**Anna Plust, 36147**

**Duże zbiory danych**

**Lab 1, Sprawozdanie**

1. **Wykonać program dla platformy Spark w trybie interaktywnym.**

Wykonanie programu lokanie w trybie interaktywnym:

Text

Description automatically generated

Text

Description automatically generated

Dane wynikowe można wydrukować i zobaczyć w konsoli, ponieważ wszystkie komendy są wykonywane lokalnie.

Wykonanie programu w trybie interaktywnym uruchomionego w trybie klastra: spark-shell --master spark://hadoop1:7077

Text

Description automatically generated

W trybie klastra wynik nie jest widoczny w konsoli, ponieważ program (w tym metoda println) jest wykonywany na serwerze. Wynik można zobaczyć za standardowym wyjściu workerów, które zajmowały się wykonaniem programu.

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

**2) Stworzyć i uruchomić w języku Java aplikację dla platformy Spark**

public class Main {  
 public static void main(String[] args) {

SparkConf conf = new SparkConf();  
 JavaSparkContext sc = new JavaSparkContext(conf);

JavaRDD<String> text = sc.textFile("hdfs://hadoop1:9000/students/st36148/plik.txt");  
 JavaRDD<String> words = text.flatMap(line -> Arrays.asList(line.split("[ ,;\\.]")).iterator());  
 JavaPairRDD<String, Integer> pairs = words.mapToPair(word -> new Tuple2<>(word, 1));  
 JavaPairRDD<String, Integer> counts = pairs.reduceByKey((a, b) -> a + b);  
 counts.foreach(c -> System.out.println(c));  
 }  
}

**Tryb lokalny**

Porgram uruchomiony komendą:

spark-submit --class Main lab1-1.0-SNAPSHOT.jar

Wynik jest widoczy w konsoli:

Text

Description automatically generated

**Tryb klastra, manager spark**

spark-submit --class Main --master spark://hadoop1:7077 lab1-1.0-SNAPSHOT.jar

Graphical user interface, application

Description automatically generated with medium confidence

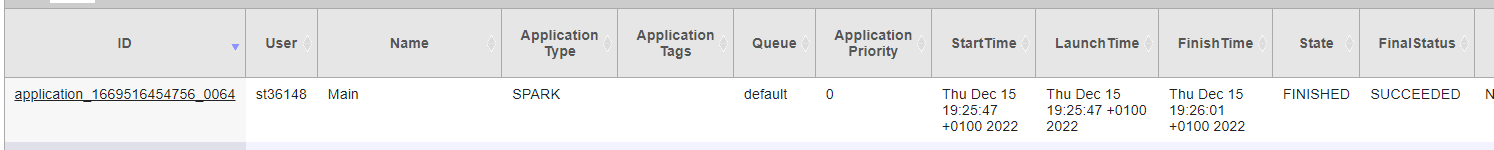
Wynik nie jest widoczny w konsoli, ale na standardowym wyjściu workera:

Graphical user interface, application

Description automatically generated

**Tryb klastra, manager yarn, client**

spark-submit --class Main --master yarn --deploy-mode client lab1-1.0-SNAPSHOT.jar

****

**Graphical user interface, text, application

Description automatically generated**

Driver program jest uruchomiony po stronie klienta, dlatego węzeł nie zstał przydzielony.

Wynik nie jest widoczny w konsoli.

**Tryb klastra, manager yarn, cluster**

spark-submit --class Main --master yarn --deploy-mode cluster lab1-1.0-SNAPSHOT.jar

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Driver program jest uruchomiony na nodzie klastra.

Wynik nie jest widoczny w konsoli.