Performance Comparison for Mining Large Data from the Internet and Learning using ID3 Algorithm in a Docker versus Virtual Machine Environment

Abstract

Every day, 2.5 quintillion bytes of data are generated. A sizeable portion of the data is available through the internet. The efficacy of the decisions being made revolves around the extent to which analysis is performed on the procured data. Containers provide Operating System Virtualization and Linux Containers present secure execution environments by independently executing processes.[1] This paper aims at proving that the performance of Docker Container in mining large data from the internet and learning using ID3 algorithm to generate a decision tree to predict useful results is much better than the performance in a Virtual Machine Environment.

References

7. https://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm
8. Jeffrey Dean and Sanjay Ghemawat, “MapReduce: Simplified Data Processing on Large
9. Carl Boettiger, An introduction to Docker for reproducible research, with examples from
   the R environment, (2015) ACM SIGOPS Operating Systems Review, Special Issue on
   Repeatability and Sharing of Experimental Artifacts. 49(1), pp.71-79.

Index Terms

Computer Science

Algorithms

Keywords

Docker, Container, Virtual Machine.