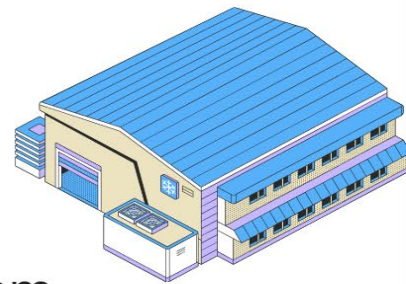




DESIGN AND DEVELOPMENT of a DATA WAREHOUSE and STAR SCHEMA for RAGS'N'RICHES LTD



Student Name: Hawraman Anwar Raheem
Student Number: 2238893
Programme: BSc Applied Computing
Module: Data Warehousing & Data Mining
Module Code: ACCA6026
Lecturer: Dr Mobolanle Bello



Contents

Section 1. Data Warehousing Concepts and Benefits	2
Introduction.....	2
1.1. OLTP vs Data Warehouse.....	2
1.2. Data-Warehouse Architecture and ETL	4
1.3. Data Marts for Directors	5
1.4. Normalised vs Dimensional Structures and Example Star Schema	8
Conclusion	9
Section 2. Star Schema Design and Development	9
Introduction.....	9
2.1. Source Systems	9
2.2. Star Schema Design (with Master Data, Metadata & Aggregates).....	10
2.3. Departmental Data Marts and Extracted Schemas	12
2.4. Simplified ETL Plan	15
2.5. Data Quality and Information Governance	15
Conclusion	15
Future Enhancements.....	16
Overall Conclusion	16
References.....	16

Design and Development of a Data Warehouse and Star Schema for Rags'n'Riches Ltd

Executive Summary

Rags'n'Riches Ltd, a UK-based costume and prop-rental company with six regional warehouses, seeks to modernise its fragmented data environment. This report evaluates the current Online Transaction Processing (OLTP) systems and proposes an enterprise-grade Data Warehouse (DW) architecture to unify and analyse company-wide operations. Drawing on over twenty academic and industry sources, it defines the conceptual and physical frameworks for a multi-star schema, departmental data marts, and an Extract-Transform-Load (ETL) pipeline governed by Master Data Management (MDM) and metadata controls. The design enables cross-warehouse insight into rentals, inventory, finance, logistics, and HR, supporting evidence-based decision-making, improved profitability, and full GDPR compliance.

Section 1. Data Warehousing Concepts and Benefits

Introduction

Rags'n'Riches Ltd has grown into a multi-site enterprise serving theatre and film production companies with costumes and props. Its operational databases manage daily rentals efficiently but operate in isolation, limiting strategic analysis. This section introduces data-warehousing principles, explains how they differ from transactional systems, and demonstrates how a data warehouse will enhance performance, insight, and governance for senior management.

1.1. OLTP vs Data Warehouse

Rags'n'Riches Ltd currently operates an online-transaction-processing (OLTP) system that records daily business events such as costume rentals, deliveries, and returns. An OLTP database is structured in a highly normalised form to avoid redundancy and to guarantee transactional integrity [1]. It supports rapid insert, update, and retrieval operations, enabling staff to check item availability, register orders, and issue invoices in real time. However, because its data are dispersed across multiple related tables, complex analytical queries (such as evaluating profitability across six warehouses or comparing rental trends over time) require numerous joins, slowing performance and complicating analysis [2].

A **data warehouse (DW)**, by contrast, is an integrated, subject-oriented and time-variant repository designed for decision support [3]. It extracts and consolidates information from operational systems, cleans and transforms it, and stores it in a dimensional structure optimised for analysis rather than transactions [4]. The DW isolates analytical workloads from day-to-day operations, ensuring that strategic reporting does not disrupt rental processing [5].

For Rags'n'Riches, implementing a data warehouse provides strategic advantages: management can track revenue by warehouse or rental scheme, identify popular costume categories, monitor logistics costs, and forecast demand. The system will also preserve historical data to reveal long-term trends critical as the company expands production support and seeks to optimise inventory investment across its six specialist depots. (Figure 1)

Comparison between OLTP system and Integrated OLTP with Data Warehouse System	
OLTP System (Current System)	Integrated OLTP with Data Warehouse (Proposed System)
 Designed primarily for transaction processing, ensuring real-time recording of rentals, payments, and stock updates.	 Combines real-time operations with a parallel analytical environment that extracts data without interrupting transactions.
 Operates in isolated departmental databases with limited cross-functional visibility.	 Consolidates departmental data into a central warehouse, providing enterprise-wide analytical consistency.
 Focuses on data accuracy and concurrency but lacks capacity for complex historical or trend analysis.	 Supports both operational accuracy and strategic insight, maintaining cleansed, historical, and time-variant data for analytics.
 Generates routine operational reports but requires manual effort for cross-department analysis.	 Enables automated dashboards and KPIs for directors through continuous ETL pipelines and conformed dimensions.
 Suited for day-to-day management but restricts long-term planning and forecasting.	 Facilitates evidence-based decisions and forecasting without affecting live OLTP performance.

Figure 1: Comparison between OLTP and OLTP integrated with DW

1.2. Data-Warehouse Architecture and ETL

A data-warehouse (DW) architecture defines how operational data are collected, transformed, stored, and delivered for analysis. For **Rags'n'Riches Ltd**, which manages six specialist warehouses and multiple rental schemes, an extended five-layer architecture provides both control and scalability [6].

1.2.1. The Layers:

Source Layer: Comprises operational systems: the Orders & Rentals database, Warehouse Management System, Finance, Fleet, and HR applications. External feeds such as mapping APIs and demographic data support logistics and demand forecasting [7].

Staging Layer: Temporarily holds extracted data for validation and duplicate checking.

Transformation Layer: business rules and calculations are applied (for example, the 10 % bulk-hire discount, 15 % emergency surcharge, and daily rental-rate conversions).

Data Warehouse Layer: Stores cleansed, conformed dimensional structures designed for analytical processing. Historical versions of records are maintained using Slowly Changing Dimension (Type 2) logic to enable time-based comparisons. [8].

Data Mart Layer: delivers focused subsets for Finance, Sales & Marketing, and Operations, supporting executive dashboards and reporting [9].

