

Desain Data Warehouse Perusahaan Online Travel Agent (OTA)

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

VALIDATION PAGE

Prepared By :

No.	Name	Date	Signature
1	Nazmi Hakim		

Reviewed By :

No.	Name	Date	Signature
1	Irham Maulani		

Approved By :

No.	Name	Date	Signature
1	Irham Maulani		

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

REVISION HISTORY

Revision Code	Revised By	Date	Total Page	Description	Approved By
1.0	Nazmi Hakim	17 October 2025	30	Initial document creation	

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

TABLE OF CONTENT

Introduction	5
Reference	5
Definition and Terms	7
Design and Implementation Limitations	8
Background.....	9
Purpose of The Document.....	10
Project Scope	11
Estimasi Durasi Proyek.....	12
Business Requirement.....	12
Business Problems and Solution	12
Tabel Perbandingan Tanpa vs Dengan Data Warehouse	15
Projected Financial Impact (Annual)	15
Stakeholders and Data Owners	16
Budget Plans	18
Tools	19
Data Sources	21
Internal Data Sources	21
External Data Sources	23
ETL (Extract, Transform, Load) Design	25
Extract Process.....	26
Scheduling	28
Transform Process	30
Load Process.....	32
Penanganan Kegagalan (Failure Handling) dan Call-back	33
Data Warehouse Architecture	34

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Introduction

TravelNusantara adalah perusahaan yang bergerak di bidang Online Travel Agent (OTA), menyediakan layanan pemesanan tiket pesawat, hotel, dan berbagai produk perjalanan lainnya. Dalam industri yang kompetitif dan dinamis, kecepatan dan ketepatan dalam pengambilan keputusan berbasis data menjadi faktor kunci untuk perusahaan agar bertahan dan tumbuh.

Namun, kondisi data di TravelNusantara saat ini masih tersebar di berbagai sistem dan format. Data transaksi pelanggan tersimpan di database OLTP, data historis dalam file CSV, sementara data harga dan ketersediaan tiket diperoleh dari API eksternal. Fragmentasi data tersebut menyebabkan kesulitan dalam menggabungkan informasi lintas sumber untuk analisis mendalam. Proses pelaporan yang seharusnya mendukung strategi bisnis justru menjadi manual, lambat, dan rentan kesalahan.

Sebagai tim data engineering, kami mengusulkan pembuatan Data Warehouse sebagai solusi untuk mengintegrasikan seluruh sumber data ke dalam satu repositori yang konsisten dan mudah diakses. Dengan arsitektur Kimball (dimensional modeling), data warehouse ini akan menjadi “single source of truth” bagi seluruh divisi perusahaan, memungkinkan proses analitik lintas departemen dilakukan secara cepat, akurat, dan terukur.

Implementasi data warehouse ini akan:

- Mengurangi waktu pembuatan laporan dari hitungan hari menjadi menit
- Meningkatkan akurasi laporan hingga 95–99%,
- Mendukung keputusan real-time bagi divisi marketing, finance, dan operasional,
- Memungkinkan perusahaan menerapkan strategi data-driven decision making untuk meningkatkan pendapatan dan kepuasan pelanggan.

Proposal ini bertujuan untuk menjelaskan urgensi, rancangan teknis, serta manfaat bisnis dari pembangunan sistem Data Warehouse di TravelNusantara sebagai investasi jangka panjang dalam transformasi digital perusahaan.

Reference

(Abdulaziz et al., 2015; Alghamdi & Agag, 2024; Ali et al., 2020; Armbrust et al., 2020; Carey et al., n.d.; Dapkute et al., 2024; Dietz et al., 2022; Fu et al., 2024; Kumar Bansal, n.d.; Kumar Singu, 2022; M. Tank, 2015; Nishanth Reddy Mandala, 2019; Rea D'souza et al., n.d.; Romero-Chuquital & Melendres-Velasco, 2023; Ross et al., 2006; Shizuka Hutagaol et al., 2025; Vidal et al., 2024)

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Alghamdi, O. A., & Agag, G. (2024). Competitive advantage: A longitudinal analysis of the roles of data-driven innovation capabilities, marketing agility, and market turbulence. *Journal of Retailing and Consumer Services*, 76. <https://doi.org/10.1016/j.jretconser.2023.103547>

Ali, T. Z., Abdelaziz, T. M., Maatuk, A. M., & Elakeili, S. M. (2020, November 28). A framework for improving data quality in data warehouse: A case study. *Proceedings - 2020 21st International Arab Conference on Information Technology, ACIT 2020*. <https://doi.org/10.1109/ACIT50332.2020.9300119>

Armbrust, M., Das, T., Sun, L., Yavuz, B., Zhu, S., Murthy, M., Torres, J., van Hovell, H., Ionescu, A., Łuszczak, A., Świtakowski, M., Szafrański, M., Li, X., Ueshin, T., Mokhtar, M., Boncz, P., Ghodsi, A., Paranjpye, S., Senster, P., ... Zaharia, M. (2020). Delta Lake: High-Performance ACID Table Storage over Cloud Object Stores. *Proceedings of the VLDB Endowment*, 13(12), 3411–3424. <https://doi.org/10.14778/3415478.3415560>

Carey, M. J., Ceri, S., Board, E., Bernstein, P., Dayal, U., Faloutsos, C., Freytag, J. C., Gardarin, G., Jonker, W., Krishnamurthy, V., Neimat, M.-A., Valduriez, P., Weikum, G., Whang, K.-Y., & Widom, J. (n.d.). *Data-Centric Systems and Applications Series Editors*.

Dapkute, A., Siozinys, V., Jonaitis, M., Kaminickas, M., & Siozinys, M. (2024). Digital Twin Data Management: Framework and Performance Metrics of Cloud-Based ETL System. *Machines*, 12(2). <https://doi.org/10.3390/machines12020130>

Dietz, L. W., Sertkan, M., Myftija, S., Thimbiri Palage, S., Neidhardt, J., & Wörndl, W. (2022). A Comparative Study of Data-Driven Models for Travel Destination Characterization. *Frontiers in Big Data*, 5. <https://doi.org/10.3389/fdata.2022.829939>

Fu, Q., Nicholson, G. L., & Easton, J. M. (2024). Understanding data quality in a data-driven industry context: Insights from the fundamentals. In *Journal of Industrial Information Integration* (Vol. 42). Elsevier B.V. <https://doi.org/10.1016/j.jii.2024.100729>

Kumar Bansal, D. (n.d.). *ENTERPRISE DATA WAREHOUSES: TYPES, BENEFITS, AND CONSIDERATIONS*. <https://doi.org/10.56726/IRJMETS68496>

Kumar Singu, S. (2022). Impact of Data Warehousing on Business Intelligence and Analytics. *ESP Journal of Engineering & Technology Advancements*, 2(2), 101–113. <https://doi.org/10.56472/25832646/JETA-V2I2P112>

M. Tank, D. (2015). Enable Better and Timelier Decision-Making Using Real-Time Business Intelligence System. *International Journal of Information Engineering and Electronic Business*, 7(1), 43–48. <https://doi.org/10.5815/ijieeb.2015.01.06>

Nishanth Reddy Mandala. (2019). The evolution of ETL architecture: From traditional data warehousing to real-time data integration. *World Journal of Advanced Research and Reviews*, 1(3), 073–084. <https://doi.org/10.30574/wjarr.2019.1.3.0033>

Rea D'souza, R., Satam, P. M., Rao, V., & 1#, N. (n.d.). *The Role of Data Warehousing in Business Intelligence Systems to Support Rapid Decision-Making*. www.JSR.org

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Romero-Chuquital, A., & Melendres-Velasco, J. J. (2023). Use of data warehouse for business decision making: a literary review. In *Revista Cientifica de Sistemas e Informatica* (Vol. 3, Issue 2). Universidad Nacional de San Martin. <https://doi.org/10.51252/rksi.v3i2.543>

Ross, J. W., Weill, Peter., & Robertson, David. (2006). *Enterprise architecture as strategy : creating a foundation for business execution*. Harvard Business School Press.

Shizuka Hutagaol, A., Tji Beng, J., Tiatri, S., Lunzaga, E., Mulia Salsabila, T., & Nurkholiza, R. (2025). PERANCANGAN DATA WAREHOUSE UNTUK MENGOPTIMALKAN DASHBOARD DALAM MEMONITORING PEMESANAN HOTEL DESIGNING A DATA WAREHOUSE TO OPTIMIZE THE HOTEL BOOKING MONITORING DASHBOARD. *Journal of Information Technology and Computer Science (INTECOMS)*, 8(5).

Vidal, J., Carrasco, R. A., Cobo, M. J., & Blasco, M. F. (2024). Data Sources as a Driver for Market-Oriented Tourism Organizations: a Bibliometric Perspective. *Journal of the Knowledge Economy*, 15(2), 7588–7621. <https://doi.org/10.1007/s13132-023-01334-5>

Definition and Terms

- **API (Application Programming Interface):** Sumber data eksternal yang menyediakan data operasional *live*. Dalam konteks ini, merujuk pada Amadeus API (Flight Offers Search dan Hotel Search) yang mengembalikan respons data dalam format JSON.
- **Business Intelligence (BI):** Tujuan akhir dari DWH, yang mencakup *tools* visualisasi (seperti Tableau atau Power BI) untuk mengubah data menjadi *insight* dan mendukung *data-driven decision making*.
- **CDC (Change Data Capture):** Metode ekstraksi *event-driven* (real-time) untuk database OLTP. Metode ini membaca *transaction log* untuk menangkap perubahan (INSERT, UPDATE) secara instan dengan beban minimal pada sistem sumber.
- **Data Mart:** Bagian dari DWH yang dibuat untuk kebutuhan analisis khusus per divisi (misalnya Marketing, Finance).
- **Data Warehouse (DWH):** Solusi untuk mengintegrasikan seluruh sumber data (OLTP, CSV, API) ke dalam satu repositori yang konsisten dan mudah diakses. DWH bertindak sebagai “*single source of truth*” dan dirancang menggunakan Dimensional Modeling (Star Schema) untuk mendukung analitik.
- **Dead Letter Queue (DLQ):** Bagian dari strategi *failure handling*. Ini adalah lokasi (misal *bucket* S3) di mana *record* data yang gagal diproses (karena korup atau format tidak terduga) dialihkan untuk analisis manual, sehingga tidak menghentikan keseluruhan proses *batch*.
- **ETL (Extract, Transform, Load):** Proses otomatis untuk mengintegrasikan data. *Extract* adalah pengumpulan data dari sumber; *Transform* adalah proses

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

pembersihan, standarisasi, dan pengayaan data ; dan *Load* adalah proses memasukkan data yang sudah matang ke dalam Data Warehouse.

