

HALAMAN PENGESAHAN	i
STATEMENT	iii
PAGE OF DEDICATION	iv
PREFACE	v
CONTENTS	vi
LIST OF TABLES.....	x
LIST OF FIGURES	xi
NOMENCLATURE AND ABBREVIATION	xiii
INTISARI.....	xiv
ABSTRACT	xv
CHAPTER I Introduction	1
1.1 Research Background.....	1
1.2 Problem Statement.....	4
1.3 Research Objectives	4
1.4 Research Scope	5
1.5 Research Benefits	6
1.6 Structure of Thesis.....	7
CHAPTER II Literature Review and Theoretical Framework.....	8
2.1 Literature Review	8
2.1.1 Data Processing Technologies	8
2.1.2 Data Storage Technologies	13
2.1.3 Query Engine Technologies	16
2.1.4 Analytical Framework Selection Rationale	18
2.1.5 Research Question	19
2.2 Theoretical Framework	19
2.2.1 Comparative Analysis.....	19
2.2.2 Big Data	19
2.2.3 Supply Chain	21
2.2.4 Data Architecture	22
2.2.4.1 Online Transaction Processing (OLTP).....	22
2.2.4.2 Online Analytical Processing (OLAP).....	23
2.2.5 Cloud Computing	25
2.2.6 Open Source Data Architecture Tools.....	25
2.2.6.1 Hadoop Distributed File System (HDFS)	25
2.2.6.2 Apache Hive	27
2.2.7 Amazon Web Services (AWS).....	28

2.2.7.1	Amazon S3	28
2.2.7.2	AWS Glue	29
2.2.7.3	Amazon Athena	30
2.2.8	Evaluation Metrics	31
2.2.8.1	Analytics Maturity Model	31
2.2.8.2	Batch Processing Time	32
2.2.8.3	Throughput (Data Processed per Minute)	33
2.2.8.4	Cost Considerations in OLAP-Oriented Architectures ..	34
2.2.9	Statistical Tests	34
2.2.9.1	Normality Test	34
2.2.9.2	Wilcoxon Signed Rank Test	36
CHAPTER III Research Methodology		38
3.1	Research Tools and Materials	38
3.1.1	Research Tools	38
3.1.2	Research Materials	40
3.2	Research Methodology	41
3.2.1	Comparative Performance Evaluation	41
3.3	Research Workflow	42
3.3.1	Literature Review and Requirements Identification	42
3.3.2	Solution and Method Selection	43
3.3.3	Data Collection	46
3.3.4	Infrastructure Implementation and Development	47
3.3.4.1	On-Premise Cluster Setup	48
3.3.4.2	Cloud Environment Setup	53
3.3.4.3	Dataset Migration to Storage	54
3.3.5	Production Manufacturing Data Processing Code Development ...	56
3.3.6	Experimentation and Testing	60
3.3.6.1	Testing Scheme	60
3.3.6.2	Data Architecture Performance Retrieval Code Development	61
3.3.6.3	Throughput Calculation	62
3.3.6.4	Cost Estimation Calculation	62
3.3.6.5	Test Results and Analysis	66
3.4	Research Limitation	68
CHAPTER IV Results and Discussion		70
4.1	Results of Deployment and Integration within the On-premise OLAP Data Architecture	70
4.1.1	Infrastructure Configuration and Cluster Setup	70
4.1.2	HDFS Storage and Node Utilization	73

