

## Perancangan Data Warehouse Menggunakan Metode Star Schema dan ETL Untuk Menghasilkan Laporan Efektifitas Proses Rekrutmen Kandidat di Enigma Camp

Rosa Hartanto<sup>\*1</sup>, Jution Candra Kirana<sup>2</sup>, Indra<sup>3</sup>

<sup>1,2</sup>Master Ilmu Komputer, Teknologi Informasi, Universitas Budi Luhur, Indonesia

<sup>3</sup>Sistem Informasi, Teknologi Informasi, Universitas Budi Luhur, Indonesia

Email: <sup>1</sup>2311601153@student.budiluhur.ac.id, <sup>2</sup>2311601633@student.budiluhur.ac.id,

<sup>3</sup>indra@budiluhur.ac.id

### Abstrak

Enigma Camp merupakan perusahaan IT Bootcamp dan manajemen talenta yang berfokus pada pelatihan *programmer* dengan modul yang disesuaikan untuk memenuhi standar industri. Dalam hal konsolidasi data dan analisa, Enigma Camp sering kali mengalami kesulitan dan membutuhkan banyak waktu. Hal tersebut dikarenakan data rekrutmen bersumber dari banyak aplikasi *platform* rekrutmen dengan ukuran data yang besar. Perancangan *Data Warehouse* ini diharapkan mampu mempermudah konsolidasi data tersebut sehingga informasi data dapat digunakan untuk memonitor, menganalisis, serta meningkatkan efektivitas dan efisiensi keseluruhan proses rekrutmen di Enigma Camp. Model *Data Warehouse* memerlukan identifikasi kebutuhan informasi yang akurat dan pemilihan sumber data yang terpercaya. Proses perancangannya mengimplementasikan skema bintang (*Star Schema*) yang melibatkan tahapan ETL (*Extract, Transform, Load*) guna mengintegrasikan, mengekstrak, membersihkan, mentransformasi, dan memuat data ke dalam *Data Warehouse*. Penggunaan metode *Star Schema* dipilih karena kemampuannya dalam menampung beragam tabel dimensi, termasuk sub-tabel dimensi, yang berpotensi menghasilkan informasi lebih kaya untuk mendukung pengambilan keputusan. Sistem *Data Warehouse* yang dirancang ini mampu menyajikan laporan analisis rekrutmen secara efektif, sehingga mendukung pengambilan keputusan strategis yang lebih cepat dan tepat. Hal ini berkontribusi pada peningkatan efisiensi operasional dan kualitas rekrutmen kandidat di Enigma Camp.

**Kata kunci:** *Data Warehouse, Enigma Camp, ETL, Star Schema*

## ***Data Warehouse Design Using Star Schema and ETL Methods to Generate Candidate Recruitment Process Effectiveness Reports at Enigma Camp***

### ***Abstract***

*Enigma Camp is an IT Bootcamp and talent management company that focuses on programmer training with modules tailored to meet industry standards. Enigma Camp often experiences difficulties and time-consuming data consolidation and analysis. This is because recruitment data is sourced from numerous recruitment platform applications with large data sizes. The design of this Data Warehouse is expected to facilitate data consolidation so that data information can be used to monitor, analyze, and improve the effectiveness and efficiency of the entire recruitment process at Enigma Camp. The Data Warehouse model requires accurate identification of information needs and the selection of reliable data sources. The design process implements a star schema that involves ETL (Extract, Transform, Load) stages to integrate, extract, clean, transform, and load data into the Data Warehouse. The use of the Star Schema method was chosen because of its ability to accommodate various dimension tables, including sub-dimension tables, which have the potential to produce richer information to support decision-making. The designed Data Warehouse system is able to present recruitment analysis reports effectively, thereby supporting faster and more accurate strategic decision-making. This contributes to improving operational efficiency and the quality of candidate recruitment at Enigma Camp.*

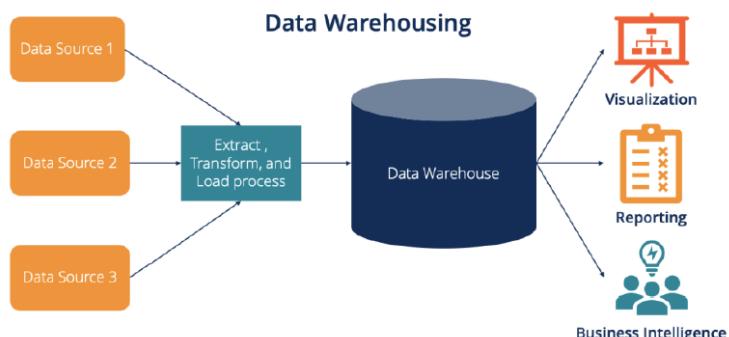
**Keywords:** *Data Warehouse, Enigma Camp, ETL, Star Schema*

## 1. PENDAHULUAN

Pemanfaatan *Data Warehouse* menjadi kebutuhan esensial bagi hampir setiap organisasi, tak terkecuali Enigma Camp, perusahaan IT Bootcamp dan manajemen talenta. Dengan adanya *Data Warehouse*, Enigma Camp dapat mengintegrasikan data dengan lebih mudah. Secara fundamental, proses tersebut adalah penggabungan data dari beragam platform rekrutmen eksternal (seperti Glints, Kalibr, LinkedIn) dan juga pencatatan internal yang dilakukan secara manual, di mana terjadi pemindahan data antar file excel yang berbeda.

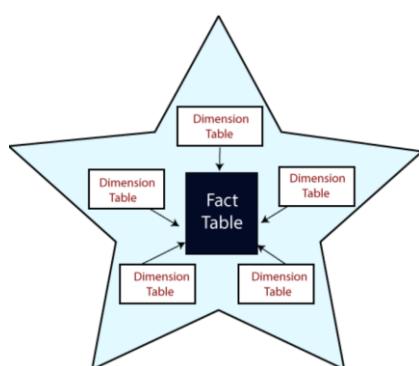
*Data Warehouse* menyediakan lingkungan informasi strategis dengan cara mengumpulkan, membersihkan, dan mentransformasi data yang relevan untuk pengambilan keputusan strategis. Informasi strategis ini diharapkan dapat meningkatkan produktivitas dan memperbaiki operasional perusahaan, memfasilitasi pertumbuhan berkelanjutan di tengah keterbatasan ekspansi. Sumber data yang digunakan dapat berasal dari dalam atau luar perusahaan [1].

Definisi *Data Warehouse* menurut [2] adalah metode yang dirancang khusus untuk mendukung sistem pengambilan keputusan (DSS) dan sistem informasi eksekutif (EIS). Meskipun wujud fisiknya adalah basis data, ada perbedaan mendasar dalam desainnya, *Data Warehouse* tidak menerapkan normalisasi seperti basis data pada umumnya. *Data Warehouse* mempunyai ciri khas berfokus pada topik, terintegrasi, mencatat riwayat waktu, mudah diakses, dan sangat mendukung proses pengambilan keputusan manajemen. Menurut [3], *Data Warehouse* adalah sekumpulan data yang berorientasi pada subjek, saling terintegrasi, *non-volatile* dan bersifat *time series* guna mendukung pengambilan keputusan bagi *executive dicisioner*. Menurut [4], *Data Warehouse* adalah gudang data yang di dalamnya menyimpan beragam informasi baik yang terstruktur ataupun tidak terstruktur. Model *Data Warehouse* dapat dilihat pada Gambar 1 berikut.



Gambar 1. Model *Data Warehouse* [5]

*Star Schema* adalah teknik pemodelan data yang umum digunakan dalam *Data Warehouse*, pemodelan ini sangat optimal untuk kueri dan pelaporan karena menyederhanakan kueri kompleks dan meningkatkan kinerja dengan mengurangi jumlah gabungan tabel [6]. Model *Star Schema* terdiri dari tabel fakta di tengah, kemudian tabel dimensi yang berada di sekelilingnya [7]. Hubungan antara keduanya dibangun melalui penggunaan kunci utama di tabel dimensi dan kunci asing di tabel fakta. Berbagai tabel dimensi, dengan tingkat detail yang berbeda, digabungkan dengan tabel fakta untuk membentuk skema bintang, sebagaimana ditemukan dalam penelitian terbaru [8]. Model *Star Schema* dapat dilihat pada Gambar 2 berikut.