- **Kimball (Dimensional Modeling):** Pendekatan arsitektur DWH yang dipilih karena fokus pada kebutuhan bisnis , mudah dipahami pengguna non-teknis , dan optimal untuk laporan/dashboard menggunakan Star Schema.
- **Medallion Architecture:** Desain penyimpanan data yang menjamin kualitas dan tata kelola, terdiri dari tiga lapisan:
 - **Bronze (Raw Zone):** Lapisan penyimpanan data mentah apa adanya dari sumber (data *immutable*).
 - **Silver (Cleansed Zone):** Lapisan tempat data dibersihkan, divalidasi, dideDuplikasi, dan dinormalisasi. Ini menjadi *single source of truth* untuk analisis.
 - **Gold (Curated Zone):** Lapisan penyajian data yang sudah diagregasi dan siap dikonsumsi bisnis , biasanya dalam bentuk Star Schema untuk *tools BI*.
- **OLAP (Online Analytical Processing):** Tipe kueri analitis (seperti kueri *ad-hoc* atau *dashboard*). Arsitektur Star Schema dan platform seperti BigQuery dioptimalkan untuk jenis kueri ini.
- **OLTP (Online Transaction Processing):** Database operasional harian perusahaan (misal MySQL/PostgreSQL) yang menyimpan data transaksi dan pelanggan. Sistem ini dioptimalkan untuk operasi harian, bukan untuk analisis historis.
- **Staging Area (Data Lake):** Tempat penyimpanan sementara (misal GCS/S3) di mana data dari semua sumber (OLTP, API, CSV) dikumpulkan (*di-land*) sebelum diproses oleh ETL Server.
- **Star Schema:** Model data dimensional yang memisahkan data menjadi satu tabel Fakta (transaksi/peristiwa, misal FactBooking) dan beberapa tabel Dimensi (konteks siapa, apa, di mana, kapan, misal DimCustomer, DimDate).
- **Surrogate Key (SK):** Kunci unik yang dibuat oleh DWH (misal customer_sk) yang berfungsi sebagai *Primary Key (PK)* di tabel dimensi untuk mengintegrasikan data dari berbagai sumber.

Design and Implementation Limitations

- **Validitas Analisis Bisnis:** Seluruh analisis kuantitatif, proyeksi *financial impact* (Total Benefit Rp 705.400.000) , dan identifikasi masalah bisnis (Problem 1-5) didasarkan pada analisis dataset yang terbatas, yaitu hanya 139 transaksi perjalanan. Performa atau masalah yang diidentifikasi mungkin berbeda saat implementasi menangani volume data produksi penuh.
- **Ketergantungan pada API Eksternal:** Implementasi sangat bergantung pada ketersediaan dan performa API Amadeus. Desain ini rentan terhadap

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

kegagalan di luar kendali tim, seperti jika API Amadeus mengembalikan *status code* error (misal 5xx), yang akan menunda *job ETL*.

- **Keterbatasan Penanganan Error:** Desain penanganan kegagalan (ETL) tidak sepenuhnya otomatis. Jika terjadi kegagalan data spesifik (misal data korup dari CSV atau format JSON API tidak terduga), data tersebut akan dialihkan ke *Dead Letter Queue* (DLQ) dan memerlukan "analisis manual" oleh tim Data Engineering untuk menyelesaiannya.
- **Risiko Implementasi CDC:** Desain utama untuk ekstraksi OLTP merekomendasikan CDC (*Change Data Capture*) untuk data *real-time*. Namun, dokumen ini mencantumkan risiko bahwa CDC "tidak memungkinkan" untuk diimplementasikan.
- **Keterbatasan Metode Fallback:** Jika CDC gagal diimplementasikan, desain bergantung pada metode *fallback* berupa *Query Inkremental* setiap 1 jam. Ini secara langsung membatasi kemampuan sistem untuk memenuhi kebutuhan bisnis akan *dashboard* operasional *near-real-time*, karena akan ada "toleransi terhadap jeda data" hingga 1 jam.
- **Batasan Anggaran dan Waktu:** Proyek ini memiliki estimasi durasi yang ketat yaitu 8 minggu (2 bulan) dengan anggaran tetap (Rp 193.775.000). Dana kontingensi yang dialokasikan hanya ±15%, sehingga setiap penyesuaian *scope* atau kebutuhan *compute* yang lebih tinggi (misal saat *backfill* data) dapat menghambat penyelesaian proyek tepat waktu dan sesuai anggaran.
- **Ketergantungan *Lookup* yang Tidak Terdefinisi:** Proses transformasi (Transform Process) memiliki aturan untuk "Standarisasi Mata Uang" yang bergantung pada "tabel *lookup* kurs harian". Namun, sumber data (Data Sources) untuk *lookup* tabel kurs harian ini tidak didefinisikan dalam desain.

Background

TravelNusantara menghadapi tantangan besar dalam mengelola data yang berasal dari berbagai sumber dan format. Volume transaksi mencapai ratusan ribu per hari, dengan jutaan interaksi pengguna dari berbagai platform digital. Namun, data yang dihasilkan tidak terpusat dan sulit digunakan untuk analisis lintas departemen.

Saat ini:

- Data pelanggan dan transaksi tersimpan di database OLTP yang dioptimalkan untuk operasi harian, bukan analisis historis.
- Data historis dikirim melalui file CSV yang diproses manual setiap minggu.
- Data harga dan ketersediaan tiket diambil dari API eksternal tanpa penyimpanan historis terstruktur.

Akibatnya:

- Tim analis memerlukan 3–5 hari hanya untuk membersihkan dan menggabungkan data sebelum membuat laporan.

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

- Divisi marketing kesulitan mengukur efektivitas kampanye karena data pelanggan tidak terintegrasi.
- Divisi finance tidak memiliki visibilitas real-time terhadap pendapatan lintas kanal.
- Tim product tidak dapat dengan cepat mengidentifikasi tren destinasi populer.

Dampak langsung dari kondisi ini adalah kehilangan peluang bisnis, pengambilan keputusan berbasis asumsi, dan menurunnya daya saing terhadap kompetitor yang telah mengadopsi analitik yang lebih modern.

Oleh karena itu, dibutuhkan sebuah sistem Data Warehouse yang mampu:

1. Mengintegrasikan data dari OLTP, CSV, dan API secara otomatis melalui proses ETL (Extract, Transform, Load).
2. Menyediakan data historis yang bersih, konsisten, dan mudah diakses oleh semua departemen.
3. Menjadi dasar bagi penerapan Business Intelligence (BI) dan data visualization tools seperti Power BI atau Tableau.

Dengan arsitektur data warehouse yang terencana, TravelNusantara akan bertransformasi menjadi organisasi yang berorientasi pada data (data-driven organization). Sistem ini akan mempercepat proses pelaporan, tetapi juga menjadi fondasi untuk strategi pemasaran, pengelolaan pendapatan, dan optimalisasi pengalaman pelanggan.

Purpose of The Document

Tujuan utama dokumen ini adalah untuk mengajukan rancangan desain dan implementasi Data Warehouse bagi PT TravelNusantara sebagai solusi strategis dalam mengatasi permasalahan integrasi data, kecepatan pelaporan, dan akurasi analitik.

Dokumen ini disusun oleh tim Data Engineering dengan tujuan sebagai berikut:

1. Menjelaskan urgensi bisnis dan teknis mengapa TravelNusantara membutuhkan Data Warehouse pada tahap transformasi digital saat ini.
2. Mendeskripsikan rancangan arsitektur teknis, termasuk alur proses ETL (Extract, Transform, Load), model data dimensional, serta sistem penyimpanan yang digunakan.
3. Menentukan kebutuhan bisnis (Business Requirement) dari setiap divisi pengguna data seperti marketing, finance, dan operasional.
4. Memberikan panduan implementasi dan estimasi sumber daya yang dibutuhkan agar proyek dapat berjalan efektif dan memberikan dampak nyata terhadap performa perusahaan.
5. Menjadi dasar komunikasi dan kesepakatan antara tim data engineer, pihak manajemen, dan stakeholder lain terkait arah pengembangan sistem analitik perusahaan ke depan.

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Dengan dokumen ini, diharapkan manajemen dapat memahami nilai strategis dari pembangunan Data Warehouse dan menyetujui implementasinya sebagai bagian dari roadmap digitalisasi perusahaan.

Project Scope

Proyek ini bertujuan untuk membangun sistem Data Warehouse terpusat di lingkungan perusahaan TravelNusantara, yang akan menjadi sumber data utama (single source of truth) bagi seluruh divisi seperti Marketing, Finance, dan Product.

