

IMPLEMENTASI DAN ANALISIS TEKNOLOGI DATAOPS UNTUK
MENINGKATKAN PERFORMA *DATA PIPELINE* DALAM SIKLUS HIDUP
DATA (STUDI KASUS PENGELOLAAN DATA INFRASTRUKTUR DI
TELKOM DBT INS-ARMI)

SKRIPSI

diajukan untuk memenuhi sebagian dari syarat untuk memperoleh gelar Sarjana
Komputer Program Studi Rekayasa Perangkat Lunak



oleh

Muhammad Alfian Nurul Yaqien

NIM 1903712

PROGRAM STUDI REKAYASA PERANGKAT LUNAK

KAMPUS DAERAH CIBIRU

UNIVERSITAS PENDIDIKAN INDONESIA

2023

IMPLEMENTASI DAN ANALISIS TEKNOLOGI DATAOPS UNTUK
MENINGKATKAN PERFORMA *DATA PIPELINE* DALAM SIKLUS HIDUP
DATA (STUDI KASUS PENGELOLAAN DATA INFRASTRUKTUR DI
TELKOM DBT INS-ARMI)

oleh

Muhammad Alfian Nurul Yaqien

diajukan untuk memenuhi sebagian syarat untuk memperoleh gelar Sarjana
Komputer Program Studi Rekayasa Perangkat Lunak

© Muhammad Alfian Nurul Yaqien
Universitas Pendidikan Indonesia
April 2023

Hak cipta dilindungi Undang-Undang
Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak
ulang, difotokopi, atau cara lainnya tanpa ijin dari penulis.

HALAMAN PENGESAHAN

Muhammad Alfian Nurul Yaqien

IMPLEMENTASI DAN ANALISIS TEKNOLOGI DATAOPS UNTUK
MENINGKATKAN PERFORMA *DATA PIPELINE* DALAM SIKLUS HIDUP
DATA (STUDI KASUS PENGELOLAAN DATA INFRASTRUKTUR DI
TELKOM DBT INS-ARMI)

disetujui dan disahkan oleh pembimbing:


Pembimbing I



Raditya Muhammad, S.T., M.T.

NIP 920190219920507101

Pembimbing II



Asyifa Imanda Septiana, S.Pd., M.Eng.

NIP 920190219920228201

Mengetahui,

Ketua Program Studi Rekayasa Perangkat Lunak



Mochamad Iqbal Ardimansyah, S.T., M.Kom.

NIP 920190219910328101

PERNYATAAN KEASLIAN SKRIPSI DAN BEBAS PLAGIARISME

Dengan ini saya menyatakan bahwa skripsi dengan judul “Implementasi dan Analisis Teknologi DataOps untuk Meningkatkan Performa *Data Pipeline* dalam Siklus Hidup Data (Studi Kasus Pengelolaan Data Infrastruktur di Telkom DBT INS-ARMI)” ini beserta seluruh isinya adalah benar-benar karya saya sendiri. Saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko/sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya saya ini.

Bandung, 27 April 2023

Muhammad Alfian Nurul Yaqien

NIM 1903712

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Implementasi dan Analisis Teknologi DataOps untuk Meningkatkan Performa *Data Pipeline* dalam Siklus Hidup Data (Studi Kasus Pengelolaan Data Infrastruktur di Telkom DBT INS-ARMI)” tepat pada waktunya. Pada kesempatan ini, penulis mengucapkan terima kasih kepada seluruh pihak yang selalu mendukung sehingga skripsi ini dapat terselesaikan. Dalam hal ini, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Solehuddin, M.Pd., MA., selaku Rektor Universitas Pendidikan Indonesia.
2. Bapak Prof. Dr. Asep Herry Hernawan, M.Pd., selaku Direktur Universitas Pendidikan Indonesia Kampus Cibiru.
3. Bapak Dr. Dede Margo Irianto, M.Pd., selaku Wakil Direktur Universitas Pendidikan Indonesia Kampus Cibiru.
4. Bapak M. Iqbal Ardimansyah, S.T., M.Kom., selaku dosen pembimbing akademik dan kepala program studi Rekayasa Perangkat Lunak yang telah banyak membantu segala bentuk administrasi penulis selama menjalani perkuliahan.
5. Bapak Raditya Muhammad, S.T., M.Kom., selaku dosen pembimbing skripsi pertama yang telah meluangkan banyak waktu untuk mengarahkan penelitian penulis.
6. Ibu Asyifa Imanda Septiana, S.Pd., M.Eng., selaku dosen pembimbing skripsi kedua yang telah memberikan masukan selama penyusunan penelitian penulis.
7. Seluruh dosen RPL yang telah memberikan banyak sekali ilmu selama penulis menjalani perkuliahan.
8. Kedua orang tua yang selalu mendukung dan mendoakan dari setiap kesulitan yang penulis hadapi.
9. Semua anggota keluarga yang telah memberikan dukungan kepada penulis selama kuliah hingga menyelesaikan skripsi ini.

