

Hurtownie danych – projekt

SZBD: **MS SQL 2016**

Oprogramowanie: **Microsoft SQL Server Management Studio, Integration Services (Visual Studio)**

Jeżeli nie potrafisz rozwiązać zadania, to próbuj podać chociaż częściowe rozwiązanie lub uzasadnienie przyczyny braku rozwiązania.

Proces tworzenia hurtowni danych powinien być poprzedzony zrozumieniem „potrzeb biznesu” oraz rzeczywistości (dziedziny problemowej) reprezentowanej przez dostępne zasoby danych. Realizacja poniższego zadania ma uzmysłowić występujące problemy w określonym (wybranym) wycinku rzeczywistości, a następnie umożliwić zidentyfikowanie (określenie) potrzeb, celu i możliwości analiz biznesowych, by wspierać procesy decyzyjne (podejmowanie właściwych decyzji biznesowych).

ZADANIE 1 – PROJEKT HD – PROPOZYCJA TEMATU

Proszę przygotować zakres realizacji projektu zgodnie z poniższą specyfikacją oraz przedyskutować propozycję projektu z osobą prowadzącą zajęcia. Poczynione uzgodnienia zarejestrować w formie wniosków

ZAKRES OPRACOWANIA PROJEKTU HD – CZ. 1.

1. Tytuł projektu
2. Charakterystyka dziedziny problemowej
3. Krótki opis obszaru analizy
4. Problemy i potrzeby
5. Cel przedsięwzięcia
 - 5.1. Oczekiwania
 - 5.2. Zakres analizy – badane aspekty (należy wskazać wielowymiarowe aspekty prowadzonej analizy)
6. Źródła danych (lokalizacja, format, dostępność)

ZAKRES OPRACOWANIA PROJEKTU HD – CZ. 2.

1. Wstępna analiza źródeł danych

L.p.	Plik	Typ	Liczba rekordów	Rozmiar [MB]	Opis
1.					
2.					
...					

Sprawozdanie bez wniosków końcowych nie będzie sprawdzane i tym samym ocena jest negatywna!

2. Analiza danych

Plik:				
L.p.	Atrybut	Typ danych	Zakres wartości	Uwagi – ocena jakości danych
1.				
2.				
3.				
...				

3. Ocena przydatności danych w pliku do tworzenia hurtowni danych

L.p.	Plik	Ocena jakości danych
1.		
2.		
3.		
...		

4. Definicja typów encji/klas (wraz z własnościami) oraz związków pomiędzy nimi
5. Propozycja wymiarów, hierarchii miar (w tym nieaddytywnych)
6. Diagram związków encji lub diagram klas – model danych utworzony na podstawie danych zgromadzonych w plikach

ZAKRES OPRACOWANIA PROJEKTU HD – CZ. 3.

1. Utworzyć bazę danych zgodnie z zaproponowanym konceptualnym modelem danych
2. Utworzone tabele w poprzednim punkcie wypełnić danymi zgodnie z ustalonymi założeniami projektowymi

Wymagania dotyczące projektu zaliczeniowego

WYMAGANIA DOTYCZĄCE PROJEKTU ZALICZENIOWEGO

Projekt końcowy powinien zawierać przynajmniej jedną kostkę Analysis Services, dotyczącą danych wybranych i przetworzonych przez studenta przy użyciu Integration Services. Utworzona kostka powinna:

- zawierać co najmniej 5 wymiarów, w tym co najmniej dwa o strukturze hierarchicznej (np. czas, miejsce, itp.),
- posiadać co najmniej 3 miary, w tym jedną nieaddytywną,
- odpowiadająca jej tabela faktów powinna posiadać co najmniej 10 000 rekordów.

ELEMENTY PROJEKTU MAJĄCE WPŁYW NA OCENĘ

PRZYGOTOWANIE KRÓTKIEGO OPISU BAZY DANYCH

- prezentacja bazy danych – krótki opis danych źródłowych.
- propozycja wymiarów, hierarchii i miar.

PROCES ETL

Przy ocenie będą brane następujące elementy pakietu(ów):

- właściwa struktura procesu ETL (odpowiednie rozbiecie procesu ETL na zadania/pakiety, dobrze dobrane nazwy poszczególnych zadań, wprowadzona automatyzacja itp.),
- stabilność i prawidłowe, bezbłędne wykonanie,
- złożoność przeprowadzonych operacji,
- dokumentacja powinna zawierać krótki opis dotyczący każdego zadania, które pozwoli zorientować się jaki jest jego cel (przykładowo wystarczy napisać: zadanie Z kopiuje dane z tabeli X i Y do tabeli T dokonując przy okazji denormalizacji).

KOSTKA

Przy ocenie będą brane następujące elementy kostki:

- prawidłowa struktura kostki – model kostki powinien analitykowi na intuicyjne i łatwe korzystanie z danych,
- miary kalkulowane,
- dokumentacja, która powinna zawierać krótki opis wszystkich wymiarów, wszystkich ich atrybutów oraz wszystkich miar.

PREZENTACJA

Prezentacja powinna zawierać 4-8 slajdów (trwać ok. 8 minut) i wyjaśniać jakie dane są przedmiotem analizy. Prezentacja powinna być zakończona, krótką demonstracją, która pokaże najciekawsze związki między danymi znajdującymi się w kostce.

Uwaga. Projekt będzie ostatecznie zaliczony po złożeniu pisemnego sprawozdania zawierającego opisy poszczególnych etapów pracy.