Sistem ini akan mengintegrasikan data dari berbagai sumber (OLTP, CSV, dan API eksternal) ke dalam repositori yang terstruktur, konsisten, dan mudah diakses untuk analisis bisnis.

Cakupan proyek pengembangan Data Warehouse mencakup beberapa tahapan utama berikut:

No	Tahap	Deskripsi
1	Analisis Kebutuhan Data (Requirement Gathering)	Mengidentifikasi kebutuhan data tiap divisi, termasuk jenis laporan dan KPI yang ingin dianalisis (penjualan, pelanggan, pendapatan, promosi, dan lain-lain).
2	Desain Arsitektur Data Warehouse	Menentukan pendekatan arsitektur menggunakan Kimball Dimensional Model (Star Schema) untuk mendukung analisis multi-dimensi.
3	Pengumpulan dan Integrasi Data (ETL)	Mengembangkan proses Extract, Transform, Load (ETL) untuk menggabungkan data dari sistem OLTP, file CSV, dan API eksternal ke staging area sebelum masuk ke warehouse.
4	Penyimpanan dan Manajemen Data	Merancang struktur penyimpanan yang efisien untuk data historis dengan dukungan partisi waktu, serta mendefinisikan kebijakan retensi dan backup data.
5	Data Mart dan Akses Analitik	Membuat data mart tematik (Marketing, Finance, Product) untuk kebutuhan analisis khusus divisi, serta menghubungkannya ke alat BI seperti Power BI atau Tableau .
6	Keamanan dan Akses Kontrol	Menetapkan otorisasi pengguna berbasis peran (role-based access) untuk menjaga kerahasiaan data antar divisi.
7	Monitoring dan Maintenance	Menyediakan sistem pemantauan ETL, logging, serta dokumentasi pipeline untuk memastikan keberlangsungan dan kualitas data.

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Hasil akhir yang diharapkan dari proyek ini adalah dokumen desain arsitektur serta skema data warehouse dalam bentuk *Star Schema*. Selain itu, akan dihasilkan *pipeline ETL* otomatis yang mengintegrasikan data dari berbagai sumber seperti OLTP, file CSV, dan API. Proyek ini juga akan menghasilkan *database* Data Warehouse yang siap digunakan, misalnya menggunakan PostgreSQL atau Google BigQuery. Sebagai pelengkap, disediakan pula panduan akses dan dokumentasi teknis yang dapat digunakan oleh tim analis maupun manajemen.

Estimasi Durasi Proyek

Tahapan	Estimasi Durasi
Analisis kebutuhan dan desain arsitektur	2 minggu
Pengembangan proses ETL dan integrasi data	3 minggu
Implementasi penyimpanan dan skema warehouse	2 minggu
Pengujian dan dokumentasi	1 minggu
Total Estimasi Durasi Proyek	8 minggu (2 bulan)

Business Requirement

Business Problems and Solution

Semua data kuantitatif dalam dokumen ini didasarkan pada analisis aktual dari dataset yang mencakup 139 transaksi perjalanan dengan 13 atribut data. Dataset tersebut adalah representasi konkret dan representatif yang mencerminkan kondisi riil operasional TravelNusantara.

Problem 1: Proses Pelaporan Manual yang Tidak Efisien

Saat ini, proses pelaporan di TravelNusantara masih dilakukan secara manual dengan tingkat efisiensi yang rendah. Tim analis membutuhkan waktu antara 4 hingga 6 jam setiap hari hanya untuk mengekstraksi data dari berbagai sumber, sementara pembuatan satu laporan lengkap bisa memakan waktu hingga 3 sampai 5 hari. Secara keseluruhan, staf menghabiskan sekitar 40 jam per minggu untuk pekerjaan manual terkait data, yang menimbulkan biaya operasional antara Rp16 hingga Rp20 juta per bulan. Dampaknya, terjadi keterlambatan signifikan dalam pengambilan keputusan strategis, staf tidak dapat fokus pada analisis bernilai tambah, dan perusahaan kehilangan hingga 87,5% produktivitas potensial. Selain itu, risiko kesalahan manual mencapai 15–20%, yang dapat berdampak langsung pada akurasi revenue.

Dengan implementasi Data Warehouse, proses ETL dapat diotomatisasi sehingga waktu ekstraksi data menurun drastis menjadi kurang dari satu jam per hari. Pembuatan laporan dapat dilakukan hanya dalam 5 hingga 15 menit, dan pekerjaan manual berkurang menjadi sekitar 5 jam per minggu. Biaya

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

operasional pun turun menjadi Rp2–3 juta per bulan. Manfaat terukur dari perubahan ini mencakup penghematan waktu sebesar 35 jam per minggu (reduksi 87,5%), penghematan biaya Rp14–17 juta per bulan atau sekitar Rp168–204 juta per tahun, serta peningkatan produktivitas hingga 700%, dengan estimasi ROI break-even dalam 6–8 bulan.

Problem 2: Kualitas Data Rendah dan Tidak Konsisten

Kualitas data yang rendah menjadi tantangan utama bagi perusahaan. Berdasarkan analisis terhadap dataset, ditemukan tingkat *missing data* sebesar 11,51% dari total dataset, dengan 26 nilai kosong dari 1.807 data poin. Format data juga tidak konsisten, misalnya terdapat sembilan variasi berbeda untuk tipe transportasi seperti “Flight”, “Airplane”, “Plane”, dan “Ferry” yang seharusnya distandardisasi. Selain itu, tingkat kesalahan manual mencapai 15–20%, sementara proses validasi data membutuhkan waktu antara 8 hingga 12 jam per minggu hanya untuk membersihkan data.

Kondisi ini menyebabkan pengambilan keputusan bisnis sering kali didasarkan pada data yang tidak akurat dan tidak lengkap, berisiko menimbulkan revenue loss akibat kesalahan harga atau pemesanan, serta menurunkan kepuasan pelanggan dan reputasi perusahaan. Dengan penerapan Data Warehouse, sistem dapat melakukan pemeriksaan kualitas data secara otomatis dengan akurasi 95–99%, menstandardisasi format dari semua sumber melalui proses transformasi ETL, dan menekan *missing data rate* hingga di bawah 2%. Proses validasi dilakukan secara real-time dengan deteksi anomali otomatis. Hasilnya, tingkat kesalahan berkurang hingga 90%, kualitas data meningkat 85%, keluhan pelanggan menurun 30–40%, dan tingkat akurasi pengambilan keputusan meningkat hingga 99%.

Problem 3: Tidak Ada Visibilitas Revenue Secara Real-Time

Perusahaan saat ini tidak memiliki sistem pelacakan pendapatan (revenue tracking) secara real-time. Dari analisis terhadap 139 transaksi, total pendapatan tercatat sebesar \$179.950, namun pelaporannya dilakukan secara manual dengan jeda 5–7 hari sebelum laporan tersedia. Meskipun destinasi dengan performa tertinggi dapat diidentifikasi seperti Paris (\$14.600), Tokyo (\$14.400), dan New York City (\$12.500), data tersebut tidak dapat diakses secara langsung saat dibutuhkan. Akibatnya, perusahaan kesulitan mengukur *Marketing ROI* dengan akurat karena data tersebar di berbagai sumber.

Dampak bisnisnya cukup besar, perusahaan tidak dapat merespons perubahan pasar secara cepat, pengeluaran pemasaran menjadi tidak efisien, dan peluang *dynamic pricing* untuk memaksimalkan pendapatan sering terlewat. Dengan adanya Data Warehouse, seluruh pendapatan dapat dipantau melalui *real-time revenue dashboard* yang menampilkan *breakdown* per destinasi, jenis akomodasi, dan transportasi. Selain itu, perusahaan dapat melakukan segmentasi pelanggan, menghitung *Customer Lifetime Value (CLV)*, serta memanfaatkan analitik prediktif untuk melakukan peramalan pendapatan. Implementasi ini berpotensi meningkatkan pendapatan sebesar 10–15%, menghasilkan tambahan revenue sekitar \$22.493,75 per tahun, meningkatkan

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

efisiensi pemasaran 25–30%, serta menurunkan *Customer Acquisition Cost* hingga 20%.

Problem 4: Data Pelanggan Tersebar di Berbagai Sistem

Data pelanggan saat ini tersebar di berbagai sistem yang tidak saling terintegrasi. Berdasarkan analisis demografi pelanggan, terdapat 139 perjalanan dari beberapa sistem berbeda, dengan 70 pelanggan perempuan dan 67 laki-laki. Kebanyakan pelanggan berasal dari Amerika (24), Korea (13), Inggris (12), dan Kanada (9), dengan usia rata-rata 33 tahun dan rentang 20–60 tahun. Namun, tidak ada *single customer ID* yang memungkinkan pandangan menyeluruh terhadap pelanggan.

Situasi ini menyebabkan perusahaan tidak dapat mempersonalisasi pengalaman pelanggan berdasarkan riwayat transaksi, kampanye pemasaran menjadi bersifat umum dan tidak terarah, serta *Customer Lifetime Value* tidak dapat dihitung untuk mendukung strategi retensi. Dengan penerapan Data Warehouse, perusahaan dapat membangun *360° customer view* dengan profil terintegrasi dari semua *touchpoints*. Sistem ini juga memungkinkan segmentasi pelanggan berdasarkan nilai, perilaku, demografi, dan preferensi, serta memberikan rekomendasi personal berdasarkan riwayat pemesanan. Hasil yang diharapkan meliputi peningkatan retensi pelanggan sebesar 20–25%, kenaikan *repeat booking rate* hingga 35%, peningkatan konversi pemasaran 40–50%, dan peningkatan *Customer Satisfaction Score* sebesar 25–30%.

