



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

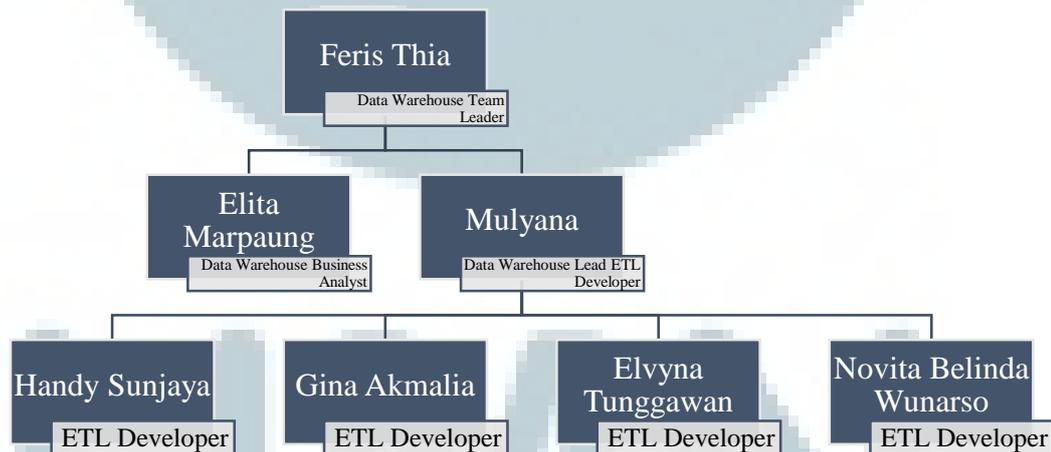
BAB III

PELAKSANAAN KERJA MAGANG

3.1. Kedudukan dan Koordinasi

3.1.1. Divisi

Dalam praktik kerja magang ini, penulis berada pada divisi *Data Warehouse*, yang dipimpin oleh seorang *team leader*, yaitu Pak Feris Thia. Penulis bekerja sebagai *ETL developer* bersama dengan anggota tim lain, dengan dipimpin oleh Pak Mulyana dan Handy Sunjaya sebagai koordinator tim.



Gambar 4 Struktur Divisi *Data Warehouse*

3.1.2. Kedudukan dan Koordinasi

Pada divisi *ETL Developer*, seluruh anggota tim memiliki ruang lingkup kerja yang berbeda. Penulis bertanggung jawab dalam penyusunan *script extract, transform, load* (ETL) untuk dokumen *rate* penjualan dan pendapatan PT X, *plan* induk perusahaan PT X, serta *proposal budget* masing-masing distrik cabang PT X. Sebelum menyusun *script ETL*, penulis berkoordinasi dengan *business analyst* dalam memahami *functional specification document* PT X sekaligus memastikan data mana saja yang dibutuhkan pada *data warehouse*. Secara umum, *ETL developer* bertugas untuk menyusun *script ETL* sesuai dengan kebutuhan PT X yang tertera pada *functional specification document* serta melakukan *testing*, validasi data, *quality checking*, dan *performance tuning* terhadap *script* yang dibuat.

Selama penyusunan *script ETL*, penulis banyak berkoordinasi dengan *lead ETL developer* dan *team leader* terkait teknik-teknik penyusunan *script*. Ruang lingkup pekerjaan penulis bersinggungan dengan ruang lingkup pekerjaan anggota tim lain, di antaranya yaitu *plan* acuan PT X per enam bulan yang dikerjakan Gina Akmalia serta data pengeluaran yang berada di ruang lingkup Novita Belinda Wunarso. Seluruh data *plan* akan disatukan pada *data warehouse* dan seluruh data pengeluaran serta pendapatan akan dihitung menggunakan *rate* mata uang yang berlaku saat itu. Oleh karena itu, susunan *script ETL* penulis juga digunakan oleh anggota tim lain.

Berikut rincian ruang lingkup pekerjaan penulis.

1. *Rate Cost*

Dokumen *rate cost* berisi tabel yang menyimpan angka *rate* dasar sebagai pengali terhadap ketercapaian aktivitas produksi yang dilakukan oleh *partner* PT X. Oleh karena perbedaan aktivitas produksi dan *partner* yang dimiliki setiap distrik, dokumen *rate cost* diunggah oleh masing-masing distrik.

2. *Rate Revenue Basic*

Lain halnya dengan *rate cost*, dokumen *rate revenue basic* berisi angka *rate* dasar sebagai pengali terhadap volume produksi yang dilakukan oleh setiap distrik.

3. *Rate Revenue Adjustment*

Dokumen ini menyimpan nilai *rate* yang digunakan dalam proses penyesuaian (*adjustment*) terhadap data yang sudah diakumulasi sebelumnya.

4. *Plan Induk Perusahaan*

PT X memiliki sejumlah induk perusahaan, yang masing-masing memberikan rencana sebagai acuan terhadap hasil produksi yang harus dicapai pada tahun berikutnya.

5. *Proposal Budget per Distrik*

Dokumen ini berisi rencana anggaran setiap distrik yang dijadikan sebagai acuan terhadap kinerja produksi pada tahun berikutnya.

3.2. Tugas-Tugas yang Dikerjakan

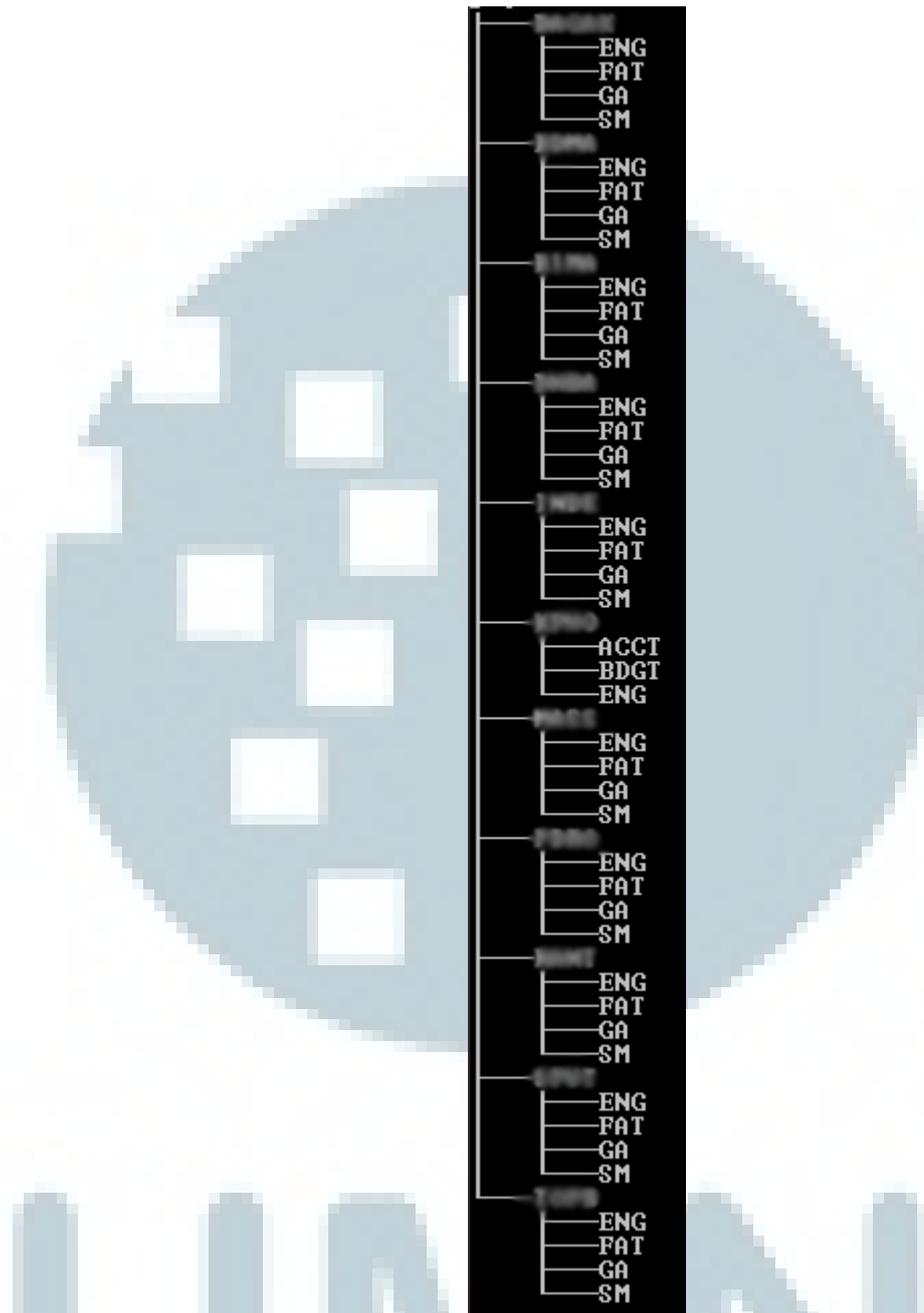
Berikut ini tugas-tugas yang dikerjakan penulis selama bekerja sebagai *ETL developer* di PT Putera Handal Indotama.

3.2.1. Penyusunan *Script ETL*

Script ETL dibutuhkan dalam mengintegrasikan data yang dimiliki oleh PT X. Data milik PT X berasal dari beberapa sumber, yaitu *database Ellipse*, *Ewacs*, dan dokumen *Microsoft Excel*. Namun demikian, ruang lingkup pekerjaan penulis hanya berkaitan dengan dokumen *Microsoft Excel*, tidak mencakup sumber lain.

Seluruh dokumen *Microsoft Excel* disimpan pada *server* PT X dengan susunan direktori seperti berikut.

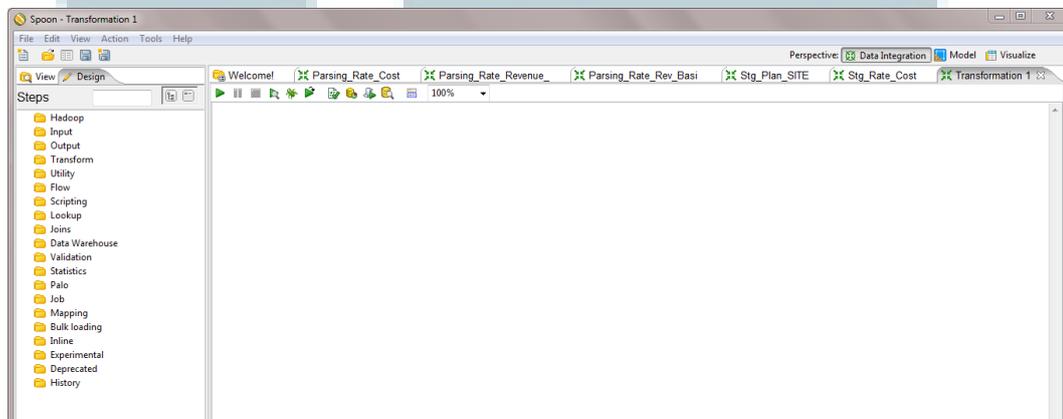
UMMN



Gambar 5 Susunan Direktori *File* pada *Server*

PT X memiliki satu kantor pusat yang memiliki divisi *Accounting*, *Budgeting*, dan *Engineering* serta sejumlah distrik. Masing-masing distrik mengunggah dokumen *Microsoft Excel* ke *server* pada tanggal dan periode tertentu.

Penulis menyusun *script ETL* dengan memanfaatkan *Pentaho Data Integration*. Pada *Pentaho Data Integration*, terdapat beberapa program yang digunakan, yaitu *Spoon*, *Pan*, *Kitchen*, dan *Carte*. *Spoon* berfungsi untuk merancang serta menjalankan *transformation* dan *job* secara visual menggunakan *Graphical User Interface (GUI)*. *Spoon* hanya digunakan pada tahap *development* dan *testing*. *Pan* berupa *script* dengan ekstensi *.bat* atau *.sh* yang digunakan untuk mengeksekusi *transformation*, sedangkan *Kitchen* digunakan untuk mengeksekusi *job*. *Carte* merupakan utilitas *cluster* yang digunakan untuk mengeksekusi *transformation* ataupun *job*.



Gambar 6 Tampilan Antarmuka *Spoon*

Transformation terdiri dari sejumlah *hop*, yaitu langkah-langkah pengerjaan dalam integrasi data, seperti *file input*, *string replacement*, *table output*, dan sebagainya. *Transformation* dapat memiliki beberapa cabang dan menghasilkan beberapa *output*. *Job* terdiri dari rangkaian *transformation*, yang akan dijalankan sesuai dengan urutan yang ditentukan.

Pada praktik kerja magang ini, penulis membuat sejumlah *transformation* dan *job*, serta memanfaatkan *Spoon* dan *Kitchen*. Berikut tahap-tahap dalam penyusunan *script ETL*.

1. Landing Zone

Tahap ini merupakan tahap awal dalam ekstraksi data, disebut juga sebagai *parsing*. Data pada dokumen *Microsoft Excel* yang sudah memenuhi *template* yang ditentukan diekstraksi dan dipindahkan ke dalam tabel *MySQL*. Pada tahap ini, tidak ada perubahan bentuk data, seluruh *field* pada dokumen *Microsoft Excel* murni hanya dipindahkan. Hal ini bertujuan untuk memudahkan *tracing* apabila ada kesalahan dalam integrasi data serta memastikan bahwa data awal PT X tidak diubah.

2. Staging

Pada tahap ini, data yang ada pada tabel *MySQL* diolah kembali dan diubah strukturnya sesuai dengan yang diinginkan pada tabel pada *data warehouse*. Pengubahan data yang dilakukan yaitu *string replacement*, denormalisasi tabel, kalkulasi data, dan sebagainya.

3. Mapping ke Data Warehouse

Tabel hasil *staging* dipetakan ke struktur tabel pada *data warehouse*. Pada tahap ini, verifikasi tipe data dilakukan agar tidak mengubah struktur tabel yang sudah disediakan dan tidak mempengaruhi penghitungan pada *dashboard* yang akan dibuat.

3.2.2. *Testing Script*

Script yang dibuat harus diuji terlebih dahulu sebelum dilanjutkan ke tahap berikutnya, yaitu *staging* dan *mapping* ke *data warehouse*. Pengujian dilakukan pada setiap *hop* dalam *transformation*.

Apabila *transformation* sudah menghasilkan *output* yang sesuai, *file parsing*, *staging*, dan *mapping* ke *data warehouse* disusun pada sebuah *file job*. Penulis kemudian memastikan kembali bahwa *output* yang dihasilkan *job* sudah sesuai. Setelah membuat *job*, penulis membuat *scheduler* berupa *shell script* yang dijalankan melalui *Kitchen*. Pembuatan *job* serta *shell script* dipandu oleh Pak Mulyana selaku *lead ETL developer*.

3.2.3. *Validasi Data dan Quality Checking*

Data yang diberikan selama tahap *development* hanya berupa data *dummy* karena data asli hanya dapat diakses pada lokasi PT X. Pada mulanya, *testing* dan validasi data dapat dilakukan secara *remote* melalui Bapak Mulyana, tetapi memasuki bulan kedua, terdapat gangguan teknis sehingga koneksi *remote* tidak dapat diandalkan. Oleh karena penulis masih dalam masa perkuliahan, validasi data dan *quality checking* terhadap data asli dilakukan oleh Bapak Mulyana, sedangkan secara rutin validasi data dan *quality checking* tetap dilakukan penulis dengan memanfaatkan data *dummy* yang dibuat. Hasil pengecekan pada data asli kemudian diberitahukan kembali kepada penulis agar dilakukan revisi terhadap *script ETL* yang dibuat.

3.2.4. Performance Tuning

Berdasarkan *feedback* yang diberikan, penulis melakukan *review* pada *transformation* dan *job* yang dibuat. Penulis memastikan bahwa tidak ada *hop* yang *redundant* maupun tidak berguna, tipe data yang didefinisikan sejak awal sudah sesuai dengan target tabel pada *data warehouse*, dan memeriksa waktu yang dibutuhkan untuk mengeksekusi *hop*. Waktu eksekusi sebisa mungkin seminimal mungkin, karena jumlah data yang digunakan pada saat *testing* terbilang sedikit jika dibandingkan dengan data asli, yang tentunya akan menghabiskan waktu eksekusi lebih besar.

3.2.5. Dokumentasi

Salah satu pekerjaan penulis dalam praktik kerja magang ini yaitu menuangkan ilmu yang diperoleh ke dalam artikel, yang akan dipublikasikan oleh PT Putera Handal Indotama dan *Indonesian Business Intelligence Community on Campus* (IBICC). Penulisan artikel dilakukan bersama dengan salah satu anggota tim, dan dalam kesempatan ini penulis bekerja sama dengan Handy Sunjaya.

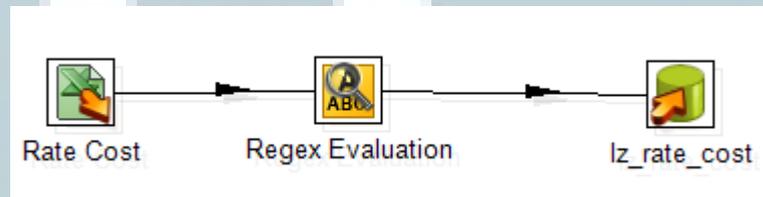
Artikel yang ditulis berkaitan dengan langkah *step by step* dalam penyelesaian masalah yang biasa ditemui saat mengolah data sehari-hari. Artikel ditulis dengan bahasa yang sederhana dengan tujuan menumbuhkan kesadaran akan pentingnya integrasi dan manajemen data serta menumbuhkan *mindset* bahwa pekerjaan ini bukanlah hal yang sangat sulit.

3.3. Hasil Kerja

Dalam praktik kerja magang ini, penulis menghasilkan sejumlah *transformation* untuk *parsing*, *staging*, dan *mapping* ke *data warehouse*; *job* untuk masing-masing ruang lingkup pekerjaan penulis, serta *shell script* sebagai *scheduler* dalam menjalankan *job*.

Berikut hasil pekerjaan yang dilakukan penulis.

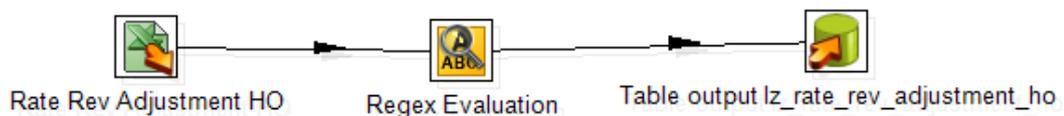
3.3.1. Parsing



Gambar 7 Parsing Rate Cost



Gambar 8 Parsing Rate Revenue Basic



Gambar 9 Parsing Rate Revenue Adjustment

Pada proses *parsing*, dokumen *Microsoft Excel* diekstraksi isinya. Seluruh proses *parsing* umumnya memiliki pola *hop* yang sama seperti contoh pada Gambar

7, di mana terdapat *hop Microsoft Excel input* dan *table output* (lihat Gambar 8 dan 9). Pada *transformation* tersebut, *hop regex evaluation* digunakan untuk mengambil nama distrik, tanggal, dan bulan yang tertera dari nama *file Microsoft Excel*. Penamaan *file Microsoft Excel* dan pola *header field*-nya sudah ditentukan terlebih dahulu, misal *Rate_Cost_YYYYMM.xlsx*. Contoh *template file Microsoft Excel* dapat dilihat pada Tabel 2, 3 dan 4.

Tabel 2 Template File Rate Cost

TANGGAL EFEKTIF (YYYY-MM-DD)	KODE_COA	KODE_SUBCONT	EQUIPMENT	EQUIPMENT_TYPE	RATE1	RATE2	CURRENCY	LIMIT_MH	LIMIT_PA	JARAK	TANGGAL AKHIR KONTRAK (YYYY-MM-DD)
2015-01-01		2301		CPSP11					300	85%	
2015-01-01		2303		WTSSP006							
2015-01-02		2303		WT005							
2015-01-03		2302		WPKSB 19							

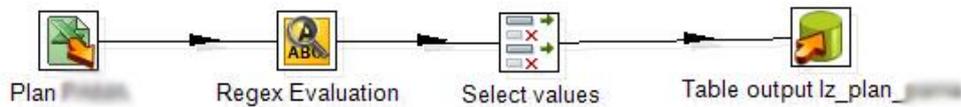
Tabel 3 Template File Plan Induk Perusahaan

COA Kode_Aktivitas	Cost Type	Description	Tahun		JAN	FEB	MAR	APR	MEI	JUNI	JULI	AUG	SEP	OKT	NOV	DES
			000.IDR	%	000.IDR	000.IDR	000.IDR	000.IDR	000.IDR	000.IDR	000.IDR	000.IDR	000.IDR	000.IDR	000.IDR	000.IDR
1000		REVENUE	677,602,500		51,036,713	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
2100	2007 F	1. Employee compensation	49,035,158		4,027,195	4,027,195	4,027,195	4,027,195	4,027,195	4,027,195	4,027,195	4,027,195	4,027,195	4,027,195	4,027,195	4,736,007
2101	2101	1. Salary	26,885,584		2,233,799	2,233,799	2,233,799	2,233,799	2,233,799	2,233,799	2,233,799	2,233,799	2,233,799	2,233,799	2,233,799	2,233,799
2102	2102	2. Overtime	10,360,000		915,000	915,000	915,000	915,000	915,000	915,000	915,000	915,000	915,000	915,000	915,000	915,000
2150	2150	3. Allowance	11,249,574		878,397	878,397	878,397	878,397	878,397	878,397	878,397	878,397	878,397	878,397	878,397	1,587,209
2200	2200 F	2. Depreciation	41,153,669		3,213,565	3,300,086	3,457,211	3,462,905	3,464,988	3,464,988	3,464,988	3,464,988	3,464,988	3,464,988	3,464,988	3,464,988
2201	2201	1. Production Equipment	20,028,691		1,669,058	1,669,058	1,669,058	1,669,058	1,669,058	1,669,058	1,669,058	1,669,058	1,669,058	1,669,058	1,669,058	1,669,058
2202	2202	2. Non Production Equipme	21,124,978		1,544,507	1,631,028	1,788,153	1,793,847	1,795,930	1,795,930	1,795,930	1,795,930	1,795,930	1,795,930	1,795,930	1,795,930
2230	2230 V	3. Consumable Goods	151,133,375		11,477,527	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	11,555,261
2231	2231 Y	1. Diesel Fuel	125,032,821		9,309,278	9,647,016	*****	*****	*****	*****	*****	*****	*****	*****	*****	9,822,471
2232	2232	2. Tyre	5,996,223		649,078	576,019	795,195	588,319	489,732	340,369	469,732	571,222	492,232	369,313	328,616	274,297
2233	2233	3. GET	1,800,000		150,000	150,000	150,000	150,000	150,000	150,000	150,000	150,000	150,000	150,000	150,000	150,000
2234	2234	4. Explosive & Accessories	14,145,301		1,051,438	1,089,735	1,091,846	1,503,555	1,567,661	1,226,740	1,104,184	1,219,852	1,147,761	1,194,424	1,052,704	976,001
2235	2235	5. Drilling Accessories	1,114,238		84,432	87,052	95,812	97,690	99,492	96,564	93,361	87,439	95,728	91,649	87,353	87,158
2236	2236	6. Material	3,084,691		233,201	242,485	267,433	269,454	275,745	267,827	260,328	270,549	260,369	250,905	242,033	244,363
2250	2250 V	4. Repair and Maintenance	75,028,263		5,454,071	5,426,303	6,041,346	7,173,305	7,164,651	5,770,494	7,250,639	7,464,003	6,698,545	6,107,283	5,539,509	5,028,122
2251	2251	1. Workshop Station	1,175,625		97,969	97,969	97,969	97,969	97,969	97,969	97,969	97,969	97,969	97,969	97,969	97,969
2252	2252	2. Spare Parts	66,015,852		4,733,456	4,680,469	5,281,472	6,327,584	6,445,869	5,096,594	6,477,654	6,710,140	5,772,643	5,393,571	4,914,221	4,325,360
2253	2253	3. Direct Cost (Cons Good	989,976		67,664	62,365	69,427	67,537	68,839	64,398	63,694	72,497	72,201	66,182	73,215	65,561
2254	2254	4. Oil & Lubricants	6,855,286		540,614	590,927	578,110	656,847	537,696	537,155	590,763	568,029	646,264	536,193	539,735	533,844
2255	2255	5. Additional Accessories	172,425		14,369	14,369	14,369	14,369	14,369	14,369	14,369	14,369	14,369	14,369	14,369	14,369
2256	2256	6. Workshop Parts	0		0	0	0	0	0	0	0	0	0	0	0	0
2280	2280 V	5. Mobilization	4,339,756		332,370	338,708	370,048	369,955	389,021	377,364	366,510	377,334	367,329	361,515	346,577	347,885
2281	2281	1. Production Equipment	180,000		15,000	15,000	15,000	15,000	15,000	15,000	15,000	15,000	15,000	15,000	15,000	15,000
2282	2282	2. Non Production Equipme	60,000		5,000	5,000	5,000	5,000	5,000	5,000	5,000	5,000	5,000	5,000	5,000	5,000
2283	2283	3. Spare Parts	155,141		12,928	12,928	12,928	12,928	12,928	12,928	12,928	12,928	12,928	12,928	12,928	12,928
2284	2284	4. Oil	0		0	0	0	0	0	0	0	0	0	0	0	0
2285	2285	5. Fuel, Tyre, Explosive	3,944,615		299,842	305,780	337,120	337,026	350,093	345,036	333,582	344,465	334,400	328,687	313,649	314,937
2286	2286	6. Including Charge	0		0	0	0	0	0	0	0	0	0	0	0	0
2287	2287	7. Others	0		0	0	0	0	0	0	0	0	0	0	0	0
2300	2300 V	6. Sub-Contractor	257,293,269		19,420,090	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
2301	2301	1. Production	55,168,269		4,139,545	4,289,681	4,725,393	4,853,856	4,930,765	4,799,676	4,726,980	4,846,246	4,732,850	4,519,900	4,300,550	4,300,627

Tabel 4 Template File Proposal Budget

HIGHLIGHT											Kurs		Fuel Price			
COA	TYPE	KODE_AKTIVI	DESCRIPTION	YTD	Jan	Feb	Mar	Apr	May	Jun	Juli	Agst	Sept	Okt	Nov	Des
			OB (BCM)	2,310,006	282,659	487,343	282,659	487,343	282,659	487,343						
			COAL MINING (TON)	515,664	77,356	102,666	77,356	102,666	77,356	77,665						
			COAL CRUSHING (TON)	1,010,326	168,488	168,488	168,488	168,488	168,488	168,488						
			COAL HAULING	325,798	154,299	154,299	154,299	154,299	154,299	154,299						
			SHIPMENT (MT)	325,798	154,299	154,299	154,299	154,299	154,299	154,299						
			Konveer BCM -> Ton	3,742,622	623,770	623,770	623,770	623,770	623,770	623,770						
			SR	26.70	3.65	4.74	3.65	4.74	3.65	6.26						
1101			OB Removal	103,544,823,588	*****	*****	*****	*****	*****	*****						
1101			Coal getting	3,876,212,541	*****	*****	*****	*****	*****	*****						
1101			Coal Hauling Pit to PDM	10,941,250,152	*****	*****	*****	*****	*****	*****						
1101			Flood Maintenance	3,201,338,054	533,556,342	533,556,342	533,556,342	533,556,342	533,556,342	533,556,342						
1101			Crushing, Rehandling, Stockpiling	12,671,963,130	*****	*****	*****	*****	*****	*****						
1101			Coal Hauling To Port	68,883,837,048	*****	*****	*****	*****	*****	*****						
1101			OHDA													
1101			Chemical	1,171,131,696	195,188,616	195,188,616	195,188,616	195,188,616	195,188,616	195,188,616						
1101			FBM													
1101			Reklamasi (non Revegetasi)													
1101			Other Revenue Pemakaian Jalan KPF													
1101			Discount	(18,356,296,181)	*****	*****	*****	*****	*****	*****						
1101			Crushing Other Source													
			TOTAL REVENUE	193,034,360,039	*****	*****	*****	*****	*****	*****						
			COST OF PRODUCTION													
2100			Salary and compensation - Lo	5,110,756,229	636,599,473	*****	615,574,755	*****	615,574,755	*****						
2101	F		Honorarium	2,638,158,352	384,477,518	543,808,328	395,468,830	543,808,328	395,468,830	543,808,328						
2102	F		Bonus & THR	1,970,150,192	268,144,563	409,131,161	237,236,073	409,131,161	237,236,073	409,131,161						
2103	F		Uang Jasa	27,287,267	3,713,933	5,886,867	3,286,667	5,886,867	3,286,667	5,886,867						
2104	F		Uang Pesangon	606,980,370	8,253,185	12,592,593	7,803,704	12,592,593	7,803,704	12,592,593						
2105	F		Insentif	433,564,348	59,010,274	90,037,037	52,221,481	90,037,037	52,221,481	90,037,037						
2150			Employee welfare - Local	3,915,686,537	657,444,389	443,445,111	663,953,407	443,445,111	663,953,407	443,445,111						

Proses *parsing* untuk *file plan* induk perusahaan dan *proposal budget* memiliki pola serupa, dengan tambahan *hop select values* (lihat Gambar 10 dan 11). *Hop* ini berfungsi untuk memilih maupun menghapus *field* yang tidak dibutuhkan, atau mengubah *metadata field*. Hal ini bertujuan agar proses pada *hop* berikutnya lebih cepat dijalankan – karena *field* yang akan digunakan sudah disaring terlebih dahulu.



Gambar 10 Parsing Plan

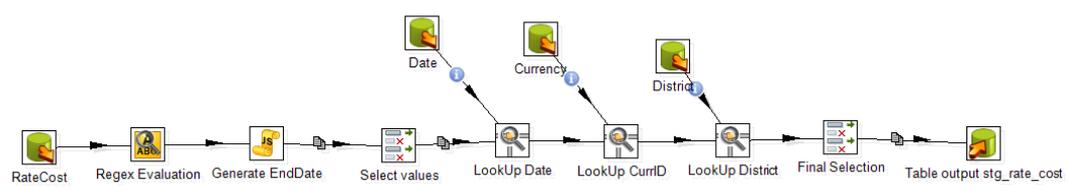
File Plan_SITE_YYYY.xlsx di setiap distrik
 Diupload tiap tanggal 28 bulan sebelumnya



Gambar 11 Parsing Proposal Budget

3.3.2. Staging

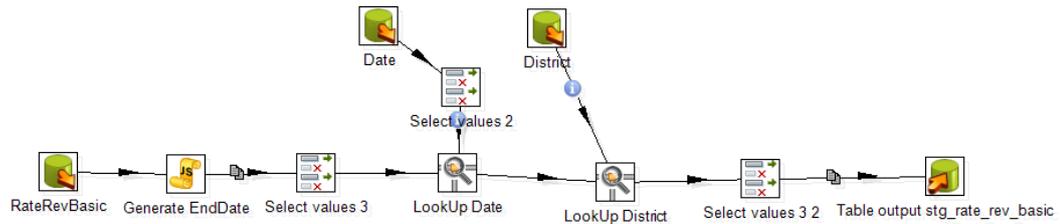
Rangkaian *hop* pada tahap *staging* lebih kompleks daripada tahap lainnya. Pada tahap ini, data yang sudah disimpan pada tabel *MySQL* diubah strukturnya dan disesuaikan dengan target tabel yang dibutuhkan pada *data warehouse*. Tidak ada pola khusus dalam *transformation* ini karena seluruhnya disesuaikan dengan data yang ada. Tahap *staging rate cost* dapat dilihat pada Gambar 12.



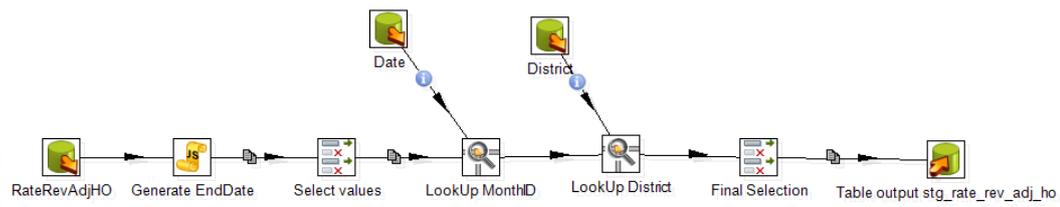
Gambar 12 Staging Rate Cost

Tabel 5 Tabel Hasil Staging Rate Cost

Month_ID	District_ID	Subcont_ID	COA_ID	Equipment_ID	Rate1	Rate2	Curr_ID	Limit_Rate	Limit_PA	Jarak
(NULL)	1	(NULL)	2301	(NULL)	(NULL)	(NULL)	(NULL)	300	0.85	(NULL)
(NULL)	1	(NULL)	2303	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
(NULL)	1	(NULL)	2303	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
(NULL)	1	(NULL)	2302	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
145	1	(NULL)	2301	(NULL)	(NULL)	(NULL)	(NULL)	300	0.85	(NULL)
145	1	(NULL)	2303	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
145	1	(NULL)	2302	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
145	4	1	2301	2	(NULL)	(NULL)	(NULL)	300	0.85	(NULL)
145	4	1	2303	2	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
145	4	1	2302	2	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)



Gambar 13 Staging Rate Revenue Basic



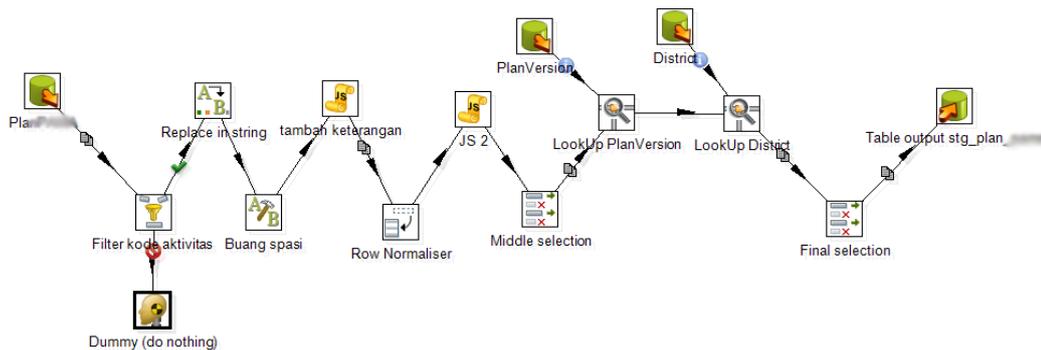
Gambar 14 Staging Rate Revenue Adjustment

Pada *staging rate cost*, *rate revenue basic*, dan *rate revenue adjustment* dilakukan beberapa proses *lookup* dengan tabel *master date*, *currency*, dan *district*. Tabel-tabel tersebut berisi daftar *key* dan deskripsi yang diperlukan, yaitu berkaitan dengan tanggal, bulan, tahun, nama mata uang yang digunakan, serta nama distrik. Proses *lookup* ini dilakukan untuk mengubah struktur tabel menjadi yang diperlukan pada *data warehouse*.

Lain halnya dengan tabel *rate*, *plan* induk perusahaan memiliki data awal yang sangat berbeda dengan target tabel akhir (lihat Tabel 3 dan Tabel 6). Oleh karena itu, proses yang dilakukan pada *staging plan* cenderung lebih kompleks daripada *staging rate*.

Tabel 6 Target Tabel *Financial Plan*

COA_Activity_ID	District_ID	Date_ID	PlanVersion_ID	AmtPlan	QtyVolume	Kettle_flag
(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)



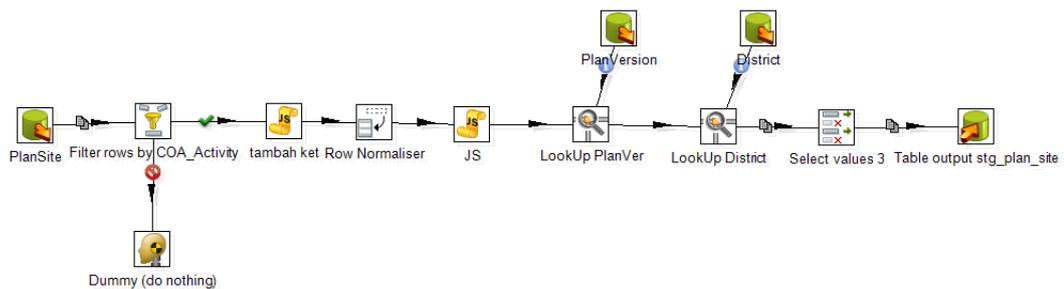
Gambar 15 *Staging Plan* Induk Perusahaan

Gambar 15 merupakan salah satu dari dua *transformation staging plan* induk perusahaan PT X. Pada *transformation* ini, sebelum proses *lookup* dilakukan, data pada tabel dirapihkan terlebih dahulu, antara lain dengan memastikan teks pada kolom deskripsi sudah tepat, menambah keterangan nama induk perusahaan pada tabel, serta melakukan normalisasi pada kolom bulan agar lebih mudah diproses. Proses ini dilakukan agar tersedia *key* yang ingin di-*lookup* pada tabel *master*. Hasil *staging plan* induk perusahaan memiliki struktur seperti pada Tabel 7.

Tabel 7 Hasil *Staging Plan* Induk Perusahaan

COA_Activity_ID	District_ID	Date_ID	PlanVersion_ID	AmtPlan
2007	1	2015-01-01 00:00:00	4	4027195.4768139226
2007	1	2015-02-01 00:00:00	4	4027195.4768139226
2007	1	2015-03-01 00:00:00	4	4027195.4768139226
2007	1	2015-04-01 00:00:00	4	4027195.4768139226
2007	1	2015-05-01 00:00:00	4	4027195.4768139226
2007	1	2015-06-01 00:00:00	4	4027195.4768139226
2007	1	2015-07-01 00:00:00	4	4027195.4768139226
2007	1	2015-08-01 00:00:00	4	4027195.4768139226
2007	1	2015-09-01 00:00:00	4	4027195.4768139226
2007	1	2015-10-01 00:00:00	4	4027195.4768139226
2007	1	2015-11-01 00:00:00	4	4027195.4768139226
2007	1	2015-12-01 00:00:00	4	4736007.2968557700
2101	1	2015-01-01 00:00:00	4	2233798.6435547000
2101	1	2015-02-01 00:00:00	4	2233798.6435547000
2101	1	2015-03-01 00:00:00	4	2233798.6435547000

Proses *staging* pada *proposal budget* menghasilkan struktur tabel yang sama dengan *staging plan* induk perusahaan, tetapi memiliki perbedaan pada kolom *PlanVersion_ID*. Seluruh data *plan* induk perusahaan dan *proposal budget* dari masing-masing distrik akan diintegrasikan ke satu tabel pada *data warehouse*. Proses *staging proposal budget* dapat dilihat pada Gambar 16.



Gambar 16 *Staging Proposal Budget* per Distrik

3.3.3. Mapping ke Data Warehouse

Tabel *output* pada tahap *staging* mengacu pada tabel target pada *data warehouse*, akan tetapi pada pengerjaannya terdapat sejumlah revisi yang diberikan. Oleh karena itu, tabel hasil *staging* merupakan “persinggahan” sebelum data benar-benar diintegrasikan ke *data warehouse*. Pada tahap ini, tabel *staging* diolah kembali jika perlu, kemudian dipetakan ke tabel pada *data warehouse* sesuai dengan *requirement* yang diberikan oleh PT X.

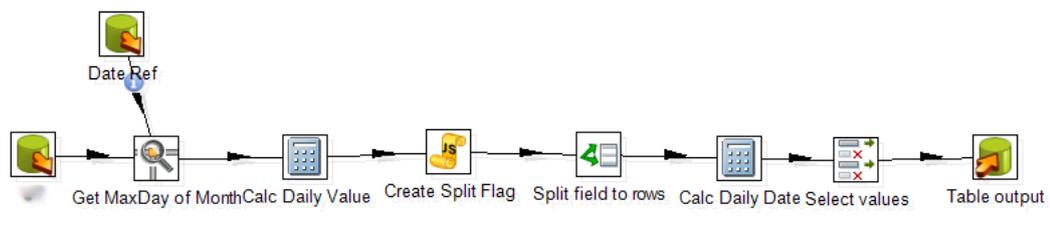


Gambar 17 Mapping Rate Cost ke Data Warehouse

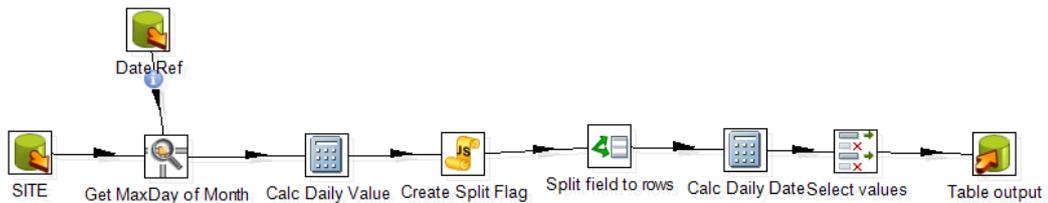
Rate cost berisi *rate* dasar yang digunakan oleh masing-masing distrik terhadap *partnernya*. Metode *insert/update* digunakan agar data yang digunakan sebelumnya selalu diperbarui apabila terdapat perubahan *rate* dan data baru akan langsung dimasukkan ke *data warehouse*. Sama halnya dengan *rate cost*, *mapping rate revenue basic* dan *rate revenue adjustment* menggunakan metode *insert/update*. Proses *update* dilakukan apabila terdapat data distrik, bulan, *charts of accounts*, atau *customer* yang sama. Hasil pemetaan tabel *rate* yang dilakukan penulis sudah sesuai dengan target pada *data warehouse*.

Oleh karena *plan* induk perusahaan dan *proposal budget* per distrik PT X memiliki target tabel yang sama dan sudah memiliki struktur yang sama setelah tahap *staging*, keduanya memiliki proses *mapping* yang serupa ke *data warehouse*.

(Lihat Gambar 18 dan 19)



Gambar 18 Mapping Plan Induk Perusahaan



Gambar 19 Mapping Proposal Budget per Distrik

Pada hasil *staging*, *field* AmtPlan yang dihasilkan berisi total rencana pendapatan dan *budget* per bulannya, sedangkan yang diinginkan pada *data warehouse* yaitu jumlah per harinya. Oleh karena itu, nilai pada AmtPlan dibagi dengan jumlah hari per bulannya. *Hop create split flag*, *split field to rows*, dan *calculate daily date* digunakan untuk membuat daftar tanggal dalam bulan tersebut, diawali dengan membuat penanda (*flag*) untuk masing-masing tanggal, memecahnya jadi sejumlah baris (sebanyak tanggal dalam bulan tersebut), kemudian menghitung tanggal masing-masing. Namun, hasil pemetaan *plan* dan *proposal budget* tidak memenuhi target tabel *financial_plan* pada *data warehouse* (lihat Tabel 6) karena kurangnya informasi mengenai sumber data pada *field QtyVolume* dan *KettleFlag*.

3.3.4. Job

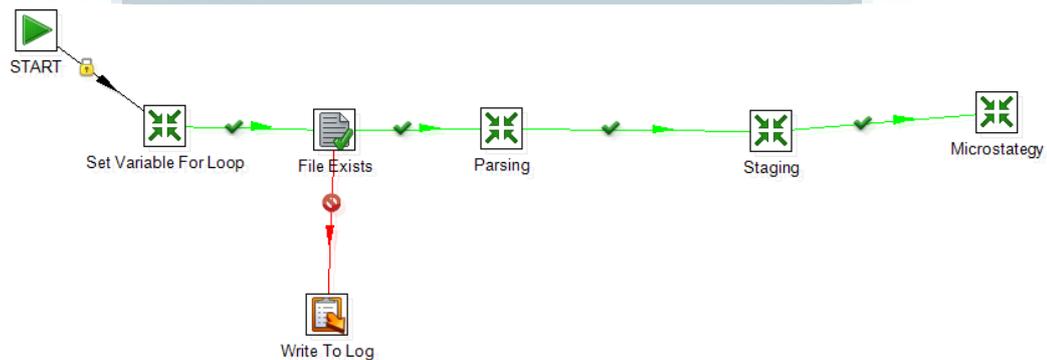
Job dibuat agar *transformation parsing*, *staging*, dan *mapping* ke *data warehouse* dapat dijalankan secara berurutan. Dalam praktik kerja magang ini, seluruh *file job* untuk masing-masing ruang lingkup memiliki pola yang sama. (Lihat Gambar 20 dan 21)

Agar tidak kompleks, *file job* dipecah menjadi dua: *job utama (main job)* dan *job proses (process job)*. Nama dokumen *Microsoft Excel* yang akan diekstraksi

ditentukan pada *main job* melalui *hop set variables*. Variabel diatur menjadi ID dokumen yang sudah ditetapkan pada tabel *files*. *Command Line Parameters* akan menangkap parameter yang diberikan. Jika parameter kosong, *transformation* ini akan menghasilkan *output* berupa tahun (YYYY), bulan (MM), dan tanggal (DD) yang diambil dari tanggal sistem saat *job* dijalankan. *Get Filename from DB* akan menghasilkan nama file sesuai dengan ID dan parameter tahun, bulan, serta tanggal yang diberikan, yang kemudian akan diekstraksi datanya. Sebelum dokumen diekstraksi, tabel *staging* dikosongkan terlebih dahulu agar tidak ada duplikasi data.



