# Zajęcia 9

Zadanie 1

Partycjonowanie.

Pobierz jeden plik z <https://dumps.wikimedia.org/other/pageviews/readme.html>

Rozpakuj lokalnie, będzie najprościej i załaduj oba to wybranej ścieżki w dbfs.

Przerób notatnik z wykładów (Partycjonowanie.dbc) i dodatkowy lab (Partycjonowanie Lab.dbc), przetestuj, jak zachowuje się Spark, kiedy zmieniasz ilość partycji. Dla porównania możesz użyć lokalnych zasobów (laptop).

Zadanie 2

Przejdź przez notatniki (wybrany język) Complex Data Types i Heigher Order Functions i przerób przykłady żeby zapoznać się z dostępnymi funkcjami.

Zadanie 3

Stwórz projekt w IntelliJ lub PyCharm, wykorzystaj podany przykład jest trochę popsuty (lekki challenge).

<https://github.com/cegladanych/spark-maven-demo.git>

Dodaj 3 metody jedną wczytującą plik, drugą robiącą jakąś transformacje i trzecią zapisującą dane.

Sprawdź czy jesteś w stanie stworzyć paczkę jar lub wheel/egg.