

Big Data Systems and Architectures

Exploring International Flights in 2017 Data

Having completed the specific assignment my feedback is positive. I achieved to learn the basic syntax of Spark and the different possibilities to create a Linear Regression model through it.

The difficulties of the task was ranked, so as each of us, without having a technical background could answer at least the first two tasks. It will be really helpful, to provide us a description document, explaining the columns of the dataset. I used the internet in order to understand the meaning of such variables like ORIGIN and CARRIER, which indicate the airport and airway respectively. The most important help that should be included in this assignment is a tutorial with steps to install the Spark. Although I am working as a software engineer and I have experience with such installations, I spent a lot of hours for debugging and installations of Spark. As it is known, three different assignments were having the same deadline, so it was a great loss of time the installation and the fixes of the problems.