

Homework 10

Introduction to Big Data Systems

N Y Hendrikx < 2019403077 >

May 24, 2019

1

This report contains the details of the homework 10. The task is to implement consistent hashing and answer questions. In the project attachment I added the file 'consistent_hashing.py'. Also the plots are attached from this document.

1.1

I plotted the load imbalance ratio per amount of nodes per machine.

Figure 1: Load imbalance ratio, step size 1. nodes 1 to 200.

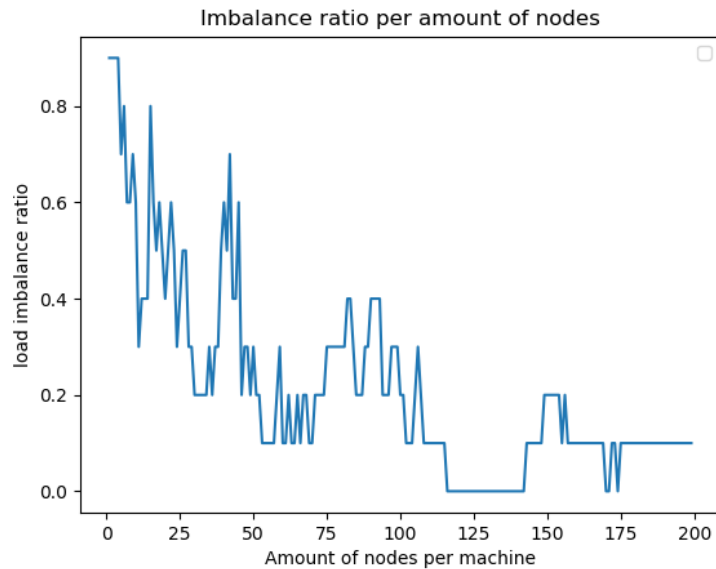


Table 1: 10 machines with each 120 nodes										
Machine	0	1	2	3	4	5	6	7	8	9
Load	11094	11419	10020	10503	9467	9152	9080	8840	11118	9307

Table 2: 11 machines with each 120 nodes											
Machine	0	1	2	3	4	5	6	7	8	9	10
Load	10372	10036	9093	9043	9059	8227	8564	7912	9372	8577	9745

1.2

Theoretically: When using 120 nodes will produce a load imbalance of 0. If all the machines are balanced it means it will be likely that it still we balanced when adding a new machine. So $0.15 * 10000 = 1500$. So that will most likely be around 10000 with a maximum of 1500. Actually this has to be lower because there are 11 machines instead of 10.

Visible in the table with 11 machines that machine 11 has a load of **9745** meaning those keys have been relocated there. This is the exact answer.

2

2.1

Read will be slow, write fast. Poor durability.

2.2

Read will be fast, write slow. Poor durability.