Muhammad Alfian Nurul Yaqien, 2023
IMPLEMENTASI DAN ANALISIS TEKNOLOGI DATAOPS UNTUK MENINGKATKAN PERFORMA DATA PIPELINE DALAM SIKLUS HIDUP DATA (STUDI KASUS PENGELOLAAN DATA INFRASTRUKTUR DI TELKOM DBT INS-ARMI)

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

10. Seluruh teman-teman kuliah yang telah memberikan dukungan serta berbagi ilmu yang bermanfaat.
11. ChatGPT yang telah berperan dalam membantu menyelesaikan skripsi ini.
12. Sahabat-sahabat MA yang telah kebersamai dalam perjuangan untuk terus menjadi lebih baik.
13. Semua orang yang pernah datang dan pergi di kehidupan penulis, yang telah memberikan banyak pelajaran yang begitu berharga.

Ucapan terima kasih penulis persembahkan kepada semua yang telah membantu penulis dalam melakukan penelitian sekaligus penyusunan skripsi ini hingga selesai, semoga kita semua selalu dalam ridho dan lindungan-Nya. Semoga semua kebaikan yang telah diberikan mendapat balasan yang setimpal dari Allah SWT. Aamiin.

Bandung, 27 April 2023

Muhammad Alfian Nurul Yaqien

**IMPLEMENTASI DAN ANALISIS TEKNOLOGI DATAOPS UNTUK
MENINGKATKAN PERFORMA *DATA PIPELINE* DALAM SIKLUS
HIDUP DATA (STUDI KASUS PENGELOLAAN DATA
INFRASTRUKTUR DI TELKOM DBT INS-ARMI)**

ABSTRAK

DataOps hadir sebagai praktik untuk mengatasi permasalahan dari proyek analitik data, seperti adanya batasan antar tim yang menyebabkan lemahnya kolaborasi dan komunikasi, tidak adanya standarisasi data, serta lambatnya proses pengiriman produk analitik data. Penelitian ini ditujukan untuk mengimplementasikan praktik DataOps dalam pengelolaan data infrastruktur perusahaan di Telkom DBT, menganalisis perbandingan performa *data pipeline* antara eksekusi manual dengan pendekatan DataOps, serta menganalisis hasil evaluasi teknologi dan alat DataOps yang digunakan. Untuk perbandingan performa *data pipeline* diukur dari waktu eksekusi yang mengacu kepada penelitian sebelumnya. Kemudian Skala Likert digunakan sebagai alat untuk mengevaluasi teknologi dan alat DataOps yang digunakan melalui kuesioner yang diisi oleh tim data dari Squad ARMI. Hasil yang didapat dari perbandingan performa *data pipeline*, eksekusi pada ETL Live Infra dihasilkan bahwa implementasi DataOps memiliki waktu eksekusi 70632,7 ms (1 menit 17,77 detik) lebih cepat dari pada Eksekusi yang dilakukan secara manual. Kemudian eksekusi ETL Static Infra dengan pendekatan DataOps memiliki selisih waktu eksekusi 21286,6 ms (35,477 detik) lebih cepat dari pada eksekusi yang dilakukan secara manual. Maka implementasi DataOps ini berhasil meningkatkan performa *data pipeline* dengan perbedaan selisih yang cukup signifikan. Sedangkan hasil dari evaluasi teknologi atau alat DataOps dengan metrik *Simplicity*, *Usability*, dan *Compatibility* yaitu rata-rata persentase skor teknologi secara keseluruhan diperoleh 80,88% dan dapat diinterpretasi bahwa teknologi yang digunakan memiliki kriteria Sangat Layak. Rata-rata persentase disetiap indikator diperoleh *Simplicity* sebesar 80,37%, *Usability* sebesar 80,40%, dan *Compatibility* sebesar 81,87% sehingga semua indikator memiliki kriteria Sangat Layak.