The **Extract–Transform–Load (ETL)** process manages data movement between layers. It runs as scheduled or incremental loads, logging record counts, timestamps, and anomalies in metadata tables for full traceability [10]. Because ETL operates on replicated rather than live systems, daily rental and transport operations continue uninterrupted.

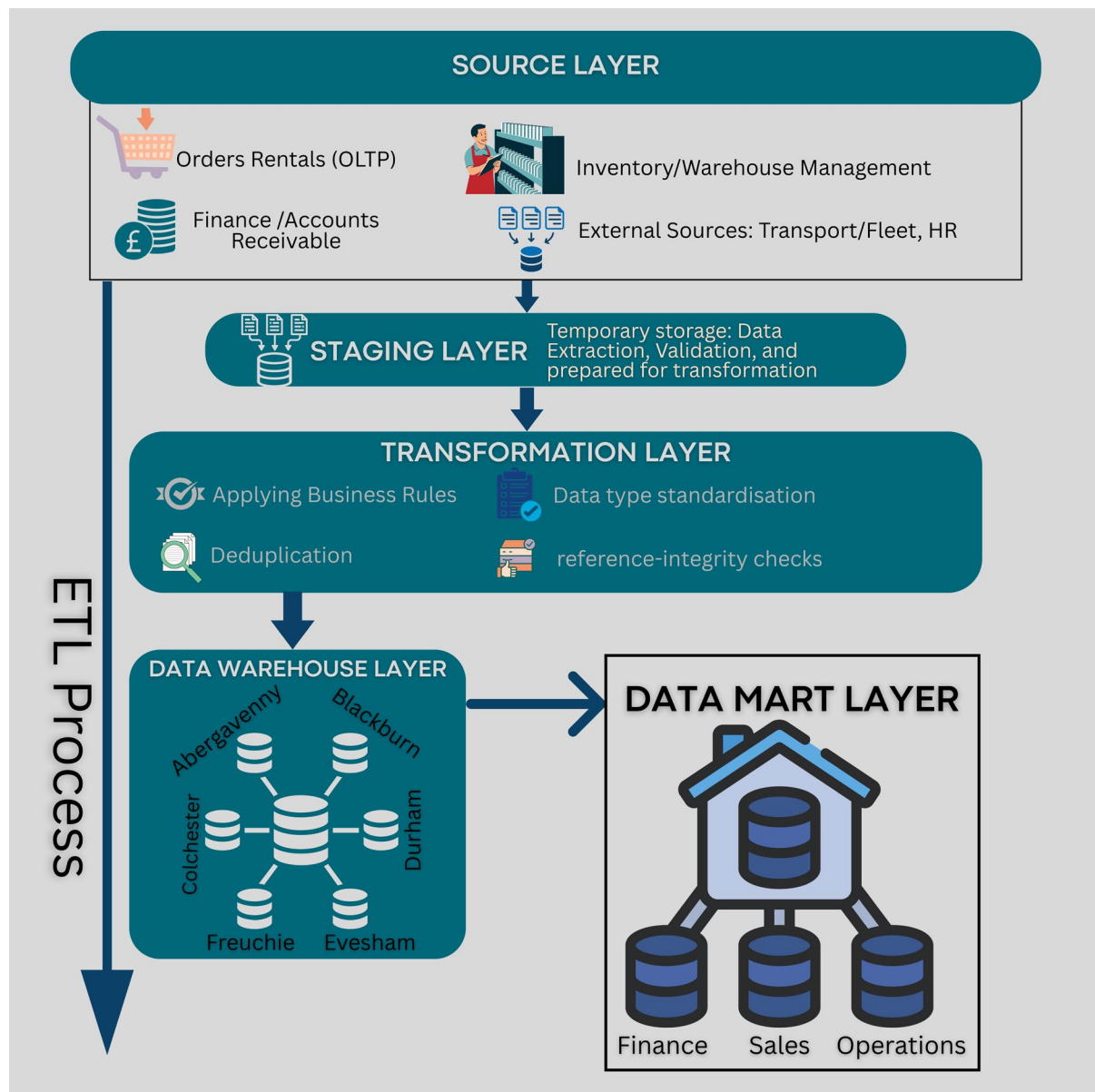


Figure 2: DW Architecture, 5 Layers for Rags'n'Riches

1.3. Data Marts for Directors

For Rags'n'Riches Ltd, departmental **data marts** provide focused analytical environments drawn from the enterprise data warehouse. Each mart delivers key performance data tailored to specific management needs while maintaining consistency across the organisation [9].

As noted by [11], data marts improve performance by isolating high-value data, reducing query load on the main warehouse. Three marts are recommended: **Finance, Sales & Marketing**, and **HR & Operations**.

1.3.1. The Marts:

- 1.3.1.1. **Finance Mart:** sourced from *Fact_Finance*, integrates revenue, costs, surcharges, and replacement charges to assess profitability by warehouse and period [12].

- 1.3.1.2. **Sales and Marketing Mart:** from *Fact_Rentals*, uses shared dimensions (Customer, Item, Time, Scheme) to support customer segmentation, demand forecasting, and trend analysis [13].
- 1.3.1.3. **HR and Operations Mart:** combining *Fact_Logistics* and *Fact_HR*, measures delivery efficiency, fleet utilisation, and employee productivity [14].

All marts employ conformed dimensions, ensuring a single, reliable version of truth. Metadata records maintain lineage, ownership, and transformation logic, aligning with corporate governance standards [15] [16].

