ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

Scuola di Ingegneria e Architettura
Laurea Magistrale in Ingegneria e Scienze Informatiche

USED CARS ANALYSIS

Elaborato in
BIG DATA

Presentata da
GABRIELE GUERRINI

Anno Accademico 2019 – 2020

Table of contents

1 Dataset and query **1**
1.1 Dataset 1
1.2 Query 1

2 MapReduce **3**
2.1 Workflow 3
2.2 Jobs 3
2.3 Performance evaluation 3

3 Spark **5**
3.1 Workflow 5
3.2 Performance evaluation 5

4 Conclusions **7**

Chapter 1

Dataset and query

1.1 Dataset

1.2 Query

Chapter 2

MapReduce

2.1 Workflow

2.2 Jobs

2.3 Performance evaluation

Chapter 3

Spark

3.1 Workflow

3.2 Performance evaluation

Chapter 4

Conclusions