

**LAPORAN TUGAS MISI KE-3
MATA KULIAH PERGUDANGAN DATA
KELAS RB**



Disusun Oleh:

- | | |
|----------------------------------|-----------|
| 1. Ibnu Farhan Al-Ghifari | 121450121 |
| 2. Raid Muhammad Naufal | 122450027 |
| 3. Berliana Enda Putri | 122450065 |
| 4. Danang Hilal Kurniawan | 122450085 |
| 5. Akmal Faiz Abdillah | 122450114 |
| 6. Fayyaza Aqila Syafitri Achjar | 122450131 |

**PROGRAM STUDI SAINS DATA
FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA
2025**

1. Latar Belakang

Dalam era digital saat ini, data menjadi aset yang sangat bernilai bagi suatu organisasi, khususnya bagi perusahaan pelayaran yang menghadapi tantangan besar dalam mengelola dan memanfaatkan data yang berasal dari berbagai sumber, seperti kapal, pelabuhan, cuaca, dan pasar global. Penyimpanan serta pengolahan data secara terstruktur sangat diperlukan untuk mengatasi permasalahan yang dihadapi perusahaan. Oleh karena itu, diperlukan adanya *data warehouse* sebagai sistem analitik yang mampu mengintegrasikan data historis dari berbagai sumber ke dalam satu sistem yang konsisten dan terorganisir.

Melalui perancangan data warehouse ini, perusahaan dapat mengoptimalkan kecepatan pelayaran, efisiensi penggunaan bahan bakar, perawatan berbasis data, serta meningkatkan transparansi dan daya saing. Tujuan dari perancangan ini adalah untuk membangun sistem penyimpanan data terpusat yang mampu mengintegrasikan data internal dan eksternal perusahaan pelayaran secara real-time, serta mendukung proses analitik untuk pengambilan keputusan strategis. Sistem ini diharapkan mampu meningkatkan efisiensi operasional, keselamatan pelayaran, serta memperkuat daya saing industri maritim secara berkelanjutan.

2. Desain Konseptual

Model konseptual gudang data sistem ini dikembangkan berdasarkan aktivitas operasional kapal, termasuk pengukuran performa teknis, kondisi navigasi, informasi cuaca, posisi kapal, dan waktu operasi. Data ini sangat penting dalam pengambilan keputusan strategis seperti perencanaan rute optimal, efisiensi bahan bakar, pemantauan performa mesin, dan evaluasi risiko cuaca.

2.1 Konsep Dimensi dan Fakta

2.1.1 Tabel Fakta

Tabel fakta merupakan inti dari *data warehouse* yang menyimpan data numerik dari peristiwa-peristiwa operasional kapal. Pada sistem ini, tabel fakta diberi nama Fact_Ship_Data, yang mencatat aktivitas aktual kapal dan mengacu ke berbagai tabel dimensi sebagai konteks informasional. Struktur tabel fakta ditampilkan pada Tabel 2.1. berikut.

Tabel 2.1. Tabel Fakta

Atribut	Tipe data	Deskripsi
Ship_id	INT	ID unik kapal
datetime_id	INT	Kunci waktu
navigation_id	INT	Kunci navigasi
position_id	INT	Kunci lokasi

route_id	INT	Kunci rute
weather_id	INT	Kunci cuaca
sensor_id	INT	Kunci sensor teknis
ship_speed	FLOAT	Kecepatan kapal (knot)
fuel_consumption	FLOAT	Konsumsi bahan bakar (liter/jam)
engine_temperature	FLOAT	Suhu mesin utama
navigation_status	STRING	Status navigasi (aktif/siaga)
estimated_time	DATETIME	Perkiraan waktu kedatangan (ETA)

2.1.2 Tabel Dimensi

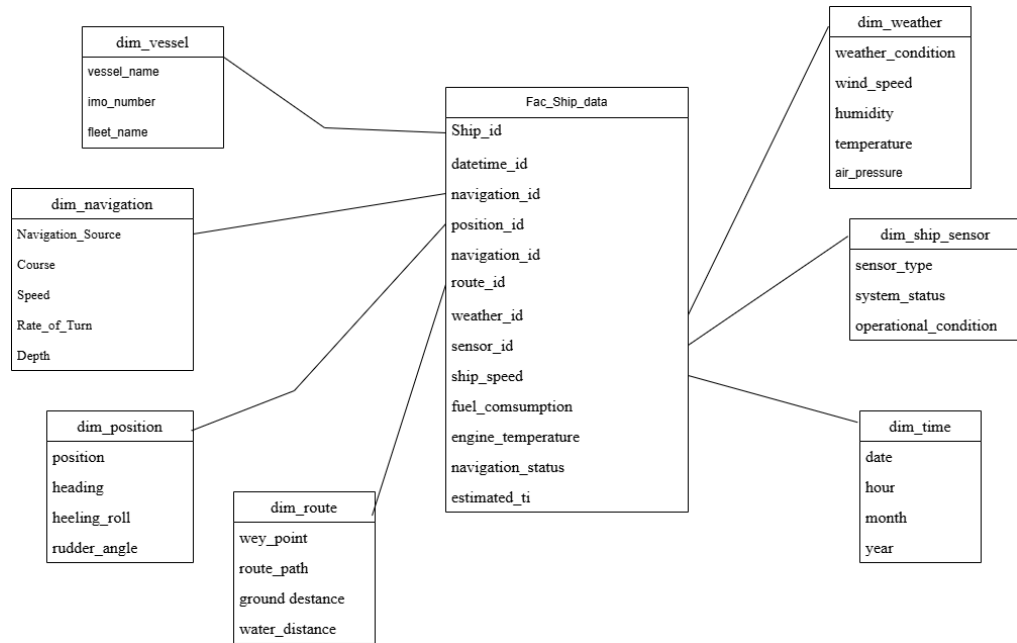
Tabel 2.2. Tabel Dimensi

Dimensi	Atribut	Hirarki
dim_navigation	<i>direction, navigation_mode, sea_condition</i>	Mode→Arah→Kondisi laut
dim_position	<i>latitude, longitude, port_origin, port_dest</i>	Pelabuhan→Kota→Negara
dim_route	<i>route_name, route_type, route_level</i>	Rute→Tipe→Kompleksitas
dim_weather	<i>wind_speed, temperature, rainfall, wave</i>	Kondisi cuaca→Risiko navigasi
dim_ship_sensor	<i>engine_temp, fuel_flow, vibration_level</i>	Sensor→Mesin→Status kinerja
dim_time	<i>date, day, month, quarter, year, season</i>	Hari→Bulan→Tahun→Musim

2.2 Skema Bintang

Desain arsitektur gudang data pada sistem ini menerapkan pendekatan skema bintang (*star schema*), yang merupakan representasi klasik dari model multidimensional. Dalam skema ini, tabel fakta berada di pusat, dikelilingi oleh tabel-tabel dimensi yang saling terhubung melalui kunci asing (*foreign key*). Skema bintang yang digunakan memungkinkan integrasi informasi operasional, teknis, temporal, dan geografis secara menyeluruh. Hal ini mendukung proses analitik untuk:

1. Mengamati performa kapal berdasarkan rute dan waktu
2. Menganalisis pengaruh kondisi cuaca terhadap konsumsi bahan bakar
3. Mengevaluasi status navigasi dan estimasi waktu kedatangan
4. Mendeteksi potensi masalah teknis melalui sensor mesin



Gambar 2. Star Schema System

2.3 Perancangan dan Implementasi Tabel Dimensi

2.2.1 Identifikasi tabel-tabel dimensi.

Tabel dimensi digunakan untuk memberikan konteks atau deskripsi lebih lanjut mengenai data yang ada pada tabel fakta, yang berisi informasi numerik atau transaksi operasional. Setiap tabel dimensi berfungsi untuk menjelaskan data dalam tabel fakta dari berbagai perspektif yang relevan dengan operasi kapal. Masing-masing tabel dimensi memuat atribut-atribut deskriptif yang memungkinkan pengguna untuk melakukan analisis berdasarkan faktor-faktor seperti navigasi, posisi kapal, rute yang ditempuh, kondisi cuaca, dan informasi teknis kapal, serta kapan itu sendiri. Tabel dimensi yang digunakan dalam sistem ini meliputi:

1. **dim_navigation:** Menyediakan informasi terkait navigasi kapal, termasuk arah, mode navigasi, dan kondisi laut yang dapat digunakan untuk menganalisis bagaimana faktor-faktor ini mempengaruhi performa operasional kapal.
2. **dim_position:** Menyediakan data posisi kapal, seperti koordinat geografi, pelabuhan asal dan tujuan, yang memungkinkan analisis berdasarkan lokasi kapal.
3. **dim_route:** Menyediakan informasi tentang rute kapal, seperti nama rute, tipe rute, dan tingkat kompleksitasnya, yang memungkinkan analisis tentang seberapa efektif dan efisien rute yang dipilih oleh kapal.
4. **dim_weather:** Menyediakan data cuaca seperti kecepatan angin, suhu, curah hujan, dan ketinggian gelombang, yang dapat digunakan untuk

menganalisis bagaimana kondisi cuaca mempengaruhi performa kapal dan efisiensi bahan bakar.

5. **dim_ship_sensor**: Menyediakan informasi terkait sensor-sensor teknis kapal, seperti suhu mesin, aliran bahan bakar, dan tingkat getaran mesin, yang memberikan gambaran mengenai kondisi teknis kapal yang dapat mempengaruhi keseluruhan operasionalnya.
6. **dim_time**: Menyediakan informasi terkait waktu operasi kapal, termasuk tanggal, hari, bulan, kuartal, tahun, dan musim, yang memungkinkan analisis trend dan performa kapal berdasarkan periode waktu tertentu.
7. **dim_vessel** : Menyediakan informasi terkait kapal itu sendiri, seperti nama kapal, tipe kapal, dan kapasitas kapal. Tabel ini membantu dalam analisis performa berdasarkan jenis dan kapasitas kapal yang digunakan.

2.2.2 Atribut-atribut penting di setiap dimensi.

Setiap tabel dimensi memiliki atribut-atribut yang memberikan deskripsi lebih lanjut mengenai data yang ada pada tabel fakta. Atribut-atribut ini memungkinkan pengguna untuk melakukan analisis mendalam dengan mempertimbangkan berbagai perspektif terkait operasi kapal. Berikut adalah atribut-atribut penting yang terdapat di setiap dimensi:

1. **dim_navigation**:
 - a. **direction**: Arah kapal (misalnya: utara, selatan, timur, barat).
 - b. **navigation_mode**: Mode navigasi kapal.
 - c. **sea_condition**: Kondisi laut
2. **dim_position** :
 - a. **latitude**: Koordinat lintang kapal.
 - b. **longitude**: Koordinat bujur kapal.
 - c. **port_origin**: Pelabuhan asal kapal.
 - d. **port_dest**: Pelabuhan tujuan kapal.
3. **dim_route** :
 - a. **route_name**: Nama rute kapal.
 - b. **route_type**: Jenis rute kapal.
 - c. **route_level**: Tingkat kompleksitas rute
4. **dim_weather**:
 - a. **wind_speed**: Kecepatan angin.
 - b. **temperature**: Suhu udara.
 - c. **rainfall**: Curah hujan.
 - d. **wave**: Ketinggian gelombang laut.
5. **dim_ship_sensor**:
 - a. **engine_temp**: Suhu mesin utama kapal.
 - b. **fuel_flow**: Aliran bahan bakar.
 - c. **vibration_level**: Tingkat getaran mesin.

6. dim_time
 - a. date: Tanggal operasi.
 - b. day: Hari dalam seminggu.
 - c. month: Bulan dalam tahun.
 - d. quarter: Kuartal dalam tahun.
 - e. year: Tahun operasi.
 - f. season: Musim.
7. dim_vessel
 - a. vessel_name: Nama kapal.
 - b. vessel_type: Tipe kapal.
 - c. vessel_capacity: Kapasitas kapal.

2.2.3 Penetapan surrogate key untuk efisiensi.

Surrogate Key adalah kunci buatan yang digunakan untuk mengidentifikasi setiap baris dalam tabel dimensi secara unik, yang lebih efisien dibandingkan dengan menggunakan kunci alami atau data yang ada di dunia nyata (seperti nama atau alamat). *Surrogate key* bersifat numerik dan tidak mengandung informasi bisnis yang bisa berubah, sehingga lebih stabil dan efisien dalam penyimpanan dan pengolahan data.

Tabel 2.3. Penetapan *Surrogate Key*

Tabel dimensi	<i>Surrogate Key</i>
im_navigation	navigation_id
dim_position	position_id
dim_route	route_id
dim_weather	weather_id
dim_ship_sensor	sensor_id
dim_vessel	vessel_id
dim_time	datetime_id

2.2.4 Perancangan dan Implementasi Tabel Fakta

1. Identifikasi Proses Bisnis Utama untuk Tabel Fakta

Tabel fakta Fact_Ship_Data dirancang untuk merekam aktivitas operasional kapal dalam satuan waktu tertentu. Proses bisnis utama yang direpresentasikan antara lain:

- Pemantauan kecepatan kapal dan konsumsi bahan bakar
- Pembacaan suhu mesin utama kapal
- Pencatatan status navigasi dan estimasi waktu kedatangan (ETA)
- Rekaman kondisi lingkungan dan teknis selama pelayaran

2. Penentuan Ukuran/Metrics dalam Tabel Fakta

Ukuran (metrics) utama dalam tabel fakta bersifat numerik dan dapat digunakan untuk analisis performa kapal, efisiensi bahan bakar, dan prediksi ETA. Atribut-atribut metrik tersebut meliputi:

- ship_speed (FLOAT) – Kecepatan kapal (knot)
- fuel_consumption (FLOAT) – Konsumsi bahan bakar (liter per jam)
- engine_temperature (FLOAT) – Suhu mesin utama
- estimated_time (DATETIME) – Estimasi waktu tiba (ETA)

Selain itu, atribut navigation_status (STRING) dapat dianalisis dalam konteks status operasional kapal (aktif, siaga, idle, dll).

3. Desain Logikal dan Fisik

Tahap ini menerjemahkan desain konseptual ke dalam model relasional/logikal dan kemudian direalisasikan menjadi struktur fisik di dalam sistem manajemen basis data (DBMS). Fokus utama adalah memastikan efisiensi akses data, integritas relasi antar-tabel, dan performa eksekusi query melalui penerapan teknik indexing yang tepat.

3.1 Perancangan dan Implementasi Index

Index dalam data warehouse memiliki peran penting dalam mempercepat proses eksekusi query, terutama dalam operasi JOIN, FILTER, dan AGGREGATE. Berikut adalah strategi penerapan index berdasarkan karakteristik data dalam sistem operasional kapal. Jenis-jenis indeks yang digunakan:

3.1.1 Clustered index

Digunakan pada kolom datetime_id di tabel fakta, yang sering menjadi dasar pengurutan data berdasarkan waktu. Karena clustered index menentukan urutan fisik penyimpanan data dalam tabel, ini cocok untuk analisis deret waktu (*time-series*).

```
CREATE CLUSTERED INDEX idx_fact_datetime  
ON Fact_Ship_Data (datetime_id);
```

3.1.2 Non-Clustered Index

Diterapkan pada kolom-kolom yang sering digunakan dalam filter, WHERE clause, atau JOIN, seperti: route_id (untuk filter berdasarkan rute) weather_id (untuk analisis kondisi cuaca) navigation_status (untuk status operasional kapal)

```
CREATE NONCLUSTERED INDEX idx_fact_route  
ON Fact_Ship_Data (route_id);
```

```
CREATE NONCLUSTERED INDEX idx_fact_weather
ON Fact_Ship_Data (weather_id);

CREATE NONCLUSTERED INDEX idx_fact_navstatus
ON Fact_Ship_Data (navigation_status);
```

3.3.3 Penggunaan Index dalam Analisis

Query yang memanfaatkan index yang telah dibuat untuk analisis performa kapal berdasarkan waktu dan rute:

```
SELECT
    t.quarter,
    r.route_name,
    AVG(f.ship_speed) AS avg_speed,
    AVG(f.fuel_consumption) AS avg_fuel
FROM Fact_Ship_Data f
JOIN dim_time t ON f.datetime_id = t.datetime_id
JOIN dim_route r ON f.route_id = r.route_id
WHERE r.route_type = 'Utama'
    AND t.year = 2024
GROUP BY t.quarter, r.route_name;
```

3.2 Strategi Desain Penyimpanan

Pada rancangan data warehouse yang mendukung aktivitas operasional kapal, penyimpanan data dirancang untuk mengoptimalkan efisiensi ruang, performa query, dan kemudahan pengelolaan data besar yang terus bertambah setiap hari. Pendekatan penyimpanan mengikuti Medallion Architecture, yang membagi data ke dalam tiga zona: Bronze, Silver, dan Gold, dengan masing-masing zona mencerminkan tingkat kualitas dan kesiapan data untuk analisis. Strategi Penyimpanan yang Diterapkan

1. Bronze Zone (Raw Data Zone)

Data mentah yang berasal dari sistem operasional kapal, seperti log pergerakan, status kapal, dan sensor pengukuran, disimpan di tabel staging (Stg_Ship_Activity). Tabel ini tidak diberi indeks maupun kompresi, agar proses ingest data dari sistem operasional dapat berjalan dengan cepat dan tanpa hambatan. Data disimpan dalam filegroup khusus bernama FG_Staging untuk memisahkan data mentah dengan data yang sudah diolah.

```
ALTER DATABASE ShipOps ADD FILEGROUP FG_Staging;
ALTER DATABASE ShipOps ADD FILE (
    NAME = 'FG_Staging_Data',
    FILENAME = 'D:\SQLData\FG_Staging.ndf'
) TO FILEGROUP FG_Staging;
```



```
CREATE TABLE Stg_Ship_Activity (
    ShipID INT,
    ActivityTimestamp DATETIME,
    Location VARCHAR(100),
    Status VARCHAR(50),
    SensorData XML
) ON FG_Staging;
```

2. Silver Zone (Clean and Integrated Data Zone)

Setelah proses ETL untuk membersihkan dan mengintegrasikan data, data dipindahkan ke tabel dimensi dan fakta utama. Tabel dimensi seperti Dim_Ship, Dim_Port, dan Dim_Time disimpan pada filegroup FG_Dimension, sedangkan tabel fakta Fact_Ship_Operation yang memuat data operasional lengkap disimpan pada filegroup FG_Fact. Pada tabel fakta, dilakukan penerapan kompresi PAGE untuk menghemat ruang penyimpanan dan mempercepat pembacaan data, mengingat volume data operasional kapal yang besar.

```
ALTER DATABASE ShipOps ADD FILEGROUP FG_Dimension;
ALTER DATABASE ShipOps ADD FILEGROUP FG_Fact;

ALTER DATABASE ShipOps ADD FILE (
    NAME = 'FG_Dimension_Data',
    FILENAME = 'D:\SQLData\FG_Dimension.ndf'
) TO FILEGROUP FG_Dimension;

ALTER DATABASE ShipOps ADD FILE (
    NAME = 'FG_Fact_Data',
    FILENAME = 'D:\SQLData\FG_Fact.ndf'
) TO FILEGROUP FG_Fact;

ALTER TABLE Fact_Ship_Operation REBUILD WITH
(DATA_COMPRESSION = PAGE);
```

Tabel fakta dipartisi berdasarkan kolom OperationYear untuk mempercepat *query* berbasis tahun dan memudahkan manajemen data historis.

```
CREATE PARTITION FUNCTION pf_OperationYear (INT)
AS RANGE RIGHT FOR VALUES (2021, 2022, 2023, 2024);

CREATE PARTITION SCHEME ps_OperationYear
AS PARTITION pf_OperationYear TO (FG_Fact, FG_Fact,
FG_Fact, FG_Fact);

CREATE CLUSTERED INDEX idx_OperationDate
ON Fact_Ship_Operation (OperationDate)
```

```
ON ps_OperationYear (OperationYear);
```

3. Gold Zone (Aggregated and Semantic Data Zone)

Pada zona ini, data sudah berbentuk agregasi untuk analisis cepat, misalnya performa kapal per rute, konsumsi bahan bakar per bulan, dan sebagainya. Data hasil agregasi disimpan dalam bentuk *indexed views* yang dioptimasi dengan *clustered index* untuk mempercepat akses *query* analitik.

```
CREATE VIEW vw_ShipPerformanceSummary
WITH SCHEMABINDING
AS
SELECT
    s.ShipID,
    r.RouteID,
    YEAR(f.OperationDate) AS OperationYear,
    COUNT_BIG(*) AS TotalOperations,
    AVG(f.FuelConsumption) AS AvgFuelConsumption
FROM dbo.Fact_Ship_Operation f
JOIN dbo.Dim_Ship s ON f.ShipID = s.ShipID
JOIN dbo.Dim_Route r ON f.RouteID = r.RouteID
GROUP BY s.ShipID, r.RouteID, YEAR(f.OperationDate);

CREATE UNIQUE CLUSTERED INDEX
idx_vw_ShipPerformanceSummary
ON vw_ShipPerformanceSummary (ShipID, RouteID,
OperationYear);
```

Pemeliharaan Penyimpanan

Untuk menjaga performa dan kapasitas penyimpanan tetap optimal, dilakukan monitoring berkala pada pertumbuhan *filegroup* dan aktivitas partisi. *Autogrowth* di setting proporsional, dan pemeliharaan indeks serta kompresi dilakukan secara rutin. Selain itu, setiap awal tahun dilakukan *split partition* untuk memasukkan data tahun baru, serta *merge partition* untuk arsip data lama agar *query* tetap efisien.

3.3 Perancangan dan Implementasi Partisi Tabel dan View

3.3.1 Jenis partisi:

- Partisi horizontal digunakan untuk memecah tabel menjadi baris-baris berdasarkan nilai tertentu dalam kolom, seperti waktu, wilayah, dan kategori. Penggunaan tabel ini diterapkan pada tabel transaksi yang dibagi berdasarkan tahun partisi yaitu 2022, 2023, dan 2024

Contoh Penerapan:

```
CREATE TABLE Fact_Ship_data_2022 PARTITION OF
```

```
Fact_Ship_data
    FOR VALUES FROM ('2022-01-01') TO
    ('2023-01-01');
CREATE TABLE Fact_Ship_data_2023 PARTITION OF
Fact_Ship_data
    FOR VALUES FROM ('2023-01-01') TO
    ('2024-01-01');
CREATE TABLE Fact_Ship_data_2024 PARTITION OF
Fact_Ship_data
    FOR VALUES FROM ('2024-01-01') TO
    ('2025-01-01');
```

- Memastikan logikal digunakan untuk pemisahan secara logis dengan menggunakan view untuk merepresentasikan subset data tanpa memecah tabel fisik. Penerapan partisi dilakukan pada view untuk data wilayah.

Contoh Penerapan:

```
CREATE VIEW vw_ship_data_region_east AS
SELECT * FROM Fact_Ship_data
WHERE position_id IN (SELECT position_id FROM
dim_position WHERE region = 'East');
```

3.3.2. Manfaat partisi

Manfaat partisi yaitu sebagai berikut :

- Percepatan Query yaitu sistem hanya membaca partisi yang relevan sehingga mempercepat eksekusi query
- Data historis dapat dikelola tanpa menggunakan data aktif
- Mengurangi beban pada sistem karena data lebih tersegmentasi

3.3.3. Perancangan Indexed Views

Indexed View adalah hasil query yang disimpan dan di indeks di database untuk mempercepat akses data dengan tujuan mempercepat analisis agregat. Manfaatnya membuat query lebih cepat dan beban sistem berkurang. Indexed View digunakan saat membutuhkan akses cepat ke data agregat yang tidak sering berubah. Indexed View digunakan dalam OLAP

4. Kesimpulan

Perancangan sistem data warehouse untuk perusahaan pelayaran yang diuraikan dalam laporan ini mencakup berbagai aspek penting untuk meningkatkan efisiensi operasional dan pengambilan keputusan strategis. Sistem ini menggunakan skema bintang (star schema) yang terdiri dari tabel fakta dan dimensi untuk menganalisis

berbagai faktor operasional kapal seperti navigasi, posisi, cuaca, dan performa teknis kapal.

Pada desain konseptual, tabel-tabel dimensi yang telah diidentifikasi (seperti `dim_navigation`, `dim_position`, `dim_route`, `dim_weather`, `dim_ship_sensor`, `dim_time`, dan `dim_vessel`) berfungsi untuk memberikan konteks yang mendalam mengenai data yang ada di dalam tabel fakta. Dengan penetapan surrogate key pada setiap tabel dimensi, sistem ini menjadi lebih efisien dalam hal penyimpanan dan pemrosesan data.

Di tahap desain logikal dan fisik, penerapan indeks dan strategi penyimpanan yang tepat menjadi sangat penting untuk memastikan efisiensi akses data, integritas relasi antar tabel, dan performa eksekusi query. Sistem ini juga memperkenalkan pendekatan Medallion Architecture untuk penyimpanan data dalam tiga zona: Bronze, Silver, dan Gold, yang masing-masing memiliki peran penting dalam mengelola data berdasarkan kualitas dan kesiapan untuk analisis.

Secara keseluruhan, desain data warehouse ini diharapkan dapat meningkatkan transparansi, efisiensi bahan bakar, dan kecepatan pelayaran, serta memfasilitasi pengambilan keputusan berbasis data yang lebih baik. Pengembangan lebih lanjut dan optimalisasi sistem ini sangat diperlukan untuk menyesuaikan dengan kebutuhan perusahaan pelayaran yang terus berkembang.