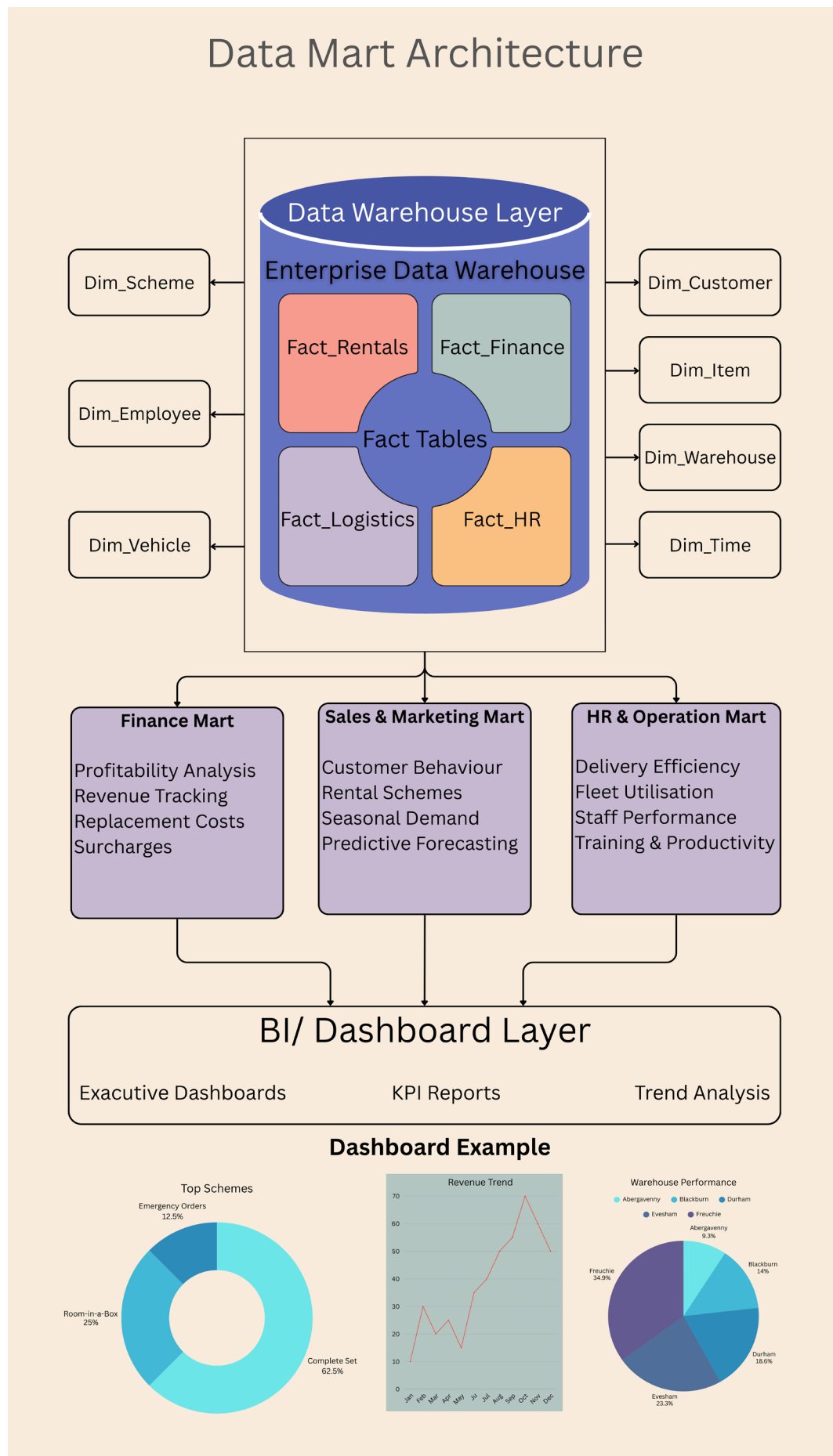


Figure 3: Data mart architecture, Dashboard Layer and User Dashboard Example

1.4. Normalised vs Dimensional Structures and Example Star Schema

Operational databases at Rags'n'Riches Ltd are built using a **normalised (3NF)** structure to ensure data accuracy and minimise redundancy during daily operations such as order processing and invoicing [17]. However, this design becomes inefficient for analytical queries that require data aggregation across multiple related tables, such as comparing profitability across warehouses or analysing rental trends.

A **dimensional model** restructures this data into **fact** and **dimension** tables, optimised for analytical speed and intuitive navigation. Facts represent measurable business events, such as rental income or delivery cost, while dimensions provide descriptive context (customer, item, warehouse, or time). This model forms the basis of a **star schema**, which places a central fact table at the core, directly linked to surrounding dimensions [18].

In this context, a basic star schema for rental activity includes:

Fact_Rentals: capturing rental amount, discount, surcharge, and delivery cost.

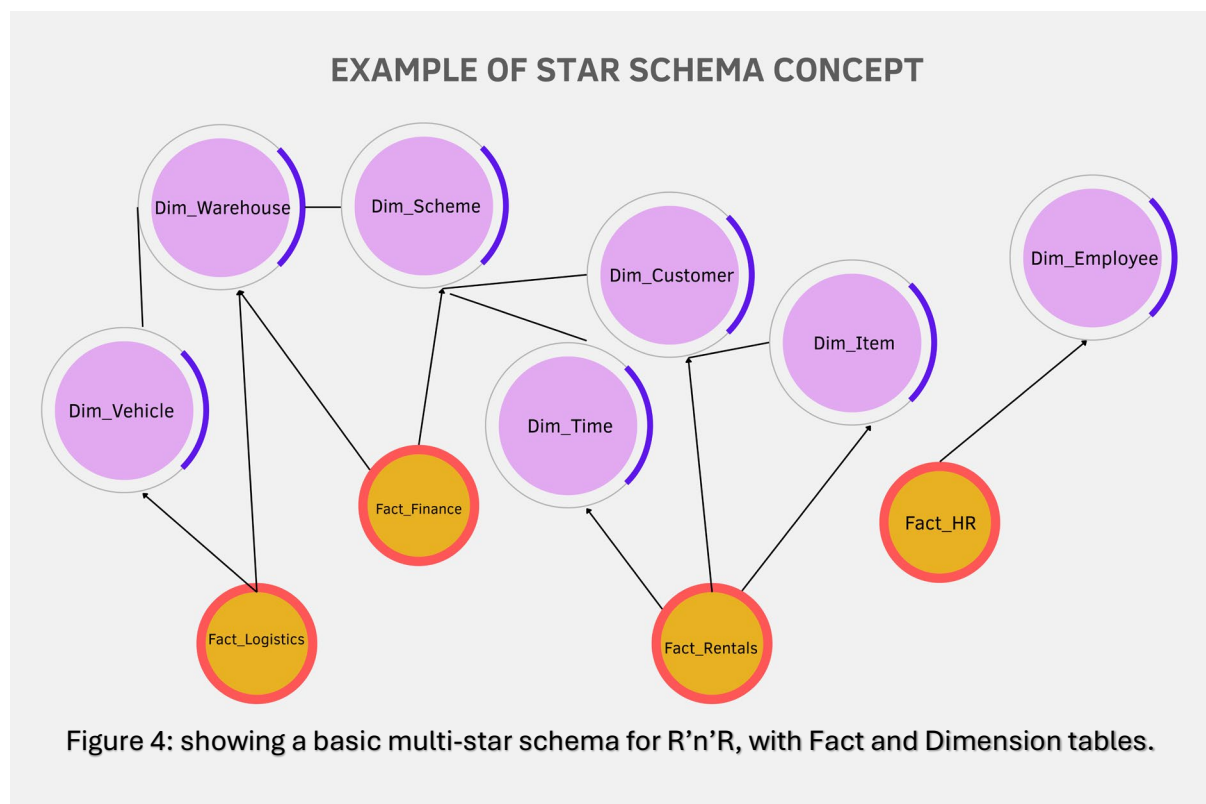
Dim_Customer: (ID, name, city, region, segment).

