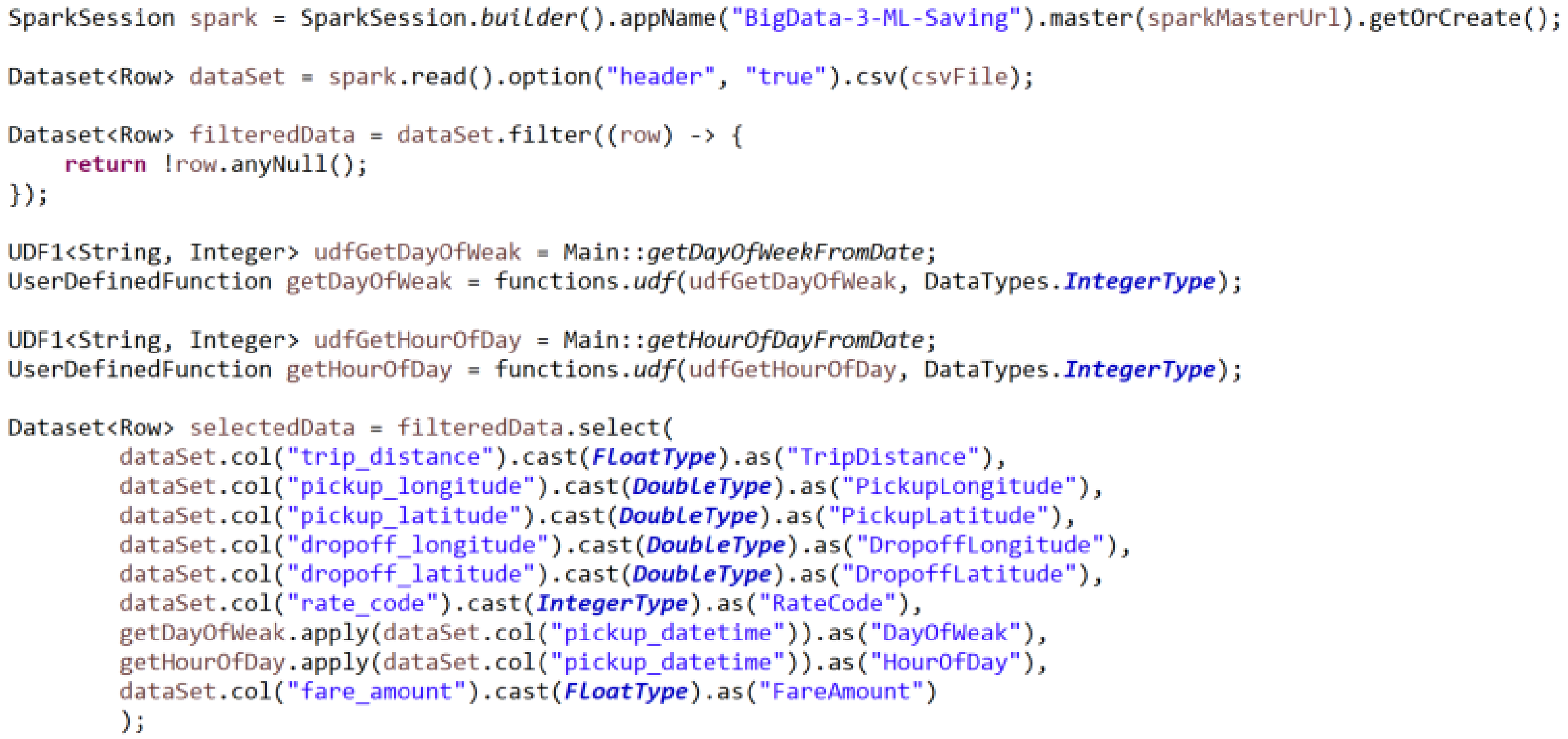
Sistemi za obradu i analizu velike količine podataka - Treći projekat

Nikola Božinović 1030

Tehnologije

- Java 8
- Hadoop 2.7
- Spark 2.4.3
- Kafka 2.4.0 (Scala 2.11)
- Docker

ML Batch

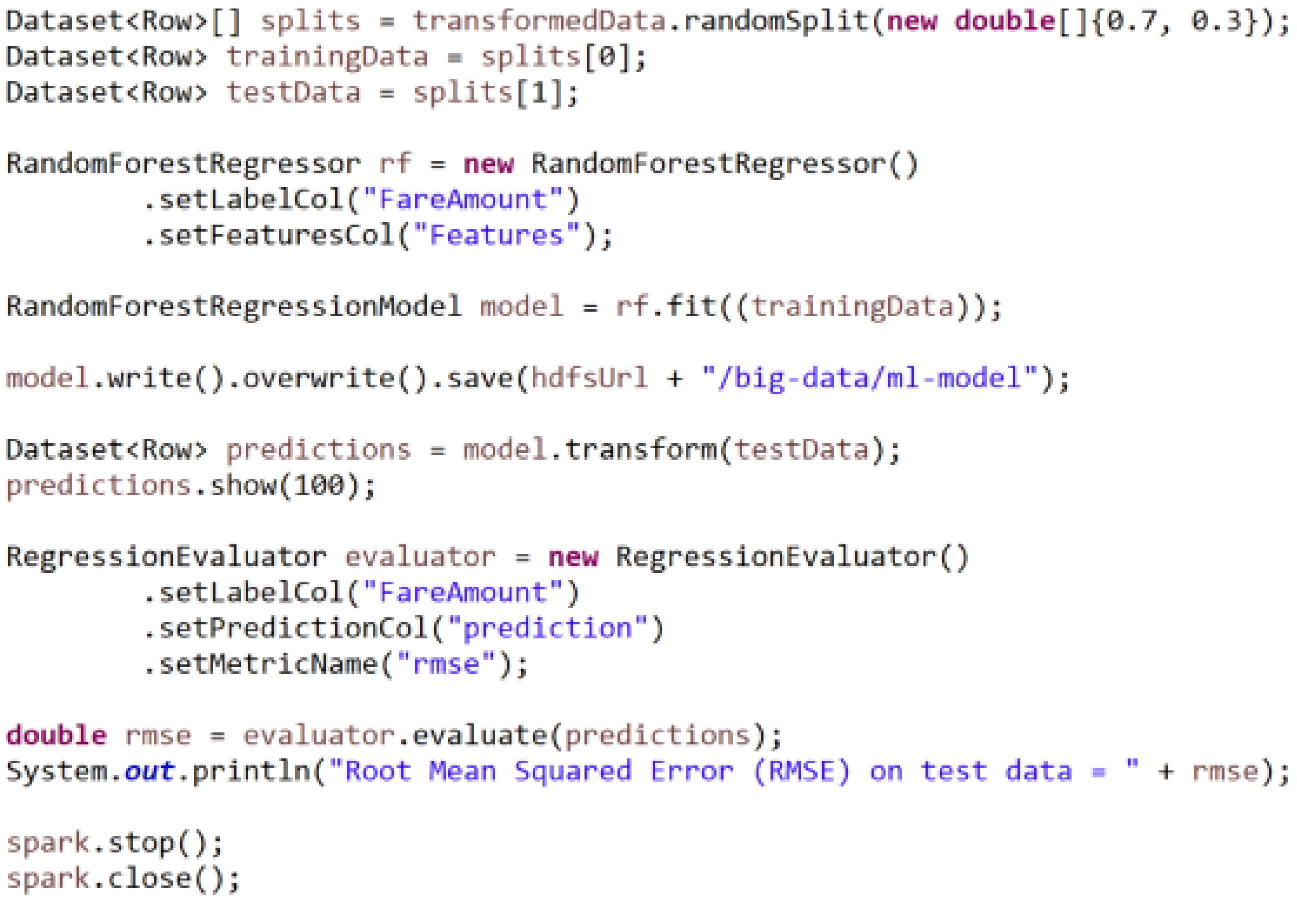


```
SparkSession spark = SparkSession.builder().appName("BigData-3-ML-Saving").master(sparkMasterUrl).getOrCreate();

Dataset<Row> dataSet = spark.read().option("header", "true").csv(csvFile);

Dataset<Row> filteredData = dataSet.filter((row) -> {
    return !row.anyNull();
});

UDF1<String, Integer> udfGetDayOfWeak = Main::getDayOfWeekFromDate;
UserDefinedFunction getDayOfWeak = functions.udf(udfGetDayOfWeak, DataTypes.IntegerType);

UDF1<String, Integer> udfGetHourOfDay = Main::getHourOfDayFromDate;
UserDefinedFunction getHourOfDay = functions.udf(udfGetHourOfDay, DataTypes.IntegerType);

Dataset<Row> selectedData = filteredData.select(
        dataSet.col("trip_distance").cast(FloatType).as("TripDistance"),
        dataSet.col("pickup_longitude").cast(DoubleType).as("PickupLongitude"),
        dataSet.col("pickup_latitude").cast(DoubleType).as("PickupLatitude"),
        dataSet.col("dropoff_longitude").cast(DoubleType).as("DropoffLongitude"),
        dataSet.col("dropoff_latitude").cast(DoubleType).as("DropoffLatitude"),
        dataSet.col("rate_code").cast(IntegerType).as("RateCode"),
        getDayOfWeak.apply(dataSet.col("pickup_datetime")).as("DayOfWeak"),
        getHourOfDay.apply(dataSet.col("pickup_datetime")).as("HourOfDay"),
        dataSet.col("fare_amount").cast(FloatType).as("FareAmount")
        );
```

