Project: Traffic Prediction

Milan Lagae

Institution/University Name: Vrije Universiteit Brussel
Faculty: Sciences and Bioengineering Sciences
Course: Big Data Processing

Contents

1 Intro

This report will discuss an implementation for the assignment "Project: Traffic Prediction" for the course: Big Data Processing. First the implementation itself will be discussed in section §2. Following that, answers to the required questions in section §3. And lastly a small section on performance benchmarks in section §4.

2 Implementation

This section will discuss the implementation (code) for the project. Full project code can be found in the associated Apache Spark project or small snippets will be placed in the text or larger ones in the Appendix section §5.

2.1 Overview

All the code can be found in the traffic package of the bdp-traffic folder. The traffic package consists of the following files:

- Traffic.scala
- TrafficLoader.scala
- TrafficJoiner.scala
- TrafficTimeSeries.scala
- TrafficTransformer.scala
- TrafficPredictor.scala

The order of the file is in which they are structured & applied to the input. Each file also has it's own logger variable set, which is used for logging, for this the build.sbt file was modified with an additional package.

2.2 Traffic.scala

This is the file that is executed when the project is ran, it executes the different steps (files) in a kind of pipeline. The complete pipeline can be seen in listing: Figure 1.

/

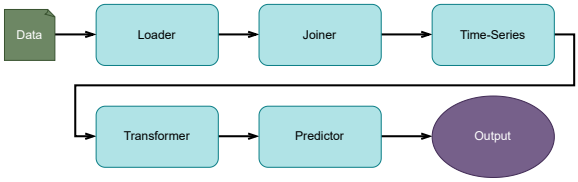

Figure 1: Execution Pipeline

2.3 Loader

2.4 Joiner

2.5 Time-Series

2.6 Transformer

2.7 Predictor

2.8 Output

3 Discussion

3.1 Question 1

3.2 Question 2

3.3 Question 3

3.4 Question 4

4 Benchmarks

5 Appendix