Kata Kunci: DataOps, *Data Pipeline*, Siklus Hidup Data, Analitik Data

**IMPLEMENTATION AND ANALYSIS OF DATAOPS TECHNOLOGY TO
IMPROVE DATA PIPELINE PERFORMANCE IN THE DATA LIFE CYCLE
(CASE STUDY IN INFRASTRUCTURE DATA MANAGEMENT AT
TELKOM DBT INS-ARMI)**

ABSTRACT

DataOps is a practice that aims to overcome issues commonly found in data analytics projects such as team boundaries causing weak collaboration and communication, absence of data standardization, and slow delivery of data analytics products. This research focuses on implementing DataOps practices for managing company infrastructure data at Telkom DBT. The study analyzes the performance comparison of data pipelines between manual execution and the DataOps approach and evaluates the technology and DataOps tools used. The research measures data pipeline performance from the execution time referred to in previous studies. The Likert Scale is used to evaluate DataOps technology and tools through a questionnaire answered by the data team from the ARMI Squad. The results show that the DataOps implementation improved the performance of the data pipeline significantly. Execution time for ETL Live Infra was 70632.7 ms (1 minute 17.77 seconds) faster than manual execution, while the execution of ETL Static Infra with the DataOps approach was 21286.6 ms (35.477 seconds) faster than manual execution. The study also shows that the technology used has Very Eligible criteria, with an average percentage score of 80.88% for the Simplicity, Usability, and Compatibility metrics. The average percentage for each indicator was 80.37% for Simplicity, 80.40% for Usability, and 81.87% for Compatibility, indicating all indicators had Very Eligible criteria.

Keywords: *DataOps, Data Pipeline, Data Lifecycle, Data Analytics*

DAFTAR ISI

HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN SKRIPSI DAN BEBAS PLAGIARISME	iv
UCAPAN TERIMA KASIH.....	v
ABSTRAK.....	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI.....	ix
DAFTAR ISTILAH	xii
DAFTAR TABEL.....	xvi
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Penelitian	1
1.2 Rumusan Masalah Penelitian	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	4
1.6 Sistematika Penulisan.....	4
BAB II KAJIAN PUSTAKA.....	6
2.1 Siklus Hidup Data	6
2.2 <i>Data Pipeline</i>	8
2.3 <i>Data Pipeline</i> di dalam Siklus Hidup Data	8
2.4 DataOps	9
2.4.1 Model Evolusi DataOps	11
2.4.2 Evolusi DataOps	12
2.4.3 Pedoman Implementasi DataOps	14
2.5 Tata Kelola Data.....	15
2.5.1 Tata Kelola Data dengan DataOps.....	15
2.6 Asal dan Silsilah Data	16

Muhammad Alfian Nurul Yaqien, 2023
IMPLEMENTASI DAN ANALISIS TEKNOLOGI DATAOPS UNTUK MENINGKATKAN PERFORMA DATA PIPELINE DALAM SIKLUS HIDUP DATA (STUDI KASUS PENGELOLAAN DATA INFRASTRUKTUR DI TELKOM DBT INS-ARM)

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

2.6.1	Asal dan Silsilah Data dengan DataOps.....	16
2.7	<i>Data Pipeline</i> Menuju <i>DataOps Pipeline</i>	17
2.8	Waktu Eksekusi dan Pengiriman Produk	18
2.9	Teknologi DataOps.....	18
2.9.1	Teknologi dan Alat <i>Data Pipeline</i>	18
2.10	Kriteria Evaluasi Alat dan Teknologi DataOps.....	22
2.11	Model Referensi	22
2.12	Penelitian Terkait (<i>State-of-the-art</i>)	23
BAB III METODOLOGI PENELITIAN.....		32
3.1	Desain Penelitian	32
3.1.1	Klarifikasi Penelitian.....	32
3.1.2	Studi Deskriptif I.....	34
3.1.3	Studi Preskriptif	34
3.1.4	Studi Deskriptif II	37
3.2	Alat dan Bahan Penelitian	38
3.2.1	Alat Penelitian.....	38
3.2.2	Bahan Penelitian.....	39
3.3	Instrumen Penelitian.....	39
3.4	Analisis Data	40
3.5	Hipotesis	41
3.6	Model Dampak	41
BAB IV TEMUAN DAN PEMBAHASAN.....		43
4.1	Desain Arsitektur Sistem DataOps.....	43
4.1.1	Sumber Data.....	44
4.1.2	Proses ETL (<i>Extract-Transform-Load</i>).....	44
4.1.3	Orkestrasi <i>Data Pipeline</i>	46
4.1.4	Data Visualisasi.....	49
4.2	Proses DataOps dalam Pengelolaan Data Infrastruktur.....	51
4.2.1	Pengumpulan Data Infrastruktur	51
4.2.2	Proses Pembangunan Sistem.....	53
4.2.3	Pengujian.....	57
4.2.4	Proses <i>Deployment</i>	58