ML Batch

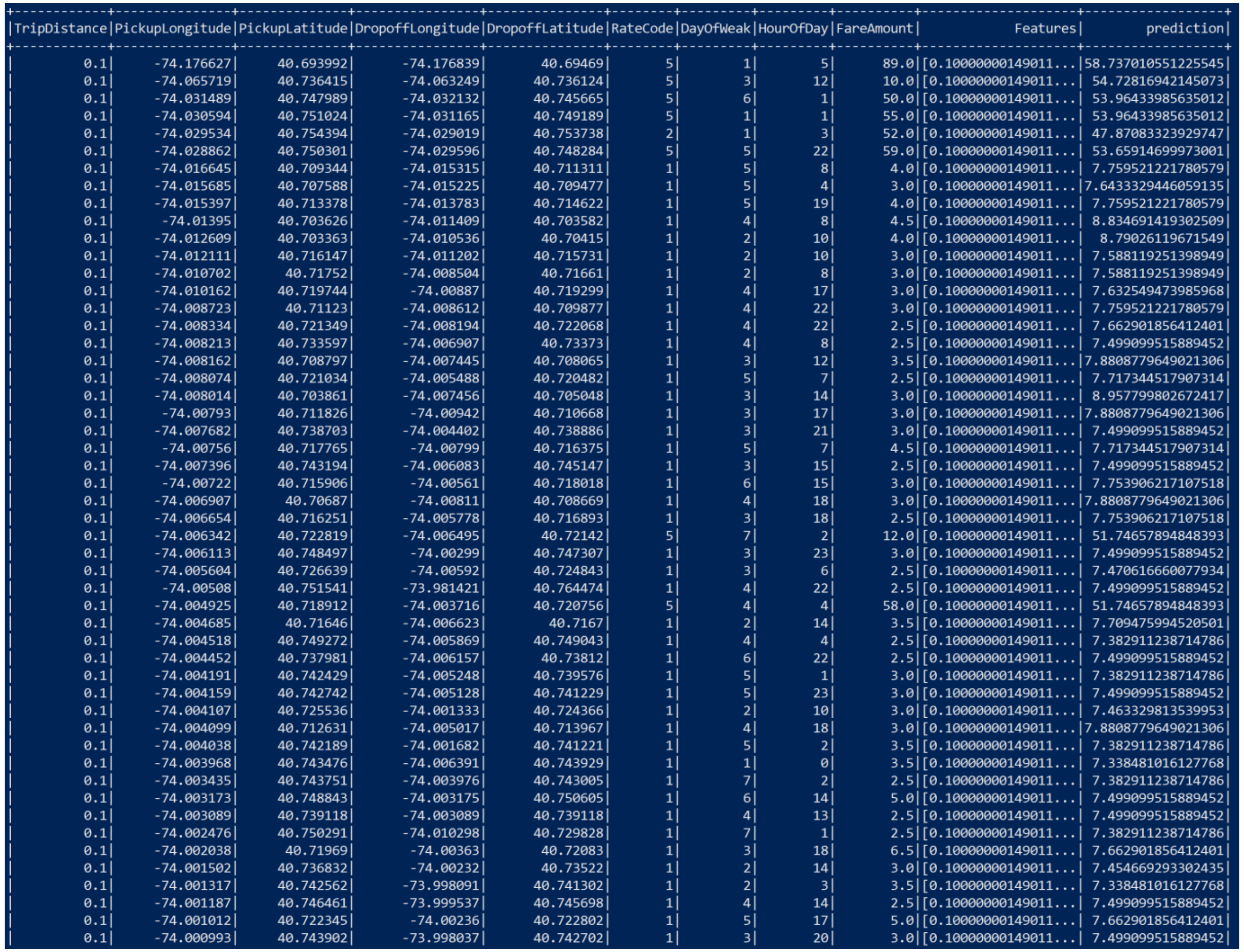
```
// nyc coordinates:
// latitude = 40.730610
// longitude = -73.935242
double nycMinLatitude = 38.730610; // -2
double nycMaxLatitude = 42.730610; // +2
double nycMinLongitude = -75.935242; // -2
double nycMaxLongitude = -71.935242; // +2

Dataset<Row> filteredAndSelectedData = selectedData.filter(selectedData.col("TripDistance").notEqual(0.0)
            .and(selectedData.col("PickupLongitude").gt(nycMinLongitude))
            .and(selectedData.col("PickupLongitude").lt(nycMaxLongitude))
            .and(selectedData.col("PickupLatitude").gt(nycMinLatitude))
            .and(selectedData.col("PickupLatitude").lt(nycMaxLatitude))
            .and(selectedData.col("DropoffLongitude").gt(nycMinLongitude))
            .and(selectedData.col("DropoffLongitude").lt(nycMaxLongitude))
            .and(selectedData.col("DropoffLatitude").gt(nycMinLatitude))
            .and(selectedData.col("DropoffLatitude").lt(nycMaxLatitude)));


VectorAssembler vectorAssembler = new VectorAssembler()
        .setInputCols(new String[]{"TripDistance", "PickupLongitude", "PickupLatitude", "DropoffLongitude",
                "DropoffLatitude","DayOfWeak", "HourOfDay","RateCode"})
        .setOutputCol("Features");

Dataset<Row> transformedData = vectorAssembler.transform(filteredAndSelectedData);
```

