

Projekt. Proces ETL

Źródło danych

1. Przedmiotem projektu jest aplikacja, która pozwoli na przeprowadzenie procesu ETL na który składa się:
 - a. pozyskanie danych ze źródeł zewnętrznych,
 - b. poddanie danych odpowiednim przekształceniom,
 - c. zasilenie danymi bazy danych.
2. Celem działania aplikacji będzie pobranie z wybranego przez zespół projektowy serwisu internetowego wszystkich opinii o wskazanym produkcie lub wybranej usłudze, wszystkich ofert dotyczących sprzedaży/wynajmu produktów spełniających określone kryteria lub wszystkich ofert pracy.
3. Dla każdej opinii/oferty/ogłoszenia należy pobrać jak najwięcej atrybutów. W przypadku, gdy brakuje którejś wartości pole należy pozostawić puste. Dla opinii dodatkowo należy przechować informację o produkcie/usłudze której dotyczy dana opinia.

Zadania do wykonania

1. Zadanie składa się z trzech części:
 - a. Aplikacji
 - b. Dokumentacji technicznej
 - c. Instrukcji obsługi

Aplikacja

2. Aplikację należy stworzyć w dowolnej, wybranej przez siebie technologii.
3. Można użyć dowolnego wybranego przez siebie systemu zarządzania bazą danych.
4. Aplikacja powinna mieć możliwość przeprowadzenia całego procesu ETL na raz (np. przycisk ETL) jak również przeprowadzenia każdej jego części oddzielnie – jedna po drugiej (np. trzy kolejne przyciski E, T, L).
5. W przypadku wykonywania procesu ETL etapami wykonanie kolejnego etapu powinno być możliwe tylko po uprzednim wykonaniu wcześniejszego etapu (np. nie powinno się dać wykonać operacji Transform, jeśli wcześniej nie została wykonana operacja Extract).
6. Po przeprowadzeniu pełnego procesu ETL lub jego ostatniego etapu (Load) produkty uboczne procesu (np. pobrane pliki) powinny być usuwane (tym samym po wykonaniu Load nie powinno być możliwości wykonania operacji Transform bez ponownego wykonania Extract).
7. Po przeprowadzeniu pełnego procesu ETL jak również po każdej z jego składowych powinny być wyświetlane informacje statystyczne np. ile plików zostało pobranych, ile rekordów zostało załadowanych do bazy danych.

8. Dane w bazie nie powinny się dublować – jeśli dana informacja zastała dodana do bazy przy poprzednim uruchomieniu procesu ETL nie powinna zostać dodana przy kolejnych.
9. Aplikacja powinna pozwalać na aktualizację danych przechowywanych w bazie.
10. Aplikacja powinna mieć przycisk pozwalający na wyczyszczenie bazy danych tak, aby możliwe było zaobserwowanie procesu ponownego zasilenia bazy danymi.
11. Aplikacja powinna umożliwiać wyświetlenie danych z bazy w sposób umożliwiający ich filtrowanie i analizę.
12. Dodatkowo aplikacja powinna mieć możliwość eksportu danych do pliku .csv (osobny przycisk lub opcja) i oddzielnych plików tekstowych dla każdej opinii/ogłoszenia/oferty.
13. Projekt powinien być uniwersalny – powinien nadal działać, gdy na stronie zostaną dodane nowe opinie/ogłoszenia/oferty lub jakaś opinia/ogłoszenie/oferta zostanie usunięte.
14. W przypadku wątpliwości co do spodziewanego sposobu działania aplikacji należy potraktować prowadzącego ćwiczenia jako klienta. Spotkania konsultacyjne i konsultacje należy traktować jako możliwość kontaktu (jedyne) z klientem w sprawie projektu. Nie przewiduje się mailowych konsultacji projektu.

Dokumentacja techniczna

15. Dokumentacja techniczna powinna zawierać:
 - a. Nazwy i wersje użytych technologii (języki programowania, system zarządzania bazą danych itp.).
 - b. Informacje na temat środowiska (minimalne wymagania sprzętowe, biblioteki itp.) potrzebnego do instalacji aplikacji.
 - c. Linki do oprogramowania, które to środowisko tworzą.
 - d. Instrukcje instalacji aplikacji
 - e. Model danych użyty w projekcie.
 - f. (jeśli obiektowo) nazwy klas, tabelkę z listą i opisem atrybutów, tabelkę z listą i opisem (wartość zwracana, przyjmowane parametry, opis działania) funkcji składowych (z konstruktorami i destruktorami włącznie);
 - g. (jeśli strukturalnie): tabelkę z listą i opisem zmiennych, tabelkę z listą i opisem (wartość zwracana, przyjmowane parametry, opis działania) funkcji w podziale na pliki.
16. Dokumentacja techniczna może zostać wygenerowana na podstawie kodu aplikacji.

Instrukcja obsługi

17. Instrukcja obsługi aplikacji powinna zawierać:
 - a. Instrukcję uruchomienia aplikacji
 - b. Opis funkcjonalności aplikacji
 - c. Opis scenariuszy użycia aplikacji
 - d. Opis menu i widoków okna aplikacji
18. Oprócz opisów instrukcja obsługi powinna zawierać zrzuty ekranu wizualizujące wygląd i działanie aplikacji.

Dodatkowe wymagania

1. Przez platformę e-learningową należy we wskazanym terminie przesłać:
 - a. Link do publicznego repozytorium Git z kodem aplikacji!
 - b. Link do publicznego repozytorium Git z kodem dokumentacji jeśli ma ona postać odrębnej aplikacji
 - c. Dokumentację techniczną (jeśli ma ona postać dokumentu tekstowego) oraz instrukcję obsługi złożone w jeden spójny dokument tekstowy
2. Jako stronę tytułową dokumentacji należy wykorzystać plik dostępny na platformie e-learningowej. Tabelkę należy wypełnić danymi członków zespołu podając dodatkowo za jakie czynności odpowiadał każdy członek zespołu i jaki to stanowiło udział procentowy we wszystkich pracach nad projektem.
3. Plikowi z dokumentacjami należy nadać nazwę według wzoru:
NumerGrupyProjektowej _Projekt.[doc | docx]

Zaliczenie projektu

1. Zespoły zaliczają projekt w trakcie zajęć konsultacyjnych w wyznaczonych indywidualnie terminach.
2. Na zaliczenie projektu zespół powinien zgłosić się w pełnym składzie.
3. Na zaliczenie projektu zespół powinien przynieść sprzęt z zainstalowaną aplikacją jeśli jest to aplikacja desktopowa lub postawiona na lokalnym serwerze.
4. Na zaliczenie projektu należy się zgłosić z wydrukowanym dokumentem zawierającym dokumentację techniczną (jeśli ma ona postać dokumentu, a nie aplikacji) i instrukcję obsługi.

