

BAB III

METODOLOGI PENELITIAN

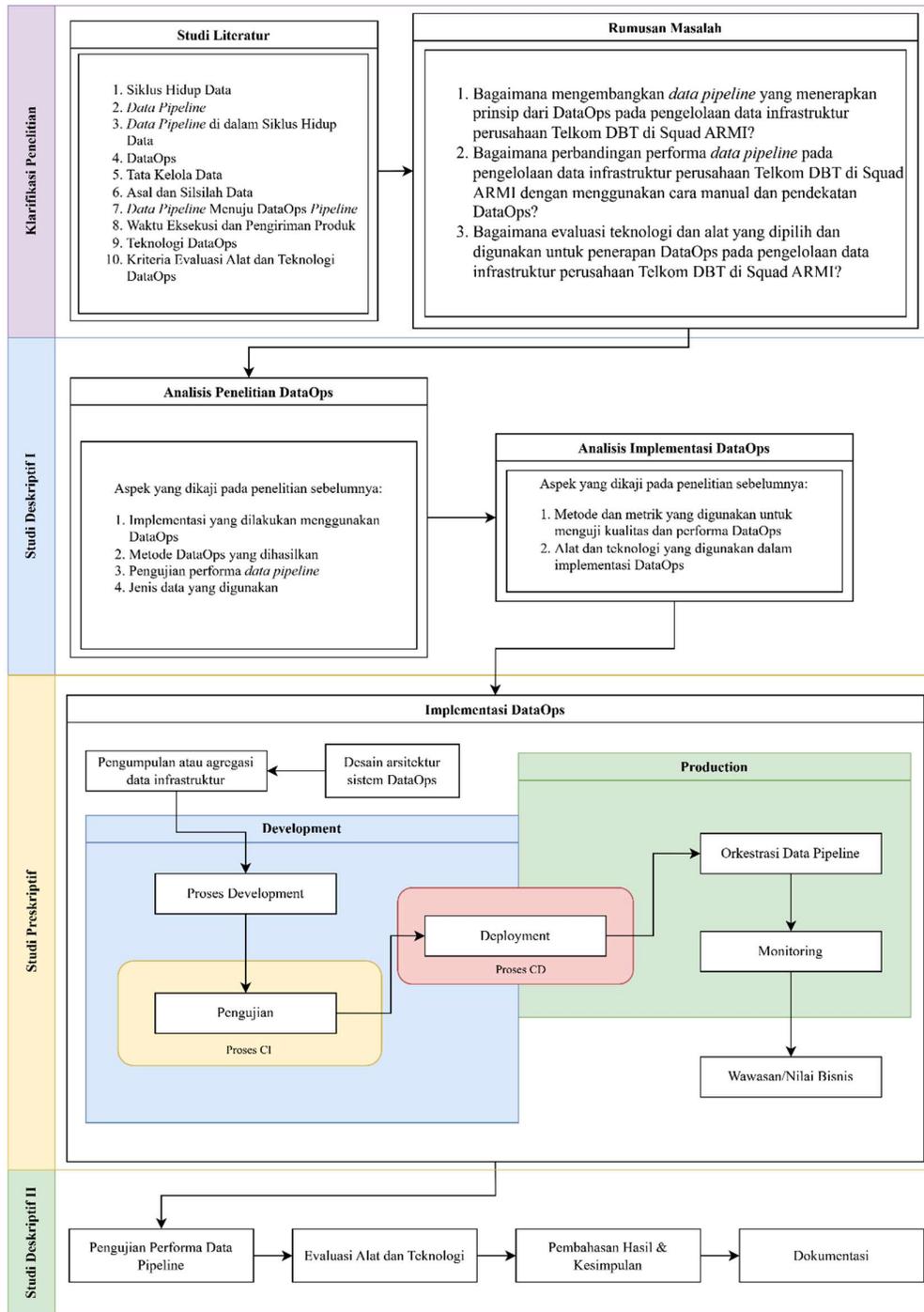
3.1 Desain Penelitian

Penelitian ini dilakukan dengan urutan langkah yang mengacu pada Design Research Methodology (DRM). Menurut Blessing & Chakrabarti (2009) DRM adalah seperangkat metode dan pedoman pendukung yang akan digunakan sebagai kerangka kerja untuk melakukan penelitian perancangan. Dijelaskan bahwa penelitian perancangan merupakan penelitian yang berhubungan dengan perumusan, validasi model dan teori tentang fenomena desain dengan segala aspeknya (orang, produk, pengetahuan, metode, alat, organisasi). Adapun desain penelitian yang dirancang pada penelitian ini terdapat pada Gambar 3.1.

3.1.1 Klarifikasi Penelitian

Klarifikasi penelitian adalah langkah awal dari DRM dalam menentukan topik yang akan diangkat. Pada bagian ini penulis mengangkat topik terkait praktik DataOps dalam pengelolaan data infrastruktur perusahaan di Telkom DBT pada Squad ARMI. Kemudian langkah berikutnya yaitu melakukan pengumpulan data dengan studi literatur guna mengumpulkan teori-teori yang dapat digunakan sebagai landasan dalam penelitian.

Beberapa teori yang dikaji dalam studi literatur diantaranya adalah Siklus Hidup Data, *Data Pipeline*, *Data Pipeline* di dalam Siklus Hidup Data, DataOps, Tata Kelola Data, Asal dan Silsilah Data, *Data Pipeline Menuju DataOps Pipeline*, Waktu Eksekusi dan Pengiriman Produk, Teknologi DataOps, serta Kriteria Evaluasi Alat dan Teknologi DataOps. Penulis mendapatkan teori-teori yang diperoleh dari berbagai sumber seperti jurnal, buku elektronik, buku, artikel ilmiah dan sejenisnya untuk dijadikan sebagai dasar penelitian. Detail penjelasan tentang teori-teori tersebut disajikan pada Bab II. Selain itu, setelah mempelajari teori-teori tersebut, terbentuk suatu rumusan masalah.



Gambar 3.1 Desain Penelitian

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

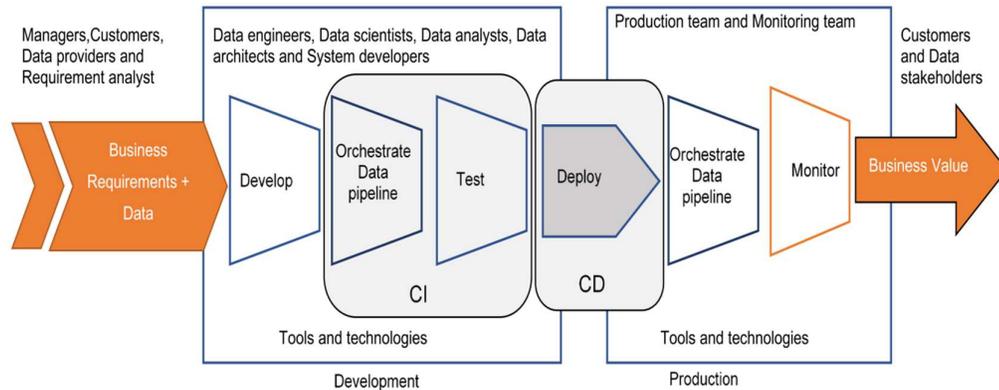
3.1.2 Studi Deskriptif I

Pada tahap ini dilakukan analisis lebih dalam seputar penelitian terdahulu yang berkaitan dengan praktik DataOps serta implementasinya. Tujuan dari tahap ini adalah memperdalam pemahaman tentang masalah yang diteliti terkait implementasi praktik DataOps serta analisis performa waktu eksekusi dari *data pipeline* yang telah dirancang dengan membandingkan pendekatan DataOps dan eksekusi manual/tradisional. Penelitian terdahulu yang berkaitan dengan DataOps dibandingkan satu sama lain untuk mengkaji *state-of-the-art* dari topik yang diangkat. Terdapat beberapa aspek yang dibandingkan yaitu implementasi, teknologi yang digunakan, metode yang dihasilkan, pengujian performa *pipeline* data, dan data infrastruktur. Dikaji juga terkait metrik pengujian untuk performa *data pipeline*, dalam kasus ini yaitu ETL *pipeline* sebagai acuan pada tahap pengujian.

3.1.3 Studi Preskriptif

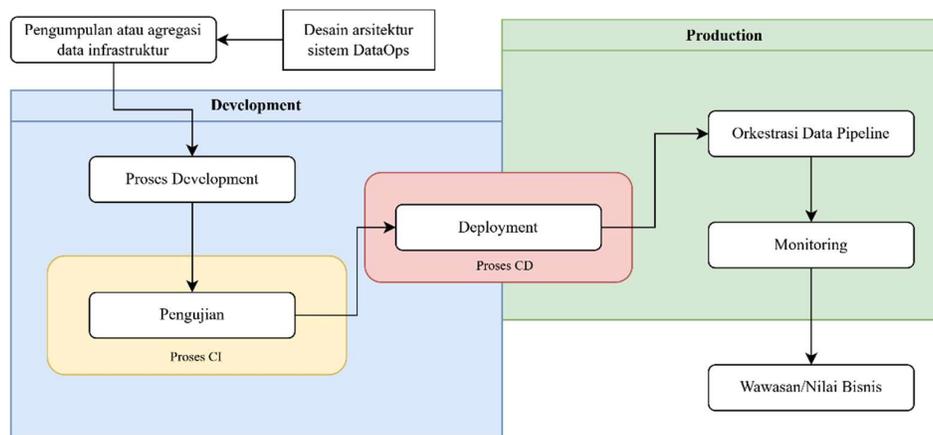
Setelah didapatkan pemahaman, selanjutnya yaitu penentuan sistem yang akan dibangun pada penelitian ini. Arsitektur sistem dibentuk memberikan gambaran terkait konsep kerangka kerja dari *pipeline* DataOps yang akan diterapkan pada studi kasus pengelolaan data infrastruktur perusahaan. Sistem ini juga menjadi acuan terhadap pengumpulan data yang dibutuhkan dalam implementasi kerangka kerja DataOps. Pada penelitian ini data yang dibutuhkan adalah data Live Infra (dinamis) dan data Static Infra (statis) sebagai bahan untuk proses *data pipeline* dan acuan dalam mempersiapkan tempat penampungan data (*data lake*), skema basis data, serta gudang data (*data warehouse*).

Pada tahap ini akan dilakukan implementasi proyek menggunakan DataOps, dengan menggunakan data infrastruktur perusahaan Telkom DBT, dan akan dilakukan tahapan-tahapan yang disebut sebagai *data pipeline* dengan pendekatan DataOps (*DataOps Pipeline*) untuk menghasilkan wawasan atau nilai bisnis sesuai dengan kebutuhan. Metode yang akan digunakan dalam implementasi ini menggunakan DataOps proses yang rancang oleh H. Atwal (2019) dan dikembangkan oleh Mainali dkk. (2021).



Gambar 3.2 Proses DataOps (Mainali et al., 2021)

Berdasarkan DataOps proses pada Gambar 3.2, penulis merancang turunan proses DataOps sesuai dengan kebutuhan studi kasus yang dilakukan pada pengelolaan data infrastruktur, sebagai berikut.



Gambar 3.3 Proses DataOps berdasarkan Studi Kasus

Sejalan dengan langkah-langkah yang dirancang oleh penulis pada Gambar 3.3, tahapan yang akan dilakukan pada pengembangan *data pipeline* dengan pendekatan DataOps adalah sebagai berikut.

1. Desain arsitektur sistem DataOps

Tahap perancangan desain arsitektur DataOps untuk studi kasus pengelolaan data infrastruktur perusahaan di Telkom DBT. Arsitektur dibangun berdasarkan hasil studi literatur dan pemahaman penulis terkait *tech stack* dan kebutuhan dalam implementasi pengelolaan data infrastruktur perusahaan.

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

2. Pengumpulan atau agregasi data infrastruktur

Tahap pengumpulan atau agregasi data infrastruktur dilakukan untuk mengumpulkan bahan yang akan diproses melalui *data pipeline* dengan konsep ETL (*Extract-Transform-Load*). Data diambil dari sumber data yaitu Datadog berupa metrik dari *Cluster* Kubernetes yang disebut sebagai Live Infra. Kemudian data infrastruktur yang kedua berupa Static Infra diambil dari Google Spreadsheet.

3. Proses pembangunan ETL

Proses ini merupakan tahapan pembangunan ETL *Data Pipeline* yang terdiri dari beberapa langkah:

1. Membuat desain ETL
2. Kode atau implementasi dengan menggunakan bahasa Python
3. Implementasi kode kedalam DAG menggunakan Airflow

4. Pengujian

Tahap pengujian ini dilakukan dengan menerapkan konsep *Continuous Integration* (CI) untuk otomatisasi dalam pengujian sesuai dengan praktik DataOps. Tujuan dari pengujian ini adalah untuk memastikan data yang masuk dan keluar sudah sesuai dengan ekspektasi dan bisa dilakukan berulang kali secara otomatis.

5. *Deployment*

Pada tahap ini akan ada proses *Continuous Deployment* (CD) yang hanya dilakukan jika kode berada pada *branch master*. Proses *code review*, *merge request*, *comment*, dan *merge* dilakukan pada *branch develop* sesuai dengan kaidah kolaborasi dan pemisahan *environment* yang diusulkan oleh DataOps. Setelah di-*merge* kode akan masuk ke *branch master*, dan akan dilakukan *deployment* kode secara otomatis menggunakan teknologi GitAutoDeploy dengan koneksi Git Webhook. Kode akan di-*deploy* menggunakan *environment* Docker Container dengan Docker Compose sebagai orkestratornya yang sudah terinstal di Google Compute Engine.

6. Orkestrasi *data pipeline*

Pipeline akan otomatis dijalankan sesuai dengan jadwal yang telah ditentukan, serta akan diberikan detail dari DAG yang telah dibuat berupa *log*, grafik, kode,

waktu berjalan, waktu eksekusi dan detail lainnya. Tujuan dari orkestrasi ini adalah untuk manajemen *data pipeline* secara otomatis serta mempermudah proses pemantauan apabila terjadi *error* atau kendala.

7. Pemantauan (*Monitoring*)

Tahap ini dilakukan untuk memantau jalannya *data pipeline* sesuai kaidah yang diusulkan oleh DataOps, agar tingkat *observability* dari *data pipeline* menjadi tinggi. Tujuan dari pemantauan ini adalah untuk menangani masalah apabila terjadi kendala atau kesalahan pada proses *data pipeline* yang telah diorkestrasi.

8. Menghasilkan Nilai Bisnis

Tahap terakhir adalah dengan menampilkan hasil dari *data pipeline* yang telah dilakukan sebelumnya kedalam *website* Satu Data INS dan Google Looker Studio. *Platform* tersebut menjadi alat visualisasi data yang sudah diproses sebelumnya untuk menghasilkan wawasan atau nilai bisnis agar memudahkan *stakeholder* seperti manajer dan pimpinan lainnya untuk memutuskan arah bisnis kedepannya. Tujuan lainnya adalah untuk mendapatkan *feedback* dan inovasi agar sistem yang telah dibangun bisa berkembang menjadi lebih baik lagi.

3.1.4 Studi Deskriptif II

Untuk mengetahui tingkat keberhasilan penelitian berdasarkan performa dari pendekatan praktik DataOps dalam *data pipeline*, diperlukan suatu pengujian dari sisi waktu eksekusi yang mengacu kepada penelitian yang dilakukan oleh Mainali (2021). Dua implementasi dilakukan untuk tujuan perbandingan. Implementasi pertama tanpa menggunakan alat orkestrasi dan penjadwalan, dilakukan di lingkungan Windows. Sebagai perbandingan, implementasi kedua menggunakan Docker Containers dengan Apache Airflow sebagai *job scheduler* yang dilakukan menggunakan mesin virtual di Google Cloud. Waktu eksekusi diukur dari kedua eksperimen dengan alat pengujian menggunakan Apache Airflow sebagai orkestra alur kerja dari *pipeline* yang sudah dibangun dan menggunakan perintah Bash yaitu *time* ketika eksekusi manual. Kemudian akan dilakukan analisis sebagai bahan pengetahuan dan pembahasan kualitas sistem atau arsitektur dengan praktik DataOps dari sisi performanya. Kemudian akan dilakukan evaluasi terhadap alat dan teknologi yang digunakan pada implementasi ini berdasarkan metrik yang dibuat oleh Mainali (2021), dengan kriteria evaluasi *Simplicity*, *Usability*, dan

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

Compatibility dan jenis aplikasi pada alat-alat yang digunakan oleh DataOps diantaranya alat orkestrasi *workflow*, alat pengujian dan pemantauan, alat automasi *deployment*, alat *versioning* kode dan artefak, alat analitik dan visualisasi, serta alat kolaborasi dan komunikasi (Mainali et al., 2021). Tujuan dari implementasi pada proyek analitik data ini adalah untuk menunjukkan perbedaan pada implementasi proyek menggunakan pendekatan DataOps dengan implementasi proyek yang ada saat ini dan menjelaskan keuntungan dalam mengikuti petunjuk implementasi DataOps.

Selanjutnya penulis akan membuat kesimpulan atas hasil yang diperoleh. Selain itu akan dipaparkan juga kelebihan dan kekurangan dalam penelitian ini. Pada bagian ini juga kesimpulan disusun untuk menjawab rumusan masalah yang sudah dibentuk. Selain itu, penulis memberikan saran kepada peneliti selanjutnya supaya penelitian ini dapat dilanjutkan dan ditingkatkan dengan lebih baik lagi. Terakhir, akan dilakukan penyusunan buku skripsi.

3.2 Alat dan Bahan Penelitian

3.2.1 Alat Penelitian

Alat penelitian yang digunakan dari sisi perangkat keras dalam penelitian ini yaitu sebagai berikut.

1. Prosesor Intel Core i3-10110U 4 CPU 2.10 GHz
2. RAM 8 GB DDR 4
3. SSD 512 GB

Sedangkan untuk perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Microsoft Windows 11 Home
2. Visual Studio Code
3. Google Cloud
4. Mesin Virtual e2-standard-2 dengan sistem operasi Ubuntu dan 50 GB penyimpanan (Google Compute Engine)
5. Docker
6. Docker Compose
7. PostgreSQL
8. Python

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

9. Python *libraries* (pip)
10. Apache Airflow 2.3
11. Gitlab CI/CD
12. GitAutoDeploy by Olipo186
13. Nginx Proxy Manager
14. Hasura
15. Datadog
16. Google Spreadsheet
17. Microsoft Excel
18. Satu Data INS (INS Console)
19. Google Cloud Storage
20. Google BigQuery
21. Google Looker Studio

3.2.2 Bahan Penelitian

Bahan penelitian yang akan digunakan diantaranya adalah jurnal, buku, buku elektronik, artikel ilmiah, dan lain sebagainya yang digunakan untuk menunjang pemahaman penulis dalam membangun sistem DataOps. Selain itu dokumentasi dari alat dan teknologi yang digunakan akan membantu dalam proses implementasi.

3.3 Instrumen Penelitian

Instrumen penelitian yang digunakan dalam penelitian ini terdiri dari 2 bagian, untuk eksperimen pengukuran waktu eksekusi dari *data pipeline* menggunakan Apache Airflow dan Bash Command *time*. Sedangkan untuk evaluasi alat dan teknologi yang pilih pada implementasi ini menggunakan instrumen kuesioner yang akan diisi oleh tim data dari Squad ARMI terhadap evaluasi dari alat dan teknologi DataOps yang digunakan untuk implementasi ini. Metrik yang diambil diantaranya adalah tingkat *Simplicity*, *Usability*, dan *Compatibility* (Mainali et al., 2021). Keduanya menggunakan metrik berdasarkan penelitian yang dilakukan oleh Mainali, Kiran (2021).

3.4 Analisis Data

Eksperimen dirancang untuk melacak waktu eksekusi setiap tahap dan menghitung total waktu eksekusi eksperimen. Total waktu eksekusi tidak mempertimbangkan waktu *buffer* antara langkah-langkah eksekusi di eksperimen 1, sedangkan di eksperimen 2, tidak ada waktu *buffer* karena telah menggunakan orkestrator alur kerja berbasis DAG untuk mengotomatiskan eksekusi. Waktu eksekusi diuji masing-masing sebanyak 10 kali dengan data yang sama, dan akan diambil dan dibandingkan rata-rata waktu eksekusi dari setiap tahap. Terakhir, kedua eksperimen ini akan diuji terhadap hipotesis yang dirancang menggunakan uji t. Untuk analisis data terhadap evaluasi alat dan teknologi DataOps berdasarkan Tabel 3.1 menggunakan Skala Likert dengan skala 1-5.

Tabel 3.1
Skor/Bobot Skala Likert (Albaum, 1997; Allen & Seaman, 2007; Joshi et al., 2015)

Jawaban Pernyataan	Skor Likert
Sangat Tidak Setuju (STS)	1
Tidak Setuju (TS)	2
Netral (N)	3
Setuju (S)	4
Sangat Setuju (ST)	5

Menurut Sugiyono (2006) Skala Likert digunakan dalam pengukuran sikap, pendapat, dan persepsi individu atau kelompok terhadap fenomena sosial yang sedang diteliti (Sugiyono, 2006). Skala Likert ini dikembangkan oleh Likert pada tahun 1932. Skala Likert terdiri dari empat atau lebih pernyataan/pertanyaan yang digabungkan menjadi sebuah nilai yang merepresentasikan sifat individu, seperti pengetahuan, sikap, dan perilaku (Pranatawijaya et al., 2019). Skala ini sering digunakan sebagai alat ukur *psikometrik* dalam kuesioner dan paling umum digunakan dalam penelitian (Joshi et al., 2015). Berikut tabel 3.2 menunjukkan interpretasi Skala Likert. Indikator yang akan dinilai pada skala ini yaitu *Simplicity*, *Usability*, dan *Compatibility* pada alat atau teknologi yang digunakan.

Tabel 3.2
Tabel Interpretasi Skala Likert (Allen & Seaman, 2007)

Skor Interpretasi	Kriteria Kelayakan
0% - 19,99%	Sangat Kurang Layak
20% - 39,99%	Kurang Layak
40% - 59,99%	Cukup Layak
60% - 79,99%	Layak
80% - 100%	Sangat Layak

Rumus yang digunakan untuk mengukur persentase rata-rata skor yang didapat dari responden dengan Skala Likert adalah sebagai berikut.

$$\text{Index (\%)} = \left(\frac{\text{Total Skor}}{\text{Total pernyataan} \times \text{Jumlah responden} \times 5} \right) \times 100\%$$

3.5 Hipotesis

Berikut merupakan hipotesis pada penelitian ini.

Hipotesis 1: DataOps meningkatkan performa *data pipeline* dalam siklus hidup data pada pengelolaan data infrastruktur perusahaan.

Hipotesis 2: Performa *data pipeline* pada praktik pendekatan DataOps dengan pendekatan tradisional.

H₀: Praktik pendekatan DataOps dengan pendekatan tradisional memiliki rata-rata performa yang sama.

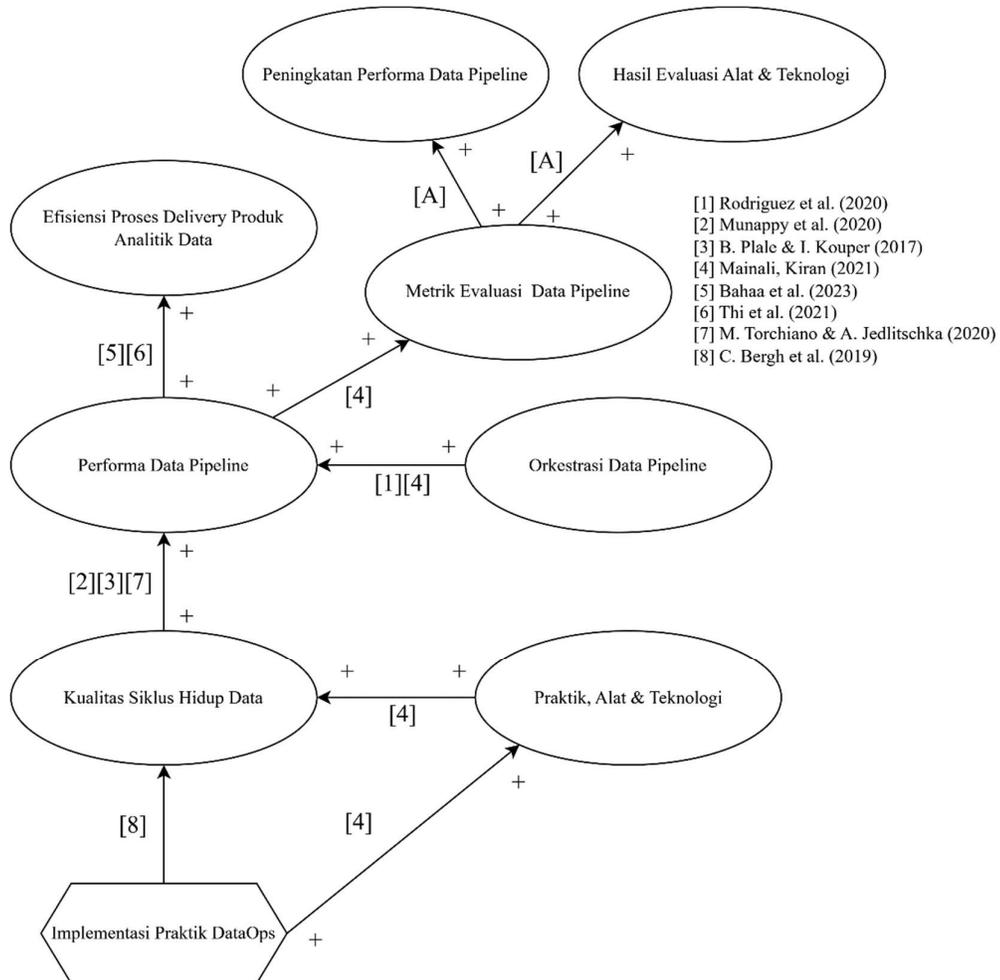
H₁: Praktik pendekatan DataOps dengan pendekatan tradisional memiliki rata-rata performa yang berbeda.

Hipotesis 3: Teknologi dan alat yang digunakan memiliki tingkat *Simplicity* yang layak, *Usability* yang layak, dan *Compatibility* yang layak.

3.6 Model Dampak

Berikut merupakan model dampak yang telah dibuat oleh Penulis sebagai acuan untuk mencapai tujuan dan menjelaskan perubahan dampak yang terjadi, serta didukung oleh hasil temuan dari penelitian sebelumnya. Pada Gambar 3.4 berdasarkan buku nya C. Bergh et al. (2019) praktik DataOps ini menjadi metode

yang dapat meningkatkan kualitas dari siklus hidup data (Bergh et al., 2019). Kemudian implementasi dari DataOps ini akan memengaruhi penggunaan dari alat dan teknologi yang sesuai dengan prinsip DataOps (Mainali et al., 2021).



Gambar 3.4 Model Dampak