

DW and DM Assignment 1

by Kyle Hudson

Submission date: 03-Apr-2024 04:22PM (UTC+0100)

Submission ID: 228458893

File name: 49924_Kyle_Hudson_DW_and_DM_Assignment_1_565908_2098604805.docx (803.32K)

Word count: 4637

Character count: 26890

Data Warehousing and Data Mining

Data Warehousing

ACCA6026

Nathan Kyle Hudson

2017102

Table of Contents

Introduction.....	3
Task 1.....	3
Online Transaction Processing (OLTP) database and Data Warehouse differences.....	3
Data Warehouse Architecture and Source System	5
Data Mart for Cambrian Real Estate and departments.....	7
Online Transaction Processing (OLTP) database and Data Warehouse structures.....	8
Task 2.....	9
Internal and External Source Systems	9
Appropriate Star Schema	10
Appropriate Data Marts.....	13
ETL Plan for Cambrian Real Estate.....	14
Data Quality and Information Governance issues	16
Bibliography	18
Figure 1. Data Warehouse structure.....	5
Figure 2. ETL plan.	6
Figure 3. Example of Multi Star Schema	9
Figure 4. Suitable Multi Star Schema for Cambrian Real Estate	11
Figure 5. Aggregate Data.....	12
Figure 6. Master Data	12
Figure 7. Meta Data	13
Figure 8. Data Mart Schema.....	14
Figure 9. ETL plan	15

Introduction

In this report, I will be addressing the tasks outlined for the company Cambrian Real Estate. This report will be tailored to their specific needs, drawing upon the provided scenario. I will ensure proper referencing throughout and incorporate relevant diagrams as necessary for each section. I will be discussing the elements of data capturing and analysis that the Cambrian Real Estate will need to create an effective system for success within their business.

Task 1

Online Transaction Processing (OLTP) database and Data Warehouse differences

Within an organization's data management infrastructure, data warehouses and online transaction processing (OLTP) databases have distinct functions. [1]. Here is an example of contrast between the two:

The purpose of each includes a transactional process, which is the focus of OLTP databases. They are designed to handle large numbers of quick, short transactions, like adding, editing, and removing records. Tasks including order processing, inventory control, and customer relationship management (CRM) are often performed by OLTP systems. According to research, OLTP systems are a modification of a database, that can be used on a day to day basis. [2].

However, research depicts that data warehouses are made for analytical processing. [3]. Within an organisation, they compile and arrange data from many sources to facilitate data analysis, business intelligence, and decision-making processes. Longer-term historical data storage is made possible by data warehouses, which provide sophisticated data analysis and querying. [4].

Author Ponniah, states that OLTP databases and data warehouses serve distinct purposes. OLTP databases are used for handling more specific and simpler data, while data warehouses are used for managing broader and more complicated data. Furthermore, the author asserts that the disparities also vary in terms of data storage, gathering, and information dissemination. According to the author, a data warehouse depends on storage to handle the substantial amounts of data from several sources, including external sources. Additionally, the data warehouse is designed to accommodate various sorts of users that will access the data. [5].

The structure of an OLTP databases is used to reduce redundancy and guarantee data integrity because they are normalised. Their data structure has been normalised, making it efficient for handling transactions and guaranteeing consistency.

A denormalized or star-schema structure is frequently used in data warehouses. By arranging the data into dimensional tables (facts and dimensions, for example), this format maximises data analysis for reporting and business intelligence needs. This facilitates querying and analysis. [6].

According to author Krishnan, the distinction between OLTP and data warehousing databases lies in the fact that data warehouses are specifically designed to display all the

accurate data that an organisation possesses, whereas an OLTP database can be utilised to gather data from external sources and then process it into the data warehouse. [7].

Each database is optimized differently to perform their query and reports. Small-scale data retrieval and manipulation can be accomplished quickly using OLTP databases because of their optimised design. Real-time transaction processing and basic queries are appropriate for this database.

However, complicated searches including joins, aggregations, and computations over huge amounts of data are best suited for complicated data warehouses. In-depth analysis, report generation, and historical trend and pattern insights are all made possible by them for companies. [8].

