

DAFTAR ISI

HALAMAN PENGESAHAN.....	iii
PERNYATAAN.....	iv
HALAMAN MOTTO DAN PERSEMBAHAN.....	v
PRAKATA.....	vii
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xv
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN	xx
INTISARI.....	xxi
ABSTRACT.....	xxi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Metodologi Penelitian	5
1.7 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA	8
BAB III LANDASAN TEORI.....	12
3.1 Data	12
3.2 Big Data.....	13
3.2.1 Hadoop	15
3.2.2 HDFS	15



3.3	Apache Spark	16
3.4	MapReduce.....	17
3.4.1	Pengertian MapReduce	17
3.4.2	Model Pemrograman	18
3.5	Audit Trail	19
3.6	User Profiling	20
3.7	Proses Agregasi Data.....	20
3.8	Visualisasi Informasi	21
3.9	Plotly	22
3.10	Python dalam Agregasi Pyspark Mapreduce.....	22
3.11	Pengujian (Testing)	22
BAB IV	ANALISIS DAN PERANCANGAN	24
4.1	Analisis Permasalahan.....	24
4.2	Rancangan Umum Penelitian	25
4.3	Analisis Dataset	29
4.4	Rancangan Prapemrosesan	31
4.5	Rancangan Validasi Program dan Pembandingan Program.....	32
4.6	Rancangan Algoritma User Transaction Profiling	32
4.7	Rancangan Pengujian	34
4.8	Pengujian Fungsional User Transaction Profiling menggunakan MapReduce.....	37
4.9	Rancangan Eksperimen Perbandingan Processing Time pada Jumlah Data dan Jumlah Node yang Berbeda	44
4.10	Rancangan Visualisasi Hasil Pengujian	45
BAB V	IMPLEMENTASI.....	47
5.1	Lingkungan Implementasi	47



5.2	Implementasi Dataset	47
5.3	Implementasi Prapemrosesan	48
5.4	Implementasi User Transaction Profiling.....	49
5.5	Implementasi Output Ke File CSV	55
5.6	Implementasi Pengujian di Cluster Hadoop 9 Node	57
5.7	Implementasi Visualisasi.....	63
BAB VI HASIL DAN PEMBAHASAN		69
6.1	Prapemrosesan.....	69
6.2	Validasi 1000 Data Antara Sum Microsoft Office Excel dengan Program. 70	
6.3	Hasil Pengujian pada Seluruh Data (164.027.463 Data).....	78
6.4	Eksperimen perbandingan processing time pada jumlah data berbeda. .	86
6.4.1	1.000 data	87
6.4.2	10.000 data	88
6.4.3	100.000 data	89
6.4.4	1.000.000 data	90
6.4.5	10.000.000 data	91
6.4.6	100.000.000 data	92
6.4.7	164.027.463 data	93
6.5	Eksperimen perbandingan processing time pada jumlah node berbeda.	94
6.5.1	1 node	94
6.5.2	2 node	95
6.5.3	3 node	96
6.5.4	4 node	97
6.5.5	5 node	99
6.5.6	6 node	100



6.5.7	7 node	101
6.5.8	8 node	102
6.5.9	9 node	103
6.6	Visualisasi	104
6.7	Pembandingan Program (Tidak Menggunakan MapReduce)	109
BAB VII	KESIMPULAN DAN SARAN	112
7.1	Kesimpulan	112
7.2	Saran	113
DAFTAR PUSTAKA	114
LAMPIRAN	116