Zajęcia 6

Zadanie 1

Pobierz notatniki Join SMJ, Generator.

**Cel zadania**: Wykonanie różnych typów jonów i sprawdzenie jaki to ma wpływ na wyniki zapytania.

Wykonaj połączenia w notatniku Joins.dbc

Dla tych jonów porównaj inner i left and right outer, Left Semi Join, Left Anti Join join

za każdym razem sprawdź .explain i zobacz jak Spark wykonuje połączenia

Zadanie 2

**Cel zadania**: usuwanie zduplikowanych kolumn

Użyj Generatora i stwórz dwie duże tabele po 1 milion wierszy i wykonaj dwa typy joinów, inner oraz left. Sprawdż plan wykonania.

Połącz tabele po tych samych kolumnach i użyj jednej metody z wykładów na na usunięcie duplikatów.