Every database has a distinct purpose. Operational staff members use OLTP databases to do daily transaction tasks. High concurrency and real-time data processing are supported by these applications.

Decision-makers, analysts, and data scientists are the main users of data warehouses when they want to extract knowledge from historical records. Forecasting, data mining, trend analysis, and strategic decision-making are all supported by them. [9].

A Data Warehouse Investment for Cambrian Real Estate Has Several Advantages

- **Improved Decision-Making:** If Cambrian Real Estate had access to a data warehouse, it could examine past information to learn more about consumer tendencies, property performance, and market trends. Determining pricing strategies, investment opportunities, and resource allocation can be made easier with the use of this information. [10].
- **Better Reporting and Analytics:** Cambrian Real Estate can discover areas for improvement, track key performance indicators (KPIs), and monitor operational efficiency by using advanced analytics and thorough reports generated by a data warehouse. [11].
- **Data Integration:** Cambrian Real Estate may combine information from multiple sources, including property listings, financial records, CRM systems, and market data, by using a data warehouse. This integrated view of the data makes thorough analysis easier and offers a comprehensive insight of the business. [12]
- **Scalability:** A data warehouse offers the potential to manage massive volumes of data effectively as Cambrian Real Estate expands and gathers more information. Without sacrificing performance, it can handle growing demands for processing and data storage. [12].
- **Data Quality and Consistency:** A data warehouse guarantees data quality and consistency by combining data into a single source of truth. By doing away with data silos and inconsistencies, it offers trustworthy information for making decisions.

Ultimately, Cambrian Real Estate would be able to drive strategic initiatives, efficiently exploit its data assets, and obtain a competitive edge in the real estate market by investing in a data warehouse and the author "Ponniah" further explains that the organisation experiences enhanced performance as a result of the different operations. This would

encompass the durations for generating reports, executing queries, and conducting data analysis. [13].

Data Warehouse Architecture and Source System

This section will illustrate the utilisation of source systems for populating the Cambrian Real Estate data warehouse. I will also examine the process of Extract, Transform, Load (ETL) and its effective utilisation for the organisation.

By utilising research sources, I have created a diagram employing suitable software to visually represent the structure and organisation of a data warehouse. [14], [15].

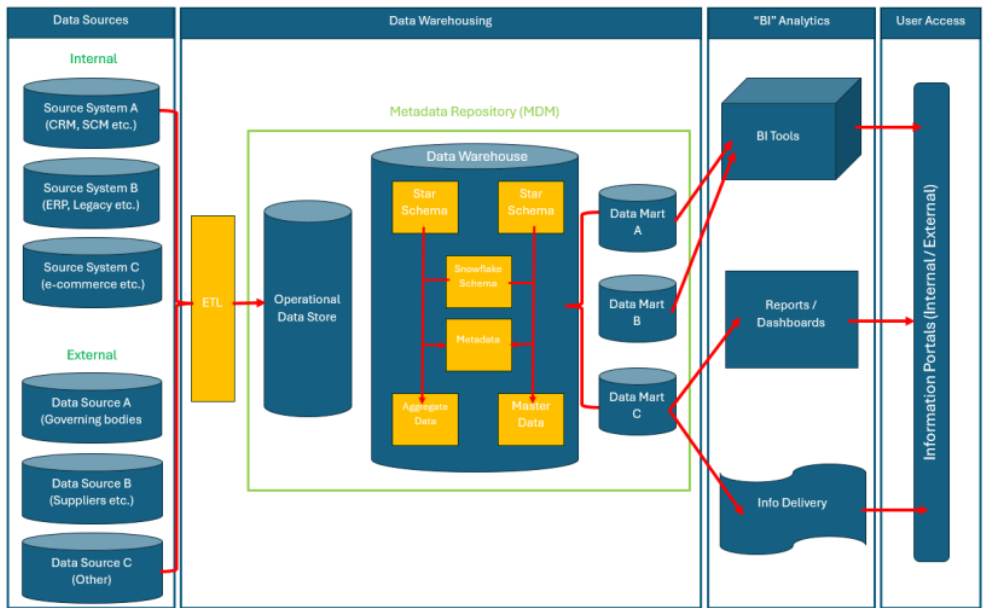


Figure 1. Data Warehouse structure

Figure 1 shows that I added the steps to the data warehouse method. During step one, the company would gather all of its data sources, both internally and externally. The second step is the ETL process, which comes after this. All of the company's data and information will be taken out of its source systems, changed or cleaned up, and then loaded into the data warehouse as part of this process.