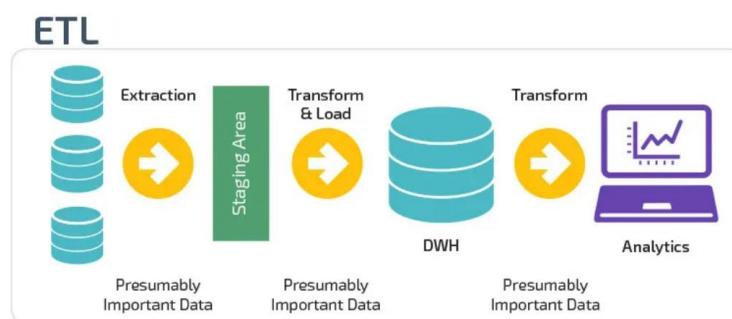


Gambar 2. Model *Star Schema* [9]

OLAP (*Online Analytical Processing*) merupakan salah satu teknik dalam analisa data yang dapat meningkatkan kualitas dalam proses pengambilan keputusan. Data dalam OLAP *Cube* disimpan dalam bentuk multidimensi. Proses OLAP dimulai dengan pengambilan data dari *Data Warehouse*. Data dari *Data Warehouse* kemudian diubah menjadi dimensi-dimensi yang dikumpulkan sedemikian rupa sehingga dapat diakses dan dianalisa [10].

Pentaho Data Integration (PDI) adalah perangkat lunak *open-source* yang berfungsi sebagai alat ETL (*Extract, Transform, Load*). PDI digunakan untuk melakukan proses pengambilan data dari berbagai sumber, mengubah format data sesuai kebutuhan, dan memuatnya ke dalam *Data Warehouse* atau sistem tujuan lainnya. Keunggulan PDI terletak pada fleksibilitasnya dalam mengintegrasikan data dari beragam *platform* dan kemudahan penggunaannya melalui antarmuka grafis [11].

ETL merupakan serangkaian proses untuk mengumpulkan dan memproses data dari berbagai sumber, lalu mengubahnya menjadi satu sumber yang baru. Ini merupakan inti dari integrasi data dan sering kali terkait erat dengan *Data Warehouse*. Melalui proses ETL, data diekstrak dari sumber terpilih diubah formatnya sesuai aturan bisnis lalu dimuat ke dalam struktur data target. ETL sangat penting pada perancangan *Data Warehouse* karena memungkinkan data operasional masuk ke dalamnya [8]. Untuk mendapatkan kualitas data yang optimal, data harus dibersihkan saat diambil. Beberapa masalah umum meliputi nomor telepon yang tidak valid, kode buku yang sudah usang, atau nilai data yang hilang (null). Tujuan ETL adalah mengumpulkan, menyaring, mengolah, dan menggabungkan data-data yang relevan dari berbagai sumber untuk disimpan ke dalam *Data Warehouse*. Hasil dari proses ETL adalah dihasilkannya data yang memenuhi kriteria data warehouse seperti data yang historis, terpadu, terangkum, statis, dan memiliki struktur yang dirancang untuk keperluan proses analisis [12]. Pendekatan konvensional ETL melibatkan tiga langkah utama: mengekstrak data dari sumbernya, meletakkannya di staging area, kemudian mengubah formatnya (transformasi) dan memuatnya ke dalam *Data Warehouse*. Gambaran detail alur proses ETL ini bisa dilihat pada Gambar 3 berikut.



Gambar 3. Alur Proses ETL [13]

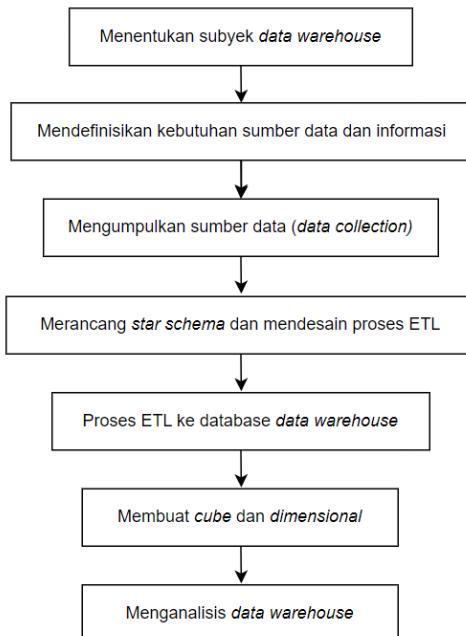
Pengimplementasian *Star Schema* dan ETL pada perancangan *Data Warehouse* telah banyak dibahas di penelitian-penelitian sebelumnya, hanya saja penelitian tersebut hanya melibatkan satu tabel fakta pada perancangan *Data Warehouse*. Contoh penelitian sebelumnya yaitu perancangan *Data Warehouse* pada lingkungan gereja dengan metode *Star Schema* dan ETL yang melibatkan satu tabel fakta yaitu tabel utama [14]. Penelitian selanjutnya yaitu implementasi *Star Schema* dan ETL dalam perancangan *Data Warehouse* pada industri *garment* yang melibatkan satu tabel fakta yaitu fact\_pesanan [15]. Selanjutnya yaitu penelitian yang dilakukan pada industri properti yaitu PT Wijaya Karya Realty, dimana metode *Star Schema* dan ETL digunakan dalam perancangan *Data Warehouse* dengan melibatkan 1 tabel fakta yaitu fact\_table [9].

Penelitian sebelumnya diatas telah berhasil diimplementasikan dengan menghasilkan satu tabel fakta dalam perancangan *Data Warehouse*. Akan tetapi, satu tabel fakta ini dirasa masih terbatas dikarenakan *insight* data yang dihasilkan kurang informatif pada skala perusahaan. Penelitian ini akan mengadaptasi dari beberapa penelitian diatas dengan fokus pada implementasi *Star Schema* dan ETL dengan menghasilkan dua tabel fakta. Dengan dua tabel fakta yang dihasilkan diharapkan *insight* dari data bisa lebih luas dan lebih informatif yang bisa dimanfaatkan dalam skala perusahaan. Penelitian ini diharapkan dapat berfungsi sebagai pedoman strategis dalam analisis dan perancangan *Data Warehouse*, terutama yang berkaitan dengan output tabel fakta lebih dari satu, seperti yang ditemukan dalam proses rekrutmen kandidat di perusahaan Enigma Camp.

## 2. METODE PENELITIAN

Penelitian ini dilakukan dengan pendekatan kuantitatif, yaitu bertujuan untuk memperoleh kumpulan data dari berberapa sumber data yang ada di perusahaan Enigma Camp sehingga hasilnya dapat dilakukan

perancangan *Data Warehouse* dengan metode *Star Schema* dan ETL. Alur proses penelitian bisa dilihat di Gambar 4 berikut.



Gambar 4. Alur Proses Penelitian

Metode penelitian melibatkan serangkaian tahapan yang terstruktur, adapun pembahasan secara rinci akan dibahas di sub-sub bab berikut.

### 2.1. Menentukan Subyek *Data Warehouse*

Pada tahap ini dilakukan identifikasi pada area bisnis atau topik utama yang menjadi fokus dalam perancangan *Data Warehouse*. Subyek yang akan dibahas adalah "Efektivitas Proses Rekrutmen". Sehingga *Data Warehouse* akan secara khusus menyimpan dan mengatur data yang berkaitan dengan berbagai aspek rekrutmen kandidat, seperti sumber lamaran, tahapan wawancara, kinerja kandidat, hasil perekrutan, waktu perekrutan, dll.

### 2.2. Menentukan Kebutuhan Sumber Data dan Informasi

Pada tahap ini dilakukan identifikasi dan pemahaman pada semua sumber data mentah yang berisi informasi relevan dengan subyek *Data Warehouse*. Selain itu, dilakukan penentuan informasi atau metrik spesifik yang dibutuhkan dari sumber-sumber tersebut untuk menjawab kebutuhan bisnis.

### 2.3. Mengumpulkan Sumber Data (*Data Collection*)

Pada tahap ini dilakukan pengumpulan data sumber (*data collection*) yaitu pengumpulan data mentah yang telah diidentifikasi dari berbagai sumber. Hal tersebut melibatkan akses ke basis data, mengekspor file, dan mengumpulkan informasi secara manual.

### 2.4. Merancang *Star Schema* dan Mendesain Proses ETL

Pada tahap ini dilakukan perancangan *Star Schema* melibatkan pembuatan model *Data Warehouse* dengan dua tabel fakta dan beberapa tabel dimensi yang terhubung langsung dengannya. Kemudian mendesain ETL yang akan menentukan bagaimana data akan dilakukan pemindahan dari kumpulan data mentah ke database *Data Warehouse*.

## 2.5. Proses ETL ke Database *Data Warehouse*

Pada tahap ini dilakukan eksekusi proses ETL berdasarkan desain yang ditentukan sebelumnya. Data yang telah ditransformasi dimuat ke dalam tabel database *Data Warehouse*. Langkah ini dilakukan secara otomatis dan dijadwalkan untuk berjalan secara teratur.

## 2.6. Membuat *Cube* dan *Dimensional*

Pada tahap ini dilakukan pembuatan *Cube* dan *Dimensional* yang mengacu pada pembangunan kubus OLAP (*Online Analytical Processing*) di atas *Data Warehouse*. Kubus ini melakukan pra-agregasi data di sepanjang berbagai dimensi, memungkinkan kinerja kueri yang sangat cepat untuk pelaporan analitis.

## 2.7. Menganalisis *Data Warehouse*

Pada tahap ini dilakukan penggunaan *Data Warehouse* untuk tujuan utamanya yaitu melakukan analisis data. Ini adalah langkah terakhir dan krusial di mana pengguna akhir (misalnya, direksi, manajer HR, pemimpin rekrutmen) akan menggunakan dashboard hasil perancangan *Data Warehouse* sebagai bahan pertimbangan dalam pengambilan keputusan atau kebijakan perusahaan.

## 3. HASIL DAN PEMBAHASAN

Pada proses perancangan *Data Warehouse* ini akan dibahas terkait proses ETL tabel dimensi, ETL tabel fakta, pembuatan *Cube* dan pembuatan dashboard. Dashboard nantinya akan menjadi hasil akhir dari perancangan *Data Warehouse* yang menunjukkan sekaligus menjadi bukti bahwa proses perancangan *Data Warehouse* telah selesai dan berhasil. Selain itu, dashboard tersebut akan menghasilkan visualisasi data yang dapat diambil *insight* untuk bahan analisa dan dasar pengambilan keputusan di Enigma Camp. Adapun pembahasan secara rinci akan dibahas di sub-sub bab berikut.

### 3.1. Proses ETL Tabel Dimensi

Pada proses *Extraction, Transformation, and Load (ETL)*, langkah pertama yang dilakukan adalah menghubungkan Pentaho Data Integration (PDI) ke database dengan koneksi database. Selanjutnya, mengambil atau mengekstrak tabel dari database menggunakan komponen tabel input. Kemudian, pada proses transformasi, data yang diinginkan akan ditampilkan dengan menggunakan *Select Values*. Selanjutnya, data akan dirapikan dengan menggunakan *Sort Rows* untuk mengurutkan data dalam tabel berdasarkan kolom tertentu. Berikut proses ETL pada beberapa tabel dimensi.



Gambar 5. ETL pada Tabel Candidate

Gambar 5 merupakan proses ETL pada tabel Candidate. proses ETL dimulai dengan membaca data menggunakan komponen tabel input. Kemudian, menggunakan komponen *Select Values* untuk memilih nilai atau nilai data yang ingin diambil untuk ditampilkan. Kemudian, proses transformasi yaitu merapikan data menggunakan *Sort Rows* untuk mengurutkan data dalam tabel berdasarkan kolom tertentu. Selanjutnya, menghilangkan baris data yang duplikat bersdasarkan kolom Email. Data Candidate yang sudah bersih, terstruktur dan unik akan dimuat kedala tabel dimensi bernama Dim Candidate di dalam *Data Warehouse*.



Gambar 6. ETL pada Tabel Location

Gambar 6 merupakan proses ETL pada tabel Location. proses ETL dimulai dengan membaca data menggunakan komponen tabel input. Kemudian, menggunakan komponen *Select Values* untuk memilih nilai atau nilai data yang ingin diambil untuk ditampilkan. Kemudian, proses transformasi yaitu merapikan data menggunakan *Sort Rows* untuk mengurutkan data dalam tabel berdasarkan kolom tertentu. Selanjutnya, menghilangkan baris data yang duplikat. Data Location yang sudah bersih, terstruktur dan unik akan dimuat kedala tabel dimensi bernama Dim Location di dalam *Data Warehouse*.



Gambar 7. ETL pada Tabel Date

Gambar 7 merupakan proses ETL pada tabel Date. proses ETL dimulai dengan membaca data menggunakan komponen tabel input. Kemudian, menggunakan komponen *Select Values* untuk memilih nilai atau nilai data yang ingin diambil untuk ditampilkan. Kemudian, dilakukan perhitunganan tanggal sesuai kebutuhan dengan menggunakan komponen *Calculator*. kemudian, proses transformasi yaitu merapikan data menggunakan *Sort Rows* untuk mengurutkan data dalam tabel berdasarkan kolom tertentu. Selanjutnya, menghilangkan baris data yang duplikat. Data Location yang sudah bersih, terstruktur dan unik akan dimuat kedalam tabel dimensi bernama Dim Date di dalam *Data Warehouse*.



Gambar 8. ETL pada Tabel Interviewer

Gambar 8 merupakan proses ETL pada tabel Interviewer. proses ETL dimulai dengan membaca data menggunakan komponen tabel input. Kemudian, menggunakan komponen *Select Values* untuk memilih nilai atau nilai data yang ingin diambil untuk ditampilkan. Kemudian, Menambahkan kolom *Role* pada tabel. kemudian, proses transformasi yaitu merapikan data menggunakan *Sort Rows* untuk mengurutkan data dalam tabel berdasarkan kolom tertentu. Selanjutnya, menghilangkan baris data yang duplikat. Data Location yang sudah bersih, terstruktur dan unik akan dimuat kedala tabel dimensi bernama Dim Interviewer di dalam *Data Warehouse*.



Gambar 9. Intruksi ETL Tabel Dimensi

Gambar 9 merupakan proses intruksi ETL pada tabel dimensi sebagai salah satu instruksi untuk mengelola alur kerja transformasi data. Alur transformasi bersifat berurutan dari proses ETL tabel Candidate, tabel Location, tabel Date dan terakhir tabel Interviewer. Kemudian, proses intruksi ETL tabel dimensi akan dijalankan jika kondisi tertentu terpenuhi. Setelah transformasi dilakukan dengan membersihkan data, maka tabel hasil transformasi akan diunggah ke database.

### **3.2. Hasil Proses ETL Tabel Dimensi**

Setelah proses *Extraction, Transformation, and Load (ETL)* selesai dilakukan maka output yang muncul berupa tabel dimensi berdasarkan masing-masing proses ETL tersebut. Berikut merupakan output dari masing-masing tabel dimensi.

	student_id	student_email	student_name	phone_number	date_of_birth	gender	grades	registering
1	Amanayur@gmail.com	Ariti	8571575537	2000-07-19	L	SPK	Hakasee Perpanjangan Lunas	
2	Ibrahim@gmail.com	Suhail Iman Pratama	8571348818	2000-08-19	L	SI	Sistan Interwasi	
3	Safira@gmail.com	Amelia Nurul Wardhani	8571348817	2000-08-19	L	SPK	Hakasee Perpanjangan Lunas	
4	Syuraini@gmail.com	Eshvarini Amorewa	8562087793	2000-09-19	L	SPK	Teknik Komputer dan Jaringan	
5	Elvira@gmail.com	Akbarilah Saleh	9653155963	1997-05-27	L	SI	Teknik Interwaksi	
6	Abdullah@gmail.com	Alexander Abdullah	8228285959	1997-05-20	L	SI	Teknik Geofisika	
7	Asya@gmail.com	Ayush Ravi Kirat	8562087793	2000-01-17	L	SPK	Hakasee Perpanjangan Lunas	
8	Aminah@gmail.com	Azizah Syaiful Kirat	8562087793	2000-01-17	L	SPK	Hakasee Perpanjangan Lunas	
9	Anechandra@gmail.com	Anissa Faiz Santosa	8461640581	1999-02-20	L	SI	Pendidikan Matematika	
10	Anehtindra@gmail.com	Akhmat Ziyun Sipang	8571110037	2000-07-20	L	SI	Informasi	
11	Annesyariah@gmail.com	Adi Prayoga Nugraha	8193483203	2000-01-19	L	SI	Teknik Geodesi	
12	Arifin@gmail.com	Aldi Arifin	8571348818	2000-08-19	L	SPK	Teknik Geodesi	
13	Azizah@gmail.com	Vicky Adriana	8562087793	2000-02-20	L	SPK	Teknologi Informasi	
14	Aldianurah@gmail.com	Akhila Farrah Pernada	8373211547	2000-07-19	L	SI	Informasi	
15	Alwanud@gmail.com	Muhamad Qabry Anandi	9653492428	2000-02-20	L	SI	Teknik Arkeologi	
16	Alfarrizal@gmail.com	Marina	89612087793	2000-01-17	L	SPK	Hakasee Perpanjangan Lunas	
17	Alisya@gmail.com	Firdausi Alisya	8571221294	2000-01-20	L	SPK	Hakasee Perpanjangan Lunas	
18	Alwynpratiwi@gmail.com	Umarina Durgawita Natalea	9556168337	2000-01-16	L	SPK	Hakasee Perpanjangan Lunas	
19	Angga@gmail.com	Rangga Triantika	8571097081	1999-01-19	L	SPK	Teknik Komputer dan Jaringan	
20	Anggraini@gmail.com	Rangga Saman	8571097081	1999-01-19	L	SPK	Teknologi Informasi	
21	Angga@gmail.com	Rangga Saman	8571097081	1999-01-19	L	SPK	Teknik Komputer dan Jaringan	
22	Anggun@gmail.com	Rangga Mardiana	8773367295	2000-01-19	L	SI	Teknik Informatika	
23	Asnawati@gmail.com	Rendy Elisa	9335325504	2000-01-19	L	SI	Teknik Interwaksi	
24	Ashadisa@gmail.com	Muhammad Zidni Ilman	8571348818	2000-01-19	L	SI	Teknik Telekomunikasi	
25	Augustina@gmail.com	Randy Werdina	857122087793	2000-01-20	L	SPK	Hakasee Perpanjangan Lunas	
26	Aulia@gmail.com	Ridwan Aulia	8571221294	2000-01-20	L	SPK	Hakasee Perpanjangan Lunas	
27	Azharah@gmail.com	Muhammad Alauddin Muhamad	8571097082	2000-01-20	L	SPK	IPB	
28	Blaam@gmail.com	Ria Dewanti	89612087793	2000-02-20	L	SI	Informasi	
29	Blaam@gmail.com	Ria Alis Khasim	82281519963	2000-02-20	L	SI	Sistan Interwasi	
30	Blaam@gmail.com	Ria Alis Khasim	82281519963	2000-02-20	L	SI	Sistan Interwasi	
31	Blaam@gmail.com	Ria Alis Khasim	2174484	2000-02-20	L	SPK	IPB	

## Gambar 10. Output ETL Tabel Candidate

Gambar 10 merupakan hasil proses ETL untuk tabel Candidate, dimana hasilnya berupa tabel dengan kolom candidate id, email, fullname, phone number, date of birth, gender, grades dan majoring. Tabel

Candidate disorder berdasarkan kolom `candidate_id` dan hasil datanya tidak ada yang duplikat sesuai konfigurasi pada saat proses ETL.

Gambar 11. Output ETL Tabel Location

Gambar 11 merupakan hasil proses ETL untuk tabel Location, dimana hasilnya berupa tabel dengan kolom location\_id, town\_origin, province\_origin dan training\_center. Tabel Location diorder berdasarkan kolom location\_id dan hasil datanya tidak ada yang duplikat sesuai konfigurasi pada saat proses ETL.

Gambar 12. Output ETL Tabel Date

Gambar 12 merupakan hasil proses ETL untuk tabel Date, dimana hasilnya berupa tabel dengan kolom date\_id, date\_value, day, month, quarter, year dan day\_of\_week. Tabel Date diorder berdasarkan kolom date\_id dan hasil datanya tidak ada yang duplikat sesuai konfigurasi pada saat proses ETL.

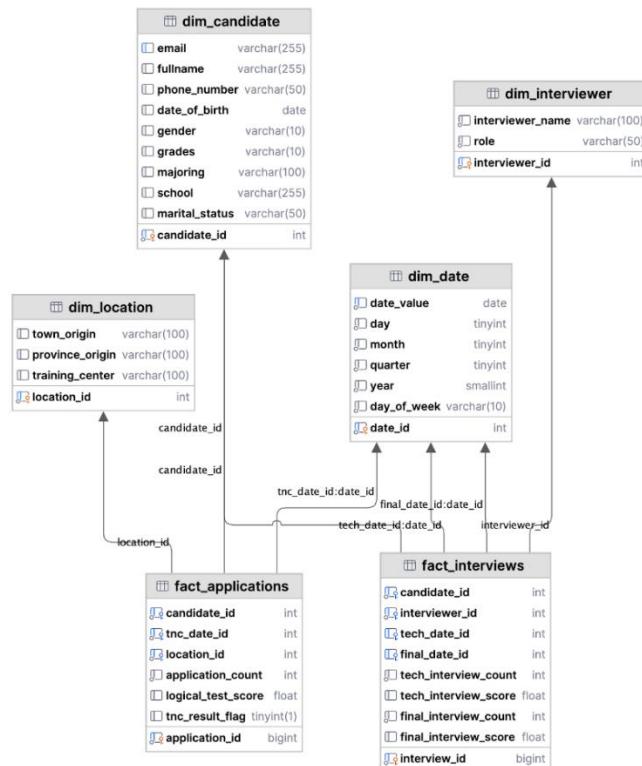
#	Entrenier_Id	Entrenier_Nama	relasi
1	1	Azam Alizi	"TEKNIS"
2	2	Benedic Nur	"TEKNIS"
3	3	Bima Agusti	"TEKNIS"
4	4	D. Suciati	"TEKNIS"
5	5	Julian Candra Kirana	"TEKNIS"
6	6	K. Dwi	"TEKNIS"
7	7	R. Faizal Ramadhan	"TEKNIS"
8	8	R. Erfagi Ramadhan	"TEKNIS"
9	9	S. Stevens Suciati	"TEKNIS"
10	10	S. Tomy	"TEKNIS"
11	11	S. Irawi Agusti	"TEKNIS"
12	12	S. Iman Yessi Kristianti	"TEKNIS"

## Gambar 13. Output ETL Tabel Interviewer

Gambar 13 merupakan hasil proses ETL untuk tabel Interviewer, dimana hasilnya berupa tabel dengan kolom interviewer\_id, interviewer\_name dan role. Tabel Interviewer diorder berdasarkan kolom interviewer\_id dan hasil datanya tidak ada yang duplikat sesuai konfigurasi pada saat proses ETL.

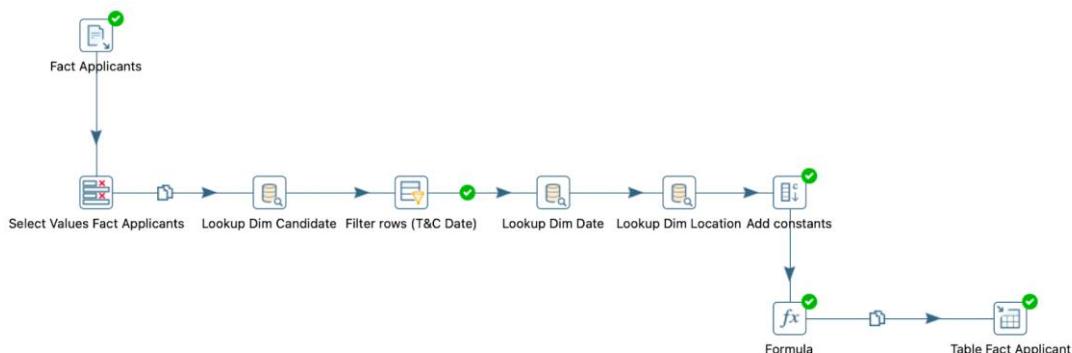
### 3.3. Proses ETL Tabel Fakta dengan *Star Schema*

Setelah dimensi-dimensi dibangun, tahap selanjutnya dalam arsitektur *Data Warehouse* di Enigma Camp adalah memproses dan memuat data ke dalam tabel fakta melalui serangkaian tahapan ETL (*Extract, Transform, Load*) dengan *Star Schema*. Proses ini krusial untuk mengintegrasikan data operasional menjadi metrik bisnis yang siap dianalisis. Model *Star Schema* pada tabel fakta bisa dilihat di Gambar 14 berikut.



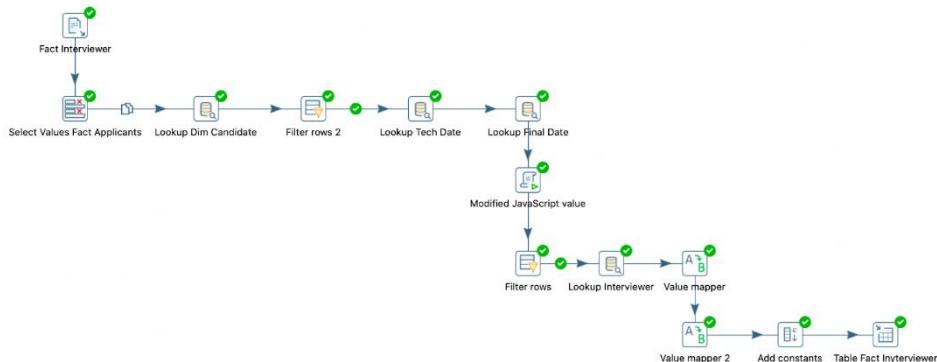
Gambar 14. Model *Star Schema* Pada Tabel Fakta

Dalam *Star Schema* ini, dapat dilihat hubungan antara beberapa tabel dimensi (dim\_candidate, dim\_location, dim\_date, dan dim\_interviewer) dan dua tabel fakta yang dihasilkan yaitu fact\_applications dan fact\_interviews. Tabel dimensi dapat dihubungkan karena tabel fakta memiliki kunci utama dari setiap tabel dimensi. *Star Schema* ini dapat dibentuk untuk meningkatkan kinerja kueri, sehingga memudahkan analisis data. Berikut proses ETL pada dua tabel fakta.



Gambar 15. Tabel Fakta Applicant

Gambar 15 merupakan gambaran proses ETL pada tabel fakta fact\_applicant yang mana ekstraksi tabel dari database MySQL dilakukan dengan menggunakan komponen tabel input. Kemudian menggunakan komponen *Select Values* untuk memilih nilai atau nilai data yang ingin diambil untuk ditampilkan. Selanjutnya, proses transformasi yaitu *lookup* dan *filtering* data menggunakan *Lookup* dan *Formula* untuk *filter* data dari hasil relasi antar tabel dimensi sehingga terbentuk tabel fakta baru.



Gambar 16. Tabel Fakta Interviewer

Gambar 16 merupakan gambaran proses ETL pada tabel fact\_interviewer yang mana ekstraksi tabel dari database MySQL dilakukan dengan menggunakan komponen tabel input. Kemudian menggunakan komponen *Select Values* untuk memilih nilai atau nilai data yang ingin diambil untuk ditampilkan. *Modified JavaScript value* memungkinkan transformasi data yang lebih kompleks atau kustom menggunakan *script* JavaScript. Ini bisa digunakan untuk menghitung metrik baru, memformat ulang data, atau menerapkan logika bisnis yang tidak dapat dicakup oleh langkah-langkah PDI standar. Selanjutnya, proses transformasi yaitu *lookup* dan *filtering* data menggunakan *Lookup* dan *Formula* untuk *filter* data dari hasil relasi antar tabel dimensi sehingga terbentuk tabel fakta baru.



Gambar 17. Proses Intruksi ETL Tabel Fakta

Gambar 17 merupakan proses intruksi ETL tabel fakta dengan *Star Schema* pada PDI. Proses intruksi ETL tersebut terdiri dari dua tabel fakta yaitu fact\_applicants dan fact\_interviewer yang berisi data yang hasil relasi beberapa tabel dimensi yaitu tabel dimensi, yaitu dim\_candidate, dim\_location, dim\_date, dan dim\_interviewer. Proses intruksi ETL tabel fakta berhasil dilakukan dengan daftar centang hijau yang menunjukkan proses yang benar.

### 3.4. Hasil Proses ETL Tabel Fakta

Setelah proses *Extraction, Transformation, and Load (ETL)* selesai dilakukan maka output yang muncul berupa tabel fakta berdasarkan masing-masing proses ETL tersebut. Berikut merupakan output dari masing-masing tabel fakta.

The screenshot shows a database interface with a query results table. The columns are labeled: application\_id, candidate\_id, tnc\_date\_id, location\_id, application\_count, logical\_test\_score, and tnc\_result\_flag. The data consists of 10 rows of numerical values.

	application_id	candidate_id	tnc_date_id	location_id	application_count	logical_test_score	tnc_result_flag
1	16	188	14	56	1	80	0
2	17	41	17	56	1	82	0
3	18	16	25	56	1	90	0
4	19	31	26	56	1	80	0
5	20	68	32	56	1	80	0
6	21	129	32	56	1	62	0
7	22	228	54	56	1	65	0
8	23	235	33	56	1	90	0
9	24	216	33	56	1	70	0
10	25	29	36	56	1	70	0

Gambar 18. Tabel Fakta Fact\_Applicants

Gambar 18 merupakan hasil proses ETL untuk tabel fact\_applicants, dimana hasilnya berupa tabel dengan kolom application\_id, candidate\_id, tnc\_date\_id, location\_id, application\_count, logical\_test\_score, dan tnc\_result\_flag. Kolom application\_count, logical\_test\_score, dan tnc\_result\_flag merupakan kolom baru dimana nilai pada kolom-kolom tersebut merupakan hasil dari formula yang ada di konfigurasi pada saat proses ETL.

