Zajęcia 6

Zadanie 1

Pobierz notatnik Cwiczenia\_6.

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

Wykonaj połączenia dla tabel w schemacie **SalesLT**

Product, ProductModel, ProductDescription, ProductModelProductDescription

Address, CustomerAddress, Customer

SalesOrderDetail, SalesOrderHeader

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 3

**Cel zadania**: usuwanie zduplikowanych kolumn

Połącz tabele po tych samych kolumnach (Jeśli nie ma takich samych to dodaj kolumnę) i użyj metody na usunięcie duplikatów.

Zadanie 3

**Cel zadania**, Przygotować się na pytania interview. Móc opisać najważniejsze narzędzia analityczne Big Data w Azure/AWS/GCP/IBM/Alibaba.

Przykład

1. Robisz PoC na wykrycie anomalii z linii produkcyjnej (jakieś IoT). Jakich narzędzi użyjesz np. w Azure. Jak będzie wyglądało PoC.
2. Zaproponuj architekturę Big Data dla Rockstar Games przetwarzanie danych w celu wykrycia i zapobiegania błędom w grze, satysfakcji użytkowników. Jakie narzędzia, gdzie będzie compute ect…

Przejrzyj co jest wymagane do certyfikacji DP-203, i DP-100

<https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE3VEHg>

<https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4MbYT>

Stwórz listę najważniejszych narzędzi analitycznych/Big Data (3) i przygotuj notatki (linki do dokumentacji ect).