Technical University of Berlin(TU Berlin)

Master Thesis

An analysis of side-effects of MapReduce optimizations in the data-center

Author:
Khwaja Zubair Sediqi

Supervisor:
Prof. Anja Feldman

A thesis submitted in fulfilment of the requirements
for the degree of Master in Computer Science

in the

Intelligent Network Group(IN)
Faculty IV

October 2013

Declaration of Authorship

I, Khwaja Zubair Sediqi, declare that this thesis titled, 'An Analysis of Side Effects of Hadoop's Optimization in Data Center' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

"Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism."

Dave Barry

TECHNICAL UNIVERSITY OF BERLIN (TU BERLIN)

Abstract

Faculty of Electrical Engineering and Computer Science
Department of Telecommunication Systems

Master of Computer Science

An Analysis of Side Effects of Hadoop's Optimization in Data Center

by Khwaja Zubair Sediqi

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Acknowledgements

The acknowledgements and the people to thank go here, don't forget to include your project advisor...

Contents

List of Figures

List of Tables

Abbreviations

LAH **L**ist **A**bbreviations **H**ere

Physical Constants

Speed of Light $c = 2.997\ 924\ 58 \times 10^8\ \mathrm{ms}^{-\mathrm{S}}$ (exact)

Symbols

a	distance	m
P	power	W (Js^{-1})
ω	angular frequency	$rads^{-1}$

For/Dedicated to/To my...

Chapter 1

Introduction

1.1 Introduction

The number of internet users grows rapidly and today's most popular applications are internet based applications such as, social media, e-commerce,etc. As the Internet applications serves millions of users around the globe, so the amount generated data is also huge. Users that interacts with internet generates various data such as click-stream data,crawled web documents, web requests, logs etc. The greater the number of users interact with system the more data is going to generate,since generated data is very larger the term Big Data is used for such set of data. For example, the number of monthly active users on Facebook; the largest social network in the world has been 1.15 billion users as on June 2013.[**?**] The interaction of such huge number of users with facebook application generates Big Data. Such Big Data is is potential gold mine for the companies to understand access pattern and ad revenue of the companies.[**?**] Traditional database systems have difficulty to process big data, hence new data management and processing techniques is required to process big data.[**?**]

MapReduce is programming paradigm used to process big data in two phases of map and reduce. Hadoop is an open source implementation of MapReduce in java, used to process big data. Hadoop performance is optimized by optimization in its scheduler.

1.2 Objective

The process of big data requires more computational resources to be performed. There are cloud services that provided computational resources to the companies, where companies can use the resources and use it according to their need. An example of such

cloud computing resources is AMAZON EC2 cloud computing.

1.3 Contribution

1.4 Results

1.5 Thesis Overview

The thesis consist of six chapters as follow. Chapter1 is about introduction of thesis and topic covered in this thesis. In chapter2 the information about background and related work is covered. Chapter3 explains the methodology used to complete the experiment and how to answer to research questions.Chapter4 provides detail information about experimental environment including the tools software used for experiments. Chapter5 is about evaluation of the experiment where detail of each experiments along with short discussion about result of experiment is discussed. Chapter6 provides analysis of cloud computing service for Afghanistan market. The last chapter, chapter7 is wrap up of the work with conclusion of thesis and future work. The bibliography and appendixes is attached at the end of the thesis.

Appendix A

Appendix Title Here

Write your Appendix content here.