# Zajęcia 9

**Zadanie 1**

Użycie funkcji **dropFields** i danych Nested.json

Usuń następujące atrybuty z columny:

"attribute1", "endGradeSeparation", "elevationAgainstDirection","formsPartOfPath"

**1a**

Sprawdź, jak działa ‘foldLeft’ i przerób dwa przykłady z internetu.

**1b challenge**

Konfiguracja:

val = excludedNestedFields = Map(”column1” -> Array(”Field1”, ” Field2”), “columm” -> Array())

Napisz metodę, która usunie wybrane elementy z zagnieżdżonej kolumny.

Ex excludedNestedFields.foldleft( (df) (k, v))

val fields = Array(“”,””)

withColumn(„k.\_1l”, $”konlj”.dropfiels(fields: \_\*)

**Zadanie 2**

Notatnik zadania.dbc

Dane w [GitHub](https://github.com/databricks/Spark-The-Definitive-Guide/tree/master/data/retail-data)

Zapisz dane w formacie delta, stwórz dużo małych plików (kilkaset) i wykonaj jakąś akcję typu count(). Potem wykonaj `OPTIMZIE` i sprawdź, ile jest plików po optymalizacji, porównaj czas.

**Zadanie 3**

Wykorzystaj notatnik i użyj funkcji ZORDER .   
Możesz użyć dane z notatnika bądź inne i zademonstruj jak działa ZORDER.

**Zadanie 4**

Wykonaj, VACUUM na plikach z zadań 3 i 4. Zobaczy czy zaobserwujesz jakąś zmianę. Jeśli nie to spróbuj nadpisać istniejące pliki. (zapisz df kilka krotnie z różną ilością partycji).