ML Batch

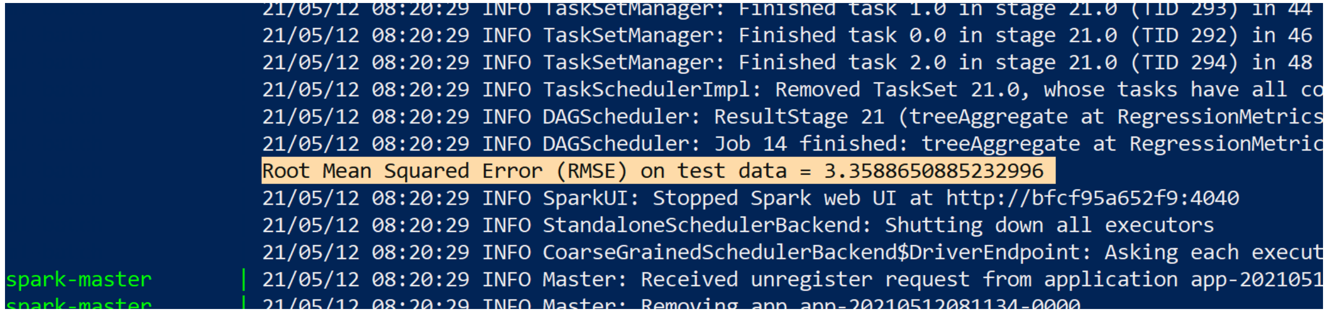


```
Dataset<Row>[] splits = transformedData.randomSplit(new double[]{0.7, 0.3});
Dataset<Row> trainingData = splits[0];
Dataset<Row> testData = splits[1];

RandomForestRegressor rf = new RandomForestRegressor()
        .setLabelCol("FareAmount")
        .setFeaturesCol("Features");

RandomForestRegressionModel model = rf.fit((trainingData));

model.write().overwrite().save(hdfsUrl + "/big-data/ml-model");

Dataset<Row> predictions = model.transform(testData);
predictions.show(100);

RegressionEvaluator evaluator = new RegressionEvaluator()
        .setLabelCol("FareAmount")
        .setPredictionCol("prediction")
        .setMetricName("rmse");

double rmse = evaluator.evaluate(predictions);
System.out.println("Root Mean Squared Error (RMSE) on test data = " + rmse);

spark.stop();
spark.close();
```

