

Przetwarzanie i analiza dużych zbiorów danych - zadanie 2

Paweł Ciupka 234048

Bartosz Łuniewski 234086

Filip Woźniak 234131

Wyniki

ID użytkownika	ID przyjaciół
924	439, 2409, 6995, 11860, 15416, 43748, 45881
8941	8943, 8944, 8940
8942	939, 8940, 8943, 8944
9019	9022, 317, 9023
9020	9021, 9016, 9017, 9022, 317, 9023
9021	9020, 9016, 9017, 9022, 317, 9023
9022	9019, 9020, 9021, 317, 9016, 9017, 9023
9990	13134, 13478, 13877, 34299, 34485, 34642, 37941
9992	9987, 9989, 35667, 9991
9993	9991, 13134, 13478, 13877, 34299, 34485, 34642, 37941

Wykorzystanie Apache Spark

Apache Spark został wykorzystany do inicjalizacji kontekstu, wewnątrz którego uruchomiony jest cały algorytm. Dzięki skorzystaniu z tej funkcjonalności, główny kontekst może tworzyć w obrębie aplikacji kilka klastrów, a wewnątrz nich “workerów” i przetwarzać oddzielnie wiele zadania, co pozwala uzyskać dużą efektywność zaimplementowanego algorytmu.