Spark UI

Jobs – pokazuje listę zadań. Widać, czy się udały, ile trwały, ile miały etapów. Po kliknięciu są szczegóły.

Stages – każdy job ma etapy. Widać, ile było tasków, jaki był czas, ile danych (shuffle). Można sprawdzić, które taski trwały najdłużej.

Storage – pokazuje, co jest w pamięci. Widać, ile partycji i ile miejsca zajmują.

Executors – info o maszynach. Widać użycie CPU, pamięci, ile danych przeszło, ile tasków.

SQL / DataFrame – widać zapytania, czas, plan działania, joiny, shuffle itd.

Dystrybucja danych – w SQL/DataFrame w Physical Plan są Exchange, Repartition, Join. W Stages i Tasks są metryki. W Storage widać dane i podział na partycje.