ML Batch

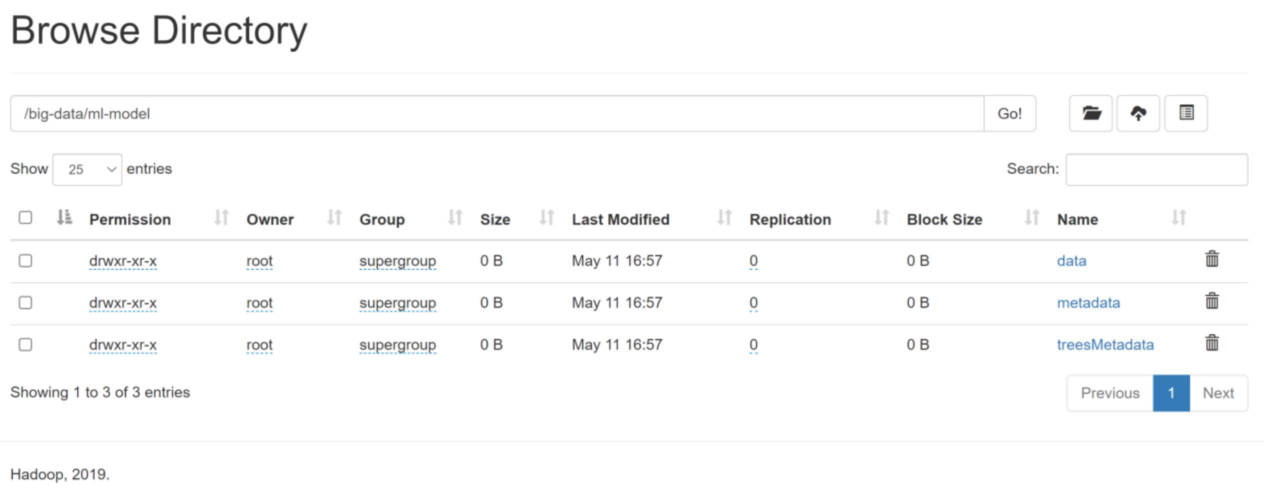


```
+------------+---------------+--------------+----------------+---------------+--------+---------+---------+----------+--------------------+------------------+
|TripDistance|PickupLongitude|PickupLatitude|DropoffLongitude|DropoffLatitude|RateCode|DayOfWeak|HourOfDay|FareAmount|            Features|        prediction|
+------------+---------------+--------------+----------------+---------------+--------+---------+---------+----------+--------------------+------------------+
|         0.1|     -74.176627|     40.693992|      -74.176839|       40.69469|       5|        1|        5|      89.0|[0.10000000149011...|58.737010551225545|
|         0.1|     -74.065719|     40.736415|      -74.063249|      40.736124|       5|        3|       12|      10.0|[0.10000000149011...| 54.72816942145073|
|         0.1|     -74.031489|     40.747989|      -74.032132|      40.745665|       5|        6|        1|      50.0|[0.10000000149011...| 53.96433985635012|
|         0.1|     -74.030594|     40.751024|      -74.031165|      40.749189|       5|        1|        1|      55.0|[0.10000000149011...| 53.96433985635012|
|         0.1|     -74.029534|     40.754394|      -74.029019|      40.753738|       2|        1|        3|      52.0|[0.10000000149011...| 47.87083323929747|
|         0.1|     -74.028862|     40.750301|      -74.029596|      40.748284|       5|        5|       22|      59.0|[0.10000000149011...| 53.65914699973001|
|         0.1|     -74.016645|     40.709344|      -74.015315|      40.711311|       1|        5|        8|       4.0|[0.10000000149011...| 7.759521221780579|
|         0.1|     -74.015685|     40.707588|      -74.015225|      40.709477|       1|        5|        4|       3.0|[0.10000000149011...|7.6433329446059135|
|         0.1|     -74.015397|     40.713378|      -74.013783|      40.714622|       1|        5|       19|       4.0|[0.10000000149011...| 7.759521221780579|
|         0.1|      -74.01395|     40.703626|      -74.011409|      40.703582|       1|        4|        8|       4.5|[0.10000000149011...| 8.834691419302509|
|         0.1|     -74.012609|     40.703363|      -74.010536|       40.70415|       1|        2|       10|       4.0|[0.10000000149011...|  8.79026119671549|
|         0.1|     -74.012111|     40.716147|      -74.011202|      40.715731|       1|        2|       10|       3.0|[0.10000000149011...| 7.588119251398949|
|         0.1|     -74.010702|      40.71752|      -74.008504|       40.71661|       1|        2|        8|       3.0|[0.10000000149011...| 7.588119251398949|
|         0.1|     -74.010162|     40.719744|       -74.00887|      40.719299|       1|        4|       17|       3.0|[0.10000000149011...| 7.632549473985968|
|         0.1|     -74.008723|      40.71123|      -74.008612|      40.709877|       1|        4|       22|       3.0|[0.10000000149011...| 7.759521221780579|
|         0.1|     -74.008334|     40.721349|      -74.008194|      40.722068|       1|        4|       22|       2.5|[0.10000000149011...| 7.662901856412401|
|         0.1|     -74.008213|     40.733597|      -74.006907|       40.73373|       1|        4|        8|       2.5|[0.10000000149011...| 7.499099515889452|
|         0.1|     -74.008162|     40.708797|      -74.007445|      40.708065|       1|        3|       12|       3.5|[0.10000000149011...|7.8808779649021306|
|         0.1|     -74.008074|     40.721034|      -74.005488|      40.720482|       1|        5|        7|       2.5|[0.10000000149011...| 7.717344517907314|
|         0.1|     -74.008014|     40.703861|      -74.007456|      40.705048|       1|        3|       14|       3.0|[0.10000000149011...| 8.957799802672417|
|         0.1|      -74.00793|     40.711826|       -74.00942|      40.710668|       1|        3|       17|       3.0|[0.10000000149011...|7.8808779649021306|
|         0.1|     -74.007682|     40.738703|      -74.004402|      40.738886|       1|        3|       21|       3.0|[0.10000000149011...| 7.499099515889452|
|         0.1|      -74.00756|     40.717765|       -74.00799|      40.716375|       1|        5|        7|       4.5|[0.10000000149011...| 7.717344517907314|
|         0.1|     -74.007396|     40.743194|      -74.006083|      40.745147|       1|        3|       15|       2.5|[0.10000000149011...| 7.499099515889452|
|         0.1|      -74.00722|     40.715906|       -74.00561|      40.718018|       1|        6|       15|       3.0|[0.10000000149011...| 7.753906217107518|
|         0.1|     -74.006907|      40.70687|       -74.00811|      40.708669|       1|        4|       18|       3.0|[0.10000000149011...|7.8808779649021306|
|         0.1|     -74.006654|     40.716251|      -74.005778|      40.716893|       1|        3|       18|       2.5|[0.10000000149011...| 7.753906217107518|
|         0.1|     -74.006342|     40.722819|      -74.006495|       40.72142|       5|        7|        2|      12.0|[0.10000000149011...| 51.74657894848393|
|         0.1|     -74.006113|     40.748497|       -74.00299|      40.747307|       1|        3|       23|       3.0|[0.10000000149011...| 7.499099515889452|
|         0.1|     -74.005604|     40.726639|       -74.00592|      40.724843|       1|        3|        6|       2.5|[0.10000000149011...| 7.470616660077934|
|         0.1|      -74.00508|     40.751541|      -73.981421|      40.764474|       1|        4|       22|       2.5|[0.10000000149011...| 7.499099515889452|
|         0.1|     -74.004925|     40.718912|      -74.003716|      40.720756|       5|        4|        4|      58.0|[0.10000000149011...| 51.74657894848393|
|         0.1|     -74.004685|      40.71646|      -74.006623|        40.7167|       1|        2|       14|       3.5|[0.10000000149011...| 7.709475994520501|
|         0.1|     -74.004518|     40.749272|      -74.005869|      40.749043|       1|        4|        4|       2.5|[0.10000000149011...| 7.382911238714786|
|         0.1|     -74.004452|     40.737981|      -74.006157|       40.73812|       1|        6|       22|       2.5|[0.10000000149011...| 7.499099515889452|
|         0.1|     -74.004191|     40.742429|      -74.005248|      40.739576|       1|        5|        1|       3.0|[0.10000000149011...| 7.382911238714786|
|         0.1|     -74.004159|     40.742742|      -74.005128|      40.741229|       1|        5|       23|       3.0|[0.10000000149011...| 7.499099515889452|
|         0.1|     -74.004107|     40.725536|      -74.001333|      40.724366|       1|        2|       10|       3.0|[0.10000000149011...| 7.463329813539953|
|         0.1|     -74.004099|     40.712631|      -74.005017|      40.713967|       1|        4|       18|       3.0|[0.10000000149011...|7.8808779649021306|
|         0.1|     -74.004038|     40.742189|      -74.001682|      40.741221|       1|        5|        2|       3.5|[0.10000000149011...| 7.382911238714786|
|         0.1|     -74.003968|     40.743476|      -74.006391|      40.743929|       1|        1|        0|       3.5|[0.10000000149011...| 7.338481016127768|
|         0.1|     -74.003435|     40.743751|      -74.003976|      40.743005|       1|        7|        2|       2.5|[0.10000000149011...| 7.382911238714786|
|         0.1|     -74.003173|     40.748843|      -74.003175|      40.750605|       1|        6|       14|       5.0|[0.10000000149011...| 7.499099515889452|
|         0.1|     -74.003089|     40.739118|      -74.003089|      40.739118|       1|        4|       13|       2.5|[0.10000000149011...| 7.499099515889452|
|         0.1|     -74.002476|     40.750291|      -74.010298|      40.729828|       1|        7|        1|       2.5|[0.10000000149011...| 7.382911238714786|
|         0.1|     -74.002038|      40.71969|       -74.00363|       40.72083|       1|        3|       18|       6.5|[0.10000000149011...| 7.662901856412401|
|         0.1|     -74.001502|     40.736832|       -74.00232|       40.73522|       1|        2|       14|       3.0|[0.10000000149011...| 7.454669293302435|
|         0.1|     -74.001317|     40.742562|      -73.998091|      40.741302|       1|        2|        3|       3.5|[0.10000000149011...| 7.338481016127768|
|         0.1|     -74.001187|     40.746461|      -73.999537|      40.745698|       1|        4|       14|       2.5|[0.10000000149011...| 7.499099515889452|
|         0.1|     -74.001012|     40.722345|       -74.00236|      40.722802|       1|        5|       17|       5.0|[0.10000000149011...| 7.662901856412401|
|         0.1|     -74.000993|     40.743902|      -73.998037|      40.742702|       1|        3|       20|       3.0|[0.10000000149011...| 7.499099515889452|
```



```
                     21/05/12 08:20:29 INFO TaskSetManager: Finished task 1.0 in stage 21.0 (TID 293) in 44
                     21/05/12 08:20:29 INFO TaskSetManager: Finished task 0.0 in stage 21.0 (TID 292) in 46
                     21/05/12 08:20:29 INFO TaskSetManager: Finished task 2.0 in stage 21.0 (TID 294) in 48
                     21/05/12 08:20:29 INFO TaskSchedulerImpl: Removed TaskSet 21.0, whose tasks have all co
                     21/05/12 08:20:29 INFO DAGScheduler: ResultStage 21 (treeAggregate at RegressionMetrics
                     21/05/12 08:20:29 INFO DAGScheduler: Job 14 finished: treeAggregate at RegressionMetric
                     Root Mean Squared Error (RMSE) on test data = 3.3588650885232996
                     21/05/12 08:20:29 INFO SparkUI: Stopped Spark web UI at http://bfcf95a652f9:4040
                     21/05/12 08:20:29 INFO StandaloneSchedulerBackend: Shutting down all executors
                     21/05/12 08:20:29 INFO CoarseGrainedSchedulerBackend$DriverEndpoint: Asking each execut
spark-master       | 21/05/12 08:20:29 INFO Master: Received unregister request from application app-2021051
spark-master       | 21/05/12 08:20:29 INFO Master: Removing app app-20210512081134-0000
```