As a result, in step three, the appropriate information is transformed into suitable schemas in order to streamline the retrieved data. Subsequently, this will be generated as the output in the fourth phase, which involves the building of a data mart. This would facilitate the access of important data by different departments inside the company.

Finally, the fifth step is the information presentation. At this stage, the collected information will be presented and examined using data analysis reports. Additionally, there is another

step that includes users accessing all of this information, through portals and/or relative displays.

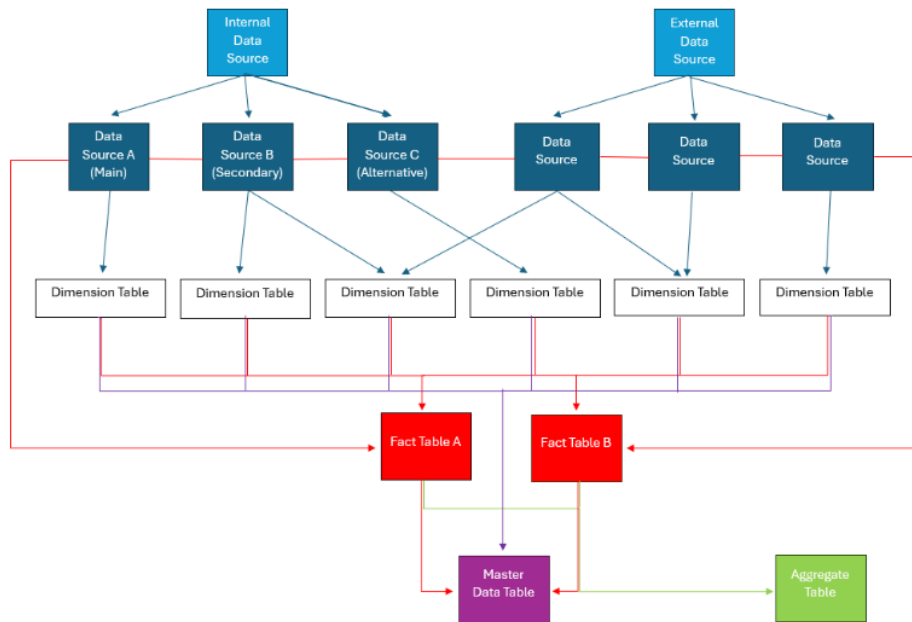


Figure 2. ETL plan.

An ETL incremental plan can demonstrate the process of transferring data from the source systems of Cambrian Real Estate to the facts and dimensions of a new data warehouse. In figure 2, we observe a comparable process in the architecture. We can also observe the movement of data from the source to the dimensions, which then enter the fact table, leading to the generation of metadata and aggregated data.

The author, Ponniah, defines an ETL process as the use of many software tools by organisations to extract data from their source systems, followed by the transformation, cleaning, and insertion of the data into a data warehouse. [16].

The data-to-day operations of Cambrian Real Estate will remain unaffected by this process, as the update to the data warehouse will occur during its offline period. Ponniah provides a description of the strategies and methodology employed in an ETL process when applying loading. These options include of appending, destructive merge, and constructive merge. The following are techniques for incorporating new entries, updating existing entries with more recent versions, and creating and defining duplicate entries of the same records. [16]

According to author Kimball, it is recommended to have an ETL manager for an ETL process, as the best practice. They will oversee the day-to-day operations of an ETL team and manage the maintenance of the data warehouse in relation to the ETL process. [17].

Data Mart for Cambrian Real Estate and departments

Cambrian Real Estate needs data marts to give tailored information to its divisions. They meet the needs of the managing, finance, and sales and marketing directors. [18].