Problem 5: Tidak Dapat Melacak Performa Destinasi dan Produk

Perusahaan juga menghadapi kesulitan dalam melacak performa destinasi dan produk secara efisien. Berdasarkan analisis terhadap data performa produk, destinasi dengan pendapatan tertinggi adalah Paris (\$14.600), Tokyo (\$14.400), dan New York City (\$12.500). Akomodasi paling populer adalah hotel dengan 31 pemesanan senilai \$90.250, disusul Airbnb dengan 14 pemesanan (\$36.050) dan hostel dengan 13 pemesanan (\$23.350). Durasi rata-rata perjalanan adalah 7,6 hari, sementara waktu analisis untuk membuat laporan performa produk mencapai 2–3 hari.

Kondisi ini menyebabkan perusahaan tidak dapat mengoptimalkan *product mix* berdasarkan profitabilitas, mengelola inventori secara efisien, ataupun menetapkan strategi harga berbasis data. Dengan implementasi Data Warehouse, perusahaan dapat mengakses *real-time destination performance dashboard*, melakukan analisis profitabilitas produk berdasarkan biaya dan margin, mengidentifikasi tren musiman untuk *forecasting*, serta melakukan *competitor benchmarking*. Dampak yang dihasilkan mencakup pengurangan inefisiensi inventori hingga 20%, peningkatan pendapatan dari optimasi harga sebesar 10–12%, peningkatan profit produk 18–22%, dan peningkatan *partnership ROI* hingga 30%.

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Tabel Perbandingan Tanpa vs Dengan Data Warehouse

Aspek	Tanpa Data Warehouse (Saat Ini)	Dengan Data Warehouse	Improvement
Waktu Pembuatan Laporan	3-5 hari per laporan	5-15 menit per laporan	99% lebih cepat
Akurasi Data	80-85% (error rate 15-20%)	95-99% (error rate <1%)	+15-19% akurasi
Biaya Operasional Bulanan	Rp 16-20 juta	Rp 2-3 juta	Hemat Rp 14-17 juta/bulan
Produktivitas Staff	40 jam/minggu untuk manual work	5 jam/minggu untuk manual work	+700% produktivitas (35 jam saved/week)
Visibilitas Revenue	Delay 5-7 hari, tidak real-time	Real-time dashboard & analytics	Instant access (100% faster)
Customer Insight	Data tersebar, tidak ada unified view	360° customer view dengan segmentasi	Complete visibility
Data Integration Time	8-12 jam manual per minggu	Automated ETL <1 jam/hari	92% lebih cepat

Projected Financial Impact (Annual)

Kategori	Annual Value (IDR)	Keterangan
Cost Savings	Rp 168,000,000	Hemat biaya operasional manual (Rp 14M/bulan × 12)
Revenue Increase	Rp 337,400,000	Revenue increase 10-15% dari baseline \$179,950
Marketing Efficiency	Rp 120,000,000	Marketing ROI improvement & reduced CAC
Operational Efficiency	Rp 80,000,000	Inventory optimization & pricing optimization
TOTAL BENEFIT	Rp 705,400,000	Total projected annual benefit dari Data Warehouse

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Stakeholders and Data Owners

Division	Stakeholder	Data Required	Data Owner	Purpose
Finance dan Accounting	CFO, Finance Manager, Revenue Analyst	Revenue per trip, Accommodation cost, Transportation cost, Payment status, Profit margin, Financial reports	Database OLTP (Booking System), Payment Gateway API, CSV	Monitor revenue streams, calculate profitability, generate financial reports, ensure accurate billing
Marketing & Sales	CMO, Marketing Manager, Sales Team	Customer demographics (age, gender, nationality), Destination trends, Booking patterns, Marketing campaign effectiveness	Database OLTP (Customer table), Marketing Platform, Google Analytics, Social Media APIs	Analyze customer segments, optimize marketing campaigns, identify high-value customers, track conversion rates
Product dan Operations	COO, Product Manager, Operations Manager	Trip details (destination, duration, dates), Accommodation & transportation type, Booking volume, Operational efficiency metrics	Database OLTP (Trip, Accommodation, Transportation tables), Amadeus Flight API, Hotel Partner CSV files	Optimize product offerings, manage inventory (flights/hotels), improve operational efficiency, identify popular destinations
Customer Service	Customer Service Head, Support Team	Customer information (name, age, nationality), Trip history, Booking records, Customer complaints/feedback	Database OLTP (Customer & Booking tables), CRM System, Support Ticket System	Provide better customer support, track customer journey, resolve issues faster, improve satisfaction
Business Intelligence	BI Manager, Data Analyst,	All integrated data from OLTP, CSV, API; Historical	Data Warehouse (integrated from all sources)	Create dashboards and reports,

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

dan Analytics	Business Analyst	trends, KPI dashboards, Predictive analytics data		perform trend analysis, support data-driven decision making, forecast business metrics
IT dan Data Engineering	CTO, Data Engineer, System Administrator	Database schema, ETL logs, Data quality metrics, System performance, API integration status	All data sources (OLTP, CSV, APIs), ETL Pipeline, Data Warehouse Infrastructure	Maintain data infrastructure, ensure data quality, optimize ETL performance, manage data security

Key Benefits per Stakeholder:

Finance Division akan mendapatkan visibilitas real-time terhadap revenue streams dari 139 transaksi senilai \$179,950, dengan breakdown detail per destination (Paris: \$14,600, Tokyo: \$14,400), accommodation type (Hotel: \$90,250, Airbnb: \$36,050), dan transportation. Ini memungkinkan accurate forecasting dan financial planning.

Marketing Division akan memiliki akses ke customer demographics yang lengkap (70 Female, 67 Male, top nationalities: American 24, Korean 13, British 12) untuk targeted campaigns, customer segmentation berdasarkan CLV, dan measurement ROI dari setiap marketing initiative.

Product & Operations dapat mengoptimalkan product mix dengan data konkret bahwa average trip duration adalah 7.6 hari, accommodation preference (Hotel 44%, Airbnb 20%, Hostel 19%), dan mengidentifikasi seasonal trends untuk inventory management.

Customer Service mendapatkan 360° customer view untuk provide personalized support, track customer journey dari 139 unique travelers dengan age range 20-60 tahun (mean 33 tahun), dan proactively address issues.

BI & Analytics Team memiliki single source of truth untuk semua analytical needs, dengan data quality 95-99% accuracy (naik dari 80-85%), memungkinkan predictive analytics dan advanced business intelligence.

IT & Data Engineering mendapatkan centralized data infrastructure yang mudah di-maintain, dengan automated ETL reducing manual work dari 40 jam/minggu menjadi 5 jam/minggu, dan comprehensive logging untuk troubleshooting.

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Budget Plans

Kategori	Item	Estimasi Biaya (IDR)	Keterangan
1. Personil (Tim Inti Proyek)	2x Data Engineer (Selama 2 bulan)	Rp 80.000.000	Biaya utama untuk pengembangan pipeline ETL ²⁵ , desain skema ²⁶ , dan implementasi DWH.
	1x BI Analyst (Selama 2 bulan)	Rp 30.000.000	Fokus pada pembangunan <i>dashboard</i> awal di Tableau/Power BI dan validasi data untuk kebutuhan <i>stakeholder</i> ²⁷ .
	1x Project Manager (Paruh waktu, 2 bulan)	Rp 25.000.000	Mengelola linimasa proyek 8 minggu ²⁸ , koordinasi dengan <i>stakeholder</i> ²⁹ , dan memastikan <i>scope</i> ³⁰ terpenuhi.
Subtotal Personil		Rp 135.000.000	
2. Infrastruktur Cloud (Estimasi 2 Bulan)	Google Cloud Storage / AWS S3	Rp 3.000.000	Biaya penyimpanan untuk <i>data ingestion</i> awal (data historis CSV ³¹) dan <i>staging</i> (Bronze/Silver Layers ³²).
	Google BigQuery	Rp 10.000.000	Biaya untuk penyimpanan (Gold Layer ³³) dan <i>compute</i> kueri selama proses pengembangan, pengujian, dan UAT.
	Layanan ETL (AWS Glue / VM Airflow)	Rp 8.000.000	Biaya <i>compute</i> untuk menjalankan <i>job</i> ETL selama pengembangan dan pengujian (misal: <i>backfill</i> data historis).
Subtotal Infrastruktur		Rp 21.000.000	
3. Perangkat Lunak & Lisensi	Lisensi Tableau/Power BI (Misal: 1 Creator, 3 Viewer)	Rp 5.000.000	Lisensi untuk BI Analyst (Creator) dan <i>stakeholder</i> kunci (Viewer) selama 2 bulan.
	Biaya Akses API Eksternal (Amadeus)	Rp 7.500.000	Alokasi biaya untuk akses <i>rate limit</i> yang lebih tinggi ke API Amadeus ³⁴ selama masa pengembangan dan pengujian.
Subtotal Perangkat Lunak		Rp 12.500.000	
4. Lain-lain	Dana Kontingensi ($\pm 15\%$)	Rp 25.275.000	Alokasi untuk biaya tak terduga, seperti kebutuhan <i>compute</i> yang

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

			lebih tinggi saat <i>backfill</i> data atau penyesuaian <i>scope</i> .
Subtotal Lain-lain		Rp 25.275.000	
TOTAL ESTIMASI PROYEK		Rp 193.775.000	Investasi awal ini sangat sebanding dengan proyeksi keuntungan tahunan sebesar Rp 705.400.000 ³⁵ dan penghematan biaya operasional bulanan Rp 14-17 juta ³⁶ .

