monitorowania, uruchamiania, strojenia procesów i elementów hurtowni danych
- Poznanie ogólnych założeń hurtowni danych tj. najpopularniejsze schematy relacyjnych hurtowni danych
- Wybór właściwego zestawu narzędzi i technik dla swoich projektów
- Skupienie się na stosie produktowym wybudowanym wokół Pentaho Business Intelligence Suite

Zalety

- Program obejmuje zarówno ogólne wprowadzenie w tematykę ETL, DWH, OLAP, jak i całościowe przedstawienie stosu produktowego Pentaho Business Intelligence
- Szkolenie jest unikalne, gdyż tematyka poruszana w jego trakcie nie jest wyczerpująco ujęta w dostępnej literaturze, a wiedza na ten temat jest mocno rozproszona
- Program jest ciągle uaktualniany ze względu na szybki rozwój rozwiązań, których dotyczy szkolenie

Dla kogo?

- Szkolenie jest adresowane do programistów, architektów oraz administratorów aplikacji, którzy chcą tworzyć lub utrzymywać hurtownie danych (DWH) z wykorzystaniem Pentaho Business Intelligence Suite

Wymagania

- Od uczestników wymagana jest podstawowa znajomość baz danych, podstawowa umiejętność programowania w języku Java



Program

1. Wprowadzenie do hurtowni danych i platformy Pentaho BI Suite
 - a. OLTP, OLAP, bazy danych, hurtownie danych, data marty
 - b. ROLAP, MOLAP, HOLAP
 - c. Normalizacja, agregacja, fakty, wymiary
 - d. SQL, MDX, XML/A
 - e. ETL i integracja offline
 - f. BigData, BigTable, NoSQL, nierelacyjne hurtownie danych
 - g. Architektura i składowe Pentaho BI Suite
 - h. Business intelligence a hurtownie danych
2. Fizyczne struktury danych w hurtowni danych
 - a. Tabele faktów i tabele wymiarów
 - b. Indeksy, widoki, widoki zmaterializowane
 - c. ROLAP w szczegółach: schemat gwiazdy, płatka śniegu, konstelacji faktów
 - d. Testy wydajności ROLAP: The TPC Benchmark H
 - e. ROLAP vs MOLAP
 - f. Metody optymalizacji szczególnie przydatne w przypadku hurtowni danych
 - g. Wykorzystanie Pentaho Aggregate Designer do optymalizacji hurtowni danych
 - h. Znaczenie wymiaru czasu w hurtowniach danych
 - i. Dziedziczne hurtownie danych a dedykowane struktury danych np.: GeoSpatial DWH
3. Zasilanie hurtowni danych z wykorzystaniem Pentaho Data Integration (PDI)
 - a. ETL
 - Ekstrakcja danych
 - Transformacja, oczyszczenie i uzupełnianie danych
 - Ładowanie
 - Wyznaczniki i ocena jakości danych
 - Staging
 - Hurtownie czasu rzeczywistego
 - Kwestie wydajnościowe w kontekście procesów ETL
 - Przegląd narzędzi ETL dostępnych na rynku
 - b. Pentaho Data Integration
 - Architektura
 - Składowe
 - c. Praca ze Spoon
 - Instalacja, uruchomienie, Look & Feel (perspektywy, zakładki etc.)
 - Zmienne
 - Hopy
 - Praca z repozytorium i plikami XML
 - Współdzielenie obiektów tj. połączenia z bazami danych, partycje, klastry
 - Czytanie i zapisywanie danych z różnych źródeł danych
 - Transformacje
 - Walidacja danych
 - Uzupełnianie danych



- Transformacja danych
- Optymalizacja
- Wykorzystanie własnego kodu w transformatach
- Utilities
- Monitorowanie wydajności i analiza logów
- Wersjonowanie i praca grupowa
- Zadania
 - Zadania (kjb) a transformacje (ktr)
 - Tworzenie złożonych zadań
 - Wykorzystanie własnego kodu w zadaniach
 - Przepływ pracy
 - Praca z plikami
 - Monitorowanie wydajności i analiza logów
 - Wersjonowanie i praca grupowa

d. Praca z Kitchen i Pan

- Uruchamianie zadań i transformat z plików i z repozytorium
- Scheduling
- Obsługa błędów
- Przekierowanie wejścia/wyjścia

e. Praca z Cartle

4. Analityka z wykorzystaniem OLAP oraz Pentaho Analysis Services (Mondrian)

a. Definiowanie kostek OLAP na hurtowni danych

b. Pentaho Schema Workbench (PSW)

c. Model logiczny i model fizyczny

- Kostki
- Miary
- Wymiary, hierarchie, levele
- Tabele
- Relacje
- Agregaty
- Wyrażenia

d. MDX w szczegółach

e. Operacje Slice, Dice, Drill na kostkach

f. Optymalizacja na potrzeby OLAP

5. Raportowanie danych zgromadzonych w hurtowni danych z wykorzystaniem Pentaho Report Designer (PRD)

a. Definiowanie raportów za pomocą PRD

- Źródła danych
- Modelowanie zapytań
- Wydobywanie danych
- Zawężanie i filtrowanie wyników
- Wizualizacja danych na raporcie

b. Osadzanie raportów

c. Eksport do PDF, HTML, RTF etc.



- d. Report Wizard
- 6. Raportowanie i analiza ad-hoc
 - a. Pentaho BI Platform/Portal (BA Platform)
 - b. Pentaho Interactive Reporting (PIR)
 - c. Pentaho Metadata Editor (PME)
 - d. Pentaho Analyzer (PAZ)
- 7. Kokpity i dashboardy managerskie na danych zgromadzonych w hurtowni danych z wykorzystaniem Pentaho Dashboard Designer (PDD)
 - a. Definiowanie dashboardów za pomocą PDD
 - Praca z wykresami - simple bar, line, area, pie, dial charts
 - Praca z tabelami
 - Praca z raportami
 - Parametry
 - Szablony
 - b. Osadzanie dashboardów
- 8. Eksploracja danych zgromadzonych w hurtowni danych z wykorzystaniem Pentaho Data Mining (WEKA)
 - a. Architektura
 - b. Praca z Explorerem
 - c. Preprocessing i praca z danymi
 - Format danych ARFF
 - Przygotowanie danych do analizy
 - Odpowiedni dobór atrybutów np.: korelacja atrybutów a wyniki eksploracji danych etc.
 - Filtrowanie i rodzaje filtrów w WEKA np.: filtrowanie, dyskretyzacja, normalizacja etc.
 - Wizualizacja
 - Przetwarzanie dużych zbiorów danych, ograniczenia JVM 32bit
 - Przetwarzanie strumieni oraz uczenie przyrostowe
 - d. Eksploracja danych
 - Klasyfikacja
 - Grupowanie
 - Odkrywanie reguł asocjacyjnych
 - Ograniczanie i transformacja przestrzeni atrybutów
 - Pozostałe algorytmy i techniki eksploracji danych dostępne w WEKA
 - e. Rozbudowa możliwości WEKA
 - f. Wykorzystanie możliwości w połączeniu z innymi produktami Pentaho
- 9. Pentaho Mobile BI

