

BAB III

PELAKSANAAN KERJA MAGANG

3.1 Kedudukan dan Koordinasi

Pelaksanaan program kerja magang untuk mahasiswa kali ini, mahasiswa mendapatkan pekerjaan dengan posisi sebagai *Data Quality Analyst Intern*, dan bergabung dalam tim IT Data. Tim ini terdiri dari 5 orang sebagai *data quality analyst*, 15 orang sebagai *data engineer*, dan 9 orang sebagai *data analyst*. Sebagai seorang *Data Quality Analyst*, mahasiswa harus memastikan dataset yang didapatkan memenuhi beberapa kriteria. Beberapa kriteria tersebut seperti *accuracy, completeness, validity, consistency, uniqueness, timeliness*, dan *fitness* [2]. Kriteria tersebut harus terpenuhi untuk memastikan data tersebut dapat digunakan untuk memenuhi tujuan bisnis. Sebagai *Data Quality Analyst* mahasiswa juga harus memastikan dataset yang dimiliki sesuai dengan yang diinginkan user, ketika terdapat suatu update pada suatu database.

Dalam pelaksanaan program kerja magang ini mahasiswa memiliki supervisi. Mahasiswa di supervisi oleh Arif H Solomon yang memiliki posisi sebagai *Deputy Chief of Data Platform*. Selama magang pak Arif mengajarkan banyak hal baik dari sisi *softskill* maupun *hardskill*. Pak Arif selalu berusaha untuk menuntun mahasiswa untuk mengatasi masalah yang tidak dimengerti, dan juga selalu memberi kritik, saran, serta motivasi untuk kedepannya.

Dalam koordinasi dengan tim, peran *data quality analyst* yaitu untuk menerima permintaan yang diberikan oleh user. Permintaan tersebut contohnya seperti permintaan untuk dibuatkan suatu *dashboard*. Setelah mendapatkan permintaan, maka seorang *data quality analyst* akan menyiapkan data yang dibutuhkan, tabel-tabel yang perlu dibentuk, dan menyiapkan *query* untuk tabel yang akan dibentuk. Data dan *query* yang disediakan akan diberikan kepada *data engineer* untuk pembuatan *job* pembuatan tabel dan pengisian data secara *daily*. Setelah *job* dan tabel sudah terbentuk dengan datanya, *data quality analyst* akan

mengecek tabel tersebut untuk memastikan data yang dibutuhkan sudah sesuai dengan permintaan, dan sumber datanya. Jika tabel sudah sesuai maka *data analyst* dapat membuat *dashboard* dengan data tersebut.

3.2 Tugas yang Dilakukan

Di Dalam bagian tim data terdapat beberapa tugas yang mahasiswa kerjakan. Tugas tersebut diberikan untuk jangka waktu 5 hari (diluar Sabtu dan Minggu). Setiap tugas akan dimulai pada hari Selasa hingga hari Senin. Tabel 3.1 merupakan timeline kerja magang yang dilakukan mahasiswa sebagai *Data Quality Analyst*.

Tabel 3.1 Ringkasan Pekerjaan

Minggu	Pekerjaan Yang Dilakukan	Tanggal Mulai	Tanggal Selesai
1	Orientasi karyawan baru, mempelajari dokumentasi <i>database</i> perusahaan dan hal seputar <i>data warehouse</i> , dan melakukan update query dan validasi pada suatu tabel	8 Januari 2024	15 Januari 2024
2	Mempelajari query pembentuk dari sebuah tabel, dan melakukan validasi data pada sebuah tabel	16 Januari 2024	22 Januari 2024
3	Melakukan validasi data pada suatu pada suatu tabel yang mendapatkan pembaruan <i>logic query</i>	23 Januari 2024	29 Januari 2024
4	Membuat query untuk memvalidasi tabel yang mendapatkan update kolom baru, memperbaiki sebuah query untuk mengambil suatu value dari table	30 Januari 2024	5 Februari 2024
5	Membuat query untuk mengecek tabel yang mendapatkan pembaruan kolom	6 Februari 2024	12 Februari 2024
6	Membuat query untuk memvalidasi pembaruan kolom pada suatu tabel, validasi update tabel di datamart,	13 Februari 2024	19 Februari 2024

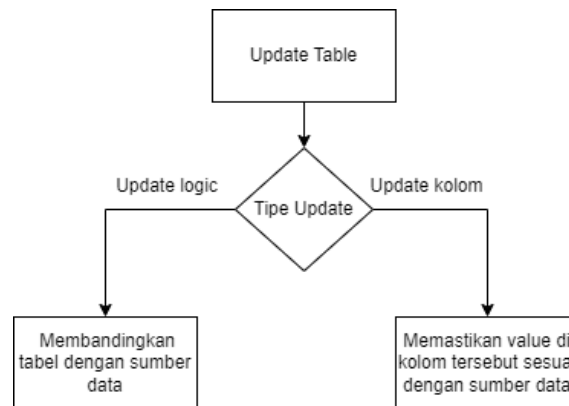
	mempelajari great expectation		
7	Mempelajari config dan fungsi-fungsi great expectation, dan membuat config great expectation untuk beberapa tabel yang sudah di list	20 Februari 2024	26 Februari 2024
8	Membuat daftar config great expectation apa saja yang dibutuhkan untuk tabel-tabel yang terdapat di <i>data warehouse</i>	27 Februari 2024	4 Maret 2024
9	Membuat config great expectation untuk tabel-tabel yang sudah didaftarkan, membuat tabel <i>fact</i> untuk keperluan dashboard	5 Maret 2024	8 Maret 2024

3.3 Penjelasan Tugas

Selama 3 bulan sebagai *data quality analyst*, tugas-tugas yang umumnya dikerjakan yaitu melakukan pengecekan data, membuat data yang sesuai untuk kebutuhan user, dan membantu pembuatan sebuah projek. Berikut penjelasan tugas yang mahasiswa lakukan.

3.3.1 Pengecekan Data

Dalam pengecekan data terdapat beberapa cara untuk melakukan pengecekan. Setiap cara pengecekan tergantung dengan apa yang diperlukan. Misal jika terdapat sebuah tabel atau pergantian logika *query* pada suatu tabel, cara mengeceknya dapat dengan membandingkan data yang ada di tabel tersebut dengan data yang diambil dari sumber. Jika terdapat penambahan tabel pada suatu kolom, maka harus memastikan kolom tersebut memiliki data yang sesuai dengan sumbernya.



Gambar 3.1 Alur Pengecekan Data

Dalam pengecekan data dapat dilakukan dengan membuat query yang membandingkan sumber data dan tabel. Variabel yang sering digunakan untuk membandingkan antara sumber data dan tabel yang baru diupdate yaitu jumlah baris dan value dari suatu tabel. Jika kedua variabel tersebut sudah sama, maka dapat dipastikan tabel tersebut sudah sesuai. Berikut contoh query yang digunakan untuk membandingkan data.

```

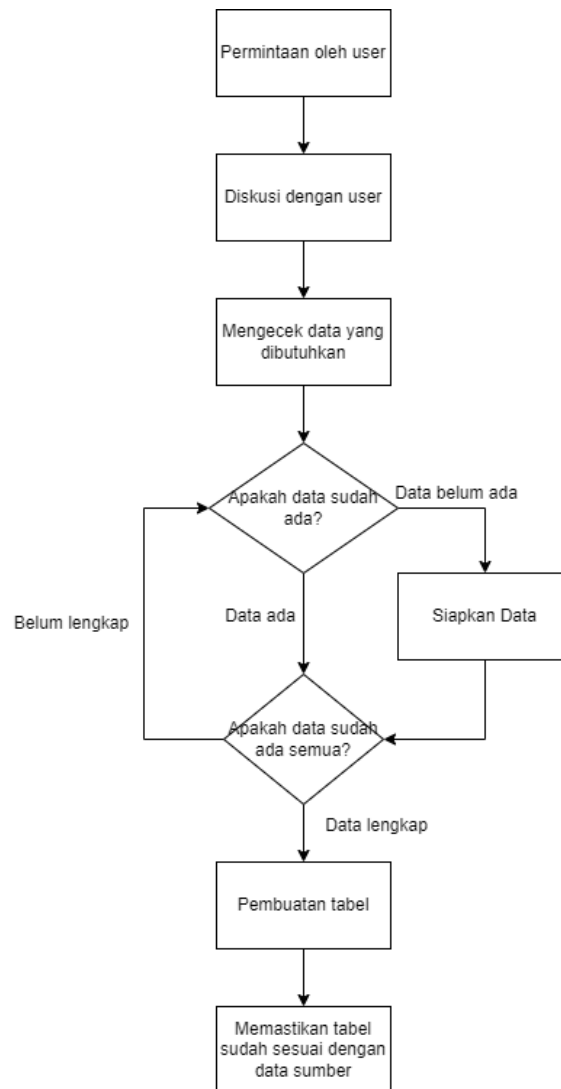
with sumber(
select sum(value) as value, count(*) as jumlah_row
from sumber1
join sumber 2 on sumber1.id = sumber2.id
)
, tabel(
select sum(value) as value, count(*) as jumlah_row
from tabel
)

select 1.value, 2.value, 1.jumlah_row, 2.jumlah_row
from sumber as 1
join tabel on 1.value = 2.value and 1.jumlah_row = 2.jumlah_row
  
```

Jika query tersebut berhasil mengeluarkan output maka dipastikan jumlah value dan row antara tabel dan sumber sudah sesuai.

3.3.2 Pembuatan Data

Dalam proses pembuatan data yang diinginkan oleh user, tahap pertama yang dilakukan yaitu harus mendiskusikan apa yang dibutuhkan oleh user terlebih dahulu. Setelah mendapatkan apa yang dibutuhkan, *data quality analyst* memastikan terlebih dahulu apakah data-data tersebut sudah ada atau belum ada. Jika belum ada maka perlu disiapkan terlebih dahulu data yang masih kurang, jika ada maka dapat dilanjutkan ke tahap berikutnya. Setelah data sudah siap maka dapat dilakukan pembuatan tabel sesuai yang sesuai dengan yang diminta berdasarkan kebutuhan user. Setelah tabel dibuat *data quality analyst* akan mengecek tabel tersebut untuk memastikan data yang ada di tabel tersebut sudah sesuai dengan data yang ada di sumber.



Gambar 3.2 Alur Pembuatan Data

Contoh terdapat sebuah kasus user memberi permintaan sebuah dashboard yang menampilkan rating yang dimiliki suatu product. Pada dashboard tersebut terdapat rating dari suatu product, dan dapat di filter berdasarkan tanggal, area manager, store manager, store, receipt, dan product. Setelah permintaan tersebut dapat dirancang bahwa data yang dibutuhkan untuk dashboard tersebut yaitu rating dari product, dan data-data sesuai filter agar filternya berfungsi.

```

SELECT
COALESCE(CAST(ort.transaction_date AS DATE), CAST(osi.ori_created_at AS DATE), CAST(ort.ori_created_at as DATE)) AS period,
COALESCE(ortreceipt_id, osireceipt_id, osi.sap_so_number) AS receipt_id,
COALESCE(osi.store_code, ort.store_code) AS store_id,
orp.sku AS product_id,
orp.rating
FROM
[redacted] orp
LEFT JOIN
[redacted] osi ON TRIM(osi.invoice_no) = TRIM(orp.invoice_no)
LEFT JOIN
[redacted] ortd ON ortd.customer_receipt_transaction_offline_detail_id = orp.customer_receipt_transaction_offline_detail_id
LEFT JOIN
[redacted] ort ON ort.customer_receipt_transaction_offline_id = ortd.customer_receipt_transaction_offline_id
WHERE
orp.company_code = 'AHI'
UNION DISTINCT
SELECT
DATE(rd.transdate) AS period,
COALESCE(TRIM(UPPER(orporeceipt_id)), TRIM(UPPER(rd.receive_no))) AS receipt_id,
CASE
WHEN orpo.s_store_code = '0' THEN rd.site_code
ELSE orpo.s_store_code
END AS store_id,
orp.sku AS product_id,
orp.rating
FROM
[redacted] orpo
LEFT JOIN
(
select distinct transdate, receive_no, mc.site_code from `gcp-kl-dev-dwh.klg_d1.ahi_receive` r
left join
mc on mc.company_code = r.company_code
union all
select distinct transdate, receive_no, mc.site_code from `gcp-kl-dev-dwh.klg_d1.ahi_umreceive` ur
left join
mc on mc.company_code = ur.company_code
) AS rd ON TRIM(UPPER(orporeceipt_id)) = TRIM(UPPER(rd.receive_no))
WHERE
orpo.company = 'AHI'

```

Gambar 3.3 Contoh Query Pembentuk

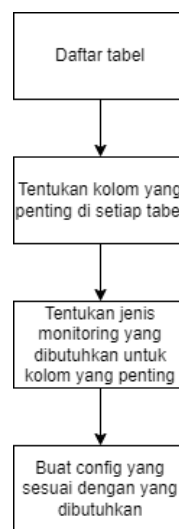
Row	period	receipt_id	store_id	product_id	rating
1	2022-11-09	33.2.20221109.3			5
2	2024-02-22	20.2.20240222.2			5
3	2024-03-09	H0.1.20240309.12			5
4	2022-11-09	33.2.20221109.3			5
5	2022-10-23	12.78.20221023.5			5

Gambar 3.4 Output Dari Query

Data pada gambar 3.4 sudah memenuhi apa yang dibutuhkan oleh user, karena data tersebut sesuai dengan dashboard yang digunakan. Data area manager dan store manager tidak ada pada tabel tersebut karena, data tersebut sudah terhubung dengan store_id dan dapat dijoin ke tabel dim_store yang berisikan informasi area manager dan store manager.

3.3.3 Great Expectation

Terdapat proyek yang sedang dikembangkan juga saat ini, yaitu great expectation. Proyek great expectation ini berguna untuk membantu pengawasan data. Dalam proyek ini *data quality analyst* berperan untuk menentukan tabel-tabel apa yang dibutuhkan untuk diawasi, menentukan kolom-kolom apa saja yang butuh diawasi dari tabel-tabel yang sudah ditentukan, dan membuat config yang sesuai untuk kebutuhan pengawasan data.



Gambar 3.5 Alur Pembuatan Config

Contoh kasus, dari suatu tabel pasti terdapat kolom yang merupakan primary key. Kolom primary key tersebut pasti perlu dijaga agar datanya selalu ada sehingga saya membuat config untuk kolom tersebut. Kemudian terdapat juga beberapa tabel yang datanya saling berhubungan sehingga perlu dijaga agar data di kedua tabel tersebut selalu ada, dan akan memberikan pengumuman error jika terdapat data yang tidak ada.

3.4 Kendala yang Dihadapi

Selama proses kerja magang ini, terdapat kendala beberapa kendala yang dihadapi, yaitu:

1. Luasnya *database* yang dimiliki oleh perusahaan ini membuat mahasiswa perlu waktu untuk beradaptasi dan memahami setiap fungsi tabel dan kolom yang ada pada *database*.
2. Terdapat beberapa tools yang baru mahasiswa gunakan seperti bigquery, SAP, dan DBeaver. Hal tersebut membuat mahasiswa perlu mempelajarinya terlebih dahulu.
3. Banyaknya projek yang sudah berjalan disaat mahasiswa memulai magang, sehingga mahasiswa perlu memahami setiap tugas yang diberikan.

3.5 Solusi atas Kendala

Beberapa masalah yang terdapat selama program kerja magang ini dapat diatasi dengan beberapa solusi, yaitu:

1. Pada kendala luasnya *database* yang dimiliki perusahaan, mahasiswa dapat mempelajari dokumentasi yang diberikan mentor kepada mahasiswa. Mahasiswa juga dapat melihat secara langsung struktur dari tabel tersebut beserta isinya. Mahasiswa juga dapat menanyakan kepada mentor jika terdapat *database* yang kurang dipahami.
2. Pada kendala tools yang baru digunakan mahasiswa, solusi yang dapat dilakukan yaitu dengan langsung mencoba menggunakan tools-tools tersebut.
3. Pada kendala projek yang sudah berjalan, mahasiswa dapat menanyakan kepada mentor yang tentang masalah apa yang terdapat didalam tugas tersebut, dan apa tujuan dari tugas tersebut.