Tools

Kategori	Perangkat/Tools	Justifikasi Penggunaan (Berdasarkan Dokumentasi Resmi & Konteks Dokumen)
Data Lake dan Staging Storage	Google Cloud Storage (GCS) / AWS S3	Dokumentasi resmi GCS/S3 menggambarkannya sebagai object storage yang sangat skalabel dan hemat biaya. Ini sangat ideal untuk Arsitektur Medallion
Orkestrasi ETL/ELT	Apache Airflow	Dokumentasi resmi Airflow menjelaskannya sebagai platform untuk membuat, menjadwalkan, dan memantau alur kerja (workflow) secara terprogram.
Containerization dan Deployment	Docker / Kubernetes	Docker digunakan untuk membungkus (containerize) layanan ETL (seperti Apache Airflow) dan dependensinya. dan memastikan konsistensi lingkungan antara <i>development</i> dan <i>production</i> . Kubernetes digunakan untuk orkestrasi <i>container</i> tersebut, mengelola <i>scaling</i> , <i>auto-healing</i> , dan <i>deployment</i> pipeline data untuk memastikan <i>high availability</i> .
Proses Transformasi (ETL)	Python (dengan library Pandas & SQLAlchemy)	Dokumentasi resmi Python dan Pandas menyoroti kemampuannya yang kuat untuk manipulasi dan pembersihan data. Ini akan menjadi 'mesin' utama untuk semua aturan di Transform Process, seperti: <ul style="list-style-type: none"> Standarisasi kategori ("Flight", "Airplane" -> "UDARA"). Parsing dan format tanggal. Konversi mata uang.
Proses Transformasi (Serverless)	AWS Glue / Google Cloud Dataflow	Dokumentasi resmi menggambarkannya sebagai layanan ETL serverless (tanpa server). Ini ideal karena: <ul style="list-style-type: none"> Mengurangi beban manajemen infrastruktur. Terintegrasi penuh dengan S3 (Storage) dan Redshift/BigQuery (Warehouse).

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

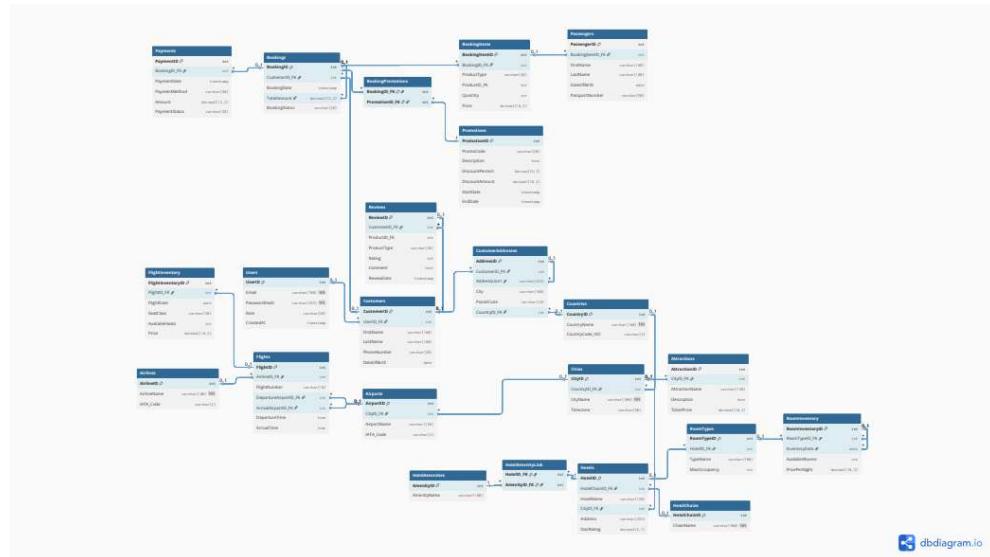
		<ul style="list-style-type: none"> Dapat menangani proses pembersihan dan transformasi data (ETL/ELT) dari Silver ke Gold layer.
Data Warehouse (Gold Layer)	Google BigQuery	<p>Dokumentasi resmi BigQuery menjelaskannya sebagai data warehouse cloud yang serverless, sangat skalabel, dan dioptimalkan untuk kueri analitis (OLAP) yang super cepat. Ini sempurna untuk:</p> <ul style="list-style-type: none"> Menyimpan Star Schema Anda (Layer Gold). Menangani kueri ad-hoc dari tim analis. Menjadi sumber data bagi dashboard BI tanpa perlu manajemen server.
Business Intelligence (BI)	Tableau / Power BI	<p>Dokumentasi resmi Tableau/Power BI menggambarkannya sebagai tools visualisasi terdepan untuk mengubah data mentah menjadi insight. Ini secara langsung memenuhi kebutuhan bisnis untuk:</p> <ul style="list-style-type: none"> Mengganti laporan manual 3-5 hari dengan dashboard real-time. Memberikan visibilitas kepada stakeholders (Finance, Marketing).

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Data Sources

Internal Data Sources

OLTP :



Contoh data di OLTP :

Customers :

Custo merID	UserID	FirstName	LastName	Email	Phone
1	101	Hakim	Nonchalant	hakim@example.com	081234567890
2	102	Arjun	Blake	arjun@example.com	082233445566
3	103	Hasbie	Whitie	hasbie@example.com	087711223344

Airlines :

AirlineID	AirlineName	IATACode
1	Garuda Indonesia	GA
2	AirAsia	QZ
3	Singapore Airlines	SQ

Title	Desain Data Warehouse Online Travel Agent (OTA)		
Document Number	BI-DW-001		
Revision	1.0		
Controlled Status	Controlled		
Classification	Internal		

Airports :

AirportID	AirportName	IATACode	CityID
1	Soekarno-Hatta Int'l	CGK	1
2	Ngurah Rai Int'l	DPS	2
3	Changi Airport	SIN	3

Flights :

FlightID	AirlineID	FlightNumber	DepartureAirportID	ArrivalAirportID	DepartureTime	ArrivalTime
201	1	GA408	1	2	2025-11-10 08:00	2025-11-10 10:50
202	2	QZ7510	1	3	2025-11-11 09:30	2025-11-11 12:15
203	3	SQ950	3	1	2025-11-15 14:00	2025-11-15 15:35

Hotels :

HotelID	HotelName	CityID	Address
701	The Ritz-Carlton	2	Jl. Raya Nusa Dua, Bali
702	Marina Bay Sands	3	10 Bayfront Ave, Singapore

Bookings :

BookingID	CustomerID	BookingDate	TotalAmount
1001	1	2025-10-20	2150000
1002	2	2025-10-21	6500000
1003	1	2025-10-21	1800000

BookingItems :

BookingItemID	BookingID	ProductType	ProductID	Price
5001	1001	Flight	201	2150000
5002	1002	Flight	202	1500000
5003	1002	Hotel	702	5000000
5004	1003	Flight	201	1800000

Payments :

PaymentID	BookingID	PaymentDate	PaymentMethod	Amount	PaymentStatus
901	1001	2025-10-20	Credit Card	2150000	Success

Title	Desain Data Warehouse Online Travel Agent (OTA)				
Document Number	BI-DW-001				
Revision	1.0				
Controlled Status	Controlled				
Classification	Internal				

902	1002	2025-10-21	VA	6500000	Success
903	1003	2025-10-21	Credit Card	1800000	Pending

External Data Sources

Untuk arsitektur Data Warehouse TravelNusantara, tim Data Engineering telah mengidentifikasi dan merencanakan integrasi untuk dua kategori utama sumber data eksternal.

Sumber pertama adalah dataset historis pemesanan statis, yang disediakan dalam file csv. File ini berisi *snapshot* data transaksional historis yang mencakup detail perjalanan, demografi pelanggan, dan rincian biaya. Data ini akan diproses melalui pipeline *batch* ETL awal untuk mengisi data historis dalam tabel dimensi dan fakta. Kegunaan utamanya adalah untuk menyediakan fondasi data yang solid bagi analisis tren jangka panjang dan pelaporan BI.

Contoh data pada file csv :

Trip ID	Destination	Start date	End date	Duration (days)	Traveler name	Traveler age	Traveler gender	Traveler nationality	Accommodation type	Accommodation cost	Transportation type	Transportation cost
1	London, UK	5/1/2023	5/8/2023	7.0	John Smith	35.0	Male	American	Hotel	1200	Flight	600
2	Phuket, Thailand	6/15/2023	6/20/2023	5.0	Jane Doe	28.0	Female	Canadian	Resort	800	Flight	500
3	Bali, Indonesia	7/1/2023	7/8/2023	7.0	David Lee	45.0	Male	Korean	Villa	1000	Flight	700

