



**SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI**

**IMPLEMENTASI DAN ANALISIS KINERJA HDFS SEBAGAI  
INFRASTRUKTUR PEMBANGUNAN BIG DATA**

**TUGAS AKHIR**

**YUNITA SURAHMAN  
0110215029**

**PROGRAM STUDI TEKNIK INFORMATIKA  
JAKARTA  
FEBRUARI 2019**



**SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI**

**IMPLEMENTASI DAN ANALISIS KINERJA HDFS SEBAGAI  
INFRASTRUKTUR PEMBANGUNAN BIG DATA**

**TUGAS AKHIR**

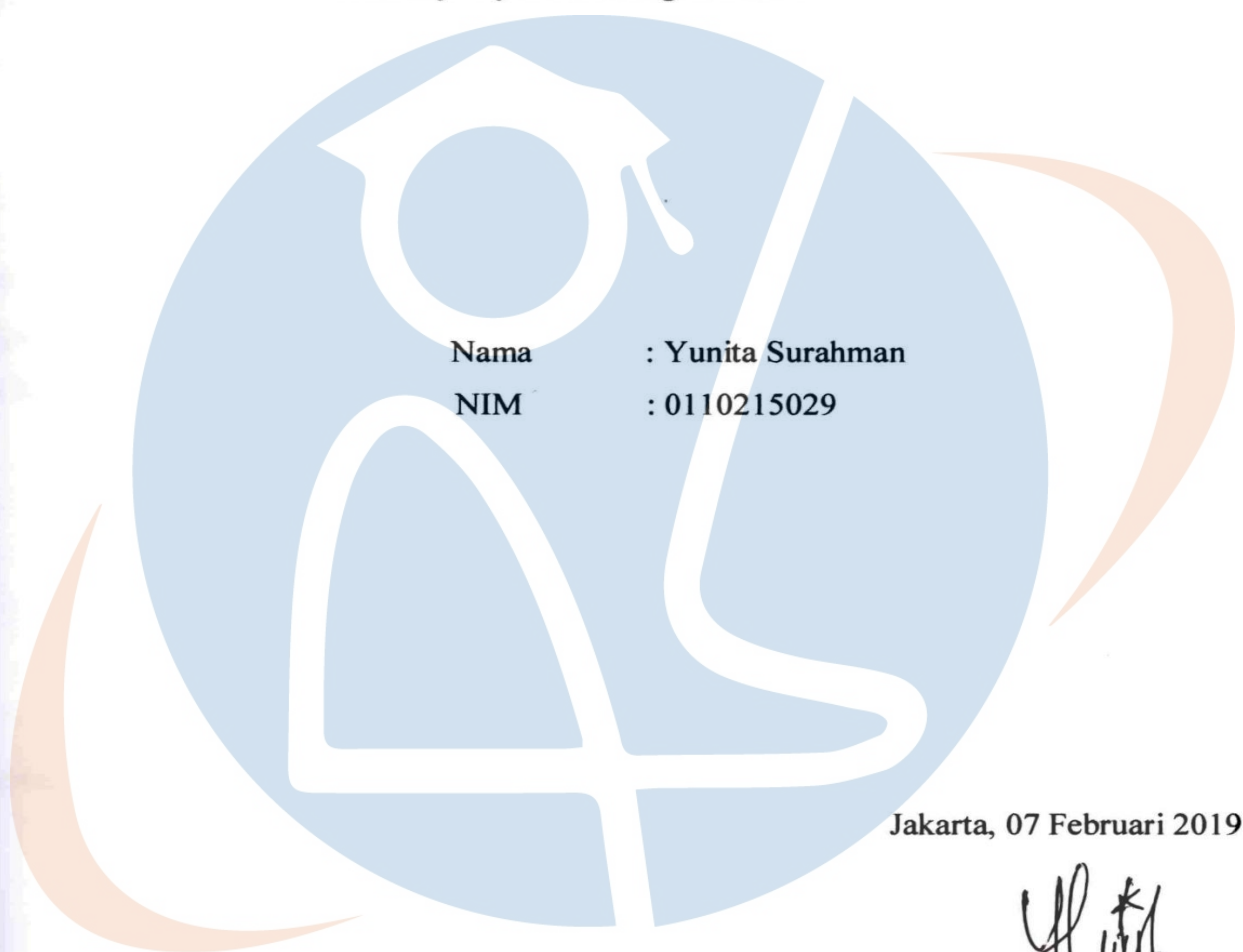
**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer**