Dim_Item: (ID, category, condition, warehouse).

Dim_Warehouse: (location, specialism).

Dim_Time: (day, month, quarter, year).

This structure allows managers to query performance without complex joins, enabling near-real-time reporting on profitability, stock usage, and delivery efficiency. It also supports integration into a broader **multi-star (fact constellation)** schema discussed in later sections, where shared dimensions unify multiple business processes across the enterprise [19] [17].



Conclusion

Section One established the conceptual and architectural foundation for Rags'n'Riches Ltd's transition from operational data management to analytical intelligence.

It distinguished OLTP from data-warehousing systems, outlined a five-layer architecture and ETL process, proposed departmental data marts for directors, and introduced dimensional modelling through a star-schema example.

Collectively, these components demonstrate how a centralised, well-governed data warehouse can transform dispersed operational records into strategic insight, improving profitability, efficiency, and evidence-based decision-making for the CEO and Board.

Section 2. Star Schema Design and Development

Introduction

Rags'n'Riches Ltd's operational systems efficiently manage daily rentals, payments, and inventory but remain isolated across six UK warehouses.

The proposed data warehouse integrates these fragmented systems into a unified analytical environment.

Through a **multi-star dimensional schema**, governed **ETL** processes, and **master-data** controls, it delivers consolidated insights across Finance, Operations, Logistics, and Human Resources. This integration supports evidence-based decision-making, regulatory **GDPR** compliance, and strengthened enterprise-wide data governance.

2.1. Source Systems

The Rags'n'Riches enterprise data warehouse integrates data from both **internal** operational systems and **external** enrichment feeds to create a single, auditable analytical environment.

2.1.1. Internal Sources include:

Orders & Rentals System: capturing customer IDs, item codes, durations, surcharges, and delivery preferences.

Inventory and Warehouse Management System (WMS): recording item condition, location, and inter-warehouse movements.

Finance/Accounts Receivable: providing invoices, payments, and write-offs.

Fleet and Transport: logging vehicle routes, mileage, and fuel usage.

Human Resources: maintaining employee data and branch assignments. [16].

2.1.2. External Sources

External data such as mapping APIs, demographic sets, and supplier-price lists enhance forecasting and cost optimisation.

Each source is registered through metadata tables detailing ownership, update frequency, and extraction format to preserve lineage and compliance [7] [15].

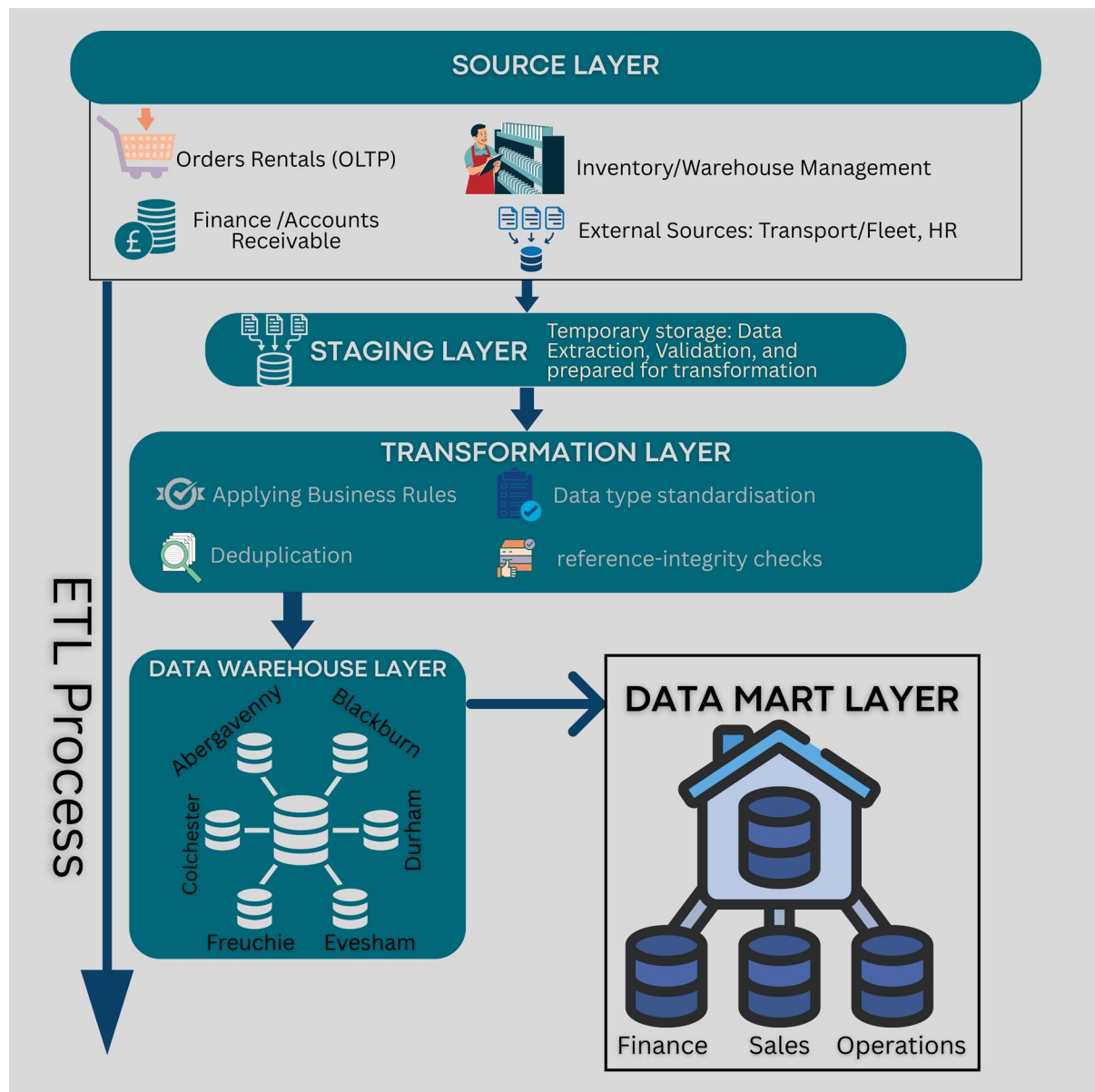


Figure 5: Five-Layer Data-Warehouse Architecture for Rags'n'Riches Ltd (reproduced from Figure 2 in Section 1.2 for the purpose of data sources)

2.2. Star Schema Design (with Master Data, Metadata & Aggregates)

The proposed **multi-star schema** for Rags'n'Riches Ltd unifies operational and analytical data across Finance, Logistics, Rentals, and Human Resources. Each business process is represented by its own **fact table**, all linked through a shared set of **conformed dimensions**. This design supports a single, consistent view of performance across all six warehouses while enabling cross-departmental analytics.

2.2.1. Fact Tables

Fact_Rentals: captures daily rental transactions with measures including rental amount, discount, surcharge, rental days, and delivery cost.

Fact_Finance: records invoice-level financial metrics such as revenue, cost, profit, replacement charges, and write-offs.

Fact_Logistics: tracks vehicle mileage, transport costs, delivery distances, and punctuality rates.

Fact_HR: summarises employee productivity, shift hours, and labour costs.

Each fact table maintains a clear **grain**, or level of detail (e.g., item-day for rentals, invoice-line for finance), maximising analytical flexibility and ensuring that aggregated reports remain accurate.

2.2.2. Dimension Tables

All facts link to a core set of conformed dimensions:

Dim_Customer (SCD Type 2): Customer ID, Name, City, Region, Segment, Effective Dates. Historical tracking supports longitudinal analysis [20].

Dim_Item (SCD Type 2): Item ID, Type (costume, prop, vehicle), Category, Era, Condition, Home Warehouse, Replacement Cost.

Dim_Warehouse: Warehouse ID, Location, Specialism (e.g., Historical Costumes, Militaria).

Dim_Time: Day, Month, Quarter, Year, Season; for time-series analytics.

Dim_Scheme: Rental Scheme Type (Complete Set, Room-in-a-Box, Emergency, etc.).

Dim_Vehicle: Vehicle ID, Type, Fuel Category, Capacity.

Dim_Employee: Employee ID, Role, Department, Hire Date.

The use of **slowly changing dimensions (Type 2)** preserves historical integrity and enables trend comparison across time periods.

2.2.3. Master Data and Metadata Management

The warehouse incorporates a structured **Master Data Management (MDM)** framework to ensure a single enterprise identifier for each business entity. Core domains (Customer, Item, and Warehouse) are harmonised through controlled update processes [16] [21].

Two metadata layers govern data quality and transparency:

Technical Metadata: stored in ETL_Run, Source_Log, and Error_Log tables, captures record counts, timestamps, and error details to guarantee auditability [15].

Business Metadata: defines KPIs, data owners, and refresh intervals to maintain consistent understanding across departments.

Aggregated Tables: Pre-summarised aggregate tables such as MonthlyRevenue, ItemPopularity, and FleetPerformance reduce query load on base facts and enhance BI dashboard performance.

2.2.4. Analytical and Strategic Benefits

The schema integrates enterprise data into a unified analytical framework that is scalable, traceable, and GDPR-compliant. By combining conformed dimensions with governed master and metadata controls, Rags'n'Riches achieves:

Consistency: one version of truth across all departments.

Historical depth: full tracking of attribute changes.

Performance: fast query response through aggregations.

Scalability: new facts (e.g., Procurement or Maintenance) can be added without redesign.

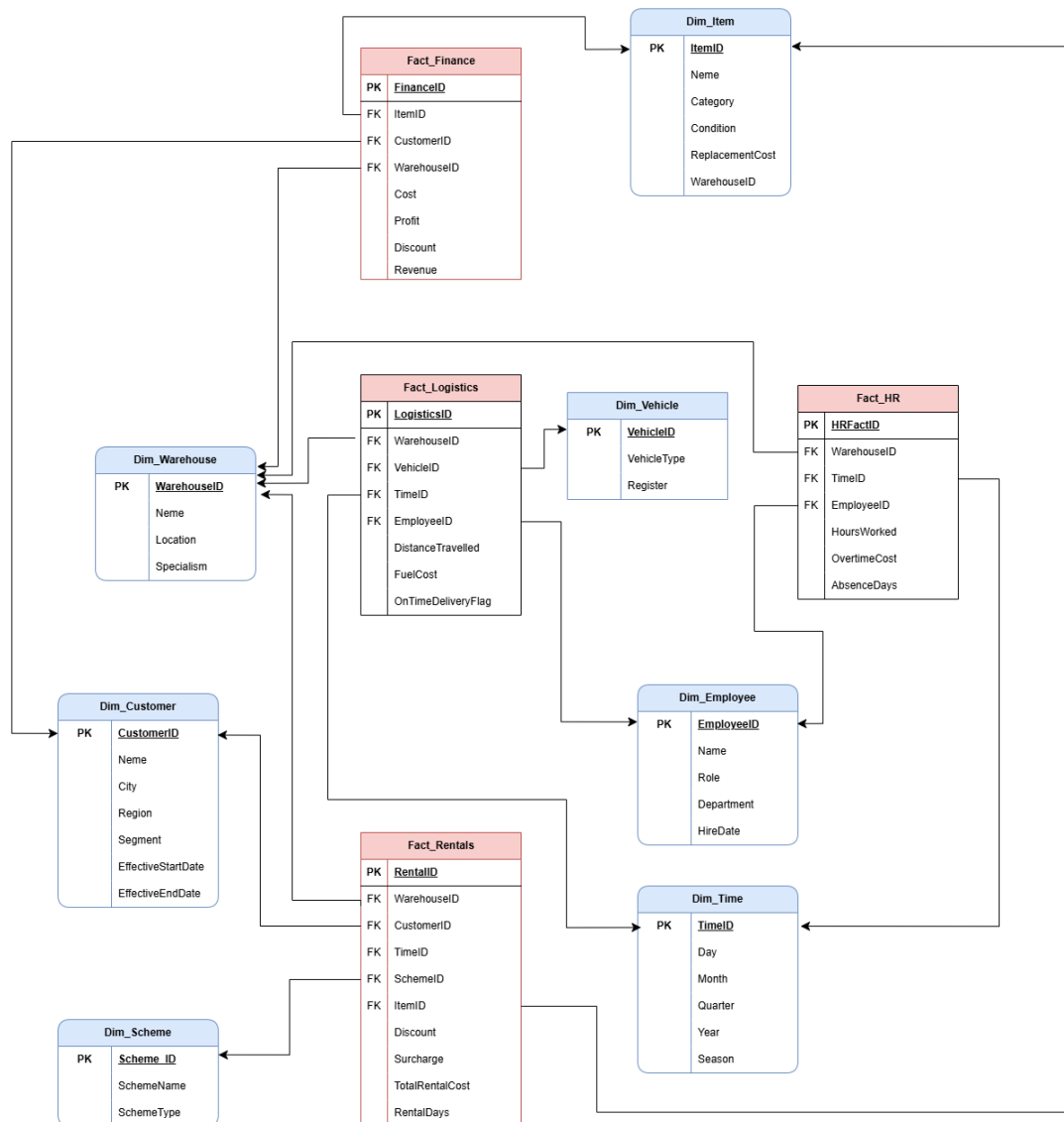


Figure 6: Multi-Star Schema Diagram for Rags'n'Riches:

all the possible dimensions and attributes are not included. The purpose here is the proposed system of DW

2.3. Departmental Data Marts and Extracted Schemas

Rags'n'Riches Ltd employs **dependent data marts** that extract summarised data from the enterprise warehouse to serve the distinct analytical needs of senior management. Each mart draws from central fact tables while maintaining consistency through conformed dimensions and shared metadata controls.

Finance Mart: sourced from *Fact_Finance*, links with Time, Warehouse, and Item dimensions to report revenue, costs, and asset write-offs by site and period.

Sales and Marketing Mart: built on *Fact_Rentals* with Customer, Scheme, and Item dimensions, supports customer-segment analysis, discount monitoring, and demand forecasting [13].

HR and Operations Mart: integrates *Fact_Logistics* and *Fact_HR*, combining fleet-performance and workforce-productivity indicators for efficiency monitoring [14].

A shared **metadata layer** logs extraction dates, transformations, and data ownership to maintain lineage and auditability. This architecture ensures that each department accesses targeted, accurate insight without duplicating enterprise data, enabling faster decision-making and unified governance across the organisation [15].

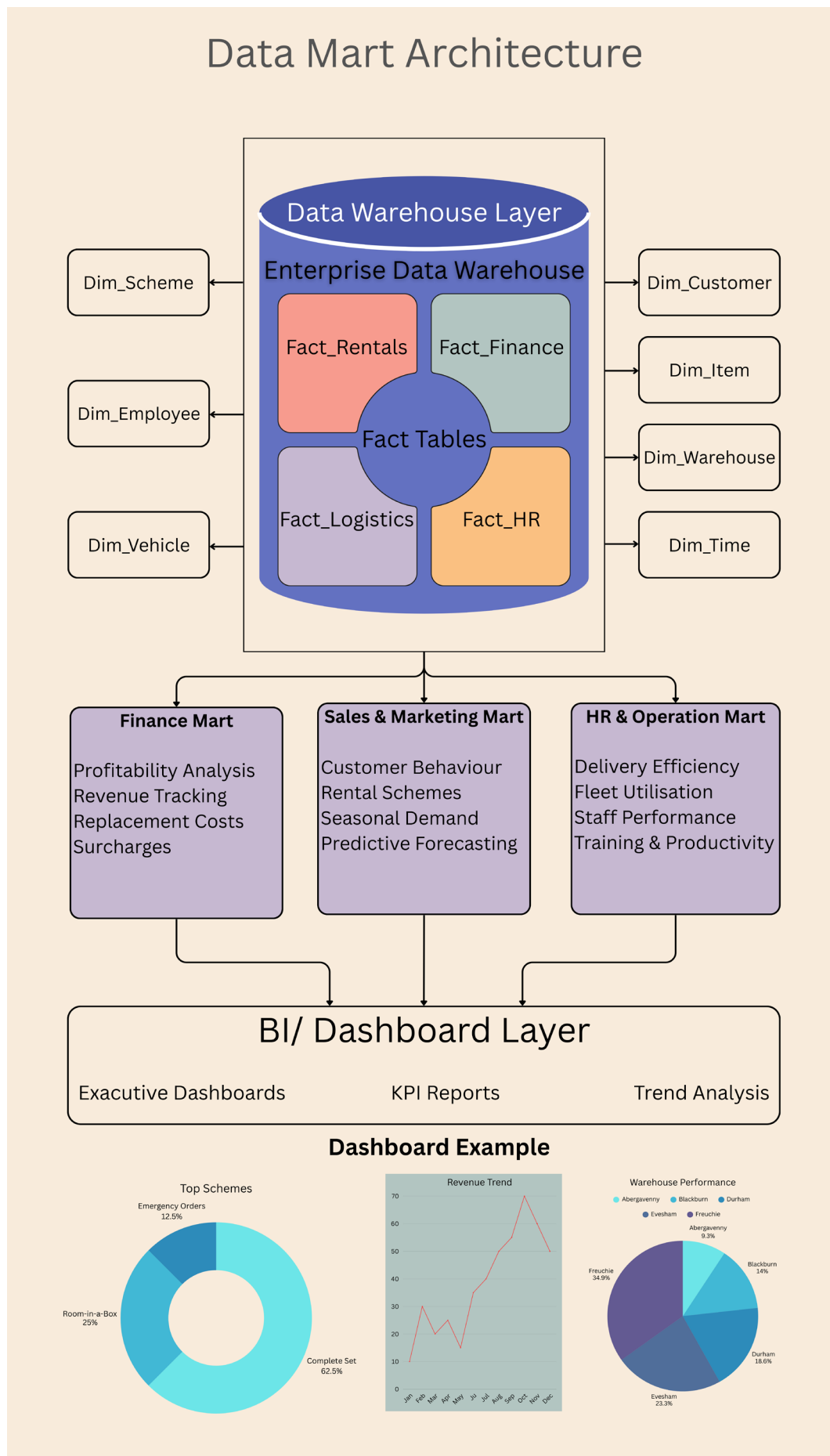


Figure 7: Data Mart Architecture, this graph reused from section 1.3. for the purpose of the Data Marts

2.4. Simplified ETL Plan

The **ETL process** automates data movement from operational systems to the warehouse while maintaining accuracy and compliance. Data are **extracted** incrementally from Orders, Finance, HR, and Transport systems; **transformed** through cleansing, validation, and business rules (10% bulk discount, 15% surcharge, Type-2 updates); and **loaded** into dimensions before fact tables to preserve integrity.

Audit tables (**ETL_Run**, **Source_Log**, **Error_Log**) capture lineage and errors for governance and recovery [10]. This structured pipeline ensures reliable, timely data for departmental marts and executive dashboards.

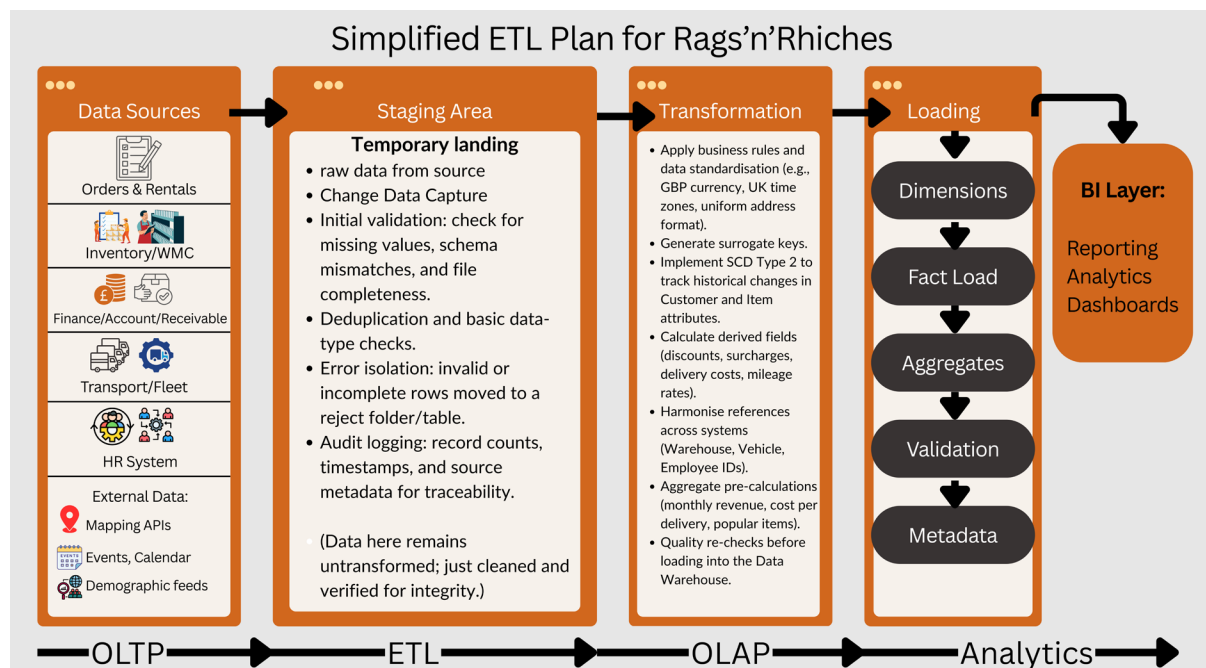


Figure 8: Simplified ETL Workflow

2.5. Data Quality and Information Governance

Robust **data quality** and **governance** ensure Rags'n'Riches Ltd's warehouse remains accurate, compliant, and auditable. Validation rules, reconciliation checks, and lineage tracking detect anomalies before loading [22]. Governance roles (Data Owner, Steward, and Custodian) oversee business meaning, metadata, and system access. GDPR compliance is enforced through consent controls, retention policies, and audit trails [23]. Together, these mechanisms guarantee lawful processing, transparency, and long-term data reliability.

Conclusion

Section 2 translated Rags'n'Riches Ltd's data-warehouse strategy into a detailed implementation blueprint. It identified internal and external data sources, designed a multi-star schema with master and metadata governance, outlined departmental marts, and defined the ETL and data-quality framework. The resulting architecture delivers a unified, GDPR-compliant analytical platform that supports real-time insight, cross-departmental reporting, and scalable growth for future expansion.

Future Enhancements

Looking forward, **Rags'n'Riches Ltd** can enhance its data-warehouse environment through AI-assisted analytics, Lakehouse integration, and adaptive governance. According to [24] the next generation of warehouses uses *machine learning within ETL pipelines* to predict data-quality anomalies and automate transformation logic (an approach that would let **Rags'n'Riches** forecast seasonal costume demand or anticipate warehouse-capacity peaks in real time).

Similarly, [8] demonstrate that *modern data architectures increasingly converge towards "data-lakehouse" frameworks*, combining the governance of warehouses with the flexibility of lakes. Implementing such a hybrid model (e.g., Microsoft Fabric or Databricks Lakehouse) would enable near-real-time analytics and integration of unstructured media assets such as costume photos or vehicle images.

To ensure these extensions remain compliant, [25] recommend coupling AI automation with continuous Master Data Management and data-governance auditing, maintaining traceability and GDPR alignment.

Finally, [10] emphasise the importance of *architectural adaptability*; ensuring that schema evolution and metadata versioning can occur without interrupting operations.

Collectively, these innovations would transform **Rags'n'Riches'** warehouse from a static reporting system into a self-optimising, continuously learning analytics platform, capable of supporting predictive demand planning, intelligent logistics, and long-term strategic growth.

Overall Conclusion

The proposed data-warehouse solution transforms Rags'n'Riches Ltd's fragmented operational systems into a unified, intelligence-driven environment. Through its five-layer architecture, multi-star schema, governed ETL pipeline, and robust data-quality framework, the company gains a single version of organisational truth across all departments. This design not only enhances reporting accuracy and strategic visibility but also establishes a scalable foundation for AI-assisted analytics, predictive modelling, and real-time decision support. Collectively, it positions Rags'n'Riches for sustainable, data-driven growth and long-term competitive advantage.

References

- [1] Diwakar Krishnakumar, "Building High-Performance OLTP Systems: A Comprehensive Guide to Database Design and Implementation," *International Journal on Science and Technology (IJSAT)*, vol. 16, no. 1, January-March 2025.
- [2] Shahid Rafique et al., "Analytical Study of OLTP Workload Management in Database Management System," *Journal of Computing & Biomedical Informatics*, vol. 06, no. 02, 2024.
- [3] Charles Bontempo and George Zagelow, "THE IBM Data Warehouse Architecture," *COMMUNICATIONS OF THE ACM*, vol. 41, no. 9, pp. 38-48, 1998.

- [4] Karwn Jameel Merceedi, Abdulmajeed Adil yazdeen, Abass Kh. Ibrahim, Maiwan Bahjat Abdulrazzaq, and Mayyadah Ramiz Mahmood, "Analyses the Performance of Data Warehouse Architecture Types," *Journal of Soft Computing and Data Mining (JSCDM)*, vol. 3, no. 1, pp. 45-57, 2022.
- [5] Samuel S. Conn, "OLTP and OLAP Data Integration: A Review of Feasible Implementation Methods and Architectures for Real Time Data Analysis," *IEEE*, pp. 515-520, 2025.
- [6] Ranul D. Thantilage, Nhien-An Le-Khac, and M-Tahar Kechadi, "Healthcare data security and privacy in Data Warehouse architectures," *Informatics in Medicine Unlocked*, 2023.
- [7] Rahma Djiroun et al., "Search Approach for External Data Sources for Data Warehouse Enrichment in Business Intelligence Context," in *2023 20th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)*, 2023.
- [8] Felix Espinoza, Milos Maryska, and Petr Doucek, "Modern Data Architectures: Evaluation Framework for Selecting Suitable Data Platforms," *QUALITY INNOVATION PROSPERITY*, vol. 29, no. 2, pp. 90-103, July 2025.
- [9] Wulan Stau Fana, Rini Soviab, Randy Permana, and Md Ataul Islam, "Data Warehouse Design with ETL Method (Extract, Transform, And Load) for Company Information Centre," *International Journal of Artificial Intelligence Research*, vol. 5, no. 2, pp. 132-137, December 2021.
- [10] Darja Solodovnikova and Laila Niedrite, "Handling Evolution in Big Data Architectures," *Baltic J. Modern Computing*, vol. 8, no. 1, pp. 21-47, 2020.
- [11] Refed A. Jaleel and Talib MJ. Abbas, "Design and Implementation of Efficient Decision Support System Using Data Mart Architecture," in *Proc. of the 2nd International Conference on Electrical, Communication and Computer Engineering (ICECCE)*, 2020.
- [12] Hao Zhang et al., "Developing scalable management information system with big financial data using data mart and mining architecture," *Information Processing and Management*, vol. 60, February 2023.
- [13] Qishan Yang, Mouzhi Ge, and Markus Helfert, "Developing Reliable Taxonomic Features for Data Warehouse Architectures," in *2020 IEEE 22nd Conference on Business Informatics (CBI)*, 2020, pp. 241-249.
- [14] Alaa Khalaf Hamoud, Maysaa AbdUlkareem, Hisham Noori Hussain, Zahraa Abdulkareem Mohammed, and Gadeer Mustafa Salih, "Improve HR Decision-Making Based On Data Mart and OLAP," *Journal of Physics: Conference Series*.
- [15] Lamy Oukhouya, Anass El haddadi, Brahim Er-raha, and Hiba Asri, "A generic metadata management model for heterogeneous sources in a data warehouse," in *E3S Web of Conferences* 297, 01069 (2021), 2021.

- [16] Sonja Ahola, Developing a Customer Master Data Management Model, November 2022.
- [17] K. Rajasekaran, P. Saravanan, and P. Prabakaran, "COMPARATIVE STUDY ON DATA WAREHOUSE TABLES AND SCHEMA-AN OVERVIEW," *Adalya Journal*, vol. 9, no. 4, pp. 126-132, 2020.
- [18] Thomas C. Hammergen and Alan R. Simon, *Data Warehousing for Dummies*, 2nd ed.: WILEY Publishing, Inc., 2009.
- [19] Willsen Wijaya, Jansen Wiratama, and Santo Fernandi 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, April 2024.
- [20] Dhamotharan Seenivasan, "Effective Strategies for Managing Slowly Changing Dimensions in Data Warehousing," *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 9, no. 4, pp. 492-496, April 2022.
- [21] Ronak Ravjibhai Pansara, "Strategies for Master Data Management Integration and Their Benefits," *Scholars Journal of Engineering and Technology*, pp. 40-47, Feb. 2024.
- [22] Kanika Goel, Fahame Emamjome, and Arthur H. M. ter Hofstede, "Data Governance for Managing Data Quality in Process Mining," in *42nd International Conference on Information Systems (ICIS 2021)*, 2021.
- [23] Ze Shi Li, Colin Werner, Neil Ernst, and Daniela Damian, "GDPR Compliance in the Context of Continuous Integration," *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING*, 2020.
- [24] Hemanth Gadde, "AI-Enhanced Data Warehousing: Optimizing ETL Processes for Real-Time Analytics," *REVISTA DE INTELIGENCIA ARTIFICIAL EN MEDICINA*, vol. 11, no. 01, pp. 300-327, 2020.
- [25] Sanny Hikmawati, Paulus Insap Santosa, and Indriana Hidayah, "Improving Data Quality and Data Governance Using Master Data Management: A Review," *IJITEE*, vol. 5, no. 3, pp. 90-95, September 2021.