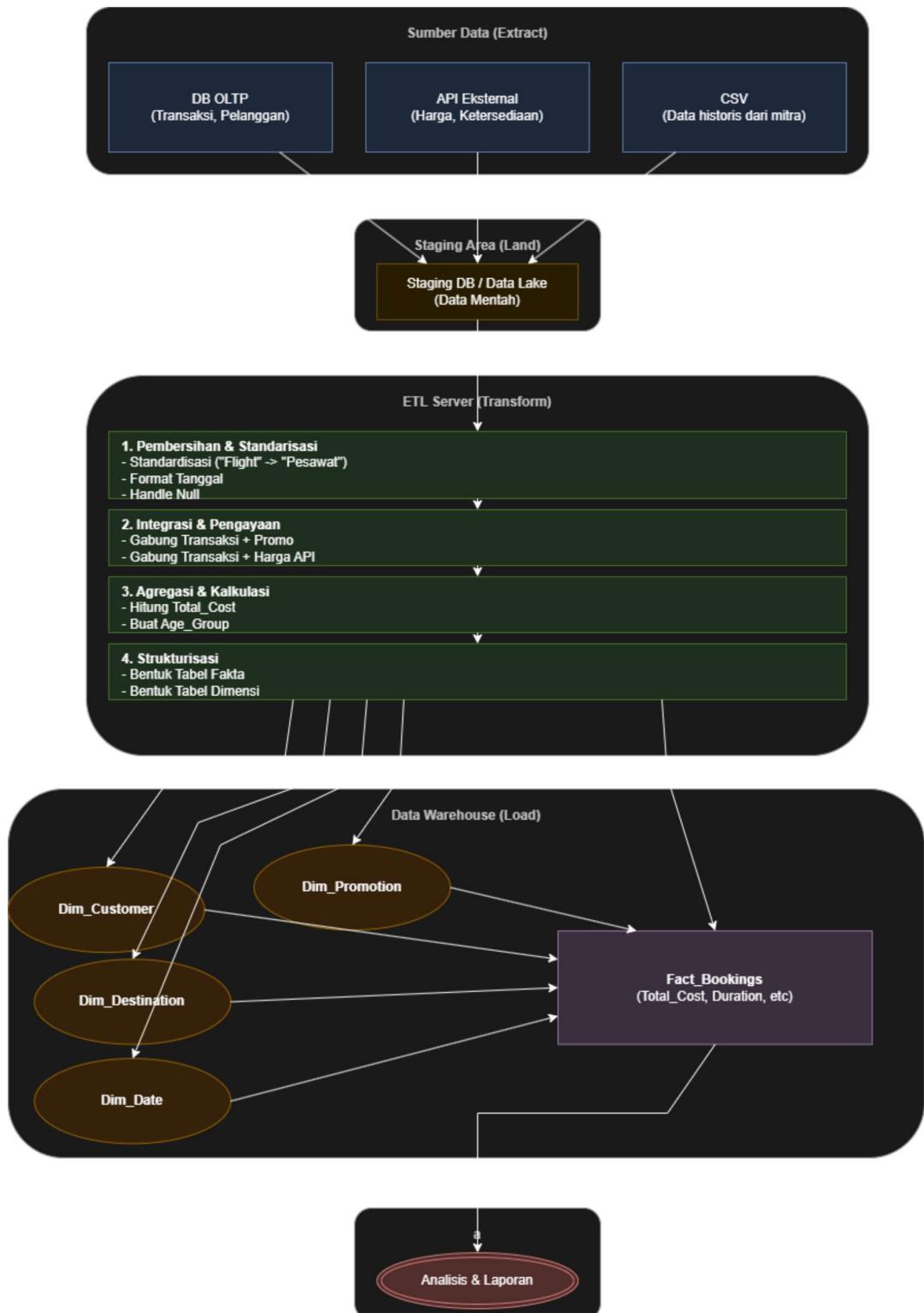
Sumber kedua, adalah API layanan perjalanan *real-time* dari Amadeus. Berdasarkan analisis kebutuhan fungsional OTA TravelNusantara, kami akan mengintegrasikan dua *endpoint* utama, yaitu Flight Offers Search API dan Hotel Search API. Kedua API ini akan menyediakan data operasional *live* yang krusial, seperti ketersediaan penerbangan, harga tiket, daftar properti hotel, dan harga kamar. Data ini akan dikonsumsi dari *endpoint* RESTful Amadeus, yang mengembalikan respons dalam format JSON. Pipeline *ingestion* data akan

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

dirancang untuk mengambil data ini secara berkala (misalnya, *micro-batch*) untuk kemudian ditransformasi dan dimuat ke dalam Data Warehouse, guna mendukung *dashboard* operasional dan analisis harga kompetitif *near-real-time*.

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

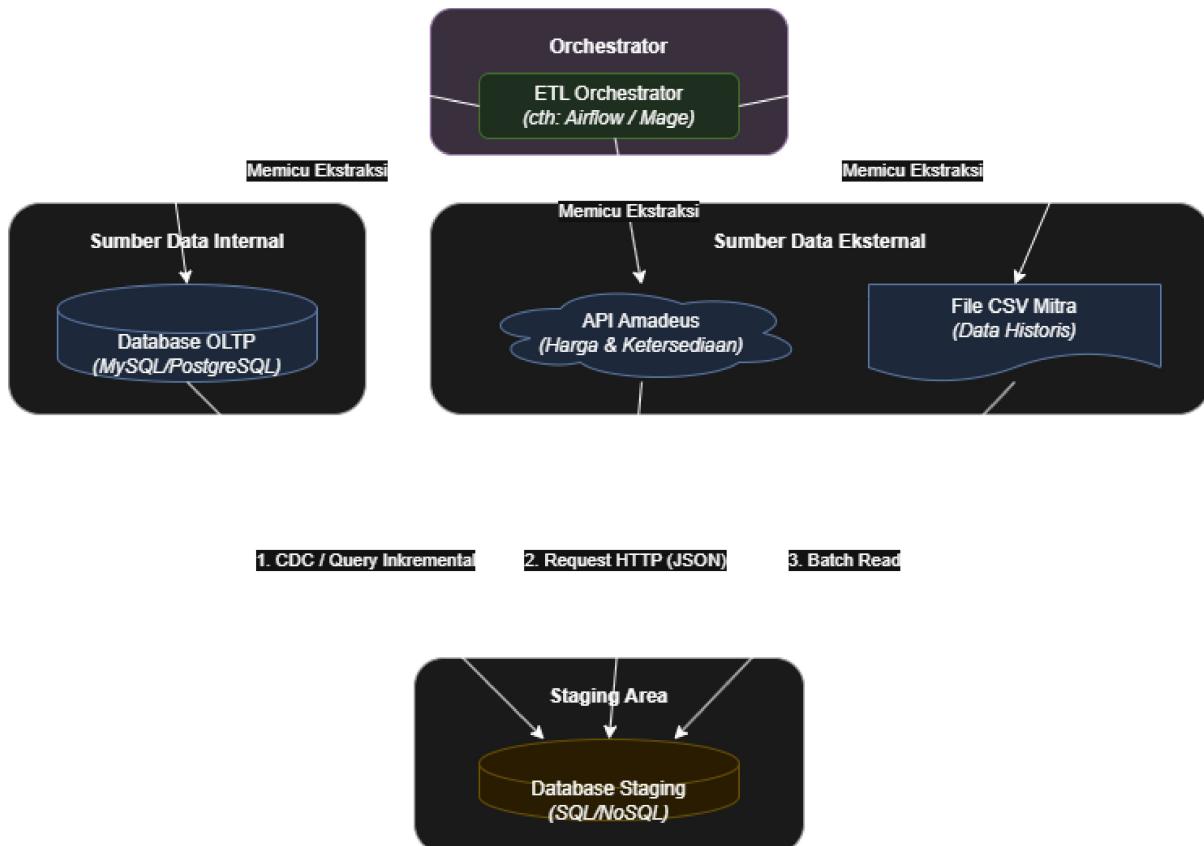
ETL (Extract, Transform, Load) Design



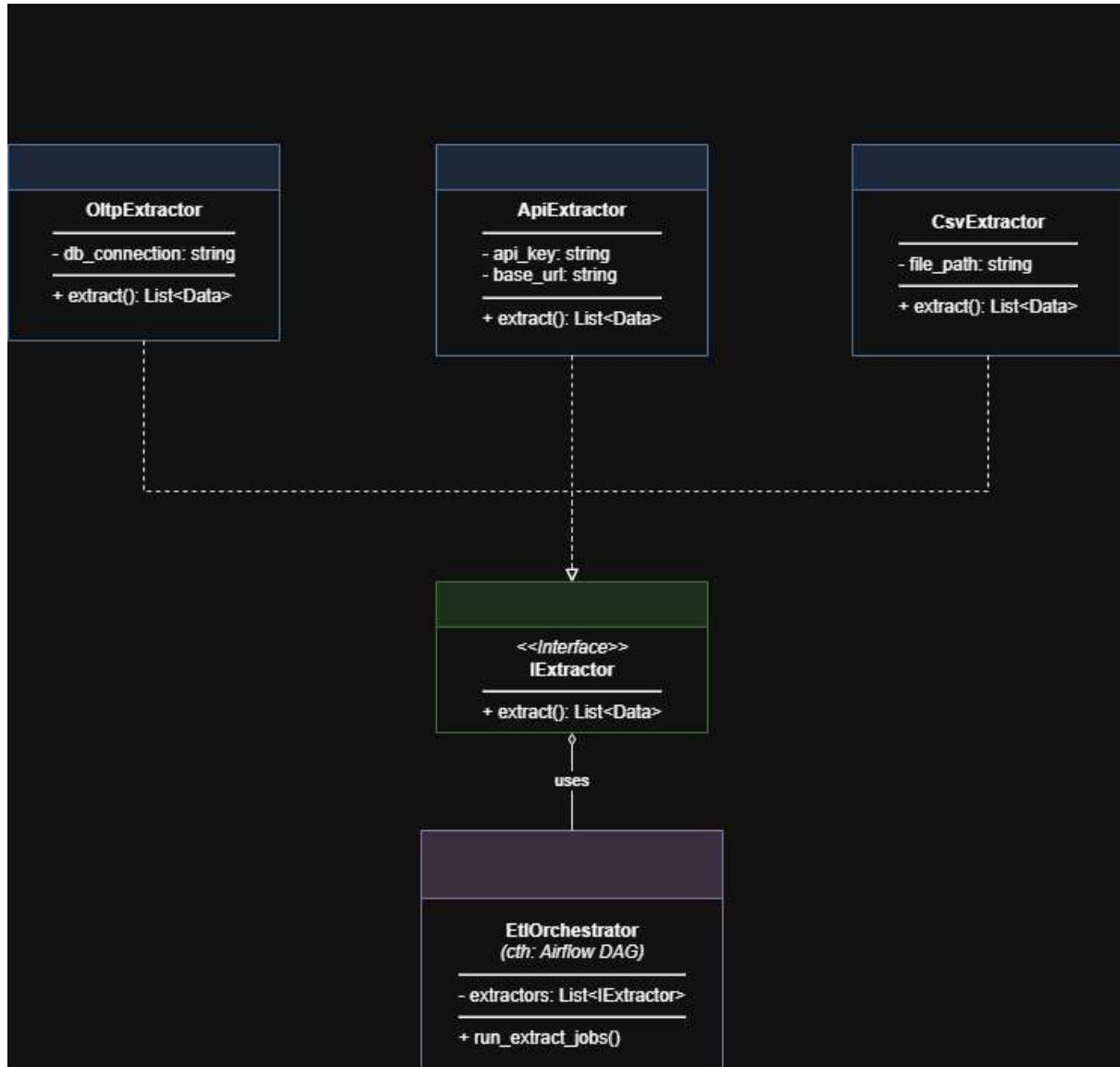
Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Extract Process

Diagram berikut mengilustrasikan bagaimana data dari berbagai sumber diekstraksi dan dikumpulkan ke dalam *Staging Area* sebelum diproses lebih lanjut.



Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal



Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Scheduling

Sumber Data	Metode Ekstraksi	Frekuensi Penjadwalan	Kebutuhan Bisnis	Tools yang Disarankan	Justifikasi Frekuensi Penjadwalan
API Amadeus	HTTP Request (JSON)	Setiap 15 Menit	Analisis harga kompetitif <i>near-real-time</i> dan ketersediaan .	Cron Job, Airflow DAG, AWS EventBridge	Data harga bersifat sangat <i>volatile</i> (mudah berubah). Pengambilan data dengan frekuensi tinggi (misal 15 menit) adalah pendekatan " micro-batch ".
Database OLTP	CDC (Change Data Capture)	<i>Event-driven</i> (Real-time)	Dashboard operasional <i>near-real-time</i> (memantau transaksi baru).	Debezium, Kafka Connect, Airflow	CDC adalah metode " event-driven " yang secara spesifik didesain untuk menangkap perubahan data (INSERT, UPDATE) saat itu juga. CDC membaca <i>transaction log</i> sehingga memiliki "beban minimal pada sistem sumber" sambil menyediakan data dalam hitungan detik.
Database OLTP	<i>Fallback:</i> Query Inkremental	Setiap 1 Jam	Update data pelanggan dan transaksi harian.	Cron Job, Airflow DAG	Jika CDC tidak memungkinkan, <i>query inkremental</i> per jam adalah metode <i>batch</i> tradisional. Frekuensi per jam dianggap cukup untuk data yang kebutuhannya "harian" (misal data pelanggan)

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

					yang jarang berubah), di mana "toleransi terhadap jeda data" (misal 1 jam) masih dapat diterima oleh bisnis.
File CSV Mitra	Batch Read (File)	1x Sehari (Pukul 02:00 Pagi)	Memuat data historis untuk analisis tren jangka panjang.	Airflow DAG, AWS Glue Job, Crontab	Data historis dari mitra umumnya statis dan berukuran besar. Praktik <i>data warehousing</i> klasik merekomendasikan pemrosesan " batch harian " untuk skenario ini. Menjalankannya pada "jam non-puncak" (<i>off-peak hours</i>) seperti pukul 2 pagi adalah strategi untuk "meminimalkan kontensi sumber daya" pada sistem. Untuk data historis yang digunakan untuk "analisis tren" jangka panjang, "prediktabilitas dan efisiensi" pemrosesan lebih penting daripada "responsivitas <i>real-time</i> ".

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Transform Process

Aturan Transformasi	Deskripsi	Contoh Data Asal (Staging)	Hasil Standarisasi (Clean)
Pembersihan Data	Menghapus baris yang kosong atau tidak valid.	Baris 128 di Travel details dataset.csv (semua kolom kosong).	Baris diabaikan/dihapus.
Standarisasi Mata Uang	Mengkonversi semua biaya (dari API dan CSV) ke mata uang tunggal (IDR) menggunakan tabel <i>lookup</i> kurs harian.	{"currency": "EUR", "total": "210.50"} (API) "Accommodation cost": 1200 (CSV, diasumsikan USD)	cost_idr: 3578500 cost_idr: 19200000
Standarisasi Kategori	Memetakan nilai-nilai yang berbeda namun bermakna sama ke satu nilai standar.	Transportation type: "Flight" (CSV), "Airplane" (CSV) Accommodation type: "Hotel", "Resort", "Villa"	transport_mode: "UDARA" accomodation_category: "HOTEL_RESORT"
Parsing & Format Tanggal	Mengubah format tanggal yang berbeda-beda (M/D/YYYY, ISO 8601) ke format standar DWH (YYYY-MM-DD).	"Start date": "5/1/2023" (CSV) "at": "2025-12-01T10:00:00" (API)	start_date_sk: 20230501 (Surrogate Key) departure_date_sk: 20251201
Pemisahan Kolom	Memecah kolom yang mengandung banyak informasi (seperti 'Destination') menjadi kolom yang lebih atomik.	"Destination": "London, UK" (CSV) "iataCode": "CGK" (API)	city: "London", country: "UK" city: "Jakarta", country: "ID"

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Integrasi & Key Generation	Mencocokkan data dari sumber berbeda (misal nama di CSV dengan customer_id di OLTP) dan membuat <i>Surrogate Key</i> (SK) untuk DWH.	Traveler name: "John Smith" (CSV) customer_id: 1001, name: "John Smith" (OLTP)	customer_sk: 1 (Menunjuk ke DimCustomer)
---------------------------------------	--	---	---

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

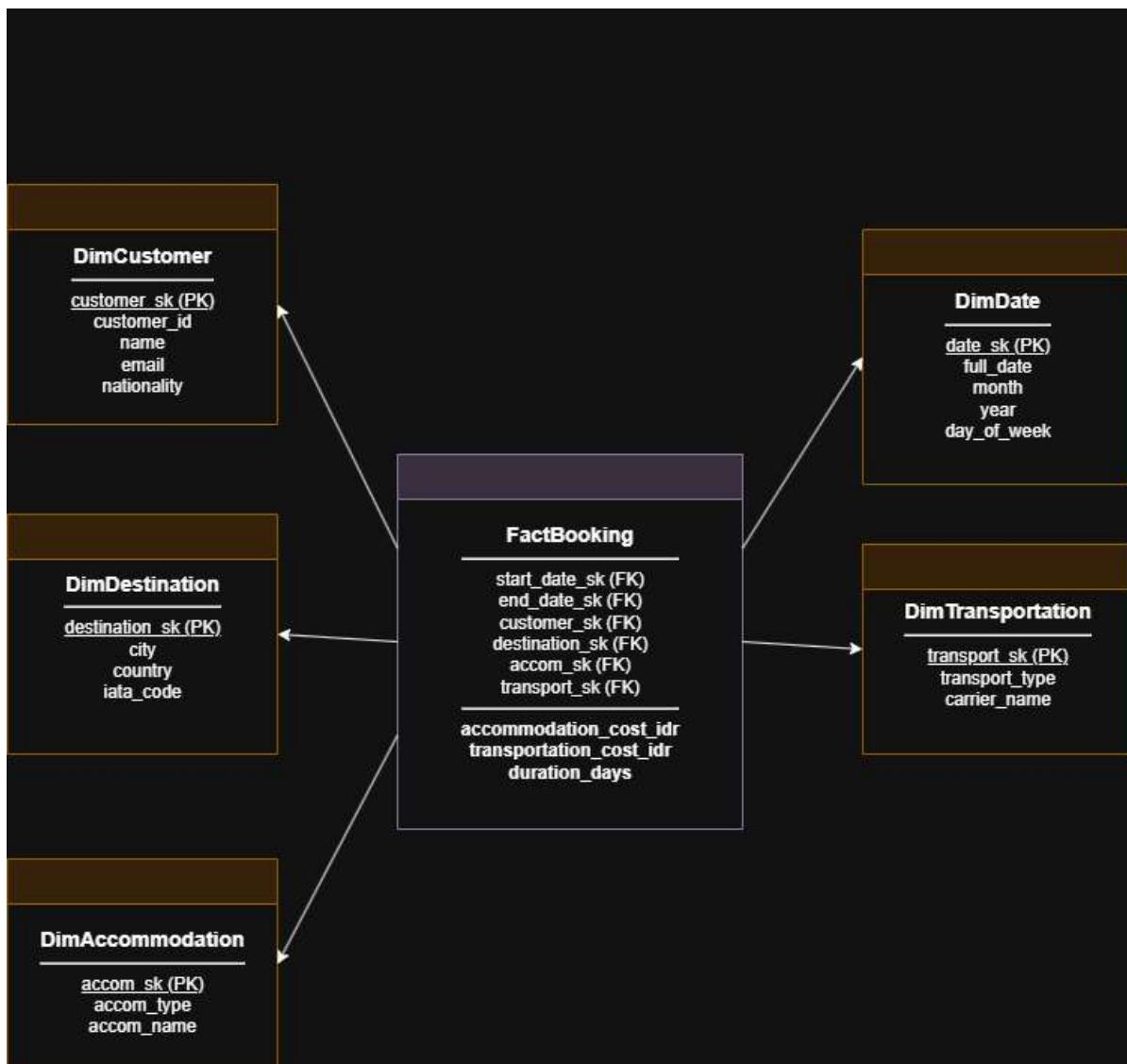
Load Process

Arsitektur: Dimensional Modeling (Kimball)

Sesuai dengan tujuan dokumen untuk analisis dan *dashboard*, arsitektur yang dipilih adalah Dimensional Modeling (Star Schema). Arsitektur ini mengutamakan kemudahan dan kecepatan *query* analitis (OLAP) dengan memisahkan data menjadi tabel Fakta (transaksi/peristiwa) dan tabel Dimensi (konteks siapa, apa, di mana, kapan).