**YUNITA SURAHMAN  
0110215029**

**PROGRAM STUDI TEKNIK INFORMATIKA  
JAKARTA  
FEBRUARI/2019**

## HALAMAN PERNYATAAN ORISINALITAS

Skripsi/Tugas Akhir ini adalah hasil karya sendiri,  
dan semua sumber baik yang dikutip maupun dirujuk  
telah saya nyatakan dengan benar.



Nama : Yunita Surahman  
NIM : 0110215029

Jakarta, 07 Februari 2019



(Yunita Surahman)

STT - NF

## HALAMAN PENGESAHAN

Skripsi/Tugas Akhir ini diajukan oleh:

Nama : Yunita Surahman  
Nim : 0110215029  
Program Studi : Teknik Informatika  
Judul Skripsi : Implementasi dan Analisis Kinerja HDFS sebagai Infrastruktur  
Pembangunan *Big data*

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Terpadu Nurul Fikri.

### DEWAN PENGUJI

Pembimbing : Henry Saptono, S.Si, M.Kom

(  )

Penguji I : Sirojul Munir, S.Si, M.Kom

(  )

Penguji II : Hilmy Abidzar Tawakal, S.T, M.Kom

(  )

## KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Skripsi/Tugas Akhir ini. Penulisan Skripsi/Tugas Akhir ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Program Studi Teknik Informatika pada Sekolah Tinggi Teknologi Terpadu Nurul Fikri. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan Skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan Skripsi/Tugas Akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Allah SWT yang telah memberikan rahmat, karunia, dan hidayahnya sehingga saya dapat menyelesaikan penyusunan skripsi/tugas akhir ini.
2. Bapak Henry Saptono, S.Si, M.Kom selaku dosen pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan dalam penyusunan skripsi/tugas akhir ini.
3. Bapak Sirojul Munir, S.Si, M.Kom dan Bapak Hilmy Abidzar Tawakal, S.T, M.Kom selaku dosen penguji yang telah bersedia menyediakan waktunya untuk menguji serta memberikan saran dalam penyusunan skripsi/tugas akhir ini.
4. Orang tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral.
5. Teman-Teman saya, Agung Prayoga, Christine Ayu, Debi Onas Saputra, dan Erick Davied Prayoga yang selalu memberikan support dan dukungannya sehingga saya bisa menyelesaikan penyusunan skripsi/tugas akhir ini.
6. Imam Jazuli

Akhir kata, saya berharap Allah SWT berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga skripsi/tugas akhir ini membawa manfaat bagi pengembangan ilmu.

STT - NF

Jakarta, 07 Februari 2019

Penulis

Yunita Surahman

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI  
TUGAS AKHIR UNTUK KEPERLUAN AKADEMIS**

Sebagai civitas akademik Sekolah Tinggi Teknologi Terpadu Nurul Fikri, saya yang bertanda tangan dibawah ini:

Nama : Yunita Surahman  
NIM : 0110215029  
Program Studi : Teknik Informatika  
Jenis Karya : Skripsi/Tugas Akhir

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada STT Nurul Fikri Hak Bebas Royalti Noneksklusif (Non-Exclusive Royalti – Free Right) atas karya ilmiah saya yang berjudul:

**Implementasi Dan Analisis Kinerja HDFS Sebagai Infrastruktur Pembangunan *Big Data***

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini STT-NF berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Dengan ini pernyataan saya buat dengan sebenarnya.

