Brunel University London

College of Engineering, Design and Physical Science
Electronic and Computer Engineering

Assignment
Java Testing and Measuring

Distributed Computing Systems Engineering Msc

Author: Christoph Gschrey

Lab-Partner: Matthias Gebert

Date: 22. September 2017

Supervisor: Prof. Dr. Peter Väterlein

College of Engineering, Design and Physical Sciences

Department of Engineering and Design

ASSIGNMENT SUBMISSION FORM

Please note: that no course work will be accepted without this cover sheet.

Please ensure: that you keep a copy of work submitted and retain your receipt in case of query.

Student Number:	**SPO ID Number** (Office use only):	
Course:		**Level:**

MODULE	
Module Code:	**Module Title:**
Lab / Assignment:	**Deadline:**
Lab group (if applicable):	**Date Stamp** (Office use only):
Academic Responsible:	
Administrator:	

Please note: that detailed feedback will be provided on a feedback form.

✂.....

RECEIPT SECTION (Office Copy)	
Student Number:	**SPO ID Number** (Office use only):
Student First Name:	**Student Last Name:**
Module Code:	**Module Title:**
Lab / Assignment:	
Lab group (if applicable):	**Deadline:**
Academic Responsible:	**Number of Days late:**

DECLARATION	
I have read and I understand the guidelines on plagiarism and cheating in the Handbook and I certify that my contribution to this report fully complies with these guidelines. I confirm that I have kept a copy of my work and that I have not lent my work to any other students.	
Signed:	**Date Stamp** (Office use only):

✂.....

RECEIPT SECTION (Student Copy)	
Student Number:	**Student Name**:
Lab / Assignment:	
Lab group (if applicable):	**Module Title:**
Academic Responsible:	**Deadline:**
Module Code:	**Date Stamp** (Office use only):

The University penalty system will be applied to any work submitted late.

IMPORTANT: You **MUST** keep this receipt in a safe place as you may be asked to produce it at any time as proof of submission of the assignment. Please submit this form with the assignment attached to theDepartment of Design Education Office in the Michael Sterling Building, room MCST 055.

Contents

1 Introduction

As the last few years and decades have seen ever-increasing amounts of data analyzed and managed in ever shorter periods of time, a challenge faced by the IT industry has been the limited computing power of individual machines and simple networks. Tasks are constantly evolving, so they need more and more computing power to solve them.
In order to solve these problems, various techniques have been developed that help to continuously improve the performance of the machines. These include supercomputers, computer clusters, and various methods and algorithms that can be summarized under the term High Performance Computing.
This workshop introduces the programming model **MapReduce** and a simulation of the model based on the **Hadoop** framework. The aim of this workshop was to use the Hadoop Simulation HSim to gain a basic understanding of the functionality of MapReduce.

2 High Performance Computing With MapReduce and Hadoop

As already mentioned in the introduction, MapReduce was introduced as a programming model to process large amounts of data (Big Data) in parallel on several machines. This reduces processing time, since the load can be distributed over multiple machines.
The following chapters briefly explain the basics of this workshop. First, MapReduce and its functionality are described. This is followed by a short introduction to Hadoop.

2.0.1 MapReduce

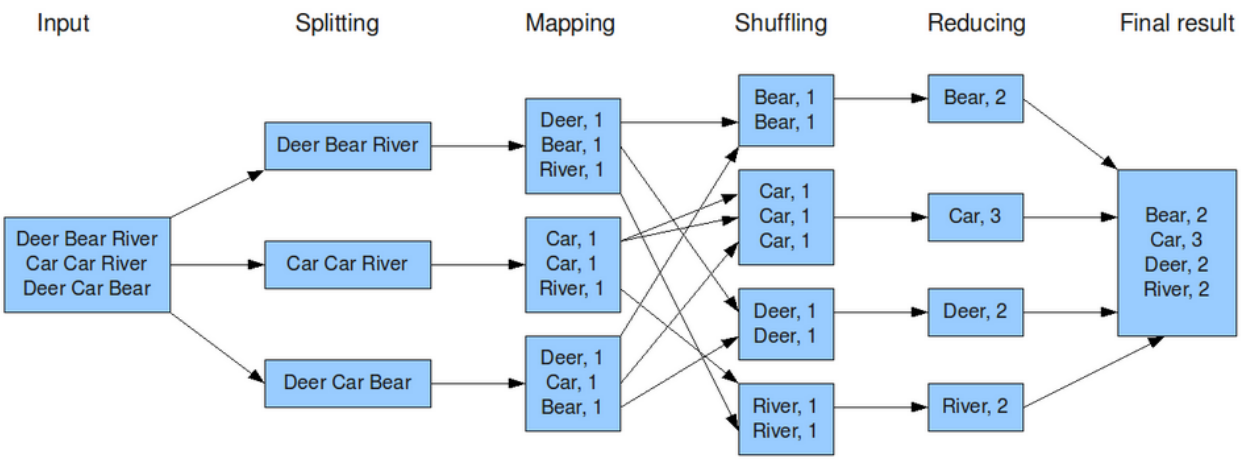

Figure 2.1: The MapReduce model[1]

2.0.2 Hadoop

[see Ap1]

[1] image-source: https://cs.calvin.edu/courses/cs/374/exercises/12/lab

3 Hadoop's Parameters

4 HSim

As already mentioned in the introduction, no complete cluster is being set up for this workshop, as the effort would have been far too much. Instead the provided software *HSim* was used, which provides a simulated Hadoop environment with several machines, routers and nodes. HSim and is based on Java and Hadoop v2 and it offers the possibility to adjust different Hadoop parameters and to display the effects on the simulation realistically.

5 Evaluation

5.0.1 Changing the dataSize

5.0.2 Changing the sortFactor

5.0.3 Changing the requiredMappers and requiredReducers

6 Conclusion

Bibliography

[Ap1] *HCFS*. URL: `https://wiki.apache.org/hadoop/HCFS` (cit. on p. 2).