This is how each department can use data marts:

Managing director data marts may include financial, sales, marketing, and operations data. Consolidated reports and dashboards will show revenue, profitability, market share, customer happiness, and corporate trends. The data mart could include external market data and industry benchmarks for decision-making. Data marts track long-term property pricing, demand, and competitor trends. With this knowledge, the managing director can foresee market moves and change corporate strategy. Data marts reveal resource utilisation, project deadlines, and employee productivity to help the managing director optimise operations and manage resources. [19].

A finance-specific data mart could incorporate ERP, accounting, and financial database data. It would include financial records, cash flow analysis, budget vs. actual comparisons, cost analysis, and forecasts. The data mart could improve strategic planning and risk management using financial modelling and scenario analysis. The finance director can get accurate financial reporting using data marts. Revenue breakdowns by property, area, or client, spending analysis, cash flow forecasts, and budget vs. reality are presented. Data marts help control risk with multi-source data. Monitor mortgage delinquencies, tenant credit risk, and regulatory compliance. Data marts examine property development, purchase, and upkeep expenses. The finance director can reduce expenses, improve supplier negotiations, and optimise capital allocation with this information. [20].

A sales and marketing data mart may include CRM, sales pipeline, marketing automation, and customer feedback. It would show sales performance metrics (conversion rates, sales velocity, customer acquisition costs), customer demographics, buying behaviour, campaign efficacy, and market segmentation. The data mart's advanced analytics could enable predictive modelling, lead scoring, and customised marketing. Data marts categorise customers by demographics, preferences, buying behaviour, and lifecycle stage. The sales and marketing director can target certain target categories with marketing and sales. Data marts may track real-time marketing campaign lead generation, conversion rates, client acquisition costs, and ROI. A data mart can assess competitors' pricing, market positioning, and promotions. Sales and marketing directors can identify and address competitive threats. [21].

Examples of Data Mart Applications:

- Sales data marts store transactional data on customer orders, product sales, and revenue performance. Sales managers may track sales, evaluate customers, and estimate opportunities with the programme.
- Marketing data marts store data on campaign performance, customer demographics, and market segmentation. It can help marketers evaluate their marketing efforts, target certain client groups, and improve their strategies.
- A finance data mart stores financial data such as income statements, balance sheets, and cash flow statements. Finance managers can use it to analyse budget discrepancies, financial performance, and department cost allocation.

Data marts foster departmental collaboration. A possible implementation is below:

- Managing Director + Sales and Marketing Director
- Managing Director + Finance Director
- Finance Director + Sales and Marketing Director

Every data mart holds crucial data for departmental managers' tasks. Data marts provide quick, accurate, and actionable data to assist managers make wise business decisions. These data marts integrate with the company's systems to improve information access and infrastructure use.

Online Transaction Processing (OLTP) database and Data Warehouse structures

The structural distinctions between an OLTP database and a data warehouse are varied. The OLTP database adheres to a normalisation structure, while a data warehouse database adheres to a facts/dimensions structure. Cambrian Real Estate currently uses their OLTP database. [22].

On the other hand, the design of a data warehouse is dimensional, with a focus on efficiently retrieving and analysing data for the goal of decision support and business intelligence.

Dimensional modelling entails the arrangement of data into two distinct categories of tables:

Fact tables are comprised of quantitative data, also known as facts, which reflect business metrics or measurements. These facts can include information such as sales income, product quantities, or customer orders. Fact tables commonly establish foreign key associations with dimension tables and function as the primary focal point for analytical queries and analysis. [23].

Dimension tables are comprised of descriptive attributes, also known as dimensions, which offer context and supplementary information about the facts included in the fact tables.

Dimensions refer to category or textual data, such as product categories, client segments, or time periods. Dimension tables function as reference tables for accessing and filtering data in fact tables. [24].