ML Batch



Browse Directory

/big-data/ml-model Go!

Show 25 entries

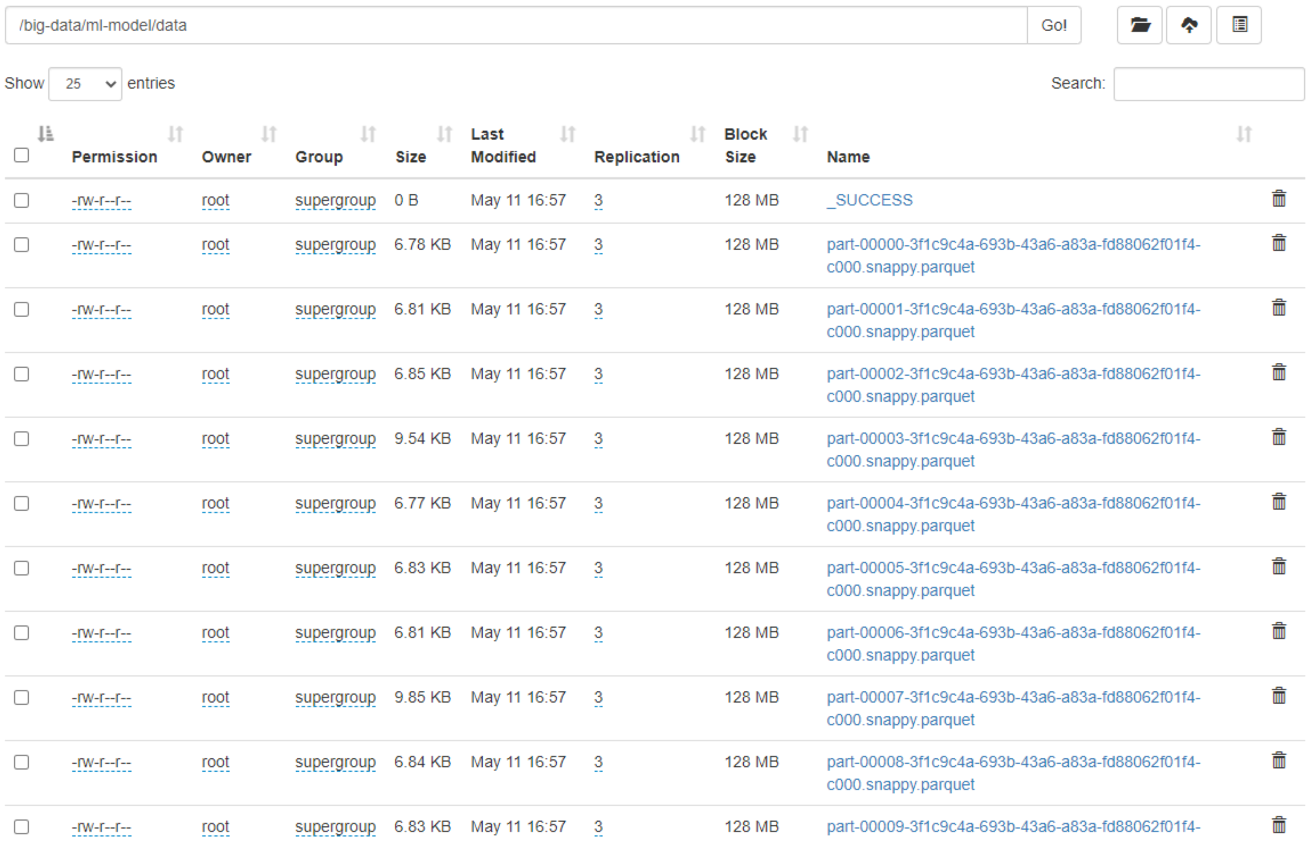
Search:

□	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
□	drwxr-xr-x	root	supergroup	0 B	May 11 16:57	0	0 B	data	🗑
□	drwxr-xr-x	root	supergroup	0 B	May 11 16:57	0	0 B	metadata	🗑
□	drwxr-xr-x	root	supergroup	0 B	May 11 16:57	0	0 B	treesMetadata	🗑

Showing 1 to 3 of 3 entries

Previous 1 Next

Hadoop, 2019.



