# Zajęcia 13

**Zadanie 1**

Wykorzystaj swoją aplikację tą z jar (wheel).

Dodaj logi do aplikacji tak aby były widoczne w Spark UI > Driver Logs > Log4j output

Przykład jest dostępny na

<https://github.com/databricks/Spark-The-Definitive-Guide/tree/master/project-templates/scala>

<https://github.com/cegladanych/spark-maven-demo>

Przykład pozwoli na dodanie logów aplikacji do logów Sparka, to jest stary kod i może wymagać modyfikacji.

Dodaj trzy typy **Infor, Error i Warn**

Log4j output

**22/06/06 16:54:18 INFO To ja napisałem** ….

**Zadanie 2**

Uruchom istniejącą aplikację/notatnik i zapoznaj się ze **Spark UI**, powinnieneś wiedzieć, gdzie znaleźć informację o Jobs > Stage > Tasks, SQL/Dataframe

**Gdzie jest informacja o dystrybucji danych?**

Przerób jakiś przykład z poprzednich ćwiczeń (agregacja, szufle, join ect) i przyjrzycie UI i logi żeby wiedzieć gdzie szukać informacji.

**Zadanie 3**

**Bucketing**

Przygotuj przykład użycia bucketing i porównaj go z partitionBy.

Dla tych samych danych wykonaj obie operacje i sprawdź czy będą różnice w plikach?

**Zadanie 4**

Stwórz jedną tabelę i pobierz statystyki dla tabeli i wszystkich kolumn (przykład na wykładach).