

Hurtownie danych – Projekt

Proces tworzenia hurtowni danych powinien być poprzedzony zrozumieniem „potrzeb biznesu” oraz rzeczywistości (dziedziny problemowej) reprezentowanej przez dostępne zasoby danych. Realizacja poniższego zadania ma uzmysłwić 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).

Projekt końcowy powinien zawierać przynajmniej jedną kostkę Analysis Services, dotyczącą **danych pochodzących z co najmniej dwóch źródeł danych** (jeden zbiór danych ma być wybrany spośród przedstawionych na zajęciach laboratoryjnych, a drugi – dowolnie wybrany przez studenta).

Utworzona kostka powinna:

- zawierać **przynajmniej 5 wymiarów**, w tym co najmniej dwa o strukturze hierarchicznej (np. czas, miejsce, itp)
- posiadać **co najmniej 3 miary**, w tym min. jedną nieaddytywną
- odpowiadająca jej tabela faktów powinna posiadać **co najmniej 10000 rekordów**.

Etap I – 19.05.2025 r.

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

1. Tytuł projektu
2. Charakterystyka dziedziny problemowej, krótki opis obszaru analizy, problemy i potrzeby
3. Cel przedsięwzięcia (oczekiwania) oraz zakres analizy – badane aspekty
4. Źródła danych (lokalizacja, format, dostępność), wstępna analiza źródeł danych

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

5. Profilowanie danych (analiza jakości danych oraz ich przydatności w projekcie)

Lp.	Atrybut	Typ danych	Zakres wartości	Uwagi – ocena jakości danych
1.				
2.				

6. Definicja typów encji/klas (wraz z własnościami) oraz związków pomiędzy nimi, diagram klas (proponując wymiary, hierarchii, miar addytywnych i nieaddytywnych)
7. **Min. 10 wielowymiarowych zestawień, które zostaną utworzone po wdrożeniu kostki**
8. Implementacja bazy danych zgodnie z zaproponowanym konceptualnym modelem danych

Wnioski:

Etap II – 26.05.2025 r.

Proces ETL

1. Utworzone tabele w poprzednim punkcie wypełnić danymi zgodnie z ustalonymi założeniami projektowymi wykorzystując zapytania SQL, dostępne narzędzia w Integration Services lub inne dostępne narzędzia.

Przy ocenie będą brane następujące elementy pakietu(ów):

- właściwa struktura procesu ETL – odpowiednie rozbitcie procesu ETL na zadania/pakiety, dobrze dobrane nazwy poszczególnych zadań
- **złożoność przeprowadzonych operacji** – przykładowo, jeżeli dane źródłowe już są w pełni zdenormalizowane proszę nie spodziewać się maksymalnej liczby punktów za ten element
- **wprowadzona automatyzacja np. przyrostowe zasilanie hurtowni, automatyzacja czyszczenia danych (metody fuzzy), obsługa błędów, itp.**
- stabilność i prawidłowe, bezbłędne wykonanie
- dokumentacja powinna zawierać krótki opis dotyczący każdego zadania, które pozwoli zorientować się, jaki jest jego cel (np. zadanie Z kopiuje dane z tabeli X i Y do tabeli T dokonując denormalizacji) oraz mapę logiczną procesu ETL.

Wnioski:

Etap III – 02.06.2025 r.

Kostka:

1. Przygotować projekt kostki, edytować wymiary, dodać miary kalkulowane. Przygotować zestawienia z etapu I (p. 7) oraz pokazać inne ciekawe zależności w analizowanych danych (analiza w głębi, a nie tylko tabele przestawne).

Przy ocenie będą brane następujące elementy kostki:

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

Wnioski:

Etap IV – 16.06.2025 r.

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.