4.2.5	Orkestrasi <i>Data Pipeline</i>	61
4.2.6	Pemantauan (<i>Monitoring</i>)	62
4.2.7	Produk Data Analitik.....	64
4.3	Pengujian Performa <i>Data Pipeline</i> (Waktu Eksekusi).....	65
4.3.1	Performa Live Infra ETL	65
4.3.2	Performa Static Infra ETL.....	67
4.4	Evaluasi Alat dan Teknologi DataOps	68
4.4.1	Apache Airflow	69
4.4.2	Gitlab.....	70
4.4.3	Gitlab CI/CD	70
4.4.4	Satu Data INS.....	71
4.4.5	Docker.....	72
4.4.6	Docker Compose.....	72
4.4.7	GitAutoDeploy.....	73
4.5	Pembahasan	74
4.5.1	Pembahasan Evaluasi Performa <i>Data Pipeline</i>	74
4.5.2	Pembahasan Evaluasi Teknologi DataOps	77
BAB V SIMPULAN, IMPLIKASI, DAN REKOMENDASI		80
5.1	Simpulan.....	80
5.2	Implikasi	81
5.3	Rekomendasi	81
DAFTAR PUSTAKA		83
LAMPIRAN.....		86

DAFTAR ISTILAH

1. DAG

DAG adalah singkatan dari *Directed Acyclic Graph*, yaitu sebuah struktur data yang terdiri dari simpul-simpul (*node*) yang dihubungkan oleh sisi-sisi (*edge*). DAG merupakan jenis graf yang memiliki arah (*directed*) pada setiap sisi-sisinya, dan tidak memiliki siklus (*acyclic*), yaitu tidak ada jalur yang memungkinkan untuk kembali ke simpul yang sama. DAG sering digunakan untuk merepresentasikan hubungan atau ketergantungan antar tugas atau pekerjaan dalam suatu alur kerja atau proses bisnis, serta digunakan sebagai dasar dalam pemrograman dinamis dan optimisasi komputasi. DAG sangat berguna dalam memodelkan proses yang dapat dijalankan secara paralel dan dalam menjaga urutan eksekusi yang benar.

2. Deployment

Deployment adalah proses untuk menjalankan atau mengirimkan aplikasi atau perangkat lunak ke lingkungan produksi atau lingkungan yang akan digunakan oleh pengguna akhir. Proses *deployment* melibatkan serangkaian langkah, mulai dari pengumpulan kode, konfigurasi, tes, hingga pemeliharaan dan *monitoring*. Tujuan dari *deployment* adalah untuk memastikan bahwa aplikasi berjalan dengan lancar dan sesuai dengan harapan pengguna. Proses *deployment* dapat dilakukan secara manual atau otomatis menggunakan alat dan *platform* seperti Jenkins, Kubernetes, atau Docker.

3. ETL

ETL adalah singkatan dari *Extract, Transform, dan Load*, suatu proses yang digunakan untuk mengambil data dari berbagai sumber data, menjalankan transformasi dan membersihkan data tersebut, lalu memuatnya ke dalam sistem lain untuk digunakan. Proses ETL umumnya digunakan untuk mengintegrasikan data dari sumber yang berbeda, seperti *database, file* teks, atau *data streaming*, dan mentransformasi data menjadi bentuk yang lebih mudah digunakan untuk analisis atau pemrosesan lebih lanjut. ETL biasanya digunakan dalam pengolahan data skala besar dan sistem analitik data.

4. **Infrastruktur**

Infrastruktur perusahaan dalam konteks IT merujuk pada kumpulan perangkat keras, perangkat lunak, dan teknologi lainnya yang digunakan untuk mendukung operasi dan kegiatan bisnis sebuah organisasi. Infrastruktur dapat mencakup server, jaringan, penyimpanan data, aplikasi, *platform*, layanan, dan alat pengembangan. Infrastruktur perusahaan biasanya dikembangkan dan dipelihara oleh tim internal atau oleh pihak ketiga, dan bertujuan untuk memastikan ketersediaan, keamanan, dan skalabilitas teknologi yang dibutuhkan oleh organisasi.

5. **INS**

INS (*Digital Infrastructure & Security*) adalah bidang/unit bisnis di Telkom Digital Business Transformation (DBT) yang bertanggung jawab untuk menyediakan solusi infrastruktur digital dan keamanan informasi bagi pelanggan perusahaan di Indonesia. Unit bisnis ini menawarkan berbagai layanan, termasuk solusi *cloud*, kolaborasi, konektivitas, keamanan siber, dan manajemen layanan IT. Dengan adanya bidang INS ini, perusahaan Telkom DBT dapat mengoptimalkan infrastruktur teknologi informasi dan meningkatkan efisiensi bisnis mereka secara keseluruhan.

6. **Live Infra**

Live Infra (infrastruktur langsung) merujuk pada sistem atau lingkungan IT yang menghasilkan data secara dinamis dalam waktu nyata. Artinya, infrastruktur ini dapat menghasilkan dan mengirimkan data secara terus-menerus selama operasionalnya berjalan, seperti sistem *monitoring* jaringan, aplikasi *streaming video*, sistem *log*, dan sebagainya. Data yang dihasilkan dari Live Infra sangat penting dalam membantu pemantauan kinerja sistem, mendiagnosis masalah, dan memberikan wawasan tentang cara meningkatkan efisiensi dan kinerja sistem secara keseluruhan.

