Scalability Limits of HDFS

Christopher Chute, David Brandfonbrener, Leo Shimonaka, Matt Vasseur

May 2, 2016

Abstract

1 Introduction

2 HDFS Architecture

2.1 Namenode

2.2 Datanode

3 Limitations of Single Namenode Architecture

3.1 Physical Memory Size

3.2 Namenode CPU Bottleneck

4 Performance Evaluation

4.1 Experiment 1: Memory Experiments

4.2 Experiment 2: CPU Experiments

4.3 Analysis

5 Improvements to Single Namenode Architecture

6 Conclusion