Diagram Desain Skema (Contoh Star Schema untuk *Booking*)

Diagram ini menunjukkan satu tabel fakta (FactBooking) di tengah yang terhubung ke berbagai tabel dimensi.



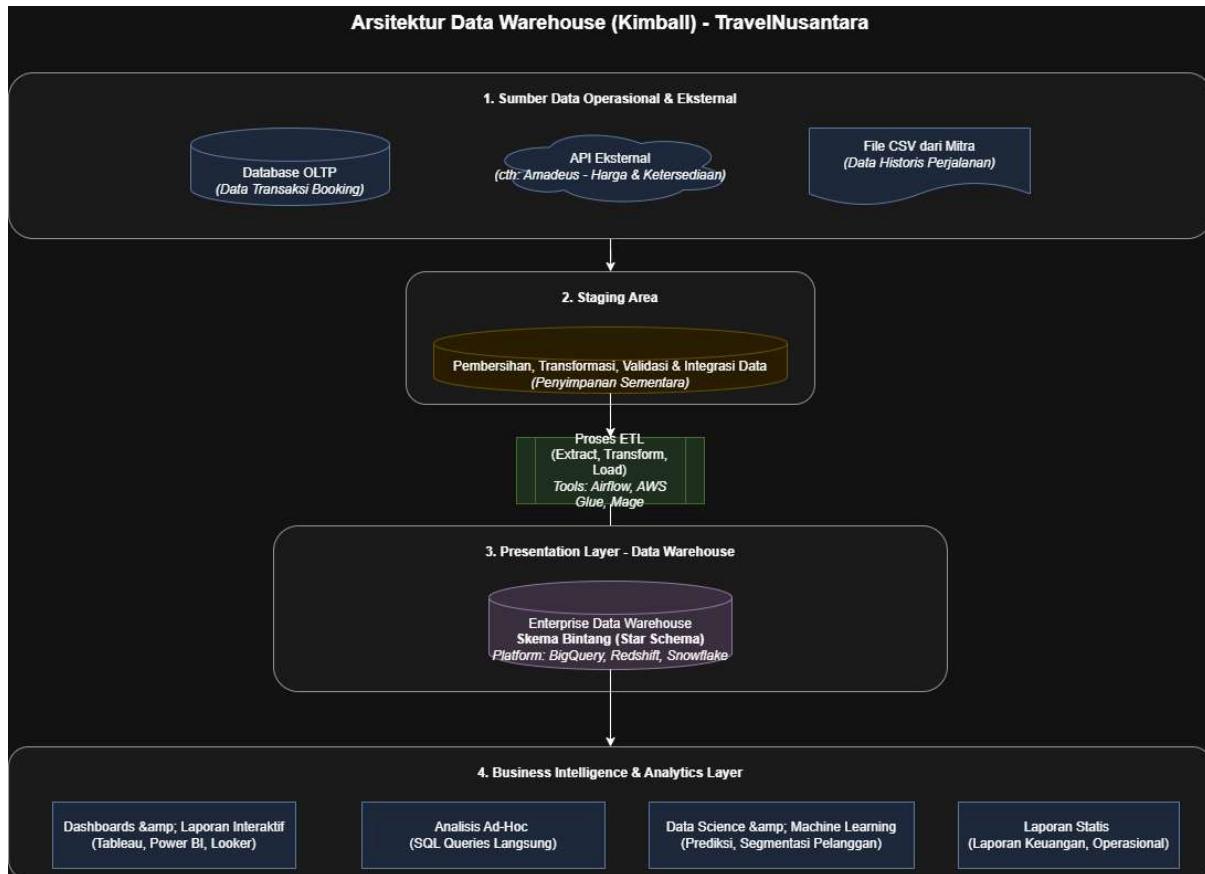
Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

Penanganan Kegagalan (Failure Handling) dan Call-back

1. **Retries dan Alerting:** Setiap task ETL (misalnya di Airflow) akan dikonfigurasi dengan kebijakan *retry* otomatis (misal: 3 kali percobaan). Jika tetap gagal, sistem akan memicu *call-back* internal untuk mengirimkan *alert* (via Email atau Slack) ke tim Data Engineering.
2. **Dead Letter Queue (DLQ):** Untuk kegagalan data spesifik (misal: data korup dari CSV atau format JSON tidak terduga dari API), *record* yang gagal diproses tidak akan menghentikan keseluruhan *batch*. Data tersebut akan dialihkan ke *Dead Letter Queue* (misalnya *bucket* GCS/S3 terpisah) untuk analisis manual, sementara data yang valid tetap lanjut diproses.
3. **Call-back Eksternal (Kegagalan API):** Jika API Amadeus mengembalikan *status code* error (misal: 5xx *Service Unavailable*), *orchestrator* akan memicu *call-back* yang menunda *job* dan mengulanginya sesuai jadwal *retry*.
4. **Transaksional Load:** Proses *load* ke Data Warehouse (Gold Layer) akan bersifat transaksional. Jika terjadi kegagalan di tengah proses *load* ke Fact_Bookings, transaksi akan di-*rollback* untuk mencegah data parsial dan menjaga konsistensi data

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal

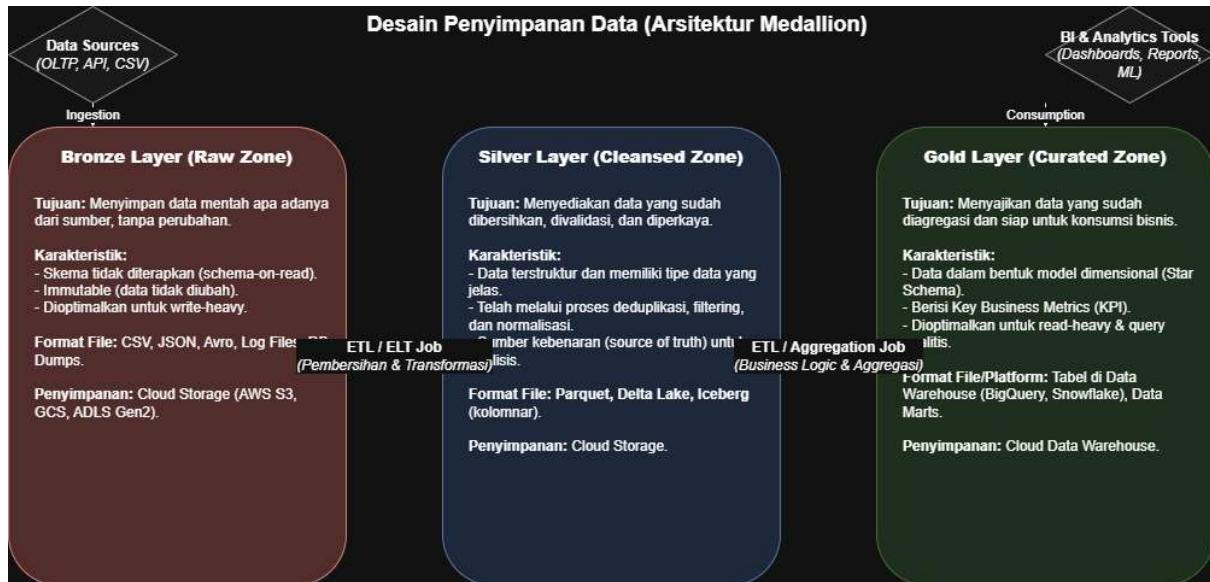
Data Warehouse Architecture



Metode Kimball dipilih karena **langsung fokus pada kebutuhan bisnis** TravelNusantara.

- **Cepat Memberi Nilai:** Pendekatan *bottom-up* memungkinkan kita membangun *data mart* per proses bisnis (contoh: pemesanan), sehingga hasilnya bisa cepat digunakan untuk analisis.
- **Optimal untuk Laporan & Dashboard:** Menggunakan **Star Schema** yang sederhana, sehingga query untuk analisis dan visualisasi data menjadi sangat cepat dan efisien.
- **Mudah Dipahami:** Model datanya (fakta dan dimensi) intuitif dan mudah dimengerti oleh pengguna bisnis non-teknis.
- **Fleksibel & Skalabel:** Dapat dikembangkan secara bertahap (inkremental) sesuai dengan pertumbuhan kebutuhan bisnis tanpa harus merombak total.

Title	Desain Data Warehouse Online Travel Agent (OTA)
Document Number	BI-DW-001
Revision	1.0
Controlled Status	Controlled
Classification	Internal



Desain Arsitektur Medallion ini dipilih untuk **menjamin kualitas dan tata kelola data** dari mentah hingga siap saji.

- **Bronze (Mentah):** Sebagai "bak penampung" data asli dari semua sumber. Ini memastikan tidak ada data yang hilang dan bisa diaudit kembali.
- **Silver (Bersih):** Di sini data dibersihkan, divalidasi, dan digabungkan menjadi satu **sumber kebenaran tunggal (single source of truth)**. Ini adalah data yang andal untuk dianalisis lebih lanjut.
- **Gold (Sajian):** Data dari Silver diubah menjadi model data yang siap pakai untuk bisnis, seperti **tabel agregat atau Star Schema**. Layer inilah yang diakses oleh tools BI untuk membuat dashboard dan laporan.