Dibuat di : Depok  
Pada tanggal : 07 Februari 2019  
Yang menyatakan



(Yunita Surahman)

## ABSTRAK

Nama : Yunita Surahman  
Program Studi : Teknik Informatika  
Judul : Implementasi Dan Analisis Kinerja HDFS Sebagai Infrastruktur  
Pembangunan *Big Data*

Perkembangan teknologi informasi yang cepat menyebabkan pertumbuhan dan peningkatan jumlah data, VCloud News mencatat pada tahun 2015, pertumbuhan data statistic per hari mencapai 2,5 quintillion (10 pangkat 18) byte atau sekitar Giga Byte (GB) perdetik. Data yang begitu banyak yang beragam dan tidak terstruktur, dengan data yang begitu banyak membuat suatu perusahaan atau organisasi mencari cara untuk mengolah atau menyimpan data tersebut. Dari permasalahan tersebut, sehingga dilakukan penelitian untuk implementasi dan analisis kinerja Hadoop File System (HDFS) sebagai infrastruktur pembangunan *big data*. Tujuan dan manfaat dari penelitian ini adalah untuk mengetahui faktor-faktor yang digunakan untuk menentukan kinerja HDFS, melakukan penilaian atau pengukuran unjuk kerja HDFS, Memahami langkah-langkah yang dilakukan, dan menghasilkan suatu karya tulis yang bisa menjadi rujukan didalam mengimplementasikan HDFS sebagai infrastruktur pembangunan *Big Data*. Impelementasi ini dilakukan pada 3 node yaitu 1 Namenode dan 2 Datanode yang akan diimplementasikan pada lingkungan virtualisasi menggunakan sistem operasi Linux Ubuntu 16.04. Pada penelitian ini, akan dilakukan beberapa pengujian yaitu, pengujian performa I/O yang dilakukan menggunakan aplikasi *benchmark* TestDFSIO dan NNBench, pengubahan *blocksize*, dan penghapusan file pada beberapa ukuran file yang berbeda, lalu akan dilakukan pengujian availability. Pada pengujian menggunakan TestDFSIO menghasilkan waktu eksekusi dan throughput yang menunjukkan semakin besar ukuran suatu file maka akan mempengaruhi waktu eksekusi dan throughputnya. Pengujian NNBench menghasilkan waktu eksekusi yang lebih lama saat jumlah file yang dieksekusi ditambah. Pengubahan *blocksize* mempengaruhi lamanya waktu eksekusi yang akan dihasilkan saat menginputkan file pada HDFS, dimana terjadi penurunan waktu eksekusi saat ukuran *blocksize*nya dirubah.

Kata Kunci : Kinerja, HDFS, TestDFSIO, NNBench, Blocksize.

## **ABSTRACT**

*Name : Yunita Surahman*  
*Study Program : Informatics Engineering*  
*Title : Implementation and Analysis of HDFS Performance as Infrastructure*  
*Big Data Development*

*The rapid development of information technology has led to growth and an increase in the amount of data, VCloud News noted in 2015, the growth of statistical data per day reached 2.5 quintillion (10 to the power 18) bytes or about Giga Byte (GB) per second. So much data that is diverse and unstructured, with so much data makes a company or organization look for ways to process or store that data. From these problems, research was carried out for the implementation and performance analysis of the Hadoop File System (HDFS) as a big data development infrastructure. The purpose and benefits of this study are to find out the factors used to determine HDFS performance, carry out assessment or measurement of HDFS performance, understand the steps taken, and produce a paper that can be a reference in implementing HDFS as Big development infrastructure Data. This implementation is carried out on 3 nodes, namely 1 Namode and 2 Datasodes which will be implemented in a virtualization environment using the Ubuntu Linux operating system 16.04. In this study, several tests will be carried out, namely, I / O performance testing using the benchmark TestDFSIO and NNbench applications, changing blocksize, and deleting files in several different file sizes, then availability testing will be performed. In testing using TestDFSIO produces execution time and throughput which shows the greater the size of a file it will affect the execution time and throughput. The NNbench test results in a longer execution time when the number of files executed is added. Changing blocksize affects the length of execution time that will be generated when inputting files on HDFS, where the execution time decreases when the blocksize size is changed.*