7. **SPC**

Statistical Process Control (SPC) adalah sebuah metode pengendalian kualitas yang menggunakan teknik statistik untuk *me-monitor* dan mengontrol proses

produksi atau manufaktur. Tujuan utama dari SPC adalah untuk menemukan ketidaksesuaian atau perbedaan dalam proses produksi yang dapat menghasilkan produk cacat, dan kemudian melakukan perbaikan pada proses tersebut agar dapat mencapai konsistensi dan kestabilan produksi yang lebih baik. SPC digunakan di berbagai bidang, seperti manufaktur, layanan, kesehatan, teknologi dan lain-lain.

8. Squad ARMI

Squad ARMI merupakan unit atau tim yang berada di bawah INS yang memiliki peran dalam bidang data. ARMI merupakan singkatan dari *Analytics, Research, and Modeling for Data Intelligence* yang saat ini berfokus pada pengelolaan data infrastruktur perusahaan dalam menghasilkan produk analitik data.

9. Static Infra

Static Infra merujuk pada infrastruktur yang tidak berubah secara dinamis (statis) selama operasi normal bisnis, seperti server, *storage*, dan jaringan. Data yang dihasilkan oleh infrastruktur ini, seperti jumlah CPU, kapasitas penyimpanan, dan jumlah memori, juga bersifat statis dan tidak berubah selama infrastruktur beroperasi. Data ini biasanya diambil pada saat infrastruktur pertama kali di-*deploy* atau saat terjadi perubahan besar pada infrastruktur, seperti penambahan server baru atau peningkatan kapasitas penyimpanan. Data ini berguna untuk memantau kesehatan dan kinerja infrastruktur serta melakukan perencanaan kapasitas untuk memastikan infrastruktur yang memadai untuk bisnis.

10. Stateful Live Infra

Stateful Live Infra merujuk pada infrastruktur yang mampu menyimpan status atau keadaan suatu sistem pada suatu waktu tertentu. *Stateful* dapat merujuk pada data atau informasi yang terkait dengan aplikasi, layanan, atau sistem yang sedang berjalan dan berubah seiring waktu. Live Infra mengacu pada infrastruktur yang dapat menghasilkan data secara dinamis dan memungkinkan operasi yang cepat dan responsif. Jadi, Stateful Live Infra adalah infrastruktur yang dapat menyimpan informasi tentang status atau keadaan sistem, dan

menghasilkan data secara dinamis dalam waktu nyata. Contoh dari Stateful Live Infra dapat berupa basis data online, sistem manajemen pesan, atau sistem jaringan yang menyimpan status koneksi atau arus lalu lintas jaringan secara *real-time*.

11. Stateless Live Infra

Stateless Live Infra merujuk pada infrastruktur yang tidak menyimpan informasi atau status sebelumnya dan tidak bergantung pada keadaan atau status sebelumnya. Setiap permintaan yang masuk dianggap sebagai permintaan baru dan diproses secara terpisah, tanpa melibatkan informasi atau status sebelumnya. Contoh dari infrastruktur Stateless Live Infra adalah server web yang melayani permintaan HTTP dari pengguna dan tidak menyimpan informasi yang berkelanjutan antara permintaan. Dalam arsitektur *modern* seperti *microservice*, Stateless Live Infra sangat populer digunakan karena fleksibilitas, skalabilitas, dan mudah dalam mengelola infrastruktur yang besar.

12. Telkom DBT

Telkom DBT (*Digital Business and Technology*) adalah divisi di PT Telkom Indonesia yang bertanggung jawab dalam memimpin transformasi digital Telkom sebagai perusahaan teknologi terdepan di Indonesia. Telkom DBT terdiri dari beberapa unit bisnis seperti Telkom Infra, Telkom Sigma, Telkom Metra, Telkom DWS, dan Telkom FSI yang masing-masing memiliki fokus pada aspek yang berbeda dalam pengembangan teknologi digital, seperti infrastruktur teknologi, solusi digital, layanan pengembangan aplikasi, dan layanan keamanan siber. Divisi Telkom DBT bertujuan untuk memberikan solusi teknologi terbaik bagi pelanggan di Indonesia, baik perusahaan maupun individu.