Gambar 20 Main Job untuk Rate Cost



Gambar 21 Process Job untuk Rate Cost

Pada *process job*, nama distrik dan departemen ditentukan terlebih dahulu melalui *transformation set variable for loop*, agar lokasi direktori dokumen dapat ditentukan. Apabila dokumen ada, *transformation parsing*, *staging*, dan *mapping* ke *data warehouse* akan dijalankan. Apabila dokumen tidak ada, akan ditulis *error message* pada *file log*.

3.3.5. Shell Script

Shell script dibuat dengan ekstensi .bat sebagai scheduler untuk menjalankan file job yang sudah dibuat. Berikut isi script yang digunakan.

```
1 |echo off
2 |cls
3 |FOR /f %a in ('WMIC OS GET LocalDateTime ^| find ".") DO set DTS=%a
4 |set cur_date=%DTS:~0,4%_DTS:~4,2%_DTS:~6,2%
5 |set module_name=Rate_Cost
6
7 |rem get current directory
8 |set originfolder="%CD%"
9 |set logs="%CD%\logs"
10 |set origindrive=%~d0
11
12 |rem check if logs folder exists, create one if does not exists
13 |if exist %logs% goto Process
14 |mkdir %logs%
15
16 |:Process
17
18 |rem get drive of Kettle Home
19 |call :GetPath ketteDrive "%KETTLE_HOME%"
20
21 |rem change drive to the drive where Kettle resides
22 |%ketteDrive%
23
24 |rem change to the Kettle home directory
25 |cd %KETTLE_HOME%
26
27 |rem call Kitchen or Pan to execute transformation
28 |call Kitchen.bat -file="%originfolder%\Main_%module_name%.kjb" -level=Detailed -logfile=%originfolder%\logs/%cur_date%_%module_name%.log %1 %2 %3
29
30 |rem change to the original drive and directory
31 |%origindrive%
32 |cd %originfolder%
33 |exit /b
34
35 |:GetPath
36 |set "%1=%~d0"
37 |exit /b
```

Gambar 22 Shell Script untuk Rate Cost

Isi script untuk setiap dokumen sama, hanya terdapat perbedaan pada variabel *module_name* yang diatur pada baris kelima. *Module_name* diisi sesuai dengan nama dokumen yang akan diproses, misalnya Rate_Cost, Rate_Revenue_Basic, Rate_Revenue_Adjustment, Plan_SITE, dan sebagainya.

3.4. Kendala yang Dihadapi

Dalam pelaksanaan praktik kerja magang ini, penulis menghadapi beberapa kendala yaitu

1. Perbedaan penggunaan terminologi antara perusahaan dengan PT X

Pada *briefing* di awal pelaksanaan kerja magang, masing-masing anggota tim diberikan ruang lingkup pekerjaan yang berbeda. Namun, setelah dokumen dan data mentah klien diberikan, terminologi yang digunakan agak berbeda sehingga anggota tim kembali memilah-milah mana yang termasuk ruang lingkup pekerjaannya dan mana yang bukan.

2. Tidak tersedianya *server* sementara pada tahap *development*

Penulis dan tim merancang *script ETL* pada komputer masing-masing dan disimpan secara terpisah. Akibatnya, perbedaan penamaan koneksi, *database*, serta *file* harus disamakan agar dapat digunakan bersama.

3. Perubahan rancangan tabel pada *data warehouse*

Pada awal praktik kerja magang, penulis dan tim hanya diberikan *functional specification document* sebagai acuan pengerjaan. Struktur tabel *data warehouse* baru diberikan pada pertengahan masa kerja magang dan tidak langsung diberikan secara lengkap, bahkan masih ada perbaikan. Akibatnya, *script ETL* direvisi berkali-kali untuk memenuhi struktur tabel yang ditentukan.

4. Kurangnya kemampuan dan pengalaman penulis dalam integrasi data

Selama perkuliahan, topik *data warehouse* hanya disinggung sekilas secara teoritis sehingga penulis tidak memiliki kemampuan praktis dalam melakukannya. Selain itu, banyak hal teknis yang kurang dialami selama perkuliahan, seperti penggunaan *regular expressions* (*RegEx*), *environment variable* pada *Windows*, dan sebagainya.

3.5. Solusi Atas Kendala

Untuk menanggulangi masalah-masalah tersebut, berikut solusi yang dilakukan.

1. *Brainstorming* ulang dengan tim

Untuk menyamakan persepsi mengenai kaitan dokumen-dokumen yang ada dengan ruang lingkup yang dimiliki, penulis bersama tim mendiskusikan kaitan *field* pada dokumen dengan ruang lingkup serta target tabel yang diinginkan pada *data warehouse*.

2. Penyimpanan *file* melalui *cloud storage*

Oleh karena tidak adanya *server* khusus untuk digunakan pada tahap *development*, seluruh *script ETL* diunggah pada *cloud storage* setelah nama koneksi, *database*, dan *filenya* disamakan.

3. Revisi *script*

Untuk menyesuaikan dengan struktur *data warehouse* PT X yang terlambat diberikan dan seringkali diperbarui, penulis memeriksa kembali seluruh *script* yang sudah dibuat dan menyesuaikannya dengan struktur terbaru yang diberikan.

4. Bimbingan dari pembimbing lapangan dan pimpinan divisi

Pembimbing lapangan dan pimpinan divisi seringkali memberikan panduan, baik secara tertulis maupun praktik, mengenai teknik-teknik yang belum dipahami penulis. Dalam praktik kerja magang, penulis juga banyak bertanya kepada pembimbing lapangan ketika menghadapi kesulitan.