*Keywords: Performance, HDFS, TestDFSIO, NNbench, Blocksize.*



## DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS.....	ii
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR .....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPERLUAN AKADEMIS.....	v
ABSTRAK.....	vi
<i>ABSTRACT</i> .....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xiii
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar belakang.....	1
1.2 Perumusan Masalah.....	2
1.3 Tujuan dan Manfaat Penelitian.....	2
1.4 Batasan Masalah.....	3
1.5 Sistematika Penulisan.....	3
BAB II KAJIAN LITERATUR.....	5
2.1 <i>Big Data</i> .....	5
2.2 Kinerja.....	6
2.3 Hadoop.....	7
2.3.1 Hadoop File System (HDFS).....	9
2.3.2 Block Size.....	13
2.3.3 MapReduce.....	13

2.3.4	YARN (Yet Another Resource Negotiator).....	13
2.4	Java Development Kit (JDK) .....	14
2.5	Secure Shell (SSH) .....	14
2.6	NameNode Benchmark (NNBench).....	15
2.7	TestDFSIO.....	15
2.8	Throughput .....	15
2.9	Penelitian Terkait.....	16
BAB III METODOLOGI PENELITIAN.....		19
3.1	Tahapan Penelitian .....	20
3.1.1	Studi Literatur .....	21
3.1.2	Analisis Kebutuhan Sistem.....	21
3.1.3	Perancangan Sistem .....	21
3.1.4	Implementasi.....	21
3.1.5	Pengujian & Analisis Hasil.....	22
3.1.6	Penarikan Kesimpulan dan Saran .....	22
3.2	Rancangan Penelitian .....	22
3.2.1	Jenis Penelitian .....	22
3.2.2	Metode Pengumpulan Data.....	23
3.2.3	Lingkungan Pengembangan.....	23
3.2.4	Alat.....	23
BAB IV .....		24
ANALISIS DAN PERANCANGAN.....		24
4.1	Analisis Kebutuhan.....	24
4.1.1	Analisis Kebutuhan Hardware .....	24
4.1.2	Analisis Kebutuhan Software .....	25

4.2	Perancangan Sistem.....	25
4.2.1	Perancangan Arsitektur Sistem.....	26
4.2.2	Rancangan Pengujian.....	28
BAB V.....		38
IMPLEMENTASI DAN PENGUJIAN.....		38
5.1	Implementasi.....	38
5.1.1	Persiapan.....	38
5.1.2	Instalasi dan Konfigurasi.....	39
5.2	Pengujian.....	58
5.2.1	Pengujian Performa I/O.....	58
5.2.2	Pengujian <i>Availability</i> .....	83
5.3	Analisis Hasil.....	88
5.3.1	Analisis Hasil Pengujian Performa I/O.....	88
5.3.2	Analisis Hasil Pengujian <i>Availability</i> .....	96
BAB VI.....		97
KESIMPULAN DAN SARAN.....		97
6.1	Kesimpulan.....	97
6.2	Saran.....	97
DAFTAR PUSTAKA.....		98

STT - NF

## DAFTAR GAMBAR

Gambar 2.1.1 Karakteristik Big Data .....	5
Gambar 2.4.1 Penyimpanan Data pada HDFS.....	10
Gambar 3.1.1 Tahapan Penelitian.....	20
Gambar 4.2.1.1. Rancangan Arsitektur Sistem.....	26
Gambar 4.2.1.2 Proses write pada HDFS .....	27
Gambar 4.2.1.3 Proses read pada HDFS.....	28
Gambar 4.2.2.1. Alur Pengujian .....	29
Gambar 5.1.2.6.1 Mengaktifkan Service HDFS .....	55
Gambar 5.1.2.6.2 Mengaktifkan Service Yarn .....	56
Gambar 5.1.2.6.3 Output JPS pada Namenode.....	56
Gambar 5.1.2.6.4 Output JPS pada Datanode1 .....	56
Gambar 5.1.2.6.5 Output JPS pada Datanode2.....	56
Gambar 5.1.2.6.6 Tampilan Hadoop melalui web browser .....	57
Gambar 5.1.2.6.7 Tampilan Hadoop melalui web browser .....	57
Gambar 5.1.2.6.8 Lives Nodes.....	58
Gambar 5.2.2.1 Space Harddisk datanode1 .....	83
Gambar 5.2.2.2 Space Harddisk datanode2 .....	84
Gambar 5.2.2.3 Lives Nodes.....	84
Gambar 5.2.2.4 Pengecekan file percobaan menggunakan web browser .....	85
Gambar 5.2.2.5 Isi file percobaan.....	86
Gambar 5.2.2.6 Pengecekan file percobaan menggunakan command prompt .....	86
Gambar 5.2.2.7 Lives Nodes.....	86
Gambar 5.2.2.8 Space Harddisk datanode pada web browser .....	87
Gambar 5.2.2.9 Space Harddisk datanode2 .....	87
Gambar 5.2.2.10 Pengecekan file pada datanode2 .....	88
Gambar 5.2.2.11 Isi dari file percobaan.....	88
Gambar 5.3.1.1.1 Grafik Hasil Pengujian Waktu Eksekusi Write TestDFSIO .....	89
Gambar 5.3.1.1.2 Grafik Hasil Pengujian throughput Write TestDFSIO.....	90
Gambar 5.3.1.1.3 Grafik Hasil Pengujian Waktu Eksekusi Read TestDFSIO .....	91

