Spark UI

Spark UI służy do monitorowania i debugowania aplikacji Spark. Składa się z kilku zakładek, z których każda dostarcza istotnych informacji:

1. **Jobs**
   * Pokazuje listę wszystkich zadań w aplikacji.
   * Zawiera status, liczbę etapów czas wykonania.
   * Klikając na job, można przejść do szczegółów etapów.
2. **Stages**
   * Przedstawia podział jobów na mniejsze jednostki – etapy.
   * Dla każdego etapu widać liczbę tasków, czas ich wykonania, rozmiar danych, liczba shuffle read/write itp.
   * Tutaj znajdziesz informację o dystrybucji danych – w szczególności w kolumnach "Shuffle Read Size" i "Shuffle Write Size".
3. **Storage**
   * Pokazuje dane przechowywane w pamięci
   * Dla każdej tabeli lub DataFrame pokazuje ilość danych, rozmiar w RAM/dysku oraz partycje.
4. **Executors**
   * Lista wszystkich executorów (procesów wykonujących obliczenia).
   * Dla każdego: użycie pamięci, ilość przetworzonych danych, tasków, czas działania.
   * Pomaga w analizie obciążenia i problemów z nierównomiernym rozłożeniem danych.
5. **SQL/DataFrame**
   * Wizualizacja zapytań SQL i operacji na DataFrame.
   * Można zobaczyć fizyczny plan wykonania, operacje typu scan, filter, join, exchange, sort, itp.
   * Pozwala analizować sposób optymalizacji zapytań przez Catalyst.