4.1.3	Spark Execution and Application Monitoring	75
4.1.4	Output Validation and File Organization.....	77
4.1.5	Integration with Hive and Visualization Layer	78
4.2	Analysis of Performance Comparison of Analytical Data Processing between Cloud and On-Premise Architectures.....	80
4.2.1	Execution Time Comparison Analysis for Batch-Based Data Ingestion	81
4.2.2	Processing Time Comparison Analysis for Batch-Based Transformation	83
4.2.2.1	Processing Time Comparison Analysis for Cleaning	84
4.2.2.2	Processing Time Comparison Analysis for Dicing.....	85
4.2.2.3	Processing Time Comparison Analysis for Grouping ...	86
4.2.2.4	Processing Time Comparison Analysis for Roll up	87
4.2.2.5	Processing Time Comparison Analysis for Summary ...	88
4.2.2.6	Processing Time Comparison Analysis for Sorting	89
4.2.3	Execution Time Comparison Analysis for Batch-Based Data Output	90
4.2.4	Throughput Comparison Analysis for Batch-Based Ingestion.....	91
4.2.5	Throughput Comparison Analysis for Batch-Based Transformation	92
4.2.5.1	Throughput Comparison Analysis for Cleaning	93
4.2.5.2	Throughput Comparison Analysis for Dicing	94
4.2.5.3	Throughput Comparison Analysis for Grouping	95
4.2.5.4	Throughput Comparison Analysis for Roll Up.....	96
4.2.5.5	Throughput Comparison Analysis for Summary	97
4.2.5.6	Throughput Comparison Analysis for Sorting	98
4.2.6	Throughput Comparison Analysis for Batch-Based Data Output ..	99
4.2.7	Statistical Analysis of Total Batch Processing Time Distributions .	100
4.2.7.1	Normality Testing Results and Interpretation	100
4.2.7.2	Statistical Distribution Analysis	100
4.3	Analysis of Cost Comparison of Analytical Data Processing between Cloud and On-Premise Architectures	101
4.3.1	Monthly Cost Analysis of On-Premise Infrastructure	102
4.3.2	Monthly Cost Analysis of Cloud Architecture	103
4.4	Expert Validation from Industry Perspective	103
4.4.1	Validation from Plant Operations Perspective	103
4.4.2	Validation from Digital Transformation Lead Perspective	104
4.4.3	Strengths and Limitations of the Study	105
CHAPTER V	Conclusion and Recommendations	107
5.1	Conclusion	107
5.2	Recommendation	108



A Comparative Cost-Performance Analysis of OLAP Data Architectures for Supply Chain Performance

Monitoring: Cloud versus On-Premise Deployment

Adzkie Khansa Meara, Dr.Eng. Silmi Fauziati, S.T., M.T.; Ir. Azkario Rizky Pratama, S.T., M.Eng., Ph.D.

UNIVERSITAS
GADJAH MADA

Universitas Gadjah Mada, 2025 | Diunduh dari <http://etd.repository.ugm.ac.id/>

REFERENCES	109
L.1 Pseudocode Processing Code.....	L-115
L.2 Full Log of Execution Time	L-118
L.3 On-premise Architecture Calculation	L-138
L.4 Supporting Assumptions and Parameters	L-139
L.5 Cloud Architecture Calculation	L-140
L.6 Post-presentation Feedback Table.....	L-141

Table 2.1	Comparison of Literature on Data Processing Tools	11
Table 2.2	Comparison of Literature on Data Storage Tools	15
Table 2.3	Comparison of Literature on Query Engine	17
Table 3.1	Key Columns in the ETO Production Dataset	40
Table 3.2	Comparison of Cloud and On-Premise Architecture Layers	45
Table 3.3	Daily File Size Summary for January 2025	46
Table 3.4	Resource Comparison: On-Premise Cluster vs. AWS Glue	48
Table 3.5	Python Library Dependencies	52
Table 3.6	Apache Spark Configuration for On-Premise IPC Cluster	53
Table 3.7	List of Low-Voltage (LV) Production Nodes Used in Dicing	58
Table 3.8	Testing Scheme Summary: Cloud vs On-Premise Environment	61
Table 3.9	Variables and Functions Used for Execution Time Recording	62
Table 3.10	Cost Estimation Summary	66
Table 4.1	Percentage Difference of Processing Time for Each Transformation Stage	83
Table 4.2	Percentage Difference of Average Throughput for Each Transformation Stage	92
Table 4.3	Tests of Normality for Total Batch Processing Time	100
Table 4.4	Wilcoxon Signed-Rank Test: Ranks Summary	101
Table 4.5	Descriptive Statistics of Total Batch Processing Time (seconds)	101
Table 4.6	Monthly Cost Summary: On-Premise Architecture	102
Table 4.7	Monthly Cost Summary: Cloud Architecture	103
Table 1	Full Log of Execution Time per Batch (January 2025)	L-118
Table 2	Compilation of Comments and Responses	L-141