Gambar 5.3.1.1.4 Grafik Hasil Pengujian throughput Read TestDFSIO..... 92  
Gambar 5.3.1.2.1 Grafik Hasil Pengujian Create NNBench ..... 93  
Gambar 5.3.1.2.2 Grafik Hasil Pengujian Read NNBench..... 93  
Gambar 5.3.1.3.1 Grafik Hasil Pengujian pengubahan blocksize ..... 94  
Gambar 5.3.1.4.1 Grafik Hasil Pengujian Penghapusan File ..... 95



## DAFTAR TABEL

Tabel 2.7.1 Penelitian Terkait.....	18
Tabel 4.1.1.1. Spesifikasi Rancangan .....	25
Tabel 4.2.2.1.1. Tabel Pengujian waktu eksekusi TestDFSIO Write .....	30
Tabel 4.2.2.1.2 Tabel Pengujian throughput TestDFSIO Read .....	31
Tabel 4.2.2.1.3. Tabel Pengujian waktu eksekusi TestDFSIO Read .....	31
Tabel 4.2.2.1.4 Tabel Pengujian throughput TestDFSIO Read .....	31
Tabel 4.2.2.2.1. Tabel Pengujian NNbench Write .....	33
Tabel 4.2.2.2.2. Tabel Pengujian NNbench Read .....	33
Tabel 4.2.2.3.1 Hasil Pengujian write file 500 MB .....	34
Tabel 4.2.2.3.2 Hasil Pengujian write file 1 GB .....	34
Tabel 4.2.2.3.3 Hasil Pengujian write file 1.5 GB .....	35
Tabel 4.2.2.3.4 Hasil Pengujian write file 2.5 GB .....	35
Tabel 4.2.2.3.5 Hasil Pengujian write file 3 GB .....	36
Tabel 4.2.2.4.1 Hasil Pengujian penghapusan file .....	36
Tabel 5.2.1.1.1 Tabel Hasil Pengujian lama waktu write TestDFSIO .....	63
Tabel 5.2.1.1.2 Tabel Hasil Pengujian Throughput write TestDFSIO .....	64
Tabel 5.2.1.1.3 Tabel Hasil Pengujian lama waktu Read TestDFSIO .....	67
Tabel 5.2.1.1.4 Tabel Hasil Pengujian throughput Read TestDFSIO .....	69
Tabel 5.2.1.2.1 Tabel Hasil Pengujian Create NNbench .....	72
Tabel 5.2.1.2.2 Tabel Hasil Pengujian Read NNbench .....	73
Tabel 5.2.1.3.1 Tabel Hasil Pengujian file 500 MB .....	75
Tabel 5.2.1.3.2 Tabel Hasil Pengujian File 1 GB .....	77
Tabel 5.2.1.3.3 Tabel Hasil Pengujian 1.5 GB .....	78
Tabel 5.2.1.3.4 Tabel Hasil Pengujian 2.5 GB .....	79
Tabel 5.2.1.3.5 Tabel Hasil Pengujian 3 GB .....	80
Tabel 5.2.1.4.1 Tabel hasil pengujian delete file .....	82