Below is an example multi star schema diagram that Cambrian Real Estate could use:

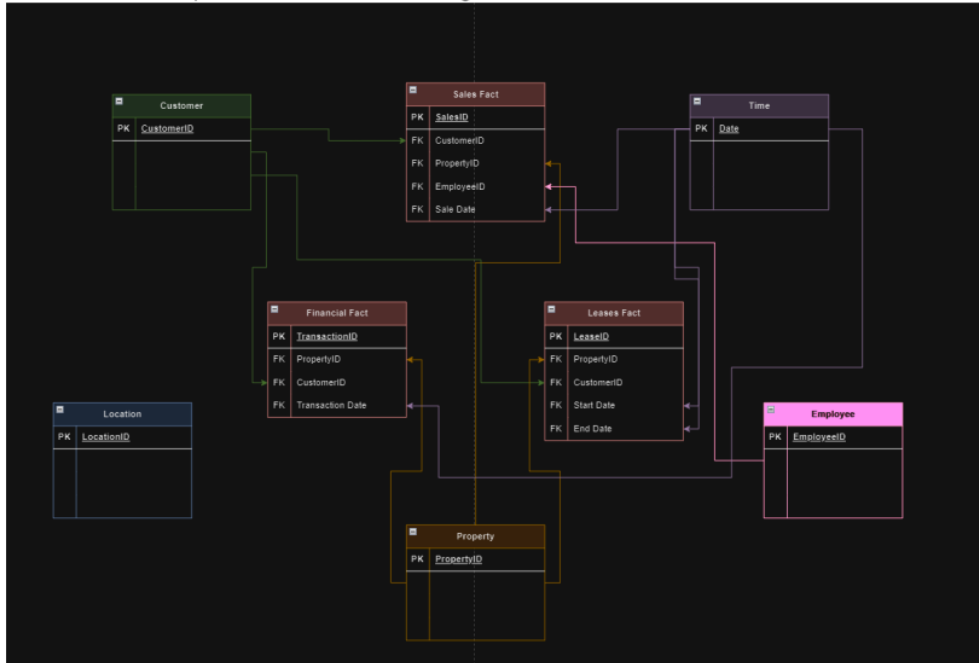


Figure 3. Example of Multi Star Schema

Task 2

Internal and External Source Systems

Upon referencing source systems, I will proceed to conduct thorough study on source systems and categorise them in accordance with the Cambrian Real Estate organisation. Research source systems refer to the systems and processes that an organisation requires in order to facilitate its growth. These systems will collect data from both internal and external sources, which will then be evaluated and incorporated into the data warehouse [25]. The sources involved will encompass orders from suppliers, distributors, and manufacturers, internal resources, physical paper files, shareholder information, as well as data obtained from sales and surveys done by the organisation. [26].

Internal source systems can be identified as follows:

- Customer Database: This system stores comprehensive information on customers, including their demographic data, contact information, and their past interactions with Cambrian Real Estate.

- The Property Listings Database is a repository that contains comprehensive information about properties that are currently on the market for sale or rent. This information includes the property's address, type, size, price, year of construction, and accompanying photographs.
- The Lease Agreement Database oversees lease agreements for properties under Cambrian's management, containing details such as terms, rental amounts, lease lengths, and renewal dates.
- The financial database stores and organises information about financial activities, such as income from property sales and rentals, expenses, wages, taxes, and other financial measurements.

External source systems identification:

- **Property Valuation Data:** External databases or services provide property valuation estimates based on factors such as location, property characteristics, market circumstances, and recent sales data.
- **Demographic Data:** External sources that offer demographic information on various regions, encompassing population demographics, income levels, employment rates, education levels, and household composition.
- **Industry Trends Data:** External data sources offer valuable information on current trends in the real estate industry, including data on average property prices, market demand, housing inventory, interest rates, and economic indicators.
- **Competitor Data:** Data pertaining to competitors in the real estate industry, encompassing their property listings, pricing techniques, market dominance, customer satisfaction ratings, and marketing strategies.