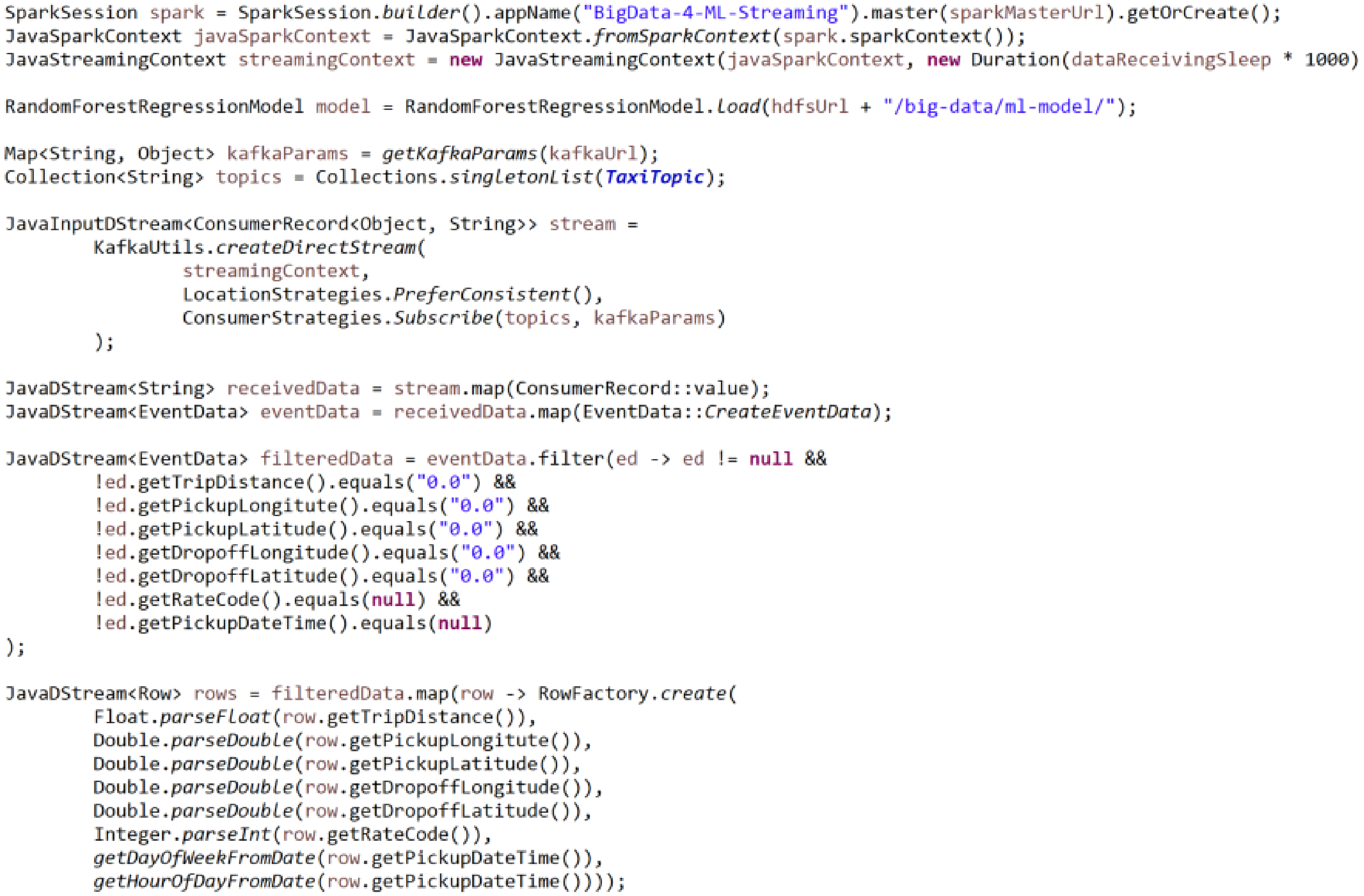
/big-data/ml-model/data Go!

Show 25 entries

Search:

□	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
□	-rw-r--r--	root	supergroup	0 B	May 11 16:57	3	128 MB	_SUCCESS	🗑
□	-rw-r--r--	root	supergroup	6.78 KB	May 11 16:57	3	128 MB	part-00000-3f1c9c4a-693b-43a6-a83a-fd88062f01f4-c000.snappy.parquet	🗑
□	-rw-r--r--	root	supergroup	6.81 KB	May 11 16:57	3	128 MB	part-00001-3f1c9c4a-693b-43a6-a83a-fd88062f01f4-c000.snappy.parquet	🗑
□	-rw-r--r--	root	supergroup	6.85 KB	May 11 16:57	3	128 MB	part-00002-3f1c9c4a-693b-43a6-a83a-fd88062f01f4-c000.snappy.parquet	🗑
□	-rw-r--r--	root	supergroup	9.54 KB	May 11 16:57	3	128 MB	part-00003-3f1c9c4a-693b-43a6-a83a-fd88062f01f4-c000.snappy.parquet	🗑
□	-rw-r--r--	root	supergroup	6.77 KB	May 11 16:57	3	128 MB	part-00004-3f1c9c4a-693b-43a6-a83a-fd88062f01f4-c000.snappy.parquet	🗑
□	-rw-r--r--	root	supergroup	6.83 KB	May 11 16:57	3	128 MB	part-00005-3f1c9c4a-693b-43a6-a83a-fd88062f01f4-c000.snappy.parquet	🗑
□	-rw-r--r--	root	supergroup	6.81 KB	May 11 16:57	3	128 MB	part-00006-3f1c9c4a-693b-43a6-a83a-fd88062f01f4-c000.snappy.parquet	🗑
□	-rw-r--r--	root	supergroup	9.85 KB	May 11 16:57	3	128 MB	part-00007-3f1c9c4a-693b-43a6-a83a-fd88062f01f4-c000.snappy.parquet	🗑
□	-rw-r--r--	root	supergroup	6.84 KB	May 11 16:57	3	128 MB	part-00008-3f1c9c4a-693b-43a6-a83a-fd88062f01f4-c000.snappy.parquet	🗑
□	-rw-r--r--	root	supergroup	6.83 KB	May 11 16:57	3	128 MB	part-00009-3f1c9c4a-693b-43a6-a83a-fd88062f01f4-	🗑

ML Streaming

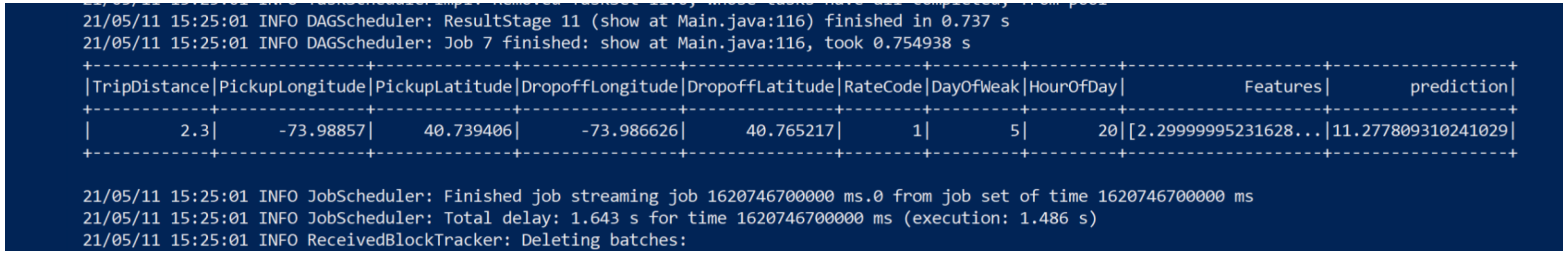


```java
SparkSession spark = SparkSession.builder().appName("BigData-4-ML-Streaming").master(sparkMasterUrl).getOrCreate();
JavaSparkContext javaSparkContext = JavaSparkContext.fromSparkContext(spark.sparkContext());
JavaStreamingContext streamingContext = new JavaStreamingContext(javaSparkContext, new Duration(dataReceivingSleep * 1000)

RandomForestRegressionModel model = RandomForestRegressionModel.load(hdfsUrl + "/big-data/ml-model/");

Map<String, Object> kafkaParams = getKafkaParams(kafkaUrl);
Collection<String> topics = Collections.singletonList(TaxiTopic);

JavaInputDStream<ConsumerRecord<Object, String>> stream =
        KafkaUtils.createDirectStream(
                streamingContext,
                LocationStrategies.PreferConsistent(),
                ConsumerStrategies.Subscribe(topics, kafkaParams)
        );

JavaDStream<String> receivedData = stream.map(ConsumerRecord::value);
JavaDStream<EventData> eventData = receivedData.map(EventData::CreateEventData);

JavaDStream<EventData> filteredData = eventData.filter(ed -> ed != null &&
        !ed.getTripDistance().equals("0.0") &&
        !ed.getPickupLongitute().equals("0.0") &&
        !ed.getPickupLatitude().equals("0.0") &&
        !ed.getDropoffLongitude().equals("0.0") &&
        !ed.getDropoffLatitude().equals("0.0") &&
        !ed.getRateCode().equals(null) &&
        !ed.getPickupDateTime().equals(null)
);

JavaDStream<Row> rows = filteredData.map(row -> RowFactory.create(
        Float.parseFloat(row.getTripDistance()),
        Double.parseDouble(row.getPickupLongitute()),
        Double.parseDouble(row.getPickupLatitude()),
        Double.parseDouble(row.getDropoffLongitude()),
        Double.parseDouble(row.getDropoffLatitude()),
        Integer.parseInt(row.getRateCode()),
        getDayOfWeekFromDate(row.getPickupDateTime()),
        getHourOfDayFromDate(row.getPickupDateTime())));
```

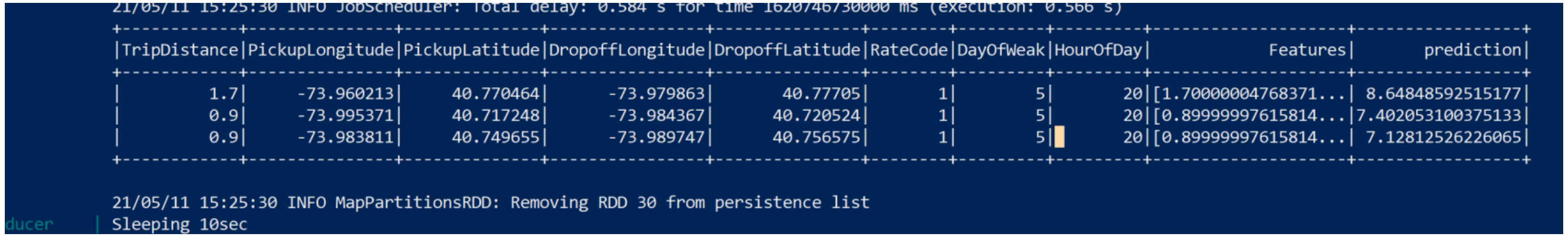
ML Streaming

```
rows.foreachRDD(d -> {
    StructType rowSchema = DataTypes.createStructType(
            new StructField[]{
                    createStructField("TripDistance", DataTypes.FloatType, false),
                    createStructField("PickupLongitude", DataTypes.DoubleType, false),
                    createStructField("PickupLatitude", DataTypes.DoubleType, false),
                    createStructField("DropoffLongitude", DataTypes.DoubleType, false),
                    createStructField("DropoffLatitude", DataTypes.DoubleType, false),
                    createStructField("RateCode", DataTypes.IntegerType, false),
                    createStructField("DayOfWeak", DataTypes.IntegerType, false),
                    createStructField("HourOfDay", DataTypes.IntegerType, false),
            });

    Dataset<Row> data = spark.createDataFrame(d, rowSchema);

    VectorAssembler vectorAssembler = new VectorAssembler()
            .setInputCols(new String[]{"TripDistance", "PickupLongitude", "PickupLatitude", "DropoffLongitude",
                    "DropoffLatitude", "RateCode", "DayOfWeak", "HourOfDay"})
            .setOutputCol("Features");

    Dataset<Row> transformed = vectorAssembler.transform(data);
    Dataset<Row> predictions = model.transform(transformed);
    predictions.show(100);
});
```

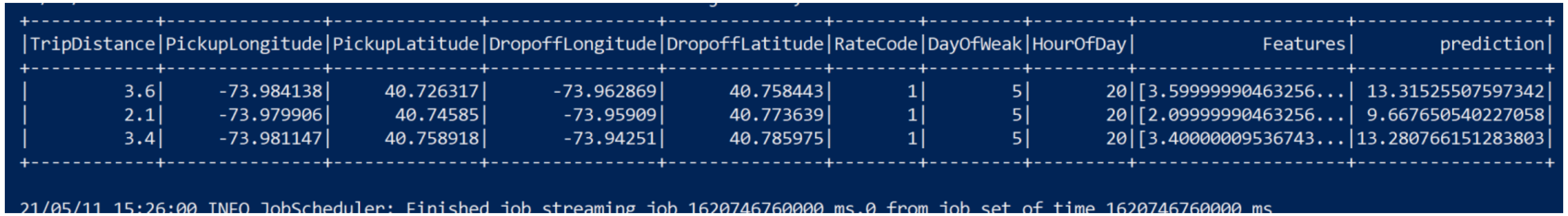
ML Streaming



```
21/05/11 15:25:01 INFO TaskSchedulerImpl: Removed TaskSet 11.0, whose tasks have all completed, from pool
21/05/11 15:25:01 INFO DAGScheduler: ResultStage 11 (show at Main.java:116) finished in 0.737 s
21/05/11 15:25:01 INFO DAGScheduler: Job 7 finished: show at Main.java:116, took 0.754938 s
+------------+---------------+--------------+----------------+---------------+--------+---------+---------+--------------------+------------------+
|TripDistance|PickupLongitude|PickupLatitude|DropoffLongitude|DropoffLatitude|RateCode|DayOfWeak|HourOfDay|            Features|        prediction|
+------------+---------------+--------------+----------------+---------------+--------+---------+---------+--------------------+------------------+
|         2.3|      -73.98857|     40.739406|      -73.986626|      40.765217|       1|        5|       20|[2.29999995231628...|11.277809310241029|
+------------+---------------+--------------+----------------+---------------+--------+---------+---------+--------------------+------------------+

21/05/11 15:25:01 INFO JobScheduler: Finished job streaming job 1620746700000 ms.0 from job set of time 1620746700000 ms
21/05/11 15:25:01 INFO JobScheduler: Total delay: 1.643 s for time 1620746700000 ms (execution: 1.486 s)
21/05/11 15:25:01 INFO ReceivedBlockTracker: Deleting batches:
```

