

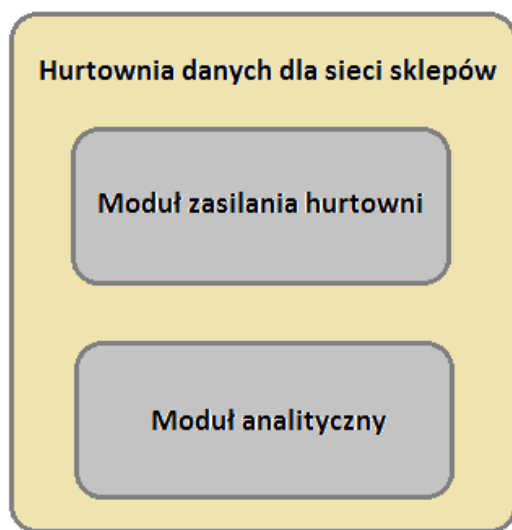
# Dokumentacja

Zakładam, że:

1. Przy każdym uruchomieniu programu w odpowiednim katalogu dostępne są jakieś pliki, które nie były jeszcze archiwizowane
2. Pliki te mają poprawną nazwę <rok><miesiąc><dzien>s<numer\_sklepu>.xlsx
3. W pliku są tabele z następującymi atrybutami: data, godzina, product\_it, ilosc i sklep\_id

Hurtownia danych dla sieci sklepów składać się będzie z 2 modułów: modułu zasilania hurtowni i modułu analitycznego.

Schemat hurtowni:



Moduł zasilania hurtowni:

1. Wczytuję do zbioru Lab6.files nazwy plików znajdujących się w katalogu "/folders/myfolders/dane\_lab6/"
2. Do tymczasowego zbioru Lab6.content wczytuję dane z plików, których nazwy znajdują się w Lab6.files, ale nie są one na liście plików które były już wcześniej zarchiwizowane – (powoduje to, że nie wczytuję 2 razy tych samych danych)
3. Sprawdzam poprawność danych znajdujących się w Lab6.content :
  - zgodność daty w obserwacji z datą w nazwie pliku
  - zgodność id produktu ze zbiorem Lab6.produkty (wcześniej wczytywane są do niego dane z pliku: „/folders/myfolders/produkty/spisProduktow.xlsx”)
  - racjonalna ilość (większa od 0 i mniejsza od 1 000 000)
  - zgodność id sklepu z nazwą pliku

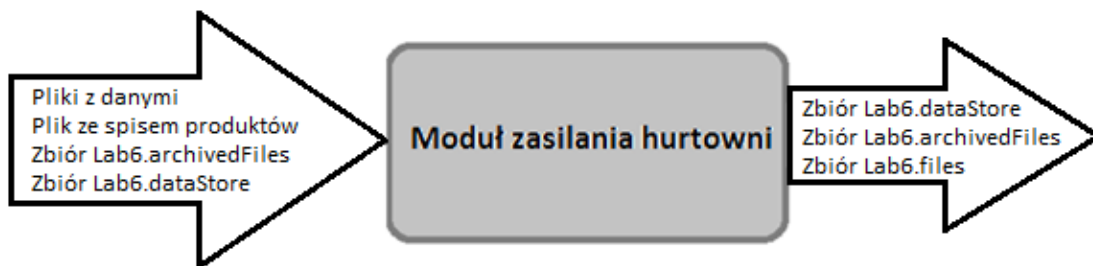
Jeśli wykryty zostanie któryś z powyższych błędów, w logu wypisywany jest stosowny komunikat.

4. Z Lab6.content usuwam dane, które pochodzą z plików, w których wykryto jakieś błędy.
5. Pozostałe dane (pochodzące z poprawnych plików) dodaję do zbioru Lab6.dataStore – w zbiorze Lab6.dataStore ostatecznie znajdują się wszystkie poprawne dane, które zostały wczytane do hurtowni.

6. W zbiorze Lab6.archivedFiles pozostają nazwy wszystkich plików, z których dane znajdują się w zbiorze Lab6.dataStore.

7. W zbiorze Lab6.files pozostają nazwy plików, które zostały wczytane na początku z podanego katalogu z danymi wraz z oznaczeniami (czy plik był już wcześniej archiwizowany- więc nie należy robić tego po raz drugi, czy plik został pomyślnie wczytany, jeśli nie to ustawiane są flagi błędów, które wystąpiły).

Schemat wejść i wyjść modułu:



#### Moduł analityczny:

Wykonywane są w nim następujące analizy:

1. Liczba produktów sprzedanych w poszczególne dni.
2. Liczba produktów sprzedanych w ostatnim dniu (bierzemy pod uwagę najpóźniejszą datę znajdującą się w zbiorze Lab6.dataStore) w zależności od sklepu.
3. Liczba produktów sprzedanych w poszczególne dni w poszczególnych sklepach.
4. Liczba produktów sprzedanych przed południem i po południu.
5. Liczba produktów sprzedanych przed południem i po południu w zależności od sklepu.
6. Liczba produktów sprzedanych przed południem i po południu w ostatnim dniu.
7. Liczba sprzedanych produktów o danym id.
8. Liczba sprzedanych produktów o danym id w zależności od sklepu.
9. Liczba produktów o danym id sprzedanych w ostatnim dniu.
10. Liczba produktów o danym id sprzedanych w ostatnim dniu w zależności od sklepu.
11. Liczba sprzedanych produktów w zależności od sklepu.

Analizy przeprowadzane są na zbiorze Lab6.dataStore., a wyniki wyświetlane są na ekranie w postaci wykresów słupkowych oraz zapisywane z zbiorach Lab8.analiza<nr>, gdzie nr to numer analizy w powyższym opisie.

Schemat wejść i wyjść modułu:

