LAPORAN PERANCANGAN DATA WAREHOUSE MISI 2 Taktika: Where Data Meets Strategy



Disusun Oleh Kelompok 23 RC

Muhammad Kaisar Firdaus	121450135
Haikal Dwi Syaputra	122450067
Aditya Rahman	122450113
Khaalishah Zuhrah A. V.	122450034
Rafa Aqilla Jungjunan	122450142

PROGRAM STUDI SAINS DATA FAKULTAS SAINS INSTITUT TEKNOLOGI SUMATERA LAMPUNG SELATAN 2025

1. Ringkasan Kebutuhan dari Misi

1.1 Ringkasan Tabel Dimensi, Hirarki, dan Ukuran

Jenis	Nama Tabel	Hirarki/Detail Atribut
Fakta	Fakta Pertandingan	Jumlah Gol, Assist, Tembakan, Akurasi, Intersep, Tackle, Rating, Penguasaan Bola (%)
Dimensi	Dim_Waktu	Tanggal → Pekan → Musim
	Dim_Pemain	Nama → Posisi → Usia → Tinggi → Kemampuan Individu
	Dim_Klub	Nama Klub \rightarrow Formasi \rightarrow Pelatih \rightarrow Liga
	Dim_Kompetisi	Nama Liga → Negara → Tahun Kompetisi
	Dim_Lokasi	Stadion → Kota → Status Home/Away
	Dim_Formasi	Formasi → Strategi Taktikal

1.2 Deskripsi Naratif Kebutuhan Bisnis

Masalah bisnis: Banyak klub kesulitan mengakses data taktis karena data tersebar, tidak real-time, dan tidak sesuai kebutuhan pelatih. DW diperlukan agar semua data terintegrasi dalam satu sistem yang cepat dan analitis.

1.3 Metadata Bisnis

Atribut	Definisi	Pengguna Utama	Sumber Data (CSV)	Catatan
home_team_goal	Jumlah gol oleh tim tuan rumah	Pelatih, Analis Klub	Match.csv	_
away_team_goal	Jumlah gol oleh tim tandang	Pelatih, Media	Match.csv	_
buildUpPlaySpe ed	Kecepatan serangan (lambat/cepat)	Analis Klub, Pelatih	Team.csv	Diganti jadi: strategi_serangan
player_api_id	ID unik pemain	Semua stakeholder teknis	Player.csv	_
date	Tanggal pertandingan	Semua stakeholder	Match.csv	_
possession_hom e	Penguasaan bola tim tuan rumah (%)	Analis Klub, Pelatih	Match.csv	_

league_name	Nama kompetisi	Analis Klub, Media	League.csv	Nama jelas
player_attributes	Gabungan kemampuan individu pemain	Rekruter, Analis Klub	Player.csv	Jika tidak digunakan → dicatat
formation_used	Formasi yang digunakan tim	Pelatih, Analis Klub	Team.csv/Matc h.csv	Nama diganti dari formation_classic

1.4 Inti Pendekatan Business-Driven

Pendekatan yang digunakan berfokus pada kebutuhan pelatih, analis, media, dan fans.

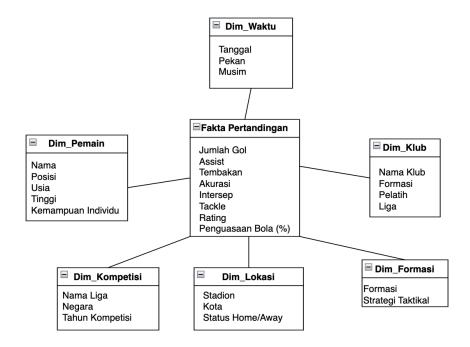
1.5 Dokumentasi Skema Sumber dan Penyesuaian

Sumber Data & File: Data diperoleh dari file seperti Match.csv, Player.csv, Team.csv, dan League.csv, bersumber dari dataset Kaggle yang di-update secara mingguan hingga triwulanan.

Penyesuaian Atribut : Atribut teknis diubah namanya menjadi istilah bisnis agar mudah dipahami (misalnya buildUpPlaySpeed → strategi_serangan).

Pendekatan yang digunakan adalah *business-driven*, yaitu fokus awal pada kebutuhan pengguna bisnis. Atribut yang tidak digunakan tetap dicatat.

2. Skema



Penjelasan Skema Bintang: Skema bintang (Star Schema) adalah model data dimensional yang umum digunakan dalam data warehouse untuk memfasilitasi analisis cepat dan efisien. Dalam skema ini:

- Tabel Fakta berada di tengah dan berisi data kuantitatif seperti gol, assist, akurasi, dan lain-lain.
- Tabel Dimensi mengelilingi tabel fakta dan menyediakan konteks analisis, seperti siapa pemainnya, kapan pertandingan berlangsung, di mana lokasinya, dan dalam kompetisi apa.

Setiap tabel dimensi terhubung langsung ke tabel fakta melalui kunci primer/asing (primary/foreign key). Bentuk "bintang" muncul dari struktur ini: fakta di tengah, dan dimensi menyebar keluar.

Keunggulan dari skema ini adalah:

- Query analisis menjadi lebih cepat
- Struktur mudah dipahami oleh analis bisnis dan pelatih
- Mendukung visualisasi dan pelaporan yang fleksibel

3. Penjelasan Tiap Komponen

Pendekatan multidimensional untuk membangun *conceptual schema data warehouse* mengorganisasi data menjadi dua jenis utama yakni fakta (facts) dan dimensi (dimensions) serta hierarki yang ada tiap dimensi. Berikut ini penjelasan tiap komponennya:

A. Fakta Pertandingan

Fakta disini adalah tabel yang berisi data numerik atau metrik yang merepresentasikan kejadian nyata yakni pertandingan sepakbola yang terdiri dari : jumlah_gol, penguasaan _bola (Persentase kendali atas bola), jumlah_operan, tembakan (ke arah gawang maupun tidak), assist (ada gol yang tidak ada assistnya) dan clean_sheet (Indikator tim dalam menjaga gawangnya tidak kebobolan). Dengan ini fakta memberikan fungsi sebagai inti dari data warehouse, menyimpan informasi kuantitatif yang dapat digunakan untuk analisis. Setiap baris dalam tabel fakta mewakili satu pertandingan. Foreign Keys yang menghubungkan tabel fakta ke masing-masing dimensi yakni : tanggal (Dim_Waktu), nama_klub (Dim_Klub), stadion (Dim_Lokasi), nama (Dim_lokasi) dan nama_liga (Dim_Kompetisi)

B. Dimensi

Dimensi disini adalah tabel yang menyediakan konteks kepada fakta. Mereka terdiri dari atribut deskriptif non-numerik pada data yang terdiri dari :

a. Dim Waktu

Dimensi ini menyediakan konteks waktu mulainya setiap pertandingan untuk analisis tren seiring waktu, seperti perubahan performa tim atau pemain selama musim berjalan dengan hierarki berikut ini:: tanggal (Data format datetime), musim (misal: "2023-2024"),pekan_ke-(misal: jika ada 20 tim maka pertandingan sebanyak 38 pekan), hari pertandingan (biasanya akhir pekan sabtu-minggu).

b. Dim Klub

Dimensi ini menyediakan informasi tentang klub sepak bola yang terlibat dalam membantu melacak performa klub secara keseluruhan, termasuk faktor-faktor seperti pengaruh pelatih terhadap tim dengan hierarki berikut ini : nama_klub (resmi), pelatih (utama), ranking (taip kompetisi yang dijalani)

c. Dim Lokasi

Dimensi ini menyediakan informasi lokasi pertandingan yang dapat membantu menganalisis faktor-faktor seperti keuntungan kandang, kondisi lapangan, atau preferensi penonton dengan hierarki berikut ini : stadion, kota (Kota tempat stadion/klub) dan status (pertandingan tandang atau kandang).

d. Dim_Pemain

Dimensi ini menyediakan informasi tentang pemain yang terlibat untuk menganalisis kontribusi individu pemain terhadap performa tim, serta faktor-faktor seperti usia atau tinggi badan yang mungkin mempengaruhi performa dengan hierarki berikut ini : nama, posisi (misal: gelandang tengah (CM), penyerang(ST)), usia, tinggi, dan kemampuan (misal: rating pertandingan (skala 0-10))

e. Dim_Kompetisi

Dimensi ini menyediakan konteks kompetitif di mana pertandingan dilangsungkan untuk membantu memahami dinamika kompetisi, seperti tingkat kesulitan lawan atau dampak kompetisi internasional terhadap performa klub dengan hierarki berikut ini: nama_liga (misal : Liga Premier Inggris, La Liga Spanyol),negara dan tingkat (misal : level nasional, internasional atau divisi).

f. Dim Formasi

Dimensi ini menyediakan konteks formasi yang dipakai dan tipe strategi dari formasi supaya sesuai antar hierarki yang terdiri dari : formasi (misal : taktik 4-3-3, 5-4-1) dan strategi (misal : menyerang,bertahan)

4. Justifikasi Desain Konseptual

4.1 Alasan Pemilihan Skema Bintang (Star Schema)

Desain konseptual menggunakan skema bintang dipilih karena memiliki keunggulan utama dalam kemudahan eksplorasi data oleh pengguna non-teknis, kinerja query yang cepat berkat struktur sederhana yang mendukung indeksasi efisien, serta kompatibilitas tinggi dengan OLAP dan visualisasi analitik. Dalam konteks industri sepak bola, skema ini sangat cocok untuk analisis performa, karena memudahkan penyajian informasi berdasarkan waktu, pemain, tim, maupun liga secara fleksibel.

4.2 Relevansi Setiap Komponen

Setiap komponen dalam skema dirancang berdasarkan pertanyaan analitik kunci dan kebutuhan stakeholder dari Misi 1:

Komponen	Justifikasi
Fakta_Performa_Pertandingan	Memuat metrik utama performa pemain yang diperlukan untuk evaluasi taktis dan manajerial.
Dim_Pemain	Menyediakan detail identitas dan posisi pemain untuk analisis individu atau peran dalam tim.
Dim_Tim	Menggambarkan konteks klub yang memungkinkan analisis performa berdasarkan asal tim.
Dim_Waktu	Diperlukan untuk analisis tren historis, performa musiman, dan evaluasi tahunan.
Dim_Liga	Berguna untuk segmentasi berdasarkan kompetisi yang diikuti, serta membandingkan performa antar liga.

4.3 Kesesuaian dengan Tujuan Analitik

Skema ini dapat menjawab kebutuhan seperti:

- Menampilkan top scorer dalam satu liga selama musim tertentu. Melihat distribusi kartu merah di setiap kuartal kompetisi. Membandingkan durasi bermain rata-rata berdasarkan posisi pemain. Menganalisis pengaruh pelatih baru terhadap performa tim.

4.4 Integrasi Sumber Data

Desain ini juga mempertimbangkan struktur data mentah dari file seperti match.csv, player.csv, team.csv, dan league.csv yang telah diuraikan pada Misi 1. Ketersediaan atribut-atribut utama dalam sumber data menjamin kepraktisan implementasi skema ini secara teknis.

5. Kesesuaian dengan sumber data

5.1 Asal Data

Data yang digunakan dalam skema konseptual dengan pendekatan multidimensional berasal dari European Soccer Database dengan platform Kaggle. Dataset tersebut merupakan sebuah dataset historis yang mengumpulkan data lebih dari 25.000 pertandingan sepak bola, 10.000 pemain dan 11 negara yang berada di Eropa. Dataset tersebut berisikan data statistik pertandingan, informasi para pemain, kompetisi yang diikuti dan detail dari klub yang dibentuk dalam format tabel. Hal tersebut menjadi dasar sistem analitik pada performa sepak bola modern.

5.2 Hubungan Skema dengan Data Asal

Desain konseptual multidimensi yang telah dirancang memiliki tingkat kesesuaian yang tinggi dengan struktur data yang tersedia dalam dataset European Soccer Database. Setiap komponen yang terdapat dalam skema, baik fakta maupun dimensi dapat dipetakan secara langsung ke dalam tabel dan atribut dalam sumber dataset. Hal tersebut memastikan bahwa proses pengisian data ke dalam skema data warehouse efisien dan realistis.

Berikut hubungan antara skema konseptual dengan data asal:

Komponen Skema	Sumber Data	Atribut Utama	Keterangan
Fakta_Performa_Pertandingan	Match.csv	home_team_goal, away_team_goal, possession, shots_on_target	Statistik performa pertandingan yang digunakan untuk mengevaluasi secara taktis dan manajerial.
Dim_Pemain	Player.csv	player_name, birthday, height, weight, potential	Identitas dan karakteristik pemain yang relevan untuk analisis tiap individu.
Dim_Tim	Team.csv	team_long_name, formation, coach_name	Informasi mengenai klub dan formasi tim yang mendukung analisis berbasis tim
Dim_Waktu	Match.csv	date, season	Informasi mengenai tanggal dan musim untuk analisis tren historis
Dim_Liga	League.cs v	league_name, country_id	Nama liga dan negara untuk segmentasi kompetisi