Figure 2.1	10 V's of Big Data	21
Figure 2.2	Modern Data Architecture Flow	22
Figure 2.3	OLAP Data Architecture	23
Figure 2.4	Apache Hadoop Logo	26
Figure 2.5	HDFS Architecture	26
Figure 2.6	Apache Hive Logo	27
Figure 2.7	Apache Hive Architecture	28
Figure 2.8	S3 Bucket Object Storage	29
Figure 2.9	AWS Glue Logo	30
Figure 2.10	Query execution workflow using Amazon Athena	31
Figure 2.11	Analytics Maturity Model Visualization	32
Figure 3.12	Research workflow illustrated in the study	43
Figure 3.13	Workflow of Literature Review and Requirements Identification ...	44
Figure 3.14	Workflow of Tool Selection, Architecture Design, and Pipeline Deployment	44
Figure 3.15	On-Premise Data Architecture	45
Figure 3.16	Cloud-Native Data Architecture	46
Figure 3.17	On-Premise Environment Implementation Workflow	49
Figure 3.18	IPC Cluster Stack Architecture	49
Figure 3.19	IPC Cluster Network Topology	50
Figure 3.20	Master Node Configuration	50
Figure 3.21	Slave Node Configuration	51
Figure 3.22	Cloud Environment Implementation Workflow	53
Figure 3.23	Workflow for Migrating Local Dataset to HDFS	55
Figure 3.24	Workflow for Migrating Local Dataset to Glue	56
Figure 3.25	Workflow for Production Manufacturing Data Processing	57
Figure 3.26	Flowchart of statistical testing procedure	67
Figure 4.27	NameNode UI for Hadoop Cluster	71
Figure 4.28	Datanode UI in Slave 1	72
Figure 4.29	Datanode UI in Slave 2	72
Figure 4.30	DataNode Disk Utilization and Node Health	73
Figure 4.31	List of Ingested Batch Files in HDFS Directory	74
Figure 4.32	DataNode Storage Utilization After Data Ingestion	75
Figure 4.33	Spark Master Dashboard with Active Workers	76
Figure 4.34	Spark Stages During Batch ETL Job Execution	76
Figure 4.35	HDFS Directory Structure in /user/hadoop/	77
Figure 4.36	Output Files in Parquet Format with Snappy Compression	78
Figure 4.37	Querying Production Data via Hive CLI	79
Figure 4.38	Production Performance Dashboard Connected to Hive	80
Figure 4.39	Comparison of Average Ingestion Time	81
Figure 4.40	Total Processing Time Comparison between Cloud and On-Premise Deployments	82
Figure 4.41	Comparison of Average Execution Time for Each Transformations	83
Figure 4.42	Batch Processing Execution Time: Cleaning	84
Figure 4.43	Batch Processing Execution Time: Dicing	85



Figure 4.44	Batch Processing Execution Time: Grouping	86
Figure 4.45	Batch Processing Execution Time: Roll up	87
Figure 4.46	Batch Processing Execution Time: Summary	88
Figure 4.47	Batch Processing Execution Time: Sorting	89
Figure 4.48	Comparison of Average Ingestion Time in Cloud vs On-Premise ...	90
Figure 4.49	Average Ingestion Throughput between Cloud and On-Premise Architectures	91
Figure 4.50	Comparison of Average Throughput per ETL Step	92
Figure 4.51	Cleaning Time Throughput Over Batches: Cloud vs On-Premise ...	93
Figure 4.52	Dicing Time Throughput Over Batches: Cloud vs On-Premise	94
Figure 4.53	Grouping Time Throughput Over Batches: Cloud vs On-Premise ..	95
Figure 4.54	Rollup Time Throughput Over Batches: Cloud vs On-Premise	96
Figure 4.55	Summary Time Throughput Over Batches: Cloud vs On-Premise ..	97
Figure 4.56	Sorting Time Throughput Over Batches: Cloud vs On-Premise	98
Figure 4.57	Comparison of Average Throughput Store in Cloud vs On-Premise	99
Figure 58	Cost Calculation Breakdown of On Premise Architecture	L-138
Figure 59	Detailed Cost Assumptions and Inputs.....	L-139
Figure 60	Cost Calculation Breakdown of Cloud Architecture	L-140