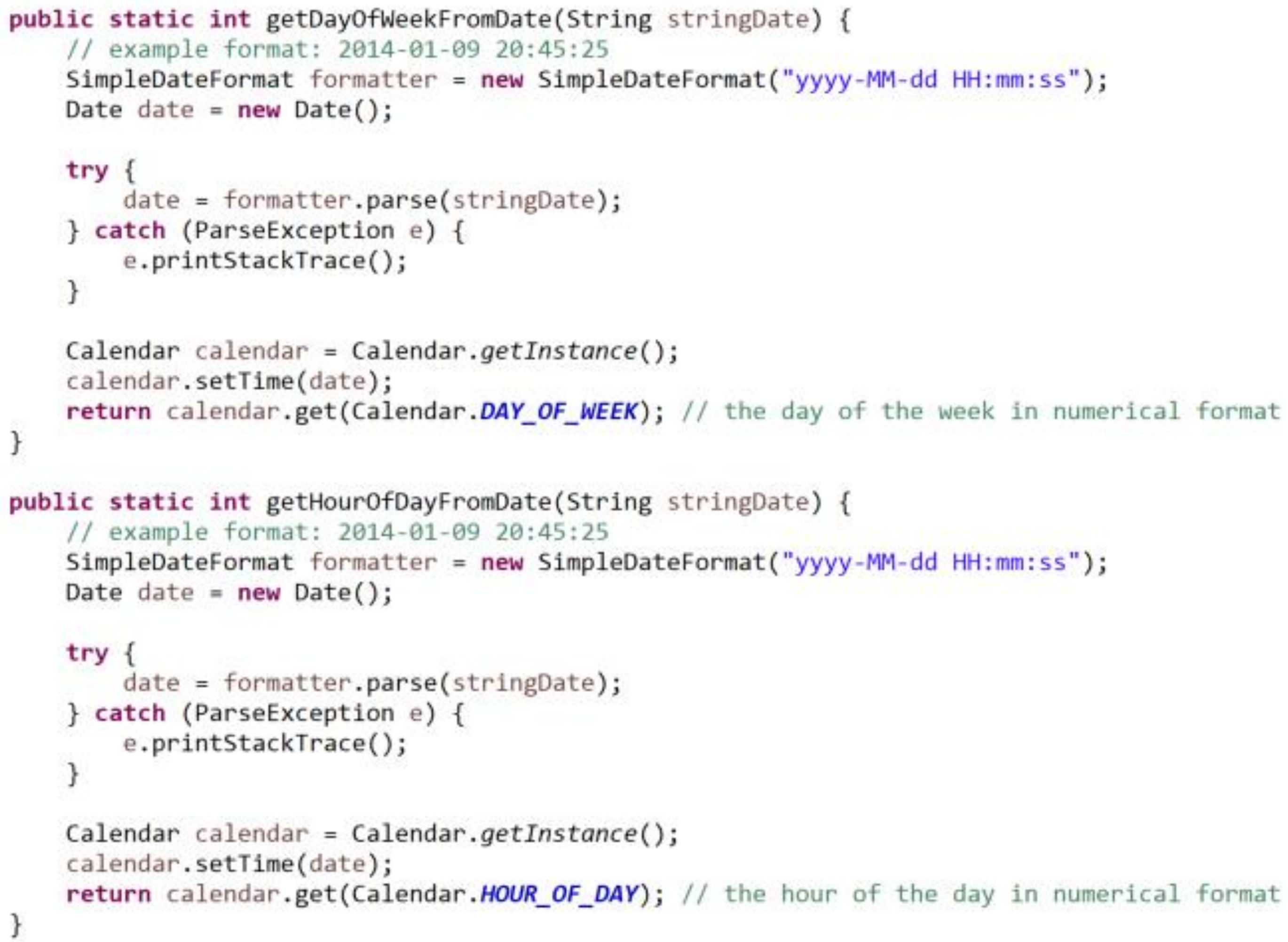


```
         21/05/11 15:25:30 INFO JobScheduler: Total delay: 0.584 s for time 1620746730000 ms (execution: 0.566 s)
         +------------+---------------+--------------+----------------+---------------+--------+---------+---------+--------------------+-----------------+
         |TripDistance|PickupLongitude|PickupLatitude|DropoffLongitude|DropoffLatitude|RateCode|DayOfWeak|HourOfDay|            Features|       prediction|
         +------------+---------------+--------------+----------------+---------------+--------+---------+---------+--------------------+-----------------+
         |         1.7|     -73.960213|     40.770464|      -73.979863|       40.77705|       1|        5|       20|[1.70000004768371...| 8.64848592515177|
         |         0.9|     -73.995371|     40.717248|      -73.984367|      40.720524|       1|        5|       20|[0.89999997615814...|7.402053100375133|
         |         0.9|     -73.983811|     40.749655|      -73.989747|      40.756575|       1|        5|       20|[0.89999997615814...| 7.12812526226065|
         +------------+---------------+--------------+----------------+---------------+--------+---------+---------+--------------------+-----------------+

         21/05/11 15:25:30 INFO MapPartitionsRDD: Removing RDD 30 from persistence list
ducer    | Sleeping 10sec
```



```
+------------+---------------+--------------+----------------+---------------+--------+---------+---------+--------------------+------------------+
|TripDistance|PickupLongitude|PickupLatitude|DropoffLongitude|DropoffLatitude|RateCode|DayOfWeak|HourOfDay|            Features|        prediction|
+------------+---------------+--------------+----------------+---------------+--------+---------+---------+--------------------+------------------+
|         3.6|     -73.984138|     40.726317|      -73.962869|      40.758443|       1|        5|       20|[3.59999990463256...| 13.31525507597342|
|         2.1|     -73.979906|      40.74585|       -73.95909|      40.773639|       1|        5|       20|[2.09999990463256...| 9.667650540227058|
|         3.4|     -73.981147|     40.758918|       -73.94251|      40.785975|       1|        5|       20|[3.40000009536743...|13.280766151283803|
+------------+---------------+--------------+----------------+---------------+--------+---------+---------+--------------------+------------------+

21/05/11 15:26:00 INFO JobScheduler: Finished job streaming job 1620746760000 ms.0 from job set of time 1620746760000 ms
```

Pomoćne funkcije



```java
public static int getDayOfWeekFromDate(String stringDate) {
    // example format: 2014-01-09 20:45:25
    SimpleDateFormat formatter = new SimpleDateFormat("yyyy-MM-dd HH:mm:ss");
    Date date = new Date();

    try {
        date = formatter.parse(stringDate);
    } catch (ParseException e) {
        e.printStackTrace();
    }

    Calendar calendar = Calendar.getInstance();
    calendar.setTime(date);
    return calendar.get(Calendar.DAY_OF_WEEK); // the day of the week in numerical format
}

public static int getHourOfDayFromDate(String stringDate) {
    // example format: 2014-01-09 20:45:25
    SimpleDateFormat formatter = new SimpleDateFormat("yyyy-MM-dd HH:mm:ss");
    Date date = new Date();

    try {
        date = formatter.parse(stringDate);
    } catch (ParseException e) {
        e.printStackTrace();
    }

    Calendar calendar = Calendar.getInstance();
    calendar.setTime(date);
    return calendar.get(Calendar.HOUR_OF_DAY); // the hour of the day in numerical format
}
```