DAFTAR TABEL

Tabel 2.1 Rangkuman <i>State-of-the-art</i> DataOps.....	27
Tabel 3.1 Skor/Bobot Skala Likert (Albaum, 1997; Allen & Seaman, 2007; Joshi et al., 2015).....	40
Tabel 3.2 Tabel Interpretasi Skala Likert (Allen & Seaman, 2007)	41
Tabel 4.1 Data <i>tribe_parent</i>	52
Tabel 4.2 <i>Script</i> Eksekusi Manual Live Infra ETL	53
Tabel 4.3 <i>Script</i> Eksekusi Manual Static Infra ETL	55
Tabel 4.4 Hasil Uji T Pada Live Infra ETL	76
Tabel 4.5 Hasil Uji T Pada Static Infra ETL.....	76
Tabel 4.6 Hasil Evaluasi Alat/Teknologi DataOps Pada Studi Kasus	78
Tabel 4.7 Hasil Evaluasi Berdasarkan Indikator	78

DAFTAR GAMBAR

Gambar 2.1 Tahapan Siklus Hidup Data (Mainali, 2020)	7
Gambar 2.2 Siklus Hidup DataOps (Bahaa et al., 2023)	7
Gambar 2.3 <i>Pipeline</i> data sederhana (Mainali et al., 2021).....	8
Gambar 2.4 <i>Pipeline</i> data di dalam siklus hidup data (Mainali, 2020).....	9
Gambar 2.5 Evolusi DataOps (Mainali, 2020).....	11
Gambar 2.6 DataOps telah berevolusi dari manufaktur ramping dan metodologi perangkat lunak (Mainali, 2020).....	14
Gambar 2.7 DataOps <i>pipeline</i> dari beberapa aspek (Mainali, 2020)	17
Gambar 2.8 Arsitektur Apache Airflow (Apache Airflow, 2020)	19
Gambar 2.9 Arsitektur Docker (Docker, 2021)	20
Gambar 2.10 Contoh Alur Kerja Gitlab CI/CD (Gitlab, n.d.).....	20
Gambar 2.11 Model Referensi	23
Gambar 3.1 Desain Penelitian.....	33
Gambar 3.2 Proses DataOps (Mainali et al., 2021).....	35
Gambar 3.3 Proses DataOps berdasarkan Studi Kasus.....	35
Gambar 3.4 Model Dampak.....	42
Gambar 4.1 Arsitektur Sistem DataOps <i>Pipeline</i>	43
Gambar 4.2 Proses ETL Live Infra.....	44
Gambar 4.3 Proses ETL Static Infra secara Garis Besar	46
Gambar 4.4 Orkestrasi <i>Data Pipeline</i> dengan Apache Airflow.....	47
Gambar 4.5 Gitlab CI/CD <i>Pipeline</i>	47
Gambar 4.6 Detail Gitlab CI/CD <i>Pipeline</i>	48
Gambar 4.7 Koneksi GitAutoDeploy ke Webhook Gitlab	49
Gambar 4.8 Data Visual dengan Satu Data INS	49
Gambar 4.9 Visualisasi Data Live Infra dengan Looker Studio	50
Gambar 4.10 Hasura sebagai Data Manager Basis Data PostgreSQL dan GraphQL <i>Query Engine</i>	50
Gambar 4.11 Penerapan Query Engine Integrasi Satu Data INS.....	51

Muhammad Alfian Nurul Yaqien, 2023
IMPLEMENTASI DAN ANALISIS TEKNOLOGI DATAOPS UNTUK MENINGKATKAN PERFORMA DATA PIPELINE DALAM SIKLUS HIDUP DATA (STUDI KASUS PENGELOLAAN DATA INFRASTRUKTUR DI TELKOM DBT INS-ARM)

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Gambar 4.12 Contoh Data Live Infra dari Datadog.....	52
Gambar 4.13 ETL Live Infra dengan Eksekusi Manual	55
Gambar 4.14 ETL Static Infra dengan Eksekusi Manual	56
Gambar 4.15 Implementasi ETL <i>Pipeline</i> dengan Pendekatan DataOps	57
Gambar 4.16 <i>Deployment</i> dengan Docker dan Docker Compose	59
Gambar 4.17 <i>Logs</i> GitAutoDeploy	60
Gambar 4.18 <i>Dashboard</i> Nginx Proxy Manager	61
Gambar 4.19 ETL <i>Pipeline</i> berbentuk DAG pada Airflow	62
Gambar 4.20 Memantau DAG yang Berjalan Sukses.....	63
Gambar 4.21 Pemantauan Sumber Daya dengan Docker	63
Gambar 4.22 Pemantauan Sumber Daya Mesin Virtual	64
Gambar 4.23 Waktu Eksekusi ETL Live Infra dengan Eksekusi Manual	66
Gambar 4.24 Waktu Eksekusi ETL Live Infra dengan Pendekatan DataOps	67
Gambar 4.25 Waktu Eksekusi ETL Static Infra dengan Eksekusi Manual	67
Gambar 4.26 Waktu Eksekusi ETL Static Infra dengan Pendekatan DataOps.....	68
Gambar 4.27 Hasil Evaluasi Penggunaan Apache Airflow	69
Gambar 4.28 Hasil Evaluasi Penggunaan Gitlab	70
Gambar 4.29 Hasil Evaluasi Penggunaan Gitlab CI/CD	71
Gambar 4.30 Hasil Evaluasi Penggunaan Satu Data INS	71
Gambar 4.31 Hasil Evaluasi Penggunaan Docker	72
Gambar 4.32 Hasil Evaluasi Penggunaan Docker Compose	73
Gambar 4.33 Hasil Evaluasi Penggunaan GitAutoDeploy	74
Gambar 4.34 Perbandingan Waktu Eksekusi Live Infra ETL	74
Gambar 4.35 Perbandingan Waktu Eksekusi Static Infra ETL.....	75

DAFTAR LAMPIRAN

Lampiran 1 Hasil Evaluasi Teknologi dan Alat DataOps.....	86
Lampiran 2 Eksekusi <i>Pipeline</i> dengan Pendekatan DataOps	92
Lampiran 3 Pembangunan Program ETL	95
Lampiran 4 Hasil Pengujian Unit dan Integrasi ETL & DAG.....	96
Lampiran 5 Konfigurasi <i>Deployment</i>	99
Lampiran 6 Konfigurasi CI/CD	104
Lampiran 7 Detail <i>Logs</i> dan <i>Monitoring</i>	107
Lampiran 8 <i>Query</i> Integrasi Satu Data INS ke Hasura	109

DAFTAR PUSTAKA

- Albaum, G. (1997). The Likert scale revisited: An alternate version. *International Journal of Market Research*, 39(2), 331–348.
- Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality Progress*, 40(7), 64–65.
- Apache Airflow. (2020). *Airflow Architecture*.
- Apache Airflow. (2023). *Apache Airflow*. <https://airflow.apache.org/docs/apache-airflow/stable/index.html>
- Arass, M. El, Tikito, I., & Souissi, N. (2017). Data lifecycles analysis: Towards intelligent cycle. *2017 Intelligent Systems and Computer Vision, ISCV 2017*.
- Arefeen, M. S., & Schiller, M. (2019). Continuous Integration Using Gitlab. *Undergraduate Research in Natural and Clinical Science and Technology (URNCSST) Journal*, 3(8), 1–6.
- Atwal, H. (2019). Practical DataOps: Delivering agile data science at scale. In *Practical DataOps: Delivering Agile Data Science at Scale*.
- Bahaa, S., Ghalwash, A. Z., & Harb, H. (2023). DataOps Lifecycle with a Case Study in Healthcare. *International Journal of Advanced Computer Science and Applications*, 14(1), 136–144.
- Belhajjame, K., Missier, P., & Goble, C. (2009). Data Provenance in Scientific Workflows. In *Handbook of Research on Computational Grid Technologies for Life Sciences, Biomedicine, and Healthcare* (pp. 46–59). IGI Global.
- Bergh, C., Benghiat, G., & Eran, S. (2019). *The DataOps Cookbook*. 133.
- Capizzi, A., Distefano, S., & Mazzara, M. (2020). From DevOps to DevDataOps: Data Management in DevOps Processes. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12055 LNCS(Cd), 52–62.
- Docker. (2021). *Docker*. <https://docs.docker.com/get-started/overview/>
- Ereth, J., & Eckerson, W. (2018). *Industrializing Data and Analytics – Strategies for Streamlining the Delivery of Insights*.
- Garriga, M., Aarns, K., Tsigkanos, C., Tamburri, D. A., & Heuvel, W. Van Den. (2021). DataOps for Cyber-Physical Systems Governance: The Airport Passenger Flow Case. *ACM Transactions on Internet Technology*, 21(2).
- Gaythorpe, K. A. M., Fitzjohn, R. G., Hinsley, W., Imai, N., Knock, E. S., Perez Guzman, P. N., Djaafara, B., Fraser, K., Baguelin, M., & Ferguson, N. M. (2023). Data pipelines in a public health emergency: The human in the
Muhammad Alfian Nurul Yaqien, 2023
IMPLEMENTASI DAN ANALISIS TEKNOLOGI DATAOPS UNTUK MENINGKATKAN PERFORMA DATA PIPELINE DALAM SIKLUS HIDUP DATA (STUDI KASUS PENGELOLAAN DATA INFRASTRUKTUR DI TELKOM DBT INS-ARMI)
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

- machine. *Epidemics*, 43, 100676. Gitlab. (n.d.). *No Gitlab CI/CD*.
<https://docs.gitlab.com/ee/ci/introduction/index.html>
- Gupta, U., & Cannon, S. (2020). Data Governance Frameworks. *A Practitioner's Guide to Data Governance*, 101–122.
- Jangla, K. (2018). Accelerating Development Velocity Using Docker. *Accelerating Development Velocity Using Docker*.
- Joshi, A., Kale, S., Chandel, S., & Pal, D. (2015). Likert Scale: Explored and Explained. *British Journal of Applied Science & Technology*, 7(4), 396–403.
- Khan, M. S., Khan, A. W., Khan, F., Khan, M. A., & Whangbo, T. K. (2022). Critical Challenges to Adopt DevOps Culture in Software Organizations: A Systematic Review. *IEEE Access*, 10, 14339–14349.
- Lohachab, A. (2020). *Bootstrapping Urban Planning: Addressing Big Data Issues in Smart Cities* (pp. 217–246).
- Mainali, K. (2020). DataOps : Towards Understanding and Defining Data Analytics Approach. *Trita-Eecs-Ex Nv - 2020:895*, 97.
- Mainali, K., Ehrlinger, L., Himmelbauer, J., & Matskin, M. (2021). Discovering DataOps: A Comprehensive Review of Definitions, Use Cases, and Tools. *DATA ANALYTICS 2021 : The Tenth International Conference on Data Analytics, October*, 61–69.
- Manifesto, D. (2020). *The DataOps Manifesto*.
<https://www.dataopsmanifesto.org/>
- Munappy, A. R., Mattos, D. I., Bosch, J., Olsson, H. H., & Dakkak, A. (2020). From Ad-Hoc data analytics to DataOps. *Proceedings - 2020 IEEE/ACM International Conference on Software and System Processes, ICSSP 2020*, 165–174.
- Olipo186. (2018). *Git-Auto-Deploy*. <https://github.com/olipo186/Git-Auto-Deploy.git>
- Palmer, A. (2022). *What is DataOps*. <https://www.tamr.com/blog/from-devops-to-dataops-by-andy-palmer/>
- Pettey, C. (2019). *Why Data and Analytics Are Key to Digital Transformation*. <https://www.gartner.com/smarterwithgartner/why-data-and-analytics-are-key-to-digital-transformation>
- Plale, B., & Kouper, I. (2017). The Centrality of Data: Data Lifecycle and Data Pipelines. In *Data Analytics for Intelligent Transportation Systems*. Elsevier Inc.
- Pranatawijaya, V. H., Widiatry, W., Priskila, R., & Putra, P. B. A. A. (2019). Penerapan Skala Likert dan Skala Dikotomi Pada Kuesioner Online. *Jurnal Sains Dan Informatika*, 5(2), 128–137.
- Rodriguez, M., De Araújo, L. J. P., & Mazzara, M. (2020). Good practices for the

Muhammad Alfian Nurul Yaqien, 2023

IMPLEMENTASI DAN ANALISIS TEKNOLOGI DATAOPS UNTUK MENINGKATKAN PERFORMA DATA PIPELINE DALAM SIKLUS HIDUP DATA (STUDI KASUS PENGELOLAAN DATA INFRASTRUKTUR DI TELKOM DBT INS-ARMI)

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

- adoption of DataOps in the software industry. *Journal of Physics: Conference Series*, 1694(1).
- Sahoo, P. R., & Premchand, A. (2019). DataOps in Manufacturing and Utilities Industries. *International Journal of Applied Information Systems (IJ AIS)*, 12(23), 1–6. www.ijais.org
- Science, O.-O. D. (2019). *DataOps and the DataOps Manifesto*. <https://odsc.medium.com/dataops-and-the-dataops-manifesto-fc6169c02398>
- Stewart, D. B. (2002). Measuring Execution Time and Real-Time Performance. *Embedded Systems Conference (ESC SF)*, September, 1–15.
- Sugiyono. (2006). *Metode Penelitian kuantitatif, kualitatif dan R & D*.
- Tamburri, D. A., Heuvel, W. J. Van Den, & Garriga, M. (2020). DataOps for Societal Intelligence: A Data Pipeline for Labor Market Skills Extraction and Matching. *Proceedings - 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science, IRI 2020*, 391–394.
- Techopedia. (2020). *What is Data Lineage?* <https://www.techopedia.com/definition/28040/data-lineage>
- Thi, N., Phuong, T., Data, B., & Program, A. (2021). *DataOps for Product Information Management : A study of adoption readiness*.
- Torchiano, M., & Jedlitschka, A. (2020). Data Pipeline Management in Practice: Challenges and Opportunities. In *Proceedings*. <http://www.springer.com/series/7408>
- Valentine, C., & Merchan, W. (2016). DataOps : An Agile Methodology for Data-Driven Organizations Data Science and Machine Learning in the Enterprise. *Oracle, Datasience.Com*. <https://www.oracle.com/cn/a/ocom/docs/oracle-ds-data-ops-map-r.pdf>
- Xu, J., Naseer, H., Maynard, S., & Fillipou, J. (2022). *Leveraging Data and Analytics for Digital Business Transformation through DataOps: An Information Processing Perspective*. 1–11.