

Investigation of Hadoop System

Adam Tóth ^a, Tamás Bérczes ^b, Ramin Karimi ^c

^aFaculty of Informatics, University of Debrecen
adamtoth102@gmail.com

^bFaculty of Informatics, University of Debrecen
berczes.tamas@inf.unideb.hu

^cFaculty of Informatics, University of Debrecen
raminkm2000@yahoo.ca

Abstract

One of the most crucial components of cluster-level infrastructure layers is resource management. In such systems it is necessary to determine the job requirements as a configuration parameter (CPU, RAM, disk I/O, network I/O) and have the scheduler translate those into an appropriate reservation and allocation of resources. In the Hadoop system YARN is an occurrent resource management, which supports only RAM and CPU reservation at present.

The novelty of this investigation is to determine the limits of the system using special data like DNA signatures in large databases such as bacterial genome database. We investigate how the different parameters like cpu usage, memory consumption, disk utilization have influence on processing of the map tasks.

Keywords: Big data, Hadoop