Appropriate Star Schema

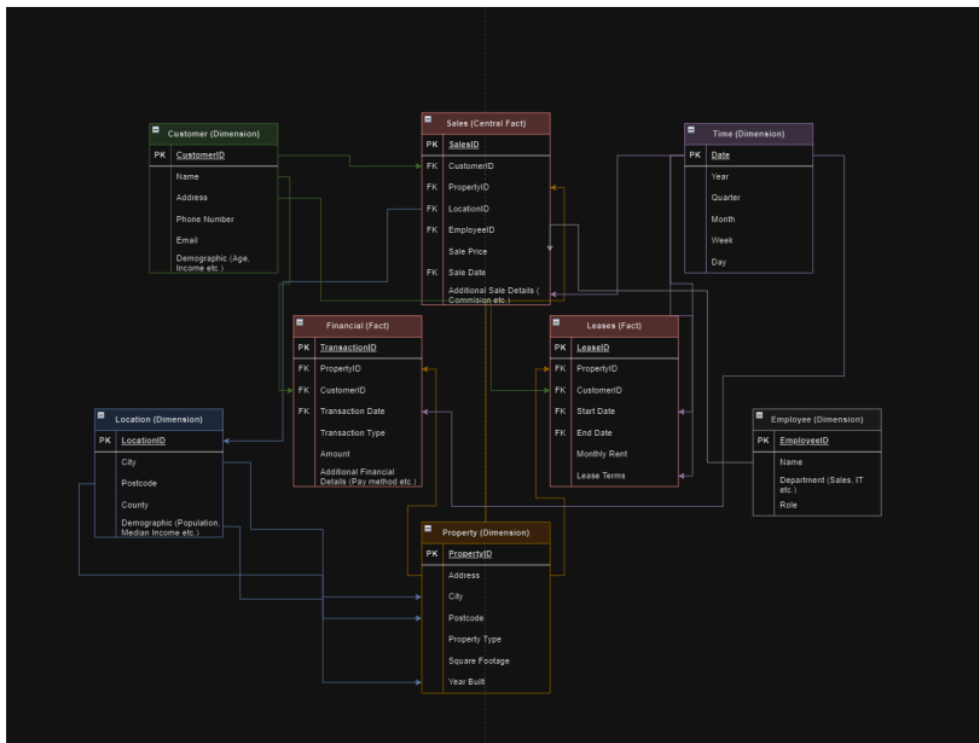


Figure 4. Suitable Multi Star Schema for Cambrian Real Estate

Building upon the appropriate star schema identified in the preceding work, I proceeded to enhance and expand the star schema. In the provided figure 4, I have specified the specific dimensions and fact tables that will constitute the data warehouse utilised by Cambrian Real Estate.

In the Star schema, the dimension tables are separate tables that are linked to the fact table just once. These objects are of a grey colour. The fact tables that are collecting all of the data from the dimensions are the tables that are tinted red. The multi-coloured tables are the tables that have been standardised and will be connected to various fact tables. In order to illustrate the connections between tables, I have employed coloured arrows to enhance the legibility of the star schema.

The star schema is developed in this manner due to the compatibility between the data union and the data transfer process in Cambrian Real Estate's point-of-sale system. Furthermore, it is well-suited for the specific data that Cambrian Real Estate gathers throughout its daily transactional activities. By utilising the star schema, the Cambrian will have the capability to generate data marts from the fact tables. This will let the departments to directly access the specific data they require. [27], [28].

Aggregate Data

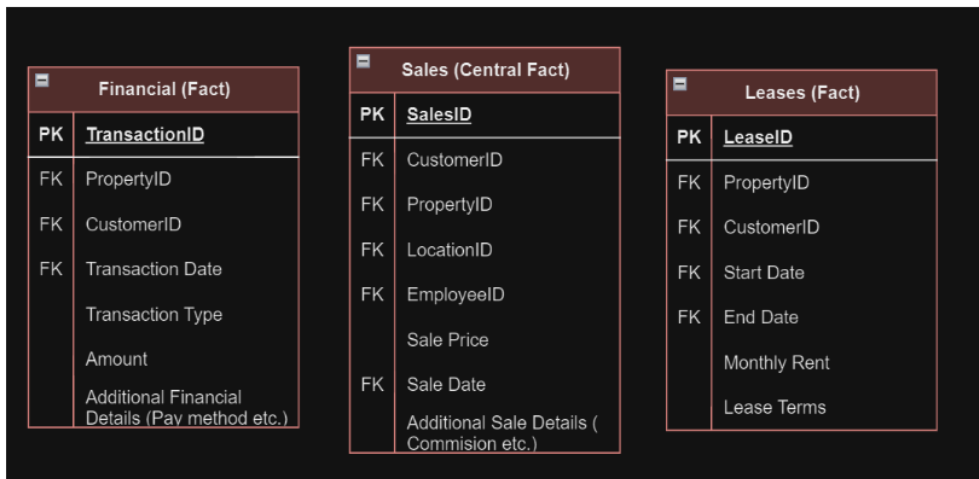


Figure 5. Aggregate Data

Master Data

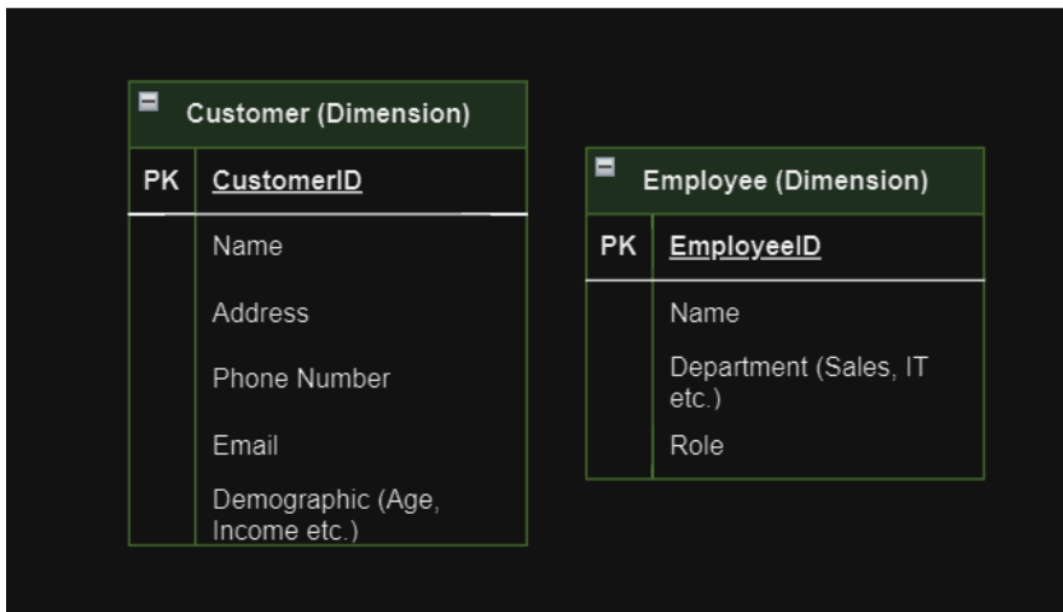


Figure 6. Master Data

Meta Data

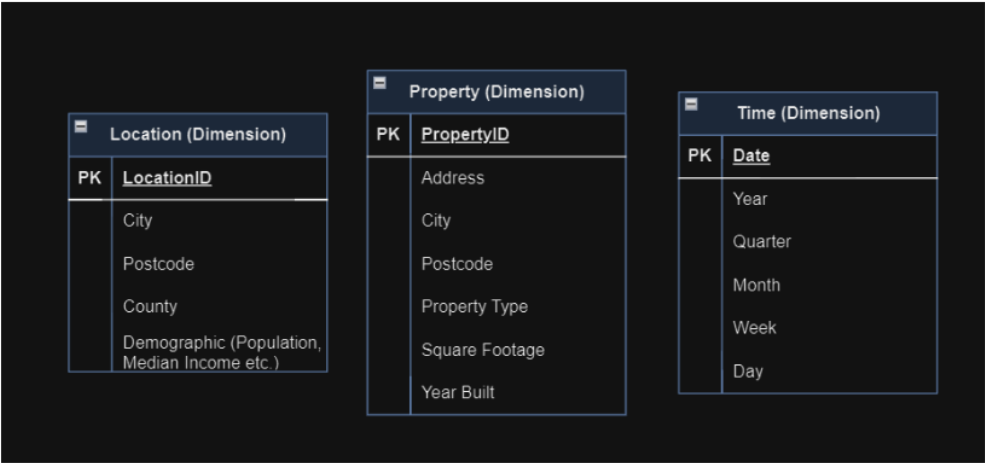


Figure 7. Meta Data

Figure 5, 6, 7 displays the metadata, aggregate data, and user data that will be gathered by each of the fact tables.

The master data, seen in figure 6, is presented in a green hue and encompasses the pertinent information for users, data sources, and tables along with their respective qualities. The rationale for this architecture of the master data is to serve as a source of useful information for end users and other queries. It includes details about table information, such as data from individuals in the 'Customer' table or members in the admission table [29].

The metadata displayed in figure 7 are highlighted in blue. They provide the necessary information to measure the modified table, identify the modified row, determine the old and new values, and identify the sources of the data. The metadata will also record the user who modified these settings and the exact timing of the modification. I selected this architecture for the metadata based on best practices and studies, which recommend it for enabling businesses to track both the front and back rooms of metadata [30].

The data shown in figure 5 is depicted in the colour red. These tables facilitate the monitoring of all the quantifiable aspects of a fact table. An illustration of this concept can be seen in the Leases aggregate table. The system will monitor the cumulative sales on a daily basis and subsequently categorise them by week, month, year, and quarter. Ultimately, the user will have the capability to monitor the exact time when the data was generated.

Appropriate Data Marts

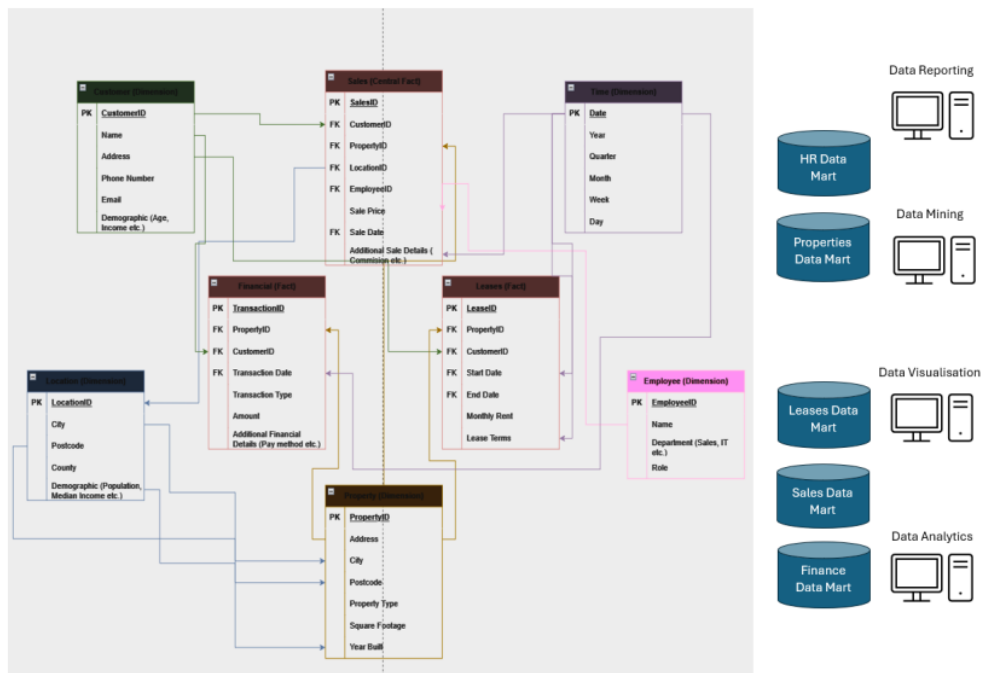


Figure 8. Data Mart Schema

Figure 8 displays the dimensions and fact tables. The fact tables will have obtained their data from the data contained in all dimension tables. This will be applicable to the Finance, Sales, Leases, Customer, and Property fact tables. The fact tables will contain independently generated data that is separate from the dimension tables. The origin of the dimension tables will be diverse. In this scenario, certain data will be created autonomously, while others will be obtained via the operational systems of the Cambrian.

ETL Plan for Cambrian Real Estate