The screenshot shows a database interface with a query results table. The columns are labeled: interview\_id, candidate\_id, interviewer\_id, tech\_date\_id, final\_date\_id, tech\_interview\_count, tech\_interview\_score, and final\_interview\_count. The data consists of 3 rows of numerical values.

	interview_id	candidate_id	interviewer_id	tech_date_id	final_date_id	tech_interview_count	tech_interview_score	final_interview_count
1	7	66	4	24	38	1	1	
2	8	182	10	32	38	1	1	
3	9	93	13	41	43	1	1	

Gambar 19. Tabel Fakta Fact\_Interviewers

Gambar 19 merupakan hasil proses ETL untuk tabel fact\_interviewers, dimana hasilnya berupa tabel dengan kolom interview\_id, candidate\_id, interviewer\_id, tech\_date\_id, final\_date\_id, tech\_interview\_count, tech\_interview\_score, dan final\_interview\_count. Kolom tech\_interview\_count, tech\_interview\_score, dan final\_interview\_count merupakan kolom-kolom baru dimana nilai pada kolom tersebut merupakan hasil dari *modified javascript value* yang ada di konfigurasi pada saat proses ETL.

### 3.5. Proses Pembuatan Cube

Dalam perancangan *Data Warehouse* di Pentaho Server, *Cube* merupakan struktur data multidimensional yang dirancang khusus untuk analisis data yang cepat dan efisien. Ini adalah inti dari kemampuan analisis di Pentaho, terutama melalui komponen *Mondrian OLAP engine*. Berikut merupakan proses pembuatan *Cube* pada perancangan *Data Warehouse* di Enigma Camp.

The screenshot shows the Pentaho Data Integration interface with the 'recruitment\_process\_cube' configuration panel. It includes a tree view of dimensions and facts, and a table for setting cube attributes like name, description, caption, cache, enabled, and visible.

Attribute	Value
name	recruitment_process_cube
description	
caption	
cache	<input checked="" type="checkbox"/>
enabled	<input checked="" type="checkbox"/>
visible	<input checked="" type="checkbox"/>

Gambar 20. Cube Rekrutmen

Gambar 20 merupakan proses pembuatan *Cube* Rekrutmen yang akan dimenghasilkan dashboard distribusi jumlah pelamar berdasarkan jenis kelamin dan lokasi asal.

The screenshot shows the Pentaho Data Integration interface with the 'candidate\_demographics\_cube' configuration panel. It includes a tree view of dimensions and facts.

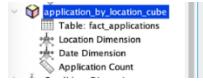
Gambar 21. Cube Candidate Demographics

Gambar 21 merupakan proses pembuatan *Cube* Candidate Demographics yang akan dimenghasilkan dashboard distribusi jumlah pelamar berdasarkan jenis kelamin dan tahun aplikasi.

The screenshot shows the Pentaho Data Integration interface with the 'interview\_analysis\_cube' configuration panel. It includes a tree view of dimensions and facts.

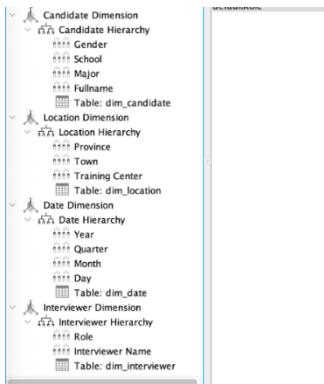
Gambar 22. Cube Interview Analysis

Gambar 22 merupakan proses pembuatan *Cube* Interview Analysis yang akan dimenghasilkan dashboard distribusi jumlah wawancara final berdasarkan kombinasi antara bulan pelaksanaan wawancara teknis dan bulan pelaksanaan wawancara final.

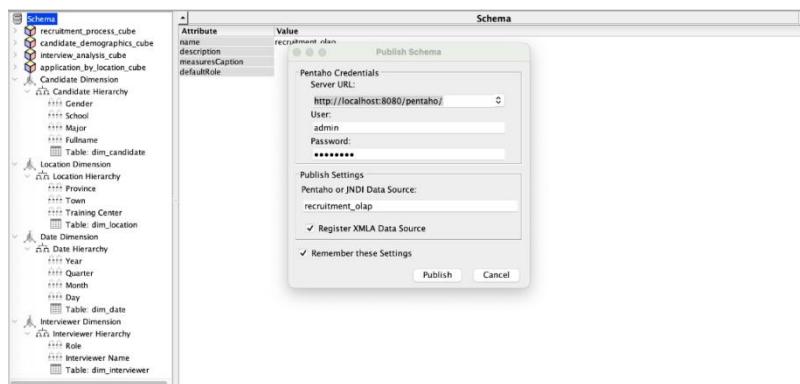
Gambar 23. *Cube Application by Location*

Gambar 23 merupakan proses pembuatan *Cube Application by Location* yang akan dimenghasilkan dashboard perbandingan jumlah aplikasi yang diterima berdasarkan pusat pelatihan.

Dalam skema OLAP, *shared dimension* adalah dimensi yang digunakan oleh lebih dari satu *Cube* untuk menjaga konsistensi dan menghindari redundansi definisi. Misalnya, *Candidate Dimension* dan *Date Dimension* digunakan bersama pada beberapa *Cube* seperti *recruitment\_process\_cube*, *candidate\_demographics\_cube*, dan *interview\_analysis\_cube*. Setiap *shared dimension* terdiri atas hierarki, level, dan atribut; contohnya, *Date Dimension* memiliki level Year, Month, dan Day, sedangkan *Candidate Dimension* mencakup Gender, School, dan Major. Struktur ini memungkinkan analisis data yang terintegrasi dan seragam di seluruh *Cube*. Proses perancangannya ditunjukkan pada Gambar 24 berikut.

Gambar 24. *Share Dimension*

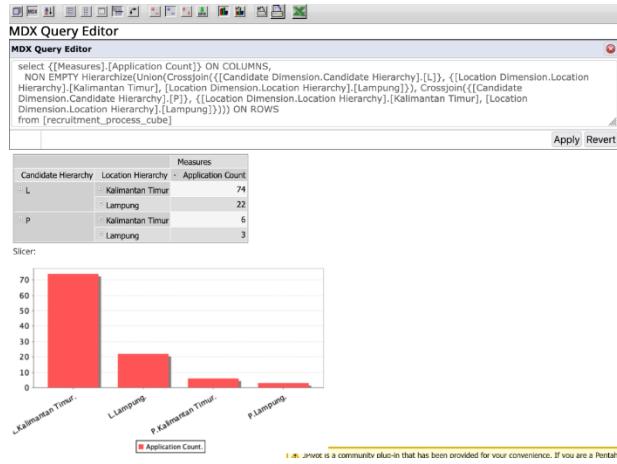
Proses *publish* dalam konteks *Schema Workbench* merujuk pada langkah penyimpanan dan pendaftaran schema OLAP agar dapat dikenali oleh server Pentaho. Tujuan utama dari proses ini adalah mengintegrasikan definisi *Cube* ke dalam lingkungan eksekusi, sehingga dapat diakses melalui Pentaho *User Console*, baik untuk eksplorasi OLAP di JPivot maupun visualisasi pada dashboard CDE. File schema disimpan dalam format .xmondrian.xml dan diletakkan pada direktori pentaho-solutions/system/olap/, dengan tambahan parameter koneksi JDBC di bagian akhir file. Setelah dilakukan refresh pada Mondrian *Schema Cache*, schema siap digunakan. Ilustrasi proses ini ditunjukkan pada Gambar 25 berikut.



Gambar 25. Proses Publish Cube

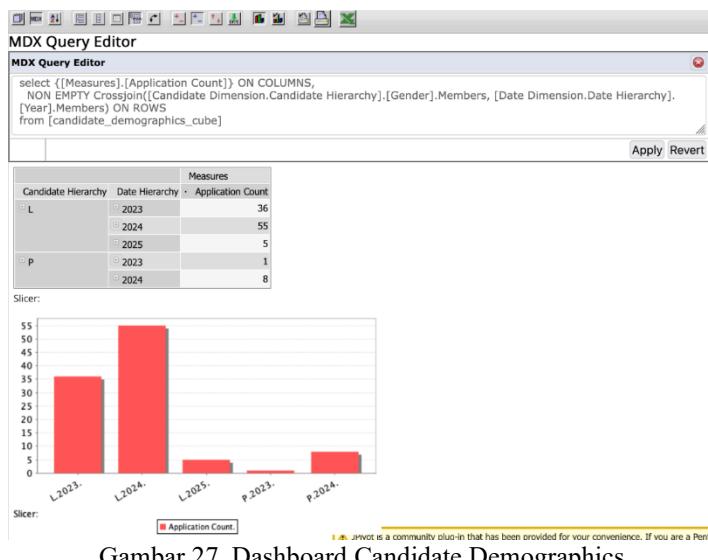
### 3.4. Analisa Dashboard Hasil Perancangan Data Warehouse

Berikut merupakan dashboard dari perancangan *Data Warehouse* berdasarkan dari masing-masing *Cube* yang telah dibuat pada pembahasan sebelumnya.



Gambar 26. Dashboard Proses Rekrutmen

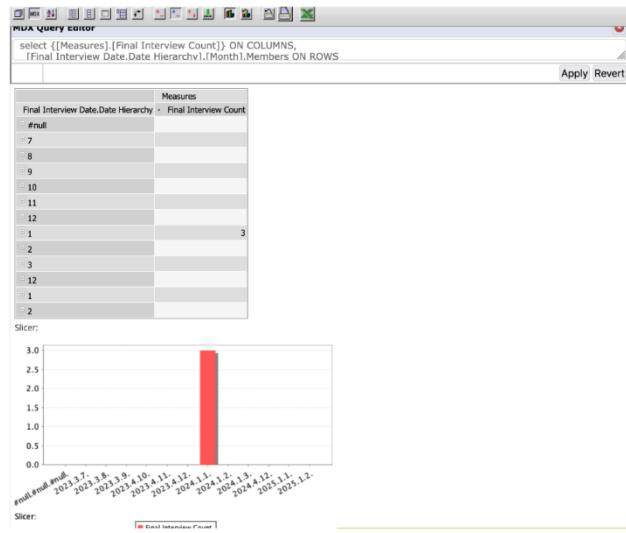
Gambar 26 menunjukkan distribusi jumlah pelamar berdasarkan kombinasi jenis kelamin (laki-laki dan perempuan) serta lokasi asal (Kalimantan Timur dan Lampung). Data diperoleh dari *recruitment\_process\_cube* menggunakan agregasi *Application Count*. Hasil analisis mengindikasikan bahwa pelamar laki-laki dari Kalimantan Timur merupakan kelompok dominan, dengan total 74 aplikasi. Pelamar laki-laki dari Lampung berada pada posisi berikutnya dengan 22 aplikasi. Sementara itu, jumlah pelamar perempuan relatif rendah, masing-masing hanya 6 dari Kalimantan Timur dan 3 dari Lampung. Temuan ini menunjukkan adanya ketimpangan distribusi pelamar berdasarkan gender dan wilayah, serta memberikan indikasi awal mengenai kecenderungan demografis pelamar yang dapat dipertimbangkan dalam pengambilan keputusan strategis dalam proses rekrutmen, misalnya apakah diperlukan adanya program lebih inklusif untuk perempuan dan promosi yang lebih aktif di wilayah yang kurang mendapat perhatian terutama Kalimantan Timur dan Lampung.



Gambar 27. Dashboard Candidate Demographics

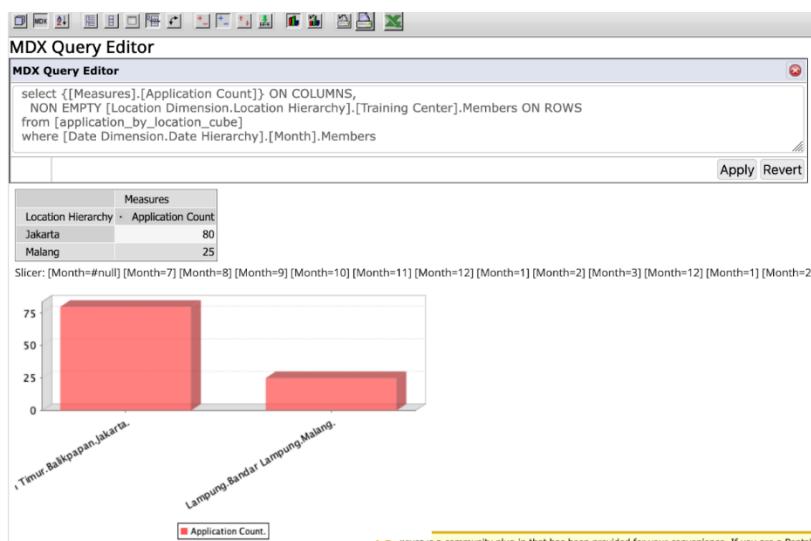
Gambar 27 menampilkan distribusi jumlah pelamar berdasarkan jenis kelamin dan tahun aplikasi, dengan data yang bersumber dari *candidate\_demographics\_cube*. Analisis ini menggunakan metrik *Application Count* yang diukur berdasarkan kombinasi antara dimensi Gender dan dimensi Tahun (Year). Hasil visualisasi menunjukkan bahwa pelamar laki-laki (L) mendominasi jumlah aplikasi pada seluruh periode, khususnya pada tahun 2024 dengan total 55 aplikasi, diikuti oleh tahun 2023 dengan 36 aplikasi, dan tahun 2025 sebanyak 5 aplikasi. Sementara itu, pelamar perempuan (P) tercatat hanya 1 aplikasi pada tahun 2023 dan 8 aplikasi pada

tahun 2024, serta tidak tercatat sama sekali pada tahun 2025. Temuan ini mengindikasikan adanya ketimpangan yang cukup signifikan dalam partisipasi pelamar berdasarkan gender, serta memberikan gambaran tren rekrutmen secara tahunan. Informasi ini dapat digunakan sebagai dasar dalam evaluasi efektivitas strategi rekrutmen yang lebih inklusif dan berimbang.



Gambar 28. Dashboard Interview Analysis

Gambar 28 menyajikan distribusi jumlah wawancara final berdasarkan kombinasi antara bulan pelaksanaan wawancara teknis dan bulan pelaksanaan wawancara final. Data ini diambil dari *interview\_analysis\_cube* menggunakan metrik *Final Interview Count*, serta memanfaatkan dua dimensi waktu yang berbeda: *Tech Interview Date* dan *Final Interview Date*. Visualisasi ini bertujuan untuk mengamati hubungan waktu antara dua tahap proses seleksi, khususnya untuk mengidentifikasi apakah pelaksanaan wawancara final umumnya dilakukan pada bulan yang sama dengan wawancara teknis, atau terdapat jeda waktu tertentu. Namun, karena hanya sebagian kandidat yang melanjutkan ke tahap final, data yang ditampilkan cenderung bersifat spars (jarang muncul), sehingga hasil analisis perlu ditafsirkan dengan hati-hati. Temuan ini dapat memberikan gambaran awal tentang efisiensi alur seleksi, serta membantu perencanaan jadwal proses rekrutmen yang lebih terstruktur dari sisi waktu.



Gambar 29. Dashboard Application by Location

Gambar 29 menampilkan perbandingan jumlah aplikasi yang diterima oleh dua pusat pelatihan utama, yaitu Jakarta dan Malang. Data diambil dari *application\_by\_location\_cube* dengan *Application Count* sebagai indikator kinerja, difokuskan pada periode rekrutmen aktif selama beberapa bulan. Hasil menunjukkan bahwa

Jakarta menerima 80 aplikasi, jauh lebih tinggi dibandingkan Malang yang hanya menerima 25 aplikasi. Penamaan label grafik mencerminkan hierarki lokasi geografis yang digunakan dalam struktur data. Temuan ini mengindikasikan bahwa Jakarta berfungsi sebagai pusat rekrutmen yang lebih produktif, kemungkinan karena faktor geografis, kepadatan populasi, dan efektivitas promosi. Sebaliknya, rendahnya angka di Malang menunjukkan perlunya evaluasi strategi untuk meningkatkan jangkauan dan daya tarik di wilayah tersebut.

#### 4. KESIMPULAN

Hasil dari proses perancangan *Data Warehouse* pada perusahaan Enigma Camp menunjukkan bahwa penerapan *Data Warehouse* dengan *Star Schema* dapat membantu merapikan dan mengintegrasikan data dalam format yang baik. *Data Warehouse* membuat data menjadi lebih terstruktur dan konsisten, sehingga memudahkan dalam menganalisis dan mengelola data. *Star Schema* dengan dua tabel fakta juga membantu perusahaan mempunyai data dan informasi lebih luas dan informatif sehingga lebih banyak *insight* dan dampak positif yang langsung bisa diambil dari hasil perancangan *Data Warehouse* yang sudah dilakukan. Seperti contohnya, ditemukan ketimpangan signifikan dalam distribusi pelamar berdasarkan gender dan wilayah. Pelamar laki-laki dari Kalimantan Timur mendominasi dengan 74 aplikasi, diikuti oleh laki-laki dari Lampung (22 aplikasi), sementara jumlah pelamar perempuan dari kedua wilayah tersebut jauh lebih rendah. Temuan ini mengindikasikan perlunya program rekrutmen yang lebih inklusif untuk perempuan dan promosi yang lebih aktif di wilayah yang kurang terwakili seperti Kalimantan Timur dan Lampung, guna mencapai keseimbangan demografis yang lebih baik.

Selain itu, penggunaan Mondrian dengan skema *workbench* meningkatkan kemampuan perusahaan dalam pengambilan keputusan berbasis data yang lebih cepat dan lebih tepat. Dimana diawali dengan proses *Extraction, Transformation, and Load (ETL)* pada tabel dimensi, perancangan ETL tabel fakta dengan *Star Schema*, pembuatan *Cube*, proses *Publish Cube*, dan menghasilkan dashboard dari masing-masing *Cube* yang dirancang sehingga menghasilkan visualisasi data yang lebih informatif.

Untuk pengembangan penelitian selanjutnya, disarankan untuk menggunakan tools *Data Warehouse* yang lebih canggih agar mempermudah dan mempercepat proses perancangan *Data Warehouse*. Seperti, pada proses ETL seharusnya bisa selesai dalam waktu yang lebih cepat. Contoh lainnya, dengan tools yang lebih canggih yaitu visualisasi dashboard bisa yang lebih menarik dan informatif dengan pilihan beberapa model *chart* dan *custom fields*. Selain menggunakan tools yang lebih canggih, disarankan menggunakan dataset yang lebih besar sehingga memperoleh hasil yang lebih valid dan relevan. Seperti, tahun penarikan data yang lebih panjang yaitu 4 sampai 5 tahun, data demografis pelamar dibagi perkota (contohnya provinsi Kalimantan Timur terdiri dari kota Samarinda, Balikpapan, Bontang dan Tarakan), dan platform rekrutmen external diperbanyak seperti menjalin kerjasama dengan perusahaan-perusahaan *headhunter* sehingga sumber datanya semakin banyak dan variatif.

#### DAFTAR PUSTAKA

- [1] A. F. Karami, ‘Perancangan Arsitektur Data Warehouse Pada Industri Perkebunan Kelapa Sawit’, *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 2, pp. 973–983, Jun. 2022, doi: 10.35957/jatisi.v9i2.1834.
- [2] D. Andriansyah, ‘Implementasi Extract-Transform-Load (ETL) Data Warehouse Laporan Harian Pool’, *Jurnal Teknik Informatika*, vol. 8, no. 2, pp. 45–49, Aug. 2022, doi: 10.51998/jti.v8i2.486.
- [3] N. S. Fitriasari, I. Ariawan, A. Rais, T. E. Ahmad, and D. R. Azhari, ‘Rancangan Dan Implementasi Modul Data Warehouse Dan Data Mining Sebagai Kritisik Sukses Faktor Pada Enterprise’, *Seminar Nasional Ilmu Komputer (SNASIKOM)*, vol. 1, no. 1, pp. 41–52, Jul. 2021, [Online]. Available: <https://proceeding.unived.ac.id/index.php/snasakiom/article/view/51>
- [4] M. F. Zulkarnaina, N. P. N. Ardiyantia, I. W. W. K. Sandia, I. Dewa, N. T. Hendrawana, and I. B. M. Mahendra, ‘Perancangan dan Implementasi Data Warehouse Penjualan (Studi Kasus: Northwind Sample Database)’, *Jurnal Elektronik Ilmu Komputer Udayana p-ISSN*, vol. 2301, p. 5373, 2021.
- [5] Q. Azizah and T. Atmojo, ‘Dirgamaya Jurnal Manajemen dan Sistem Informasi Perancangan Data Warehouse Sistem Penerimaan Siswa Baru Menggunakan Online Analytical Processing (OLAP) di TK IT Mutiara’.
- [6] A. M. Al-Rahmi, M. A. Al-Sharafi, and H. M. Al-Amrani, ‘Data Warehouse Design for Academic Performance Analysis Using Star Schema’, *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, pp. 1–8, 2022.

- 
- [7] M. Amin, A. Sutrisman, and Y. Dwitayanti, ‘Development of Star-Schema Model for Lecturer Performance in Research Activities’, *International Journal of Advanced Computer Science and Applications*, vol. 12, Jul. 2021, doi: 10.14569/IJACSA.2021.0120909.
  - [8] T. Setiadi and R. Rahutomo, *A Data Warehouse Schema for Monitoring Regional COVID-19 Case Registration*. 2022. doi: 10.1109/ICIMTech55957.2022.9915222.
  - [9] W. Wijaya, J. Wiratama, and S. F. Wijaya, ‘Implementation of Data Warehouse and Star Schema for Optimizing Property Business Decision Making’, *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 2, pp. 1242–1250, Apr. 2024, doi: 10.33379/gtech.v8i2.4091.
  - [10] D. Prastyo and A. Supriyanto, *ANALISA DAN PERANCANGAN DATA WAREHOUSE DENGAN METODE NINE STEP KIMBALL DI PT SURGANYA MOTOR INDONESIA*.
  - [11] R. M. Winarto and R. E. Sianipar, ‘Perancangan Data Warehouse Penjualan Produk Menggunakan Pentaho Data Integration (PDI) (Studi Kasus: PT. XYZ)’, *Jurnal Sistem Informasi Bisnis (JURSIBI)*, vol. 1, no. 2, pp. 1–10, 2022.
  - [12] N. Adila and A. Andri, ‘Desain Dan Implementasi Data Warehouse Pada Perpustakaan Daerah Provinsi Sumatera Selatan’, *Jurnal Nasional Ilmu Komputer*, vol. 2, no. 1, pp. 33–50, 2021.
  - [13] Purwita Sari, Lucky Indra Kesuma, Mira Afrina, and Dedy Kurniawan, ‘Pemodelan Integrasi Data Barang Milik Negara di Perguruan Tinggi Menggunakan Metode ETL (Extract, Transform, Load) dengan Pentaho’, *The Indonesian Journal of Computer Science*, vol. 13, no. 5, Oct. 2024, doi: 10.33022/ijcs.v13i5.4424.
  - [14] A. Alessandro, D. Anggoro, and A. Wijaya, ‘Perancangan Data Warehouse Umat Lingkungan Gereja XYZ Menggunakan Metode Star Schema’, *Journal Of Informatics And Busisnes*, vol. 1, no. 3, pp. 173–182, 2023.
  - [15] D. Tan, J. Wiratama, and S. Fernandi Wijaya, ‘Sales Analysis on Garment Industry with Datawarehouse and ETL Implementation on Star Schema’, *Indonesian Journal of Computer Science*, vol. 13, no. 1, Feb. 2024, doi: 10.33022/ijcs.v13i1.3770.