Apache Spark UI to interfejs graficzny, który umożliwia monitorowanie i diagnozowanie aplikacji Spark. Jest podzielony na kilka sekcji, z których każda dostarcza różnych informacji na temat działania aplikacji Spark. Oto krótki przegląd tych sekcji:

**Jobs**

W tej zakładce można znaleźć:

* Listę wszystkich uruchomionych zadań (jobs) w aplikacji.
* Status zadań (kompletny, w toku, nieudany).
* Czas trwania zadań.
* Liczbę i status etapów (stages) składających się na poszczególne zadania.
* Szczegółowe informacje o zadaniach, takie jak liczba podzadań (tasks), podział czasu na różne fazy przetwarzania (np. czas deserializacji, wykonania, serializacji).

**Stages**

Ta sekcja dostarcza szczegółowych informacji o poszczególnych etapach (stages):

* Lista wszystkich etapów z podziałem na te zakończone, w toku i nieudane.
* Liczba i status podzadań (tasks) w ramach każdego etapu.
* Szczegółowe informacje na temat wykonania podzadań, takie jak czas wykonania, liczba ponownych prób (retries) oraz dane dotyczące wejścia/wyjścia (I/O).

**Storage**

W zakładce Storage znajdziemy informacje o przechowywanych danych RDD i DataFrame:

* Lista wszystkich przechowywanych RDD i DataFrame.
* Szczegółowe dane na temat rozmiaru przechowywanych danych.
* Liczba partycji oraz ich lokalizacja w klastrze.
* Szczegółowe dane dotyczące pamięci zużywanej przez poszczególne partycje.

**Executors**

Sekcja Executors dostarcza informacji o wykonawcach (executors) w klastrze Spark:

* Lista wszystkich wykonawców z podziałem na aktywne, nieaktywne i usunięte.
* Informacje o zużyciu pamięci i zasobów CPU przez poszczególnych wykonawców.
* Dane dotyczące wejścia/wyjścia (I/O), czasów wykonania oraz liczby zadań przypisanych do każdego wykonawcy.
* Szczegółowe informacje o błędach i wyjątki występujące podczas wykonywania zadań.

**SQL/Dataframe**

Zakładka SQL/Dataframe jest szczególnie przydatna przy pracy z DataFrame i SQL w Spark:

* Lista wszystkich zapytań SQL i operacji na DataFrame.
* Szczegółowe informacje na temat planu wykonania zapytań (query execution plan), w tym plan logiczny i fizyczny.
* Dane dotyczące metryk wykonania, takie jak czas trwania operacji, liczba przetworzonych wierszy, użycie pamięci, itp.

**Informacja o dystrybucji danych**

Informacje o dystrybucji danych w klastrze Spark można znaleźć w kilku miejscach:

* **Stages**: Szczegółowe informacje o partycjach danych i ich lokalizacji w klastrze.
* **Storage**: Dane dotyczące przechowywanych RDD i DataFrame, w tym liczba partycji i ich lokalizacja.
* **Executors**: Informacje o partycjach danych przetwarzanych przez poszczególnych wykonawców.

Dzięki Spark UI można szczegółowo monitorować i analizować działanie aplikacji, co jest kluczowe dla optymalizacji wydajności oraz rozwiązywania problemów.