Mamy tu plan wykonania query z groupby (po lewej) i display(df) (po prawej). Jak widać, zostały dołożone kroki (HashAggregate z czego jeden zajął aż 9 sekund) oraz Shuffle read, czyli czytanie z różnych partycji w celu pogrupowania danych.

Diagram

Description automatically generatedDiagram

Description automatically generated