Big Data Analytics

Exercise Sheet 1

Prof. Dr.Dr. Lars Schmidt-Thieme, Mohsan Jameel

Information Systems and Machine Learning Lab University of Hildesheim

April 14th, 2016

Submission until April 20th, 2016, to mohsan.jameel@ismll.de

Exercise 1: Introduction to Big Data Platforms (5 points)

a) Elaborate with examples difference between Shared memory system and Distributed memory system. Point out the main limitations of both the systems.

b) What are the different parallel softwares available for shared and distributed memory programming?

Exercise 2: Race condition (5 points)

a) Define the determinacy race. With the help of a pseudocode (or a code snippet) show how can determinacy race occur.

Exercise 3: Multithreaded Program with Java Threads (5 points)

Convert the following serial program into a multithreaded program such that sum and min functions run in parallel (each in a separate thread). You have to provide two solutions. Solution1 will be implemented by extending Thread class and Solution 2 through implementing Runnable interface.

```
import java.util.Random;

public class MultiThread {
    public static int[] vector;
    public static int size;
    public static int sum = 0;
    public static int min;
    public static int max = 655859;
    public void findMin() {
      min = max;
      for(int i = 0; i < size; i++){
            if(min > vector[i]){
                  min = vector[i];
            }
        }
    }
    public void calSum() {
      for(int i = 0; i < size; i++){
            sum +=vector[i];
        }
    }
```

```java
    public void initVector(int[] v){
        Random rand = new Random();
        for(int i = 0; i < size; i++){
            v[i] = Math.abs(rand.nextInt()%max);
        }

    }
    public static void main(String[] args) {
        size = 1000000;
        vector= new int[size];
        initVector(vector);
        findMin();
        calSum();
        System.out.println("Main Sum=" + sum + " Min="+min);
    }
}
```