**1. Jobs**

**Opis:** Lista wszystkich zadań (jobs) uruchomionych przez Spark w ramach danej aplikacji.

**Zawartość:**

* Status zadania (np. zakończone, nieudane)
* Czas trwania
* Powiązane akcje (np. collect(), count())
* Liczba i identyfikatory Stages

**Zastosowanie:** Punkt wyjścia do śledzenia wykonania operacji.

**2. Stages**

**Opis:** Jobs są dzielone na mniejsze etapy (stages), odpowiadające jednostkom pracy bez potrzeby przetasowania danych.

**Zawartość:**

* Lista tasków wchodzących w skład stage
* Czas wykonania, liczba tasków
* Szczegóły dotyczące wejścia/wyjścia (Input/Output Size)
* Informacje o shuffle (Shuffle Read/Write)

**Zastosowanie:** Pomocne przy analizie, gdzie występują kosztowne operacje przetasowania danych.

**3. Storage**

**Opis:** Pokazuje informacje o danych przechowywanych w pamięci (cache lub persist).

**Zawartość:**

* Lista zapamiętanych RDD lub DataFrame
* Liczba partycji
* Rozmiar danych
* Rozkład przechowywanych danych między executorami

**Zastosowanie:** Główne źródło informacji o dystrybucji danych w pamięci.

**4. Executors**

**Opis:** Pokazuje informacje o wszystkich executorach działających w aplikacji Spark.

**Zawartość:**

* Ilość wykonanych tasków
* Zużycie pamięci i CPU
* Ilość danych odczytanych i zapisanych
* Ilość przetworzonych danych

**Zastosowanie:** Można tutaj sprawdzić dystrybucję obciążenia oraz danych pomiędzy różnymi executorami.

**5. SQL/Dataframe**

**Opis:** Dotyczy zapytań SQL oraz operacji na DataFrame’ach.

**Zawartość:**

* Logical Plan
* Physical Plan
* Optimized Plan
* Informacje o użytych operatorach (join, filter, aggregate)

**Zastosowanie:** Analiza, w jaki sposób Spark optymalizuje i wykonuje zapytania.

**Gdzie sprawdzić dystrybucję danych?**

* **Storage:** dystrybucja danych w pamięci i ich rozmiar
* **Executors:** rozkład danych i obciążenia między executorami
* **Stages (Shuffle Read/Write):** pokazuje, kiedy dane są przemieszczane między partycjami i węzłami