

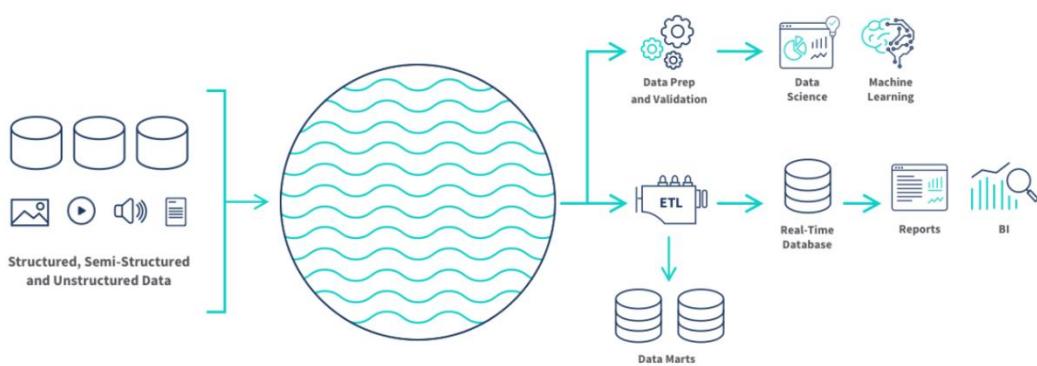


MAKALAH SUGGESTION SYSTEM

RAZIS AJI SAPUTRO

“Menurunkan Frekuensi Deviasi Data Working Hours Pada DAR(Raw Daily) dengan DAR(Raw MTD) kasus Integrasi, Menggunakan Skema Data Lake yang Terintegrasi Head Office dan Site Project (Rantau Area) Periode Juni-Agustus”

Periode 2023



**PT. HASNUR RIUNG SINERGI HEAD OFFICE
DEPARTEMEN ENGINEERING
TAHUN 2023**



DAFTAR ISI

DAFTAR ISI.....	ii
PROFIL PERUSAHAAN DAN PROFIL TIM.....	iv
STRUKTUR ORGANISASI.....	v
GLOSARIUM	vi
ABSTRAK.....	vii
LANGKAH 1 : MENENTUKAN TEMA & TARGET	1
1.1 Identifikasi Masalah.....	1
1.2 Prioritas Masalah	3
1.3 Menentukan Tema.....	3
1.4 Menentukan Sasaran Tema dengan SMART.....	3
1.5 Penguatan Sasaran Tema dengan Analisa QCDSM.....	4
1.6 Statement Stakeholder (opsional)	4
1.7 Membuat Time Frame Improvement.....	5
1.8 Pengesahan Tema	6
LANGKAH 2 : MENGIDENTIFIKASI MASALAH	7
2.1 Tabel ANAKONDA	8
2.2 Menetapkan Objek Masalah dan Analisa Sebab-Akibat	9
2.3 Validasi Akar Penyebab	10
LANGKAH 3 : MENENTUKAN SOLUSI TERBAIK.....	11
3.2 Membuat dan Menganalisis Alternatif Solusi Masalah	11
3.2 Penjelasan Pemilihan solusi	11
LANGKAH 4 : MENYUSUN RENCANA PERBAIKAN.....	14
4.1 Menyusun Rencana Perbaikan	14
4.2 Analisa Risiko dan Pencegahan	14
4.3 Pengesahan Rencana Perbaikan.....	15
LANGKAH 5 : PELAKSANAAN PERBAIKAN.....	16
5.1 Identifikasi Sumber Daya yang Dibutuhkan.....	16
5.2 Melaksanakan Perbaikan yang Telah direncanakan	16
LANGKAH 6 : EVALUASI HASIL.....	19
6.1 Analisa Hasil Improvemet.....	19
6.2 Perubahan Hasil Improvement	19
6.3 Analisa QCDSM	21
6.4 Voice of Customer.....	22
LANGKAH 7 : STANDARISASI.....	23
7.1 Mengusulkan Standard Baru	23
7.2 Pengesahan Standard Baru	23



LANGKAH 8 : Menentukan Tema Perbaikan Selanjutnya	25
8.1 Identifikasi masalah berikutnya	25
8.2 Membuat Time Frame Improvement.....	25



PROFIL PERUSAHAAN DAN PROFIL TIM

Profil Perusahaan :

PT Hasnur Riung Sinergi (HRS) merupakan Perusahaan Jasa Pertambangan yang berdiri pada tahun 2011 dan bagian dari unit usaha Hasnur Group dalam SBU Mining yang didirikan secara patungan antara PT Hasnur Jaya Energi (HJE) dengan PT Riung Mitra Lestari (RML). Sebagai mitra pemilik tambang, PT. HRS beroperasi dengan teknik dan pola penambangan yang benar dengan menerapkan Good Mining Practice untuk memberikan yang terbaik bagi stakeholder nya dalam bidang usaha yang ditekuninya Total Mining Service. PT. HRS beroperasi di Site AGM, EBL & BRE Rantau, Tapin, Kalimantan Selatan Serta anak usaha PT Fajar Anugrah Dinamika (FAD) beroperasi di Berau Coal, Kalimantan Timur.

Profil Tim :

Nama Perusahaan	:	PT. Hasnur Riung Sinergi
Nama	:	Razis Aji Saputro
Departemen	:	Engineering
Jobsite	:	Head Office
Periode Improvement	:	2023
Tanggal Pembentukan Tim	:	Januari 2023

Prestasi yang dicapai	:	
	:	1. Juara 1 SS tingkat Head Office
	:	2. Juara 3 QCP HRS tingkat Korporat
	:	3. Juara Harapan 3 QCC HRS tingkat Korporat
	:	4. Juara 2 Kompetisi video safety 2023
	:	5. Juara 2 Kompetisi poster safety 2023

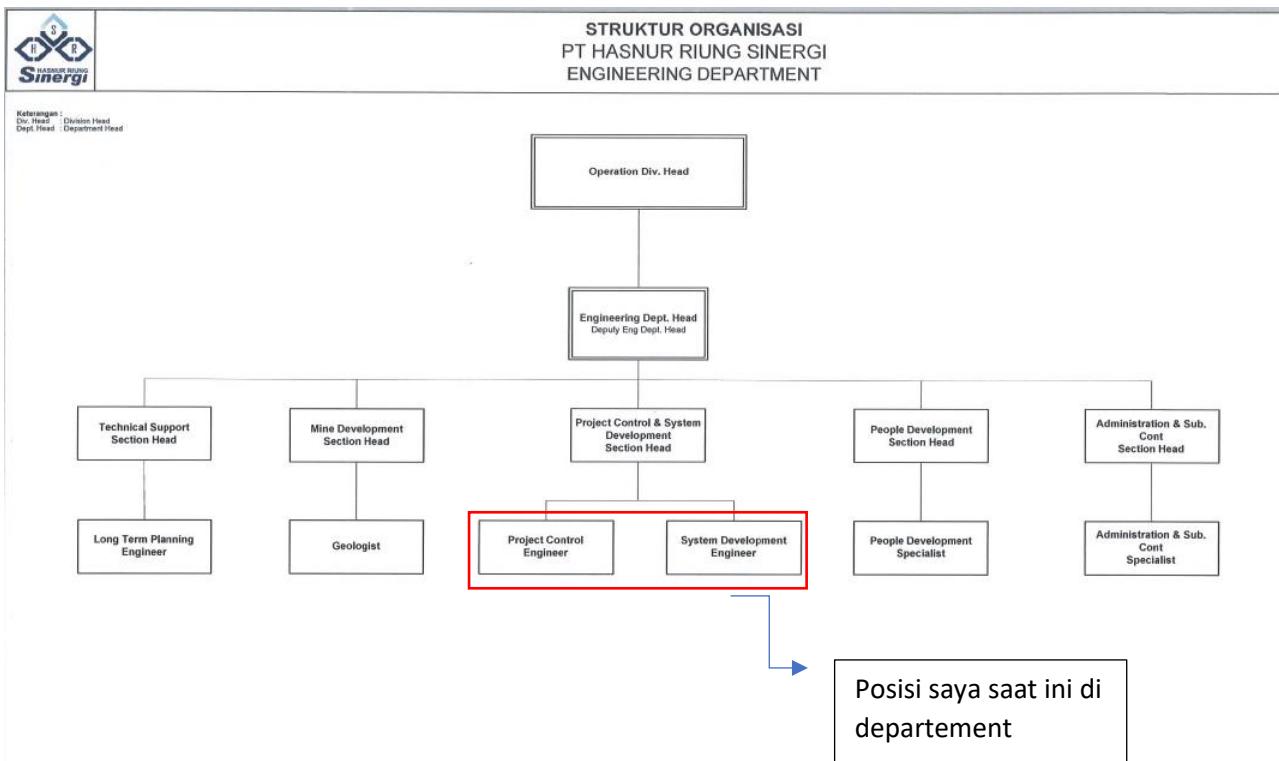
Struktur Tim

Fasilitator
 Fuad Hasan

Inovator
 Razis Aji Saputro



STRUKTUR ORGANISASI



No. Doc : INHRGS/IRS-HOK/I/2023/ISO	Prepared by : OD&PD Dept Head 	Approved by : Engineering Dept Head 	Approved by : Operation Div. Head 	Approved by : HRGS Director
Effective Date : 1 Januari 2023 Rev. : 3/10 Page :	Elly Petersandy	Wahyu Irawan	Hery Sugiyanto	Priyo Trilaksono

1. TUGAS & TANGGUNG JAWAB

(Jelaskan dengan singkat tugas pokok & tanggung jawab pemangku jabatan)

Uraian Tugas	
1. Core : Project & System Development	a. Pengembangan Project : membuat sistem pelaporan yang terpusat, update, real time dan berdasarkan data yang sama antara site dan HO. b. Membuat dan melakukan update database engineering terkait administrasi (data, pelaporan, dokumen).



GLOSARIUM

1. Data : Sekumpulan keterangan ataupun fakta yang dibuat dengan kata-kata, kalimat, simbol, angka, dan lainnya.
2. Database : Kumpulan data yang dikelola sedemikian rupa berdasarkan ketentuan tertentu yang saling berhubungan sehingga mudah dalam pengelolaannya.
3. Data Lake : Sebuah danau data atau bisa diartikan sebagai tempat penyimpanan yang memiliki skala besar dan biasa digunakan oleh perusahaan.
4. Source : Sumber data.
5. Hybrid : Penggabungan 2 metode online dan offline.
6. Scrap : Pengambilan data dari berbagai source.
7. Transformasi : Istilah untuk merubah data dalam bentuk originalnya ke bentuk data yang di inginkan.
8. Storage : Storage tempat penyimpanan data dalam bentuk berbagai ekstensi file.
9. Data Mart : Sekumpulan data yang sudah diubah bentuknya menjadi yang diinginkan dan didistribusikan untuk suatu kepentingan tertentu.
10. Pipeline : Merupakan istilah suatu proses berlaku dari hulu ke hilir.
11. Cloud : Media yang dipersiapkan oleh suatu vendor untuk memudahkan akses dimanapun dan kapanpun dalam satu sumber data center.
12. WH(Working Hours) : Jam Kerja Unit Operational, pencatatan berdasarkan segmentasi aktivitas
13. MTD (Month to Date) : Perhitungan bulan berjalan sampai tanggal terupdate
14. DAR(Daily Activity Report) : Template report untuk pengiriman data melalui email ditujukan dari site ke Head office atau owner



ABSTRAK

Departement Engineering merupakan salah satu departemen yang berada dibawah Divisi Operation. Departement Engineering sebagai salah satu garda terdepan untuk melakukan perencanaan pertambangan. Berdasarkan President Letter tahun 2023, dimana menyebutkan **Efisiensi dengan menggunakan proses digitalisasi sebagai Upaya mengikuti perkembangan teknologi.** Data merupakan salah satu tonggak utama dari PT Hasnur Riung Sinergi untuk melakukan strategi dalam meningkatkan kemajuan bisnis. Sebagai **decision making strategy** penggunaan data sangatlah penting bagi pemangku kepentingan. Site Project sebagai produsen data dan HO sebagai konsumen data serta sebagai **decision making strategy** perlu adanya **data yang terintegrasi**. Hubungan yang terjadi sekarang antara HO dan Site project, data dikirim dari site ke HO dengan menggunakan email. Tentunya ketika ada revisi data, tidak akan teridentifikasi dikemudian hari kecuali ada pengiriman kembali. Namun secara aktual pengiriman jarang dilaksanakan, sehingga menyebabkan data tidak sama antara HO dan Site Project. Integrasi data menjadi hal yang perlu dilaksanakan, dengan merubah proses bisnis yang terjadi menggunakan metode **data lake** yang tersinkronisasikan secara hybrid (offline dan cloud).

Kata Kunci : Data Lake, Tranformasi Data, Digitalisasi, Decision Making Strategy, Hybrid



LANGKAH 1 : MENENTUKAN TEMA & TARGET

1.1 Identifikasi Masalah

Departemen Engineering merupakan pioner atau bisa dibilang garda terdepan dalam perkembangan operasional kontraktor batubara bersamaan dengan department produksi. Dalam mengatur strategi untuk mencapai effisiensi dan perkembangan perusahaan, setiap tahunnya akan disusun Key Performance Indicator untuk menjadi tolak ukur keberhasilan department. Saya yang berada dalam section Project Control dan System Development, bertanggung jawab atas KPI berikut:

Monitoring Kegiatan Penambangan		Plan	Actual	Ach
3.1. Pelaporan kegiatan penambangan	Report	100%	100%	100%
3.2. Melakukan evaluasi terhadap deviasi antara aktual penambangan dengan rencana	Report & PICA	100%	100%	100%
3.3. Evaluasi Cost Penambangan	Report	100%	0%	0%
3.4. Project digitalisasi pelaporan yang terintegrasi	Management Data Terintegrasi	100%	25%	25%
3.5. Observasi parameter-parameter produktivitas alat dengan team operation	Report Observasi	100%	60%	60%

Table 1

Dari KPI diatas kemudian kami menganalisa dengan diagram pareto untuk memfokuskan perbaikan untuk meningkatkan achievement KPI pada Q2 dan Q3:

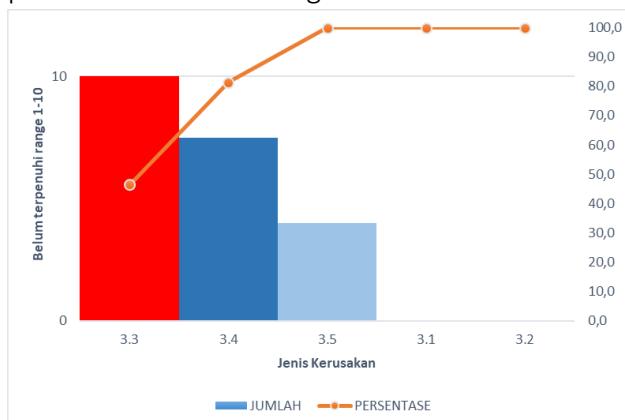


Diagram 1

Dari poin diatas pada tahun 2023 Jan-Apr departement engineering Project Control dan System Development, masih memiliki 2 PR pada poin 3.3 dan 3.4. Adapun dalam **SS ini akan berfokus kepada peningkatan point no 3.4** karena 3.3 digunakan sebagai QCC Engineering. Dimana peningkatan ini akan membantu dalam pemenuhan OTP departemen engineering.

Salah satu issue yang sering terdengar di head office adalah data email yang dikirimkan secara daily, terkadang tidak sama dengan data report secara weekly karena terdapat update data yang tidak dikirimkan kembali ketika ada revisi. Selanjutnya kami telusuri lebih jauh dan melakukan beberapa survey terdapat beberapa masalah sebagai

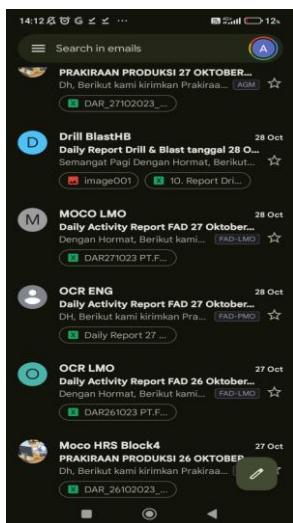


berikut yang berkaitan, 5 responden mewakili tim Head Office dan Site EBL namun belum membuktikan bahwa tuduhan ini terbukti benar.

Problem	ya	tidak	Satuan	bobot
Data terintegrasi	0	5	kali	100%
Perbedaan RAW data yang signifikan	4	1	kali	80%
Data Visibel dan akses mudah	0	5	kali	100%

Dari tabel survey diatas kami mencoba menjabarkan satu persatu:

1. Data terintegrasi



Penyebab utama adalah data Center belum terpusat dan masih menggunakan database di masing masing site project secara aktual data dikirim melalui email sebagai perantara data.

Namun beberapa kendala sering terjadi seperti:

- Data tidak terupdate
- Terdapat deviasi data karena email tidak dikirimkan revisinya
- Perbedaan formula perhitungan

2. Perbedaan deviasi RAW data yang signifikan

Hal berkaitan dengan DAR (Raw Daily) dan DAR (Raw MTD). Kami melakukan Analisa dengan membandingkan data tersebut sehingga muncul tabel perbandingan yang menyatakan deviasi data DAR (Raw Daily) dan DAR (Raw MTD). Adapun deviasi tersebut tertuang dalam Tabel 2:

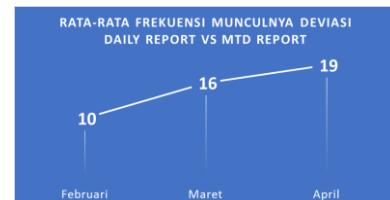
Bulan	Deviasi Rata-Rata Daily Report dengan MTD Report										
	Prod_OB	Prod_TS	Prod_MD	Prod_CO	HM	WH_Gen	WH_OB	WH_TS	WH_MD	WH_CO	STB
Februari	-	-	-	-	-	5	-	-	-	-	5
Maret	-	20.734	-	-	722	121	127	48	-	-	6 - 80
April	53.789	-	35	16	150	-	38	100	-	-	0 234

Dari hasil observasi selama 3 bulan beberapa kolom menyumbang deviasi yang signifikan, sehingga ketika pembuatan report dan dashboard di Head Office terdapat data yang tidak sama. Kemudian dari data deviasi diatas dijabarkan berapa frekuensi terjadinya deviasi tersebut (tabel 3):

Bulan	Frekuensi terjadinya Deviasi Daily Report dengan MTD Report											
	Prod_OB	Prod_TS	Prod_MD	Prod_CO	HM	WH_Gen	WH_OB	WH_TS	WH_MD	WH_CO	STB	Total
Februari	-	-	-	-	-	5	-	-	-	-	5	10
Maret	3	-	-	2	3	2	2	-	-	1	3	16
April	4	-	1	1	3	3	3	-	-	1	3	19

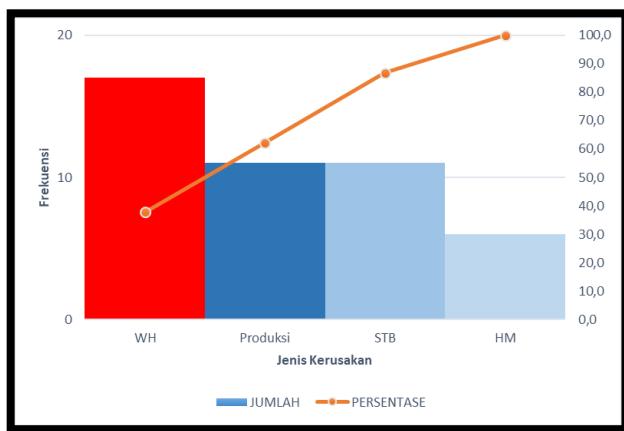
Dan dengan menyempitkan kategori dari pencatatannya diperoleh data sebagai berikut (tabel 4):

Bulan	Frekuensi terjadinya Deviasi Daily Report dengan MTD Report			
	Produksi	HM	WH	STB
Februari	-	-	5	5
Maret	5	3	5	3
April	6	3	7	3
Total	11	6	17	11



1.2 Prioritas Masalah

Permasalahan dari point 1.1 dapat dikerucutkan dengan membuat diagram pareto dari tabel 4.



Dari diagram pareto, dapat diambil bahwa frekuensi terjadinya **deviasi WH** 38% dari total frekuensi

1.3 Menentukan Tema

Tema yang kami ambil merupakan hasil dari identifikasi masalah dari hasil penjabaran f hasil survey yang tertuang dalam tabel survey. Working Hours (WH) merupakan kasus deviasi frekuensi tertinggi, dan penyebab dari deviasi tersebut adalah belum adanya integrasi data secara aktual masih menggunakan email sebagai pengiriman. Sehingga kami menentukan untuk mengambil tema "**Menurunkan Frekuensi Terjadinya Deviasi WH Data Pada DAR (Raw Daily) dengan DAR (Raw MTD) kasus integrasi**".

Hal yang memperkuat pemilihan tema tersebut adalah:

1. Project Control & System Development Section di HO merupakan representasi dari Moco Section di site, sehingga harus dapat memberikan dukungan terhadap kinerja Moco Section di site.
2. Data dan evaluasi dari Site ke HO digunakan oleh manajemen dan departemen terkait sebagai bahan pertimbangan dalam mengambil keputusan.
3. Hasil identifikasi masalah dan Survey dari HO dan Site terdapat penekanan terkait integrasi data yang menjadi problem utama.

1.4 Menentukan Sasaran Tema dengan SMART

S (Specific)	M (Measurable)	A (Achievable)	R (Reasonable)	T (Time Base)
Data HO dan Site terintegrasi	Mentargetkan frekuensi deviasi dar (raw daily) dan dengan (raw mtd) dari 17 menjadi 0	Mengimplementasikan Metode data lake secara hybrid (Offline dan Cloud)	Efisiensi dengan menggunakan proses digitalisasi sebagai Upaya mengikuti perkembangan teknologi	Dalam 2 bulan frekuensi deviasi dar (raw daily) dan dengan (raw mtd) yang muncul menjadi 0



1.5 Penguatan Sasaran Tema dengan Analisa QCDSM

Aspek	Sebelum Perbaikan	Sasaran
Quality	Data belum terintegrasi HO dan Site secara realtime	Data terintegrasi HO dan Site secara realtime
Cost	Penyimpanan file berlangganan sebagai data historis menggunakan Excel dan csv file di 51 juta / 5 tahun	Menurunkan cost tahunan menjadi 0,13x menggunakan .parquet sebagai data historis
Delivery	Proses delivery lama karena data belum terintegrasi HO dan Site	Proses delivery lebih cepat karena data sudah terintegrasi HO dan Site
Safety	Security data dalam keamanan fisik dalam kondisi terjadi bencana kebakaran atau banjir belum ada.	Terjaminnya security data ketika terjadi bencana kebakaran atau Banjir dengan menggunakan system hybrid local dan cloud.
Morale	Pengelolaan data secara berulang menyebabkan rasa malas dan bosan berakibat data tidak terupdate daily	Update data akan dilakukan dengan one klik secara otomatis sehingga menghilangkan rasa malas dan bosan

1.6 Statement Stakeholder (opsional)

No	Stakeholder	PIC	Statement
1	Dept. Head Eng HO	Wahyu Irawan	“HO dan Site perlu adanya integrasi data , karena sering terjadi perbedaan data daily Report dengan data KPI Weekly secara actual.”
2	PCSD Section Head	Fuad Hasan	“ Integrasi bisa dilakukan dengan 2 cara datawarehouse IOSDB dan datalake dengan Storage onedrive, coba bisa dimaksimalkan.”
3	Moco Section Head	M. Hanif	“ Perlu adanya integrasi data agar supaya meringankan tim site untuk mengirimkan Raw data, dan integrasi ke dashboard.”



1.7 Membuat Time Frame Improvement

Langkah			Plan /Actual	Tahun : 2023															
				Mei				Juni				July				Agustus			
				1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
P	1	Menentukan Tema	Plan																
	2	Mengidentifikasi Faktor Penyebab Masalah	Plan																
	3	Menentukan Solusi	Plan																
	4	Merencanakan Perbaikan	Plan																
D	5	Melaksanakan Perbaikan	Plan																
C	6	Evaluasi Hasil Perbaikan	Plan																
A	7	Menetapkan Standarisasi	Plan																
	8	Menentukan Tema Berikutnya	Plan																
			Actual																

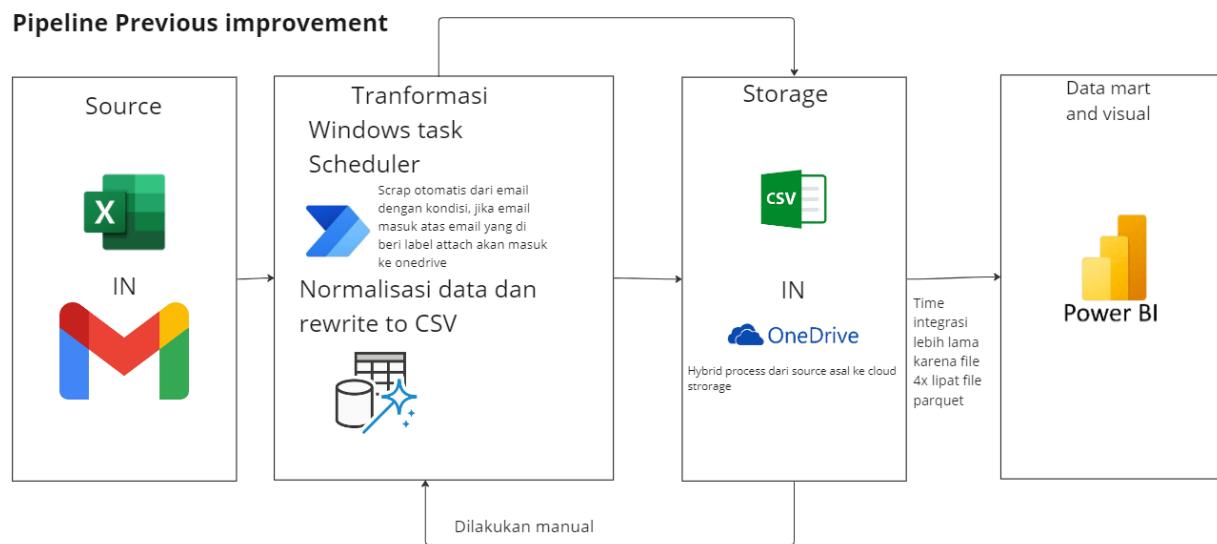


1.8 Pengesahan Tema

TEMA	JUDUL	TANGGAL
Menurunkan Frekuensi Terjadinya Deviasi WH Data Pada DAR(RAW DAILY) dengan DAR(RAW MTD) Kasus Integrasi.	Menurunkan Frekuensi Terjadinya Deviasi WH Data Pada DAR(RAW DAILY) dengan DAR(RAW MTD) Kasus Integrasi menggunakan Skema Data Lake yang Terintegrasi HO dan Site Project Rantau Juni-Agustus	5 Mei 2023
BUSINESS CASE	PROBLEM STATEMENT	PROJECT SCOPE
Perkembangan teknologi dalam dunia bisnis memberikan kemudahan kita dalam mengakses data dari suatu aplikasi. Data tersebut merupakan suatu tambang emas baru dari suatu bisnis untuk menjadi core bisnis untuk perkembangan perusahaan. Adanya data yang dihasilkan maka perlu adanya penyimpanan data, sehingga cost juga perlu dipertimbangkan dalam menyusun arsitektur database. Karena ketika cost penyimpanan terkhusus penyimpanan ini over, akan merugikan perusahaan. Merujuk pada cost dalam QCDSM, adalah cost penyimpanan data selama 5 tahun (estimasi kontrak project) secara file excel membutuhkan 51 juta/ Tahun/ Deprt/ Site, namun hal ini dapat ditekan dengan menggunakan file .parquet sehingga data history tersebut dapat ditekan menjadi 6,6 Juta/ Tahun/ Deprt/ Site.	Integrasi data yang belum terbentuk antara HO dan Site, menyebabkan adanya deviasi data DAR(Raw Daily) dengan DAR(Raw MTD) pada pengiriman email daily. Frekuensi deviasi ini dalam 3 bulan (Februari – April) rata-rata terjadi 4x dalam 1 bulan dari data yang dikirimkan site project. Akibatnya saat menganalisa data harian yang di summarykan dalam laporan weekly HO, adanya perbedaan data yang membuat management ragu dengan Analisa Prodty dan UA terkhusus pada kolom Working Hours yang memiliki deviasi tertinggi.	Project scope membatasi pengerajan secara ruang lingkup integrasi data secara hulu ke hilir atau biasa disebut data pipeline.
KOMENTAR		
Komentar Fasilitator :	→ Adanya deviasi data menyebabkan laporan yang tidak berinten. Imbasnya adalah munculnya keraguan-raguan dari manajemen atas laporan yang diterima. dan ini memberi kesan yang kurang profesional.	
Komentar Project Manager :	→ Project ini bisa dilanjutkan, karena data yang synkron antara Site dan HO bisa mempercepat proses analisa dan pengambilan keputusan	
Diajukan Oleh:	Diketahui Oleh:	Disetujui Oleh:
Razis Aji Saputro	Fuad Hasan	Wahyu Irawan
	Fasilitator	Manager Engineering

LANGKAH 2 : MENGIDENTIFIKASI MASALAH

Flow pengolahan Data di Head Office



Flow diatas merupakan flow sebelum improvement yang terjadi di head office PT hasnur Riung Sinergi. Flow tersebut terdiri dari 4 bagian, dimana masing-masing memiliki perannya masing-masing. Dari Hulu ke hilir flow diatas nantinya akan disebut sebagai data pipeline. Kemudian dari 4 bagian tersebut akan dijelaskan:

- Source** : Merupakan titik awal dimana data diambil, darimana data akan diambil. Sebagai contoh yang sering digunakan adalah sql sebagai database, excel sebagai database, dll. **Aktual sekarang data diambil dari pengiriman email dan di scraping otomatis.**
- Transformasi** : Merupakan proses pembersihan data atau merubah column dan row sesuai dengan yang dibutuhkan. **Pada aktual sekarang proses transfromasi masih manual menggunakan power query.**
- Storage** : Merupakan tempat penyimpanan file yang diambil baik sebelum dilakukan transformasi (data original) maupun data yang sudah dilakukan proses transformasi. Secara **aktual data excel banyak digunakan sebagai file penyimpanan di strorage ini dengan menggunakan aplikasi onedrive.**
- Visual report** : Merupakan visualisasi data secara business untuk mendapatkan insight dan untuk monitoring kondisi operational. Aktual menggunakan power bi dan data connecting langsung ke excel.



2.1 Tabel ANAKONDA (Analisa Kondisi yang Ada)

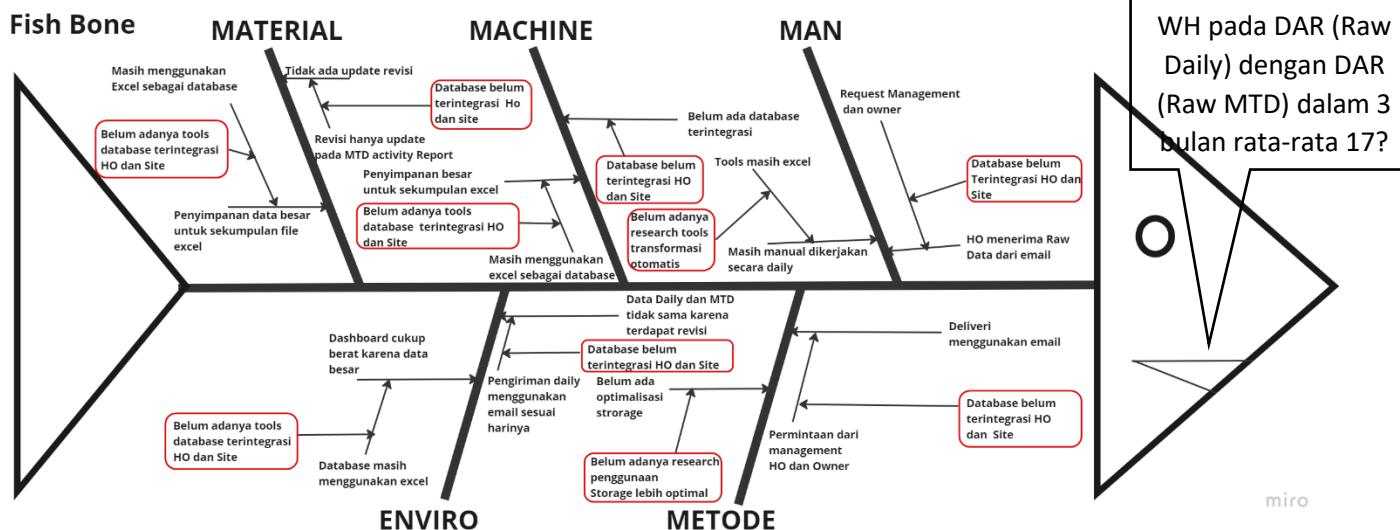
Dari flow pengolahan data di PT Hasnur Riung Sinergi. Kemudian saya breakdown dengan metode Anakonda. Berikut Analisa proses yang menjadi dasar Analisa kondisi yang ada dari alur integrasi data di HO:

No	Proses	faktor	What Should be Happen	What Actualy Happen	Validasi	validator	Kesimpulan																				
1	Scraping Source dari Email berbentuk Excel	Man	HO dapat akses untuk ke database	HO hanya menerima RAW Data dari Email	<table border="1"> <thead> <tr> <th>Problem</th><th>ya</th><th>tidak</th><th>Satuan</th><th>bobot</th></tr> </thead> <tbody> <tr> <td>Data terintegrasi</td><td>0</td><td>5 kali</td><td>100%</td><td></td></tr> <tr> <td>Perbedaan RAW data yang signifikan</td><td>4</td><td>1 kali</td><td>80%</td><td></td></tr> <tr> <td>Data Visibel dan akses mudah</td><td>0</td><td>5 kali</td><td>100%</td><td></td></tr> </tbody> </table>	Problem	ya	tidak	Satuan	bobot	Data terintegrasi	0	5 kali	100%		Perbedaan RAW data yang signifikan	4	1 kali	80%		Data Visibel dan akses mudah	0	5 kali	100%		by survey	Problem
Problem	ya	tidak	Satuan	bobot																							
Data terintegrasi	0	5 kali	100%																								
Perbedaan RAW data yang signifikan	4	1 kali	80%																								
Data Visibel dan akses mudah	0	5 kali	100%																								
		Machine	Adanya Database terintegrasi	Belum adanya database yang terintegrasi		Problem																					
		Material	Revisi Raw Data dikirimkan	Tidak ada pemberitahuan terhadap revisi	<p>FREQUENCIES OF DAILY REPORT VS MTD REPORT DEVIATIONS</p> <table border="1"> <thead> <tr> <th>Bulan</th><th>Frekuensi Deviasi</th></tr> </thead> <tbody> <tr> <td>Februari</td><td>10</td></tr> <tr> <td>Maret</td><td>16</td></tr> <tr> <td>April</td><td>19</td></tr> </tbody> </table> <p>Revisi tidak dikirimkan, karena perubahan data bisa sewaktu waktu terjadi, bisa dilihat dari deviasi Daily activity report dengan MTD activity report (Data terupdate MTD activity report)</p>	Bulan	Frekuensi Deviasi	Februari	10	Maret	16	April	19	Razis	Problem												
Bulan	Frekuensi Deviasi																										
Februari	10																										
Maret	16																										
April	19																										
		Enviro	Data HO dan site sama	Data Daily dan MTD tidak sama karena terdapat revisi yang tidak terkirim	<p>FREQUENCIES OF DAILY REPORT VS MTD REPORT DEVIATIONS</p> <table border="1"> <thead> <tr> <th>Bulan</th><th>Frekuensi Deviasi</th></tr> </thead> <tbody> <tr> <td>Februari</td><td>10</td></tr> <tr> <td>Maret</td><td>16</td></tr> <tr> <td>April</td><td>19</td></tr> </tbody> </table> <p>Revisi tidak dikirimkan, karena perubahan data bisa sewaktu waktu terjadi, bisa dilihat dari deviasi Daily activity report dengan MTD activity report (Data terupdate MTD activity report)</p>	Bulan	Frekuensi Deviasi	Februari	10	Maret	16	April	19	Razis	Problem												
Bulan	Frekuensi Deviasi																										
Februari	10																										
Maret	16																										
April	19																										
2	Transformasi Data	Man	Transformasi data daily dilakukan otomatis	Masih manual dilakukan setiap hari	Terkait pengetahuan pengolahan data yang terbatas untuk tools yang digunakan	Razis	Problem																				
		Machine	Menggunakan aplikasi transformasi data	masih menggunakan excel untuk transformasi	Terkait pengetahuan pengolahan data yang terbatas untuk tools yang digunakan	Razis	NO Problem																				
3	Storage atau tempat penyimpanan setelah proses transformasi	Man	Pengumpulan data dalam satu storage terintegrasi	HO dikirim via email atau WA dan Site di Local storage site	<p>File Explorer View of Storage Content</p> <ul style="list-style-type: none"> Ukama Promosi Sosial PRAKIRANA PRODUKSI PERIODE AGUSTUS 2023 - Berikut kami informasi Produk... 08.01 09/08/2023... <ul style="list-style-type: none"> Moco HRS BRE Fathurrohman Bagan Sisa Rugito Desawars Sisa Subcover Hauling Sisa Subcover Hauling HR Dinasaseti Simarmata WIDYA MANURUNG Tavis Widia sitho Moco Admin HRS EBL 	Razis	NO Problem																				
		Machine	Menggunakan cloud storage	HO dikirim via email atau WA dan Site di Local storage site		Razis	NO Problem																				
		Material	Penyimpanan lebih optimal dengan parquet	penyimpanan besar untuk sekumpulan excel	<p>File Explorer View of Storage Content</p> <ul style="list-style-type: none"> Name Status Date modified Type Size <ul style="list-style-type: none"> 03082023031936 Microsoft Excel Work... 03082023031938 Microsoft Excel Work... 03082023031939 Microsoft Excel Work... 03082023031940 Microsoft Excel Work... 03082023031941 Microsoft Excel Work... 03082023031942 Microsoft Excel Work... 03082023031943 Microsoft Excel Work... 03082023031944 Microsoft Excel Work... 03082023031945 Microsoft Excel Work... 03082023031946 Microsoft Excel Work... 03082023031947 Microsoft Excel Work... 03082023031948 Microsoft Excel Work... 03082023031949 Microsoft Excel Work... 03082023031950 Microsoft Excel Work... 03082023031951 Microsoft Excel Work... 03082023031952 Microsoft Excel Work... 	Razis	Problem																				
		Metode	Penyimpanan terintegrasi secara hybrid local dan cloud	Belum dimaksimalkan untuk versi local masih cloud	Pada tahun 2022, masih research tentang metode baru untuk syncronisasi secara hybrid local - cloud - local, dan masih upload atau input by google sheet	Razis	Problem																				
		Enviro	Penyimpanan data optimal	Penyimpanan data dalam skala besar cukup berat terkadang ada bug		Razis	NO Problem																				

No	Proses	faktor	What Should be Happen	What Actually Happen	Validasi	validator	Kesimpulan
4	Datamart / Report / Dashboard	Man	customer bisa mendapatkan data dan dashboard dinamis sesuai kebutuhan	customer masih melihat secara visual image dan email untuk data			NO Problem
		Machine	Adanya Datamart sebagai sarana kebutuhan data dan tersegmentasi	masih by email dan wa untuk deliveri data dan tidak tersegmentasi		Razis	Problem
		Metode	menggunakan metode datamart untuk membuat staging kebutuhan data	masih belum terbentuk datamart			NO Problem
		Enviro	Dashboard ringan ketika di akses	Dashboard cukup berat karena pengambilan data besar		Razis	Problem

2.2 Menetapkan Objek Masalah dan Analisa Sebab-Akibat

Dari anaconda yang sudah dibuat, saya kemudian menganalisa menggunakan Why Analisa dan digambarkan dengan diagram fishbone.



Berfokus pada kotak yang berwarna merah yang kemudian akan menjadi Akar Penyebab Masalah (APM). Secara garis besar terdapat 4 APM, yaitu :

1. Database belum terintegrasi HO dan Site
2. Belum adanya tools database terintegrasi HO dan Site
3. Belum adanya research penggunaan strorage lebih optimal
4. Belum adanya research tools transformasi otomatis

Dari 4 APM diatas kemudian akan divalidasi sesuai dengan kondisi dilapangan.



2.3 Validasi Akar Penyebab

Berikut adalah validasi dari akar penyebab masalah yang menjadi akar inti dari diagram fishbone. Terdapat 4 akar masalah yang ditemukan yaitu sebagai berikut:

No	Akar Masalah	Validasi						Status																				
		Hasil		Foto																								
1.	Database belum terintegrasi HO dan Site	Menggunakan hasil survei dari Tim HO dan Site untuk mengetahui statement dari karyawan PT HRS	<table border="1"> <thead> <tr> <th>Problem</th> <th>ya</th> <th>tidak</th> <th>Satuan</th> <th>Bobot</th> </tr> </thead> <tbody> <tr> <td>Data terintegrasi</td> <td>0</td> <td>5 kali</td> <td>100%</td> <td></td> </tr> <tr> <td>Perbedaan RAW data yang signifikan</td> <td>4</td> <td>1 kali</td> <td>80%</td> <td></td> </tr> <tr> <td>Data Visibel dan akses mudah</td> <td>0</td> <td>5 kali</td> <td>100%</td> <td></td> </tr> </tbody> </table>	Problem	ya	tidak	Satuan	Bobot	Data terintegrasi	0	5 kali	100%		Perbedaan RAW data yang signifikan	4	1 kali	80%		Data Visibel dan akses mudah	0	5 kali	100%						Diambil
Problem	ya	tidak	Satuan	Bobot																								
Data terintegrasi	0	5 kali	100%																									
Perbedaan RAW data yang signifikan	4	1 kali	80%																									
Data Visibel dan akses mudah	0	5 kali	100%																									
2.	Belum adanya tools database terintegrasi HO dan Site	Masih menggunakan Excel yang terpencar untuk penyimpanan data dan tidak terintegrasi untuk radius diluar site						Diambil																				
3.	Belum adanya research tools Transformasi Otomatis	Tools masih terbatas menggunakan Excel sebagai alat pengolahan data						Diambil																				
4.	Belum adanya research penggunaan Storage lebih optimal	Pada tahun 2022 masih sebatas penggunaan onedrive secara cloud, dan belum adanya research mengenai fitur-fitur yang ada di onedrive sebagai sebuah system yang bisa ditanamkan di local system						Diambil																				



LANGKAH 3 : MENENTUKAN SOLUSI TERBAIK

3.1 Membuat dan Menganalisis Alternatif Solusi Masalah

Dari 4 Akar masalah yang sudah ditemukan pada BAB 2, berikut adalah solusi alternatif yang dapat diterapkan dan dengan mempertimbangkan aspek effisiensi dan efektivitas:

No	Faktor (Akar Penyebab)	Akar Penyebab	No Ide	Alternatif Ide Perbaikan	Analisa					Keputusan
					Kemudahan	Keefektifan	Waktu Development	Biaya	Remark	
1	Machine	Belum adanya research penggunaan Storage lebih optimal	1.1.	Penggunaan Storage secara local terintegrasi dengan cloud server menggunakan onedrive dan optimalisasi file dengan .parquet	Mudah	Sangat efektif	Cepat	Rp. 2.000.000 /Tahun	Penggunaan system hybrid local -cloud-local, dan .parquet untuk memperkecil kapasitas penyimpanan	Diambil
				Pembuatan server di HO dengan pembuatan VPN sebagai alat integrasi strorage dan optimisasi file dengan .parquet	Sulit	Sangat efektif	Lama	>Rp 60.000.000 /Tahun	Membangun data center sendiri dengan ruangan terisoalasi dan dingin	Tidak diambil
2	Man	Belum adanya research tools Transformasi Otomatis	2.1.	Menggunakan Power Query	Mudah	efektif	Cepat		Lambat untuk BIGDATA dan otomatisasi sering bug	Tidak diambil
				Menggunakan Knime	Mudah	Sangat efektif	Cepat		Cepat dengan menggunakan staging proses dan dapat di otomatisasi	Diambil
			2.3.	Menggunakan Talend Data Studio	Sulit	Sangat efektif	Cepat		Cepat dengan menggunakan staging proses dan dapat di otomatisasi	Tidak diambil
3	Machine	Belum adanya tools database terintegrasi HO dan Site	3.1.	Menggunakan SQL database	Mudah	Sangat efektif	Lama	Rp. 24.000.000 / Tahun	Kekurangan : Harga sewa cloud database cukup mahal kelebihan : data sudah dipastikan rapih dengan tabulasi	Tidak diambil
				Menggunakan strorage sebagai database	Mudah	Sangat efektif	Cepat	Rp. 2.000.000 / Tahun	Kekurangan : Data lebih beragam perlu tools ETL kelebihan : Harga sewa cloud Strorage lebih miring	Diambil
4	Metode	Database belum terintegrasi HO dan Site	4.1.	Menggunakan skema Datawarehouse	Mudah	Sangat efektif	Lama		Skema tradisional menggunakan Sql / No Sql sebagai pondasi	Tidak diambil
				Menggunakan skema datalake	Mudah	Sangat efektif	Cepat		Menggunakan skema storage dimana bisa menggunakan berbagai macam type data	Diambil
			4.3.	Menggunakan Skema datalakehouse	Sulit	Sangat efektif	Lama		Gabungan antara Datawarehouse dengan datalake	Tidak diambil

Matriks Kemudahan	
Kriteria	Deskripsi
Mudah	Dapat dikerjakan 1 orang
Sulit	Dapat dikerjakan 1-3 orang
Sangat Sulit	Dikerjakan > 3 orang

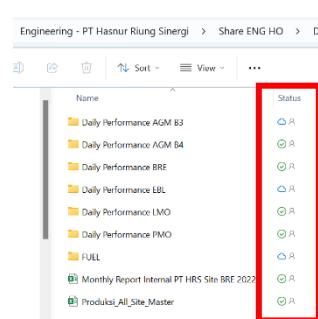
Matriks keefektifan	
kriteria	Deskripsi
Sangat Efektif	Berdampak langsung
Efektif	Dapat berdampak langsung
Tidak Efektif	Tidak berdampak langsung

Matriks Waktu	
Kriteria	Deskripsi
Cepat	Dapat Diselesaikan < 2 Minggu
Cukup Cepat	Dapat Diselesaikan 2-3 Minggu
Lama	Dapat Diselesaikan > 3 Minggu

3.2 Inpirasi dan Dasar Teori Pemilihan solusi

1. **Sistem Hybrid** merupakan penggabungan antara **onedrive offline dan cloud**, dengan menggunakan **synchronisasi otomatis dari onedrive** maka hal ini memungkinkan untuk dilakukan system secara hybrid dan terhubung antara site project dan head office. Pada

gambar disamping berlabel merah merupakan status synchronisasi. Kita dapat melakukan fleksibilitas dalam tarik ulur syncronisasi, ketika status menunjukkan awan berarti file berada di cloud sedangkan jika lingkaran hijau centang menyatakan data berada di local kita. Hal ini dapat membantu kita mengurangi beban penyimpanan di local server kita.



2. Kemudian menggunakan **file .parquet sebagai file kompresi yang dapat menekan cost dari penyimpanan di cloud storage.**

Dataset	Size on Amazon S3	Query Run time	Data Scanned	Cost
Data stored as CSV files	1 TB	236 seconds	1.15 TB	\$5.75
Data stored in Apache Parquet format*	130 GB	6.78 seconds	2.51 GB	\$0.01
Savings / Speedup	87% less with Parquet	34x faster	99% less data scanned	99.7% savings

3. **KNIME merupakan tools Extract Transfrom dan Load data**, dimana proses pembersihan dan penyusunan data dapat di develop satu kali, selanjutnya dilakukan secara otomatis dengan windows scheduler.

Workflow yang di refresh setiap hari

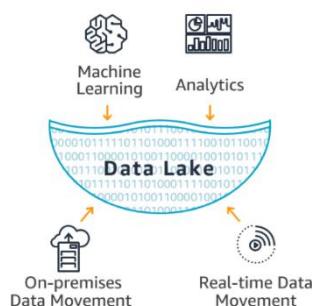
Menggunakan syntax untuk proses otomatis tanpa membuka aplikasi



4. Mempertimbangkan secara cost dan kondisi dilapangan skema SQL merupakan type data structur, namun dilapangan merupakan type data semi struktur, dengan mempertimbangkan hal itu skema strorage menjadi lebih effective untuk saat ini.

5. **Metode yang digunakan adalah data lake**, dimana proses transfromasinya adalah Extract Load dan Transform. Hal ini memungkinkan untuk menyimpan data terlebih dahulu di strorage yang kemudian dilanjutkan kepada transformasi data.

Secara teori, "Data lake adalah repositori terpusat yang memungkinkan Anda menyimpan semua data terstruktur dan tidak terstruktur pada skala apa pun. Anda dapat menyimpan data apa adanya, tanpa harus menyusun data terlebih dahulu, dan menjalankan berbagai jenis analisis—mulai dari dasbor dan visualisasi hingga pemrosesan data besar, analisis real-time, dan pembelajaran mesin untuk memandu pengambilan keputusan yang lebih baik." <https://aws.amazon.com/what-is/data-lake/>



Keuntungan dalam penggunaan datalake:

1. Cepat dalam memasukkan data
2. Mempermudah dalam penerapan machine learning
3. Tempat penyimpanan berbagai macam file dalam satu tempat
4. Format data tidak perlu disamakan
5. Secara cost lebih rendah dibandingkan datawarehouse

6. Dari Point 1-5 yang merupakan produk pakai pihak ke 3 secara logis hal ini akan mudah ketika sudah memiliki kompetensi. Namun yang menjadi tantangan adalah mendalami kegunaan penuh dari masing-masing tools secara native bukan hanya bisa pakai.



Kemudian bagaimana secara arsitektur dan pipeline tools itu dapat saling berkomunikasi. Belajar kepada beberapa pakar data engineer yang terconnect di linkedin.



Hendri Angkasa • 2nd
Data Scientist | Data Engineer | Data Enthusiast
2mo • 5

+ Follow ...

Excited to Share My Latest Project: ETL Pipeline with NYC's Taxi Data! 🚖 🚗

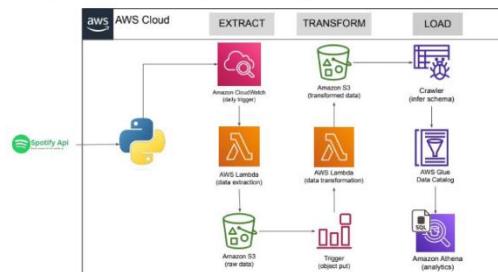
I'm thrilled to unveil my recent project that delves into the fascinating world of NYC taxi data! This pipeline integrates raw data from Google Cloud Storage, processes it using Mage VM Compute Engine, and then loads it into BigQuery for analytics. Finally, the results are visualized in Looker. I'm excited to share this project and learn from others in the field.



Onkar Ausekar • 3rd+
student
4mo • 1

Spotify Data Pipeline: Extract, Transform, and Analyze with AWS.

This project aims to build a comprehensive data pipeline for extracting ...see more

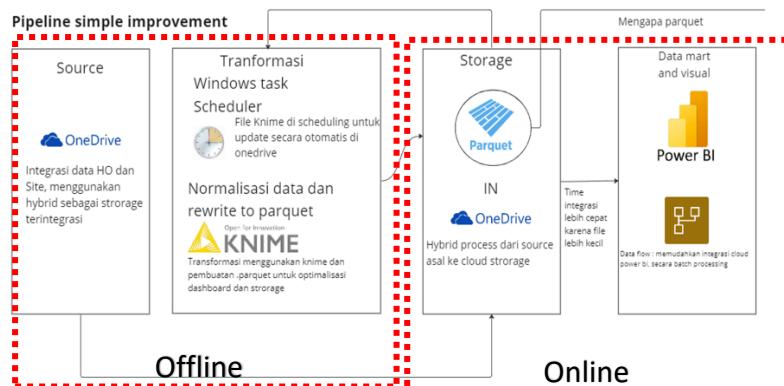


Memotivasi saya untuk dapat membuat pipeline sendiri yang mudah dimengerti oleh teman teman di site project. Uji coba penerapan di implementasikan dalam project SS ini. Langkah tersebut adalah:

1. Mempelajari fitur fitur Onedrive dan penerapannya.
2. Mempelajari dan research KNIME sebagai software ETL, dengan mencari dari komunitas maupun artikel terkait menjalankan proses scheduling otomatisasi.

03 Manipulating Big Data - Solution

3. Melakukan research terkait waktu pengiriman data dan membuat pipeline versi hybrid, agar arsitektur dapat berjalan lebih rapih dan lancar serta sesuai dengan apa yang dibutuhkan teman teman di site project.





LANGKAH 4 : MENYUSUN RENCANA PERBAIKAN

4.1 Menyusun Rencana Perbaikan

Berikut adalah rencana perbaikan yang sudah disesuaikan dengan timeframe secara actual.

No APM	Faktor (Akar)	Akar Penyebab	HOW		WHAT Sasaran	WHO Oleh	WHEN Waktu	WHERE Tempat	HOW MUCH Biaya
			No Ide	Ide Perbaikan					
1	Machine	Belum adanya research penggunaan Storage lebih optimal	1.1.	Penggunaan Storage secara local terintegrasi dengan cloud server menggunakan onedrive dan optimalisasi file dengan .parquet	Penggunaan Onedrive secara lokal, Sinkronisasi otomatis dan Cloud sebagai database untuk dashboard	Razis	Week 3 Juni	HO	Rp. 2.000.000 / Tahun
2	Man	Belum adanya research tools Transformasi Otomatis	2.2.	Menggunakan Knime	Membuat Pipeline proses data Daily Report (Cleansing Otomatis atau Oneclick)	Razis	Week 4 Juni - Week 2 Juli	HO	
3	Machine	Belum adanya tools database terintegrasi HO dan Site	3.2.	Menggunakan storage sebagai database	Onedrive Eng site, terintegrasi dengan sharepoint dengan onedrive Eng HO	Razis	Week 3 - 4 Juli	Online Cloud	Rp. 2.000.000 / Tahun
4	Metode	Database belum terintegrasi HO dan Site	4.2.	Menggunakan skema datalake	Mengubah Alur diagram dari source oleh email, menjadi source oleh sharepoint	Razis	Week 1 Agustus	HO dan Remote Site	

4.2 Analisa Risiko dan Pencegahan

Dari alternatif yang di berikan di point 3.1 kemudian saya membuat Analisa resiko dari solusi terpilih untuk resiko dan Tindakan pencegahan yang perlu dilakukan.

No Ide Alternatif Ide Perbaikan	Potensi Masalah	No	Akar Masalah	Tindakan Pencegahan	Tindakan Penanggulangan	PIC
1.1. terintegrasi dengan cloud server menggunakan onedrive dan optimalisasi file dengan .parquet	Timbul Error Sign in Ketika menanamkan Onedrive di local	1.1.1	OS Windows tidak terinstall secara benar	Memastikan kepada tim IT untuk menerapkan panduan instalasi secara benar	Melakukan Cek Sign in di windows saat instalasi selesai dilakukan	Razis
		1.2.1	OS Windows Bajakan	Memastikan Licensi OS Windows kepada tim IT	Melakukan cek berkala ketika selesai install windows	Razis
2.2. Menggunakan Knime	Kompetensi Karyawan belum terbentuk di ATMP PCSD Engineering	2.2.1	Belum adanya pelatihan yang diadakan	Mengikuti pelatihan course dari Web Community KNIME	Melakukan training yang diadakan Community International	Razis
3.2. Menggunakan storage sebagai database	Potensi Management folder yang tidak teratur	3.2.1	Belum adanya SOP Data Management	Membuat SOP Management data sesuai alur yang dibutuhkan (Dari Source ke Dashboard atau Report)	Perlu pemetaan secara berkala untuk memastikan management data benar	Razis
	Potensi Keamanan data	3.2.2	Pencurian data oleh hacker yang tidak bertanggung jawab	Memastikan service dari pihak microsoft sebagai fasilitas keamanan data	Membeli fasilitas storage dengan keamanan yang sudah teruji	Razis
4.2. Menggunakan skema datalake	Over load kapasitas dengan berbagai macam type data	4.2.1	Tidak adanya tindakan backup	Membuat reminder notifikasi untuk backup	Melakukan backup dengan terjadwal	Razis



4.3 Pengesahan Rencana Perbaikan

Tanggal	8 Juni 2023	
Komentar		
<p>1. Segera lakukan perbaikan, bisa dimulai d^eFBL, untuk kemudian terapkan di seluruh site.</p> <p>2. Optimalkan kompetensi diri yang dimiliki untuk mendapatkan cara dan tampilan.</p> <p>3. Selamat - Sukses - dan Semangat !</p>		
Diajukan Oleh :	Diketahui Oleh :	Disetujui Oleh :
Razis Aji Saputro	Fuad Hasan	Wahyu Irawan
	Fasilitator	Manager Engineering



LANGKAH 5 : PELAKSANAAN PERBAIKAN

5.1 Identifikasi Sumber Daya yang Dibutuhkan

Sumber Daya	Deskripsi
Kompetensi	<ol style="list-style-type: none"> Pengetahuan penggunaan tools onedrive dan Microsoft 365 E5 Challenge: mencari referensi melalui vendor Microsoft, internet, dan youtube. Pengetahuan menggunakan tools KNIME Challenge: Selain menggunakan tools sebagai transformasi yang sudah banyak tutorial, tantangan terbesar adalah membuat tools ini berjalan dengan oneclick atau scheduling secara otomatis tanpa membuka aplikasi Memiliki kemampuan Analisa arsitektur system Challenge: Menganalisa system yang ada dan mempertimbangkan perbaikan dengan proses otomasi, pengetahuan koding, membuat pipeline dan database juga diperlukan.
Alat	Onedrive, KNIME analytic tools, Windows Scheduler, Command Promt
Material	data .csv, data .parquet, data .xlsx

5.2 Melaksanakan Perbaikan yang Telah direncanakan

Implementasi 1 :

No	Sebelum Perbaikan	Kode Solusi	Solusi Perbaikan	Proses Perbaikan	Kendala Proses Perbaikan	Sesudah Perbaikan	Dampak																				
1	Belum adanya research penggunaan Storage lebih optimal	1.1.	Penggunaan Storage secara local terintegrasi dengan cloud server menggunakan onedrive dan optimalisasi file dengan .parquet	1 Research fitur onedrive yang dapat digunakan secara hybrid 2 perbandingan antara type file .csv dan .parquet 3 Sign in onedrive secara local yang di sematkan di local disk D: 	tidak ada		terdapat list pengetahuan baru dari penggunaan onedrive																				
Research																											
1	<table border="1"> <tr> <td>No</td> <td>Fitur</td> <td>Deskripsi</td> </tr> <tr> <td>1.</td> <td>Fitur on Demand</td> <td>Untuk managemant penyimpanan dari local dan cloud</td> </tr> <tr> <td>2.</td> <td>Document Scanning</td> <td>Pemindaihan document</td> </tr> <tr> <td>3.</td> <td>Personal vault</td> <td>Penyimpanan dengan security berganda</td> </tr> <tr> <td>4.</td> <td>Fitur sharing</td> <td>sharing document dengan keamanan berganda</td> </tr> <tr> <td>5.</td> <td>Hybrid storage</td> <td>Penyimpanan dengan synchronasi local dan cloud secara otomatis</td> </tr> <tr> <td>6.</td> <td>Integrasi terhadap microsoft partner</td> <td>memudahkan integrasi ke aplikasi microsoft yang digunakan</td> </tr> </table>	No	Fitur	Deskripsi	1.	Fitur on Demand	Untuk managemant penyimpanan dari local dan cloud	2.	Document Scanning	Pemindaihan document	3.	Personal vault	Penyimpanan dengan security berganda	4.	Fitur sharing	sharing document dengan keamanan berganda	5.	Hybrid storage	Penyimpanan dengan synchronasi local dan cloud secara otomatis	6.	Integrasi terhadap microsoft partner	memudahkan integrasi ke aplikasi microsoft yang digunakan			Penggunaan fitur yang dimaksimalkan dalam scope SS ini adalah: 1. Fitur on Demand 2. Fitur Sharing 3. Hybrid Storage 4. Integrasi terhadap microsoft partner (Power BI dan Excel)		
No	Fitur	Deskripsi																									
1.	Fitur on Demand	Untuk managemant penyimpanan dari local dan cloud																									
2.	Document Scanning	Pemindaihan document																									
3.	Personal vault	Penyimpanan dengan security berganda																									
4.	Fitur sharing	sharing document dengan keamanan berganda																									
5.	Hybrid storage	Penyimpanan dengan synchronasi local dan cloud secara otomatis																									
6.	Integrasi terhadap microsoft partner	memudahkan integrasi ke aplikasi microsoft yang digunakan																									
2	<table border="1"> <thead> <tr> <th></th> <th>Size on Amazon S3</th> <th>Query Run time</th> <th>Data Scanned</th> <th>Cost</th> </tr> </thead> <tbody> <tr> <td>Data stored as CSV files</td> <td>1 TB</td> <td>236 seconds</td> <td>1.15 TB</td> <td>\$5.75</td> </tr> <tr> <td>Data stored in Apache Parquet format*</td> <td>130 GB</td> <td>6.78 seconds</td> <td>2.51 GB</td> <td>\$0.01</td> </tr> <tr> <td>Savings / Speedup</td> <td>87% less with Parquet</td> <td>34x faster</td> <td>99% less data scanned</td> <td>99.7% savings</td> </tr> </tbody> </table>		Size on Amazon S3	Query Run time	Data Scanned	Cost	Data stored as CSV files	1 TB	236 seconds	1.15 TB	\$5.75	Data stored in Apache Parquet format*	130 GB	6.78 seconds	2.51 GB	\$0.01	Savings / Speedup	87% less with Parquet	34x faster	99% less data scanned	99.7% savings			Gambar disamping merupakan perbandingan antara .csv dan .Parquet sebagai file database. Terlihat bahwa Size volume .parquet lebih kecil daripada .csv dengan isi data yang sama, sehingga dapat menurunkan penggunaan cost penyimpanan jangka panjang untuk data structure.			
	Size on Amazon S3	Query Run time	Data Scanned	Cost																							
Data stored as CSV files	1 TB	236 seconds	1.15 TB	\$5.75																							
Data stored in Apache Parquet format*	130 GB	6.78 seconds	2.51 GB	\$0.01																							
Savings / Speedup	87% less with Parquet	34x faster	99% less data scanned	99.7% savings																							
3					Embedded storage onedrive di windows dimana file bisa tersyncronisasi otomotis.																						



Implementasi 2 :

	Sebelum Perbaikan	Kode Solusi Perbaikan	Solusi Perbaikan	Proses Perbaikan	Kendala Proses Perbaikan	Sesudah Perbaikan	Dampak
2	Belum adanya research tools Transformasi Otomatis	2.2.	Menggunakan Knime	1 Research fitur KNIME yang dapat digunakan secara hybrid 2 Instalasi KNIME 3 Pembuatan workspace bersama di Onedrive 4 Pembuatan workflow dan text otomatis untuk trasnfromasi otomatis	tidak ada tidak ada tidak ada tidak ada		terdapat list pengetahuan baru dari penggunaan KNIME analytic tool Memudahkan karyawan untuk terintegrasi dari workflow yang sudah berjalan memungkinkan untuk running pekerjaan dari Command Prompt dan dijalankan secara otomatis tanpa membuka tool
Research							
1	No Fitur Fitur	Deskripsi					Penggunaan fitur yang dimaksimalkan dalam scope SS ini adalah:
	1 tools trasnformasi data	Melakukan kegiatan daily pembersihan data, perubahan format					1. Fitur Transfromasi data
	2 tools visualisasi data	Dapat melakukan visualisasi data interaktif					2. Tools otomatisasi
	3 tools modeling prediktif dari data historis	membuat model prediktif, untuk melakukan analisa masa depan					
	4 tools otomatisasi	mengumpulkan workflow yang sering menjadi tugas harian, dengan diselesaikan secara otomatis					
2	Download di situs knime						Melakukan instalasi seperti pada umumnya dengan requirement windows dan atau linux, Ram 4 GB minimal dan strorage tersedia minimal 1 GB
3	Workflow yang di refresh setiap hari						Knime workspace berada di onedrive, untuk memudahkan sharing workflow antara tim
4	Workflow yang di refresh setiap hari			Menggunakan syntax untuk proses otomatis tanpa membuka aplikasi			1. Melakukan pembuatan workflow yang terkoneksi ke master data site, dengan disusun kebutuhan analisa tabel yang diinginkan 2. Melakukan pembuatan syntax otomatis dan dijalankan hanya melalui command prompt setiap harinya



Implementasi 3 :

No	Sebelum Perbaikan	Kode Solusi	Solusi Perbaikan	Proses Perbaikan	Kendala Proses Perbaikan	Sesudah Perbaikan	Dampak
3	Belum adanya tools database terintegrasi HO dan Site	3.2.	Menggunakan storage sebagai database	1 Penggunaan Onedrive sharing sebagai tools integrasi antara HO dan Site	Windows yang tidak terinstall dengan benar, menyebabkan tombol sign in tidak bekerja	Dilakukan install ulang untuk windows yang tidak terinstall dengan benar	Sign in berhasil dan aplikasi dapat dijalankan
Implementasi							
Remarks							
1	No	Fitur	Deskripsi		Penggunaan fitur yang dimaksimalkan dalam scope SS ini adalah:		
	1.	Fitur on Demand	Untuk managemant penyimpanan dari local dan cloud		1. Fitur on Demand		
	2.	Document Scanning	Pemindai document		2. Fitur Sharing		
	3.	Personal vault	Penyimpanan dengan security berganda		3. Hybrid Storage		
	4.	Fitur sharing	sharing document dengan keamanan berganda		4. Integrasi terhadap microsoft partner (Power BI dan Excel)		
	5.	Hybrid storage	Penyimpanan dengan syncronisasi local dan cloud secara otomatis memudahkan integrasi ke aplikasi microsoft partner				
	6.	Integrasi terhadap microsoft partner	membuatkan integrasi ke aplikasi microsoft yang digunakan				
2					<p>File sharing antara HO dan Site yang berisikan data master dari site, sebagai database utama di site dan HO (Satu Data)</p>		

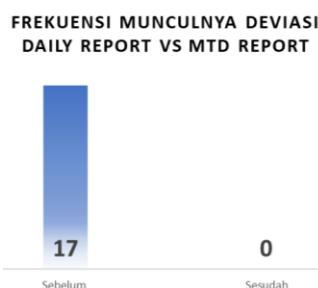
Implementasi 4 :

No	Sebelum Perbaikan	Kode Solusi	Solusi Perbaikan	Proses Perbaikan	Kendala Proses Perbaikan	Sesudah Perbaikan	Dampak
4	Database belum terintegrasi HO dan Site	4.2.	Menggunakan skema datalake	1 Penggunaan Onedrive sharing sebagai tools integrasi antara HO dan Site	Windows yang tidak terinstall dengan benar, menyebabkan tombol sign in tidak bekerja	Dilakukan install ulang untuk windows yang tidak terinstall dengan benar	Sign in berhasil dan aplikasi dapat dijalankan
Implementasi							
Remarks							
1	No	Fitur	Deskripsi		Penggunaan fitur yang dimaksimalkan dalam scope SS ini adalah:		
	1.	Fitur on Demand	Untuk managemant penyimpanan dari local dan cloud		1. Fitur on Demand		
	2.	Document Scanning	Pemindai document		2. Fitur Sharing		
	3.	Personal vault	Penyimpanan dengan security berganda		3. Hybrid Storage		
	4.	Fitur sharing	sharing document dengan keamanan berganda		4. Integrasi terhadap microsoft partner (Power BI dan Excel)		
	5.	Hybrid storage	Penyimpanan dengan syncronisasi local dan cloud secara otomatis memudahkan integrasi ke aplikasi microsoft yang digunakan				
2					<p>File sharing antara HO dan Site yang berisikan data master dari site, sebagai database utama di site dan HO (Satu Data)</p>		

LANGKAH 6 : EVALUASI HASIL

6.1 Analisa Hasil Improvemet

Dari analisa latar belakang pada langkah 1, dari diagram pareto menunjukkan frekuensi WH sebelum perbaikan adalah 17 kali. Kemudian setelah melalui Implementasi dari solusi didapatkan hasil sebagai berikut:

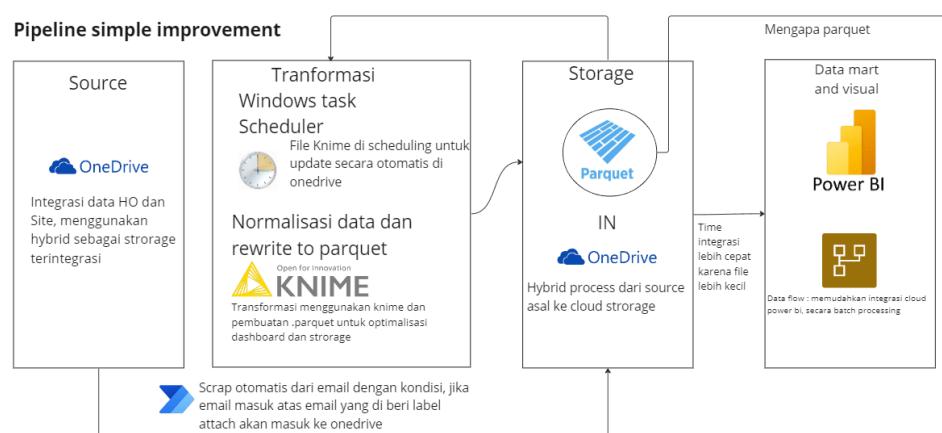


Bulan	Frekuensi terjadinya Deviasi Daily Report dengan MTD Report				Total
	Produksi	HM	WH	STB	
Juni	-	-	-	-	-
July	-	-	-	-	-
Agustus	-	-	-	-	-
Total	-	-	-	-	-

Hal ini disebabkan karena saya melakukan **rekayasa engineering** untuk mengganti source yang ada dari pengiriman email, menjadi storage terintegrasi. Hal ini berdampak kepada peminimalan terjadinya beda versi update dari data HO dan Site, dimana Ketika Site mengupdate data di local onedrive akan otomatis tersyncronisasi dengan data di cloud yang kemudian tersambung ke HO. Adapun flow proses yang berubah akan di jelaskan di point 6.2.

6.2 Perubahan Hasil Improvement

Hasil improvement merekayasa flow proses dari pengolahan data yang ada di HO. Menghapus pengiriman email dengan excel menjadikan onedrive sebagai database storage yang terintegrasi dengan skema Data Lake. Berikut adalah flow proses yang dikembangkan dengan menggunakan informasi diatas:



Perubahan dari Skema tercantum dalam table berikut:

No	Proses	Sebelum	Sesudah	Ach Implementasi	Remark
1	Source	Menggunakan data Email daily sebagai souce data yang discrap otomatis menggunakan power automate	Menggunakan system hybrid local - cloud - local sebagai bentuk mengintegrasikan data antara HO dan Site	100%	
2	Transformasi	Menggunakan proses manual dengan power query untuk mengolah dan menggabungkan data daily	menggunakan KNIME dan melakukan otomatisasi dengan file .bat kemudian di scheduling dengan windows scheduler	90%	Scheduling belum digunakan, namun otomatisasi dengan command prompt sudah dilakukan
3	Storage	Menggunakan file excel dan onedrive cloud sebagai database untuk penyimpanan	Menggunakan .parquet sebagai penyimpanan data historical dan onedrive dioptimalkan dengan local ondrive yang tersyncronisasi dengan cloud onedrive	100%	
4	Data Mart dan Report	Menggunakan excel di onedrive cloud sebagai report dashboard	Menambahkan dataflows sebagai jembatan penghubung onedrive cloud untuk mempercepat proses GET data	100%	

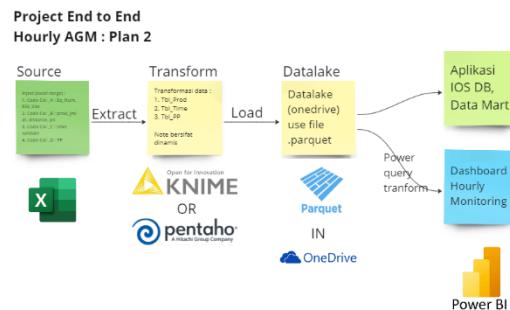
6.3 Keunggulan

Keunggulan dari metode ini dapat menjadi dasar project digitalisasi secara menyeluruh. Contoh penerapan yang sudah diterapkan dengan menganalisa secara flow proses pipeline:

1. Di site AGM Engineering replikasi project ini didasarkan dengan skema yang sama, seperti pada hasil 6.2. Hal yang dilakukan adalah:
 - Mengedukasi Tim Moco dengan membuat grup konsultasi dan melakukan sharing ilmu.



- Melakukan Pembuatan Konsep Bersama dari sisi backend



Dari hasil ini TIM Engineering Site AGM membentuk tim Jayacandra dalam bidang QCC memperoleh pencapaian

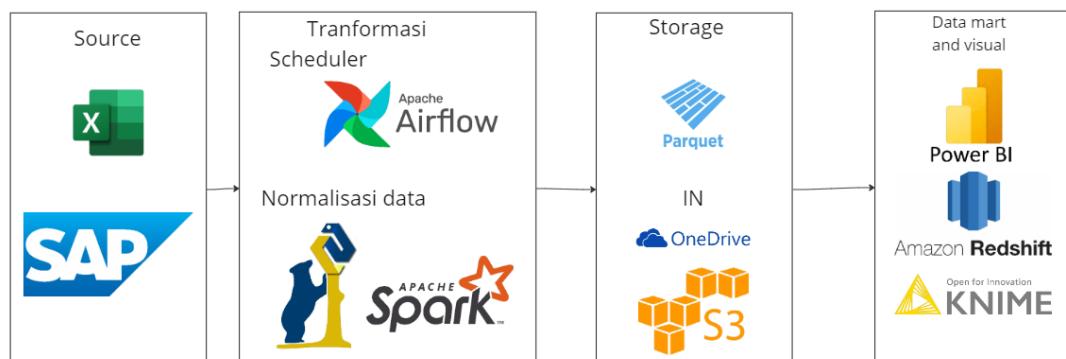
1. Juara 3 QCC Site AGM
2. Juara Harapan 2 QCC HRS Korporat
3. Melakukan bimbingan dan sharing ide terkait project improvement FGDP atas andre Rusli, membangun Report Daily EBL terkait management database yang terintegrasi Head Office dan Site EBL dengan skema yang sama datalake. Keterangan gambar





3. Lebih jauh lagi dimasa depan, akan diterapkan dan disesuaikan (Software pihak ketiga) dengan berkolaborasi dengan **HRS Satu Data**. Dengan beberapa project digitalisasi dan pembentukan tim untuk HRS Satu Data.

Pipeline simple implementasian



Lead of Digital Transformation Strategic

No	Departemen	Nama
1	Operation (OPR)	Fatih Adinanda
2	Engineering (ENG)	Razis Aji Saputro
3	Plant (PLT)	Syaiful Hidayat
4	Safety Health & Environment (SHE)	Hanifah
7	Procurement & Logistics (PLC)	Made Hariyana
8	Finance & Tax (FAT)	Deny Stio Nugroho
9	Human Resources & General Services (HGS)	Lukman Hakim
10	Management Development (MD)	Firgiawan Iksanja
11	Training Development Center (TDC) & Lembaga Sertifikasi Profesi (LSP)	Doni Saputra

Role & Responsibility

- Bertanggung jawab atas pengembangan dan pelaksanaan strategi transformasi digital perusahaan;
- Memimpin tim Digital Transformation Strategic yang terdiri dari perwakilan departemen;
- Mengkoordinasikan antara tim departemen dan tim proyek utama;
- Membantu dalam identifikasi peluang transformasi digital di berbagai bagian perusahaan;
- Memastikan bahwa strategi transformasi digital sejalan dengan visi dan tujuan perusahaan; dan
- Melakukan pemantauan perkembangan strategi dan mengadaptasinya sesuai kebutuhan.



6.4 Analisa QCDSM

Aspek	Sebelum Perbaikan	Sasaran	Sesudah Perbaikan
Quality	Data belum terintegrasi HO dan Site secara realtime	Data terintegrasi HO dan Site secara realtime	Data terintegrasi menggunakan Sharing folder onedrive, dimana site menggunakan onedrive local sebagai workspace penyimpanan data
Cost	Penyimpanan file berlangganan sebagai data historis menggunakan Excel dan csv file di 51 juta / 5 tahun	Menurunkan cost tahunan menjadi 0,13x menggunakan .parquet sebagai data historis	Data historis disimpan dalam .parquet secara berkala dan data excel dihapus sebagai pengurangan biaya berlangganan dengan estimasi 6,6 juta / 5 Tahun
Delivery	Proses delivery lama karena data belum terintegrasi HO dan Site	Proses delivery lebih cepat karena data sudah terintegrasi HO dan Site	Proses delivery akan terintegrasi dari Raw Master sehingga data akan cepat terupdate
Safety	Security data dalam keamanan fisik dalam kondisi terjadi bencana kebakaran atau banjir belum ada.	Terjaminnya security data ketika terjadi bencana kebakaran atau banjir dengan menggunakan sistem hybrid local dan cloud.	Microsoft cloud memiliki data center di berbagai negara, sehingga data akan terbackup Ketika terjadi bencana di Indonesia
Morale	Pengelolaan data secara berulang menyebabkan rasa malas dan bosan berakibat data tidak terupdate daily	Update data akan dilakukan dengan one klik secara otomatis sehingga menghilangkan rasa malas dan bosan	Mengintegrasikan data dengan tools Analisa KNIME atau power BI menjadi lebih mudah dan bisa berfokus ke Analisa bisnis

6.5 Voice of Customer

1. Manager Engineering



"Integrasi data antara site dan HO dapat meminimalisir perbedaan data, sehingga analisa dan rekomendasi bisa dilakukan dengan lebih cepat dan akurat"

2. Section Head PCSD Engineering



"Integrasi bisa dilakukan dengan 2 cara datawarehouse IOSDB dan datalake dengan Storage onedrive, coba bisa dimaksimalkan."

3. MOCO GL Engineering

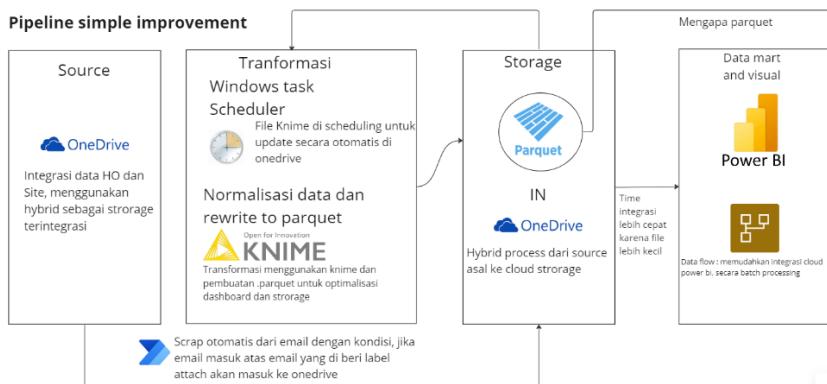


"Data management sangat perlu diimplementasikan di site, Sebagai salah satu PI yang garap dengan mas razis di EBL integrasi data di EBL sekarang sudah mulai terbentuk ."

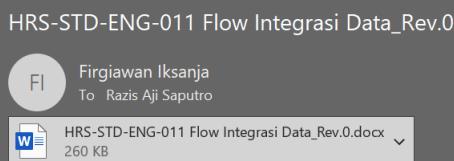
LANGKAH 7 : STANDARISASI

7.1 Mengusulkan Standard Baru

Dari Hasil improvement saya mengusulkan proses kerja (SOP) yang dapat dilakukan di Head Office PT Hasnur Riung Sinergi sebagai standar baru dalam integrasi dan pengolahan data, serta flow proses terbentuknya dashboard.



7.2 Pengesahan Standard Baru



Dear Mas Razis
Mohon maaf, karena kendala jaringan. Jadi dikirimkan melalui email.

Hormat kami,
Firgiawan

	STANDAR PARAMETER	Nomor	HRS-STD-ENG-008
		Revisi	0
		Tanggal	12/09/2023
		Halaman	1 dari 4

LEMBAR PENGESAHAN
Menunjuk Standar Parameter No. HRS-STD-ENG-008 tentang **Flow Integrasi Data**, bersama dengan ini diberitahukan kepada seluruh PIC HO dan Site terkait PT Hasnur Riung Sinergi, untuk dapat mempelajari dan melaksanakan prosedur tersebut.

Disiapkan Oleh	Nama/Jabatan	Tanda Tangan
	Wahyu Irawan Engineering Dept. Head	
Oleh	Syukra Alimanda SHE Dept Head	

	STANDAR PARAMETER	Nomor	HRS-STD-ENG-008
		Revisi	0
		Tanggal	12/09/2023
		Halaman	4 dari 4

	STANDAR PARAMETER	Nomor	HRS-STD-ENG-008
		Revisi	0
		Tanggal	12/09/2023
		Halaman	3 dari 4

1 TUJUAN

Prosedur ini bertujuan untuk :

1. Memastikan kegiatan dalam flow terkiniinya data dari hulu ke hilir.
2. Memberikan panduan dalam diagram terkiniinya data dari hulu ke hilir.

2 RUANG LINGKUP

Prosedur ini berlaku untuk penyusunan skema proses pengolahan data dari hulu ke hilir.

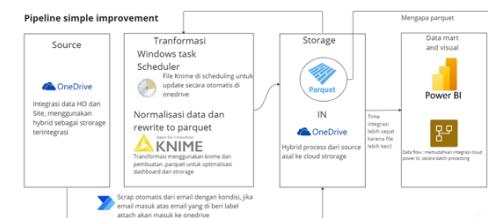
3 REFERENSI

- <https://aws.amazon.com/big-data/datalakes-and-analytics/what-is-a-data-lake/>
- <https://www.databricks.com/discover/data-lakes>
- <https://cloud.google.com/learn/what-is-a-data-lake>
- <https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-a-data-lake/>

4 DEFINISI

1. **Data Lake** adalah metode yang digunakan dalam bigdata untuk memproses data berskala besar, dengan menggunakan type data terstruktur, data tidak terstruktur dan semi struktur.
2. **Source** adalah sumber data.
3. **Hybrid** adalah penggabungan 2 metode online dan offline.
4. **Scrap** Pengambilan data dari berbagai source.

5 FLOW PROSES



6 PENJELASAN FLOW PROSES

Dari flow proses no 5, dapat diberikan penjelasan sebagai berikut:

1. **Source:** Sebagai data center yang diintegrasikan ke dalam suatu storage (metode data lake) dimana file yang akan diolah di letakkan dalam suatu storage dengan dilengkapi dengan management data.
2. **Transformasi:** sebagai prosedur dalam pengolahan data dan akan ditetapkan dalam kajian teknis perproject.
3. **Storage:** Penyimpanan data historis dan terupdate, sebagai wadah dalam proses management data.

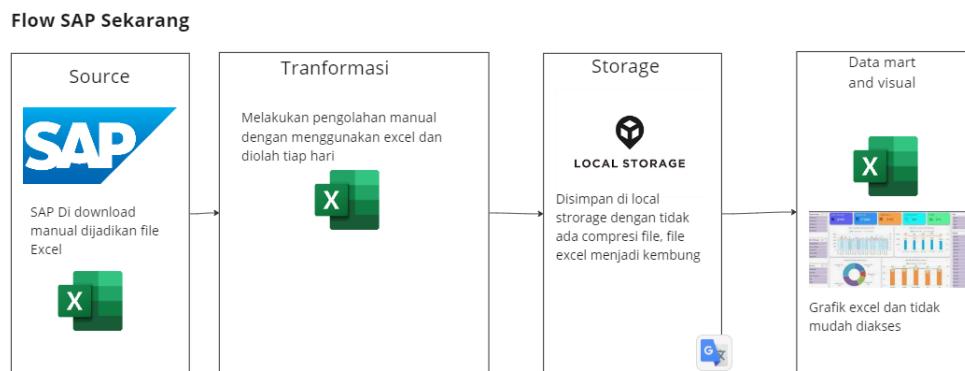


Tanggal	: 24 Agustus 2023	
Komentar		
<p>Selamat sudah dapat melakukan perbaikan. Segara ajukan standar yang dibuat agar dapat diterapkan ke semua site. Bravo!</p>		
Diajukan Oleh:	Diketahui Oleh:	Disetujui Oleh:
Razis Aji Saputro	Fuad Hasan	Wahyu Irawan
	Fasilitator	Manager Engineering

LANGKAH 8 : Menentukan Tema Perbaikan Selanjutnya

8.1 Identifikasi masalah berikutnya

Merujuk kepada proses dalam pengambilan data dari pihak ke 3, dimana SAP adalah aplikasi pihak ke 3 yang menjadi source data utama yang diakui, namun prosesnya cukup lama dan data tidak mudah diakses. Perlunya ada pengambilan atau integrasi data untuk memudahkan pihak HRS mudah dalam melakukan Analisa. Dilihat dari flow pipeline kegiatan menggunakan SAP:



Hal ini perlu di improvement karena terkait dengan effisiensi dan effektivitas kegiatan reporting ke management. Secara garis lurus, department engineering memerlukan ini untuk monitoring terkait Fuel Ratio dan Fuel konsumsi secara Daily dan Month to Date.

8.2 Membuat Time Frame Improvement

Langkah			Plan /Actual	Tahun : 2024															
				Mei				Juni				July				Agustus			
P	1	Menentukan Tema	Plan	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
	2	Mengidentifikasi Faktor Penyebab Masalah	Plan			20													
	3	Menentukan Solusi	Plan				21												
	4	Merencanakan Perbaikan	Plan					22	23	24	25								
	5	Melaksanakan Perbaikan	Plan						26	27	28	29	30	31	32	33			
D	6	Evaluasi Hasil Perbaikan	Plan							31	32	33							
A	7	Menetapkan Standarisasi	Plan																
	8	Menentukan Tema Berikutnya	Plan																



Financial Statement for InnovAction

Statement biaya langganan by AWS Cloud

Deskripsi	Value	Satuan
Harga Langganan	0,100	USD per GB
Kurs	15.000	IDR

Secara produksi data selama 1 bulan dapat diestimasikan sebagai berikut

Perbandingan	kapasitas/month/site/Dept	Satuan
Penggunaan csv	20	GB
Penggunaan parquet	2,6	GB

Hitungan berlangganan 5 Tahun

CSV atau Excel	1	2	3	4	5	6	7	8	9	10	11	12
Tahun 1	1	2	3	4	5	6	7	8	9	10	11	12
Kapasitas	20	40	60	80	100	120	140	160	180	200	220	240
Cost	30.000	60.000	90.000	120.000	150.000	180.000	210.000	240.000	270.000	300.000	330.000	360.000
Tahun 2	1	2	3	4	5	6	7	8	9	10	11	12
Kapasitas	240	260	280	300	320	340	360	380	400	420	440	460
Cost	360.000	390.000	420.000	450.000	480.000	510.000	540.000	570.000	600.000	630.000	660.000	690.000
Tahun 3	1	2	3	4	5	6	7	8	9	10	11	12
Kapasitas	460	480	500	520	540	560	580	600	620	640	660	680
Cost	690.000	720.000	750.000	780.000	810.000	840.000	870.000	900.000	930.000	960.000	990.000	1.020.000
Tahun 4	1	2	3	4	5	6	7	8	9	10	11	12
Kapasitas	680	700	720	740	760	780	800	820	840	860	880	900
Cost	1.020.000	1.050.000	1.080.000	1.110.000	1.140.000	1.170.000	1.200.000	1.230.000	1.260.000	1.290.000	1.320.000	1.350.000
Tahun 5	1	2	3	4	5	6	7	8	9	10	11	12
Kapasitas	900	920	940	960	980	1000	1020	1040	1060	1080	1100	1120
Cost	1.350.000	1.380.000	1.410.000	1.440.000	1.470.000	1.500.000	1.530.000	1.560.000	1.590.000	1.620.000	1.650.000	1.680.000

Total jumlah kapasitas dalam 5 Tahun

Perbandingan biaya langganan	kapasitas/5 Tahun /site/Depart	Satuan
Penggunaan csv	1200	GB
Penggunaan parquet	156	GB

Total biaya berlangganan dalam 5 tahun

Asumsi Kontrak owner 5 Tahun			
Perbandingan biaya langganan	kapasitas/5 Tahun /site/Depart	Cost/5 tahun/ 5 Site	Cost/5 tahun/5 Site/ 7 Dep
Penggunaan csv	51.300.000	256.500.000	1.795.500.000
Penggunaan parquet	6.669.000	33.345.000	233.415.000



	FINANCIAL STATEMENT FOR INNOVATION	

A. Initial Investment (5 Tahun/Dep/Site)

Manhours Tim	Rp
Consumable Goods / Material	Rp0
Equipment & Tools	Rp10.000.000
Trial	Rp0
Project Support : Consultancy Fee, Reference, etc	Rp15.000.000
Others	Rp
Total Initial Investment (IINV)	Rp25.000.000

B. Tangible Benefit (5 Tahun/Dep/Site)**1 Additional/Increase Revenue**

Peningkatan kapasitas produksi	Rp0
Peningkatan pendapatan terkait dengan peningkatan produk/layanan	Rp
Peningkatan pendapatan karena kenaikan harga	Rp
Others	Rp
Total Additional/Increase Revenue (AIR)	Rp0

2 Cost Reduction

Penghematan biaya tenaga kerja (Manpower Efficiency)	Rp0
Penghematan atas penggunaan material (Consumable Goods / Material Efficiency)	Rp44.631.000
Penurunan biaya depreciasi asset/sewa plant, gedung terkait dengan peningkatan utilisasi akibat perbaikan yang dilakukan	Rp0
Penurunan rework (scrap)	Rp0
Others	Rp0
Total Cost Reduction	Rp44.631.000

Total Tangible Benefit

Additional/Increase Revenue	Rp0
Cost Reduction	Rp44.631.000
Total Tangible Benefit	Rp44.631.000

C. Cost of Project Standardization (5 Tahun/Dep/Site)

Manhours Tim	Rp0
Consumable Goods / Material	Rp0
Equipment rental & Tools (non Asset)	Rp 10.000.000
Depreciation/Amortization (asumsi 5 tahun)	Rp0
Project Awareness : Pelatihan, Sosialisasi, dll	Rp 15.000.000
Others	Rp0
Total Cost of Project Standardization (CPS)	Rp 25.000.000

D. Gross Profit (5 Tahun/Dep/Site)

Total Tangible Benefit	Rp44.631.000
Cost of Project Standardization	Rp25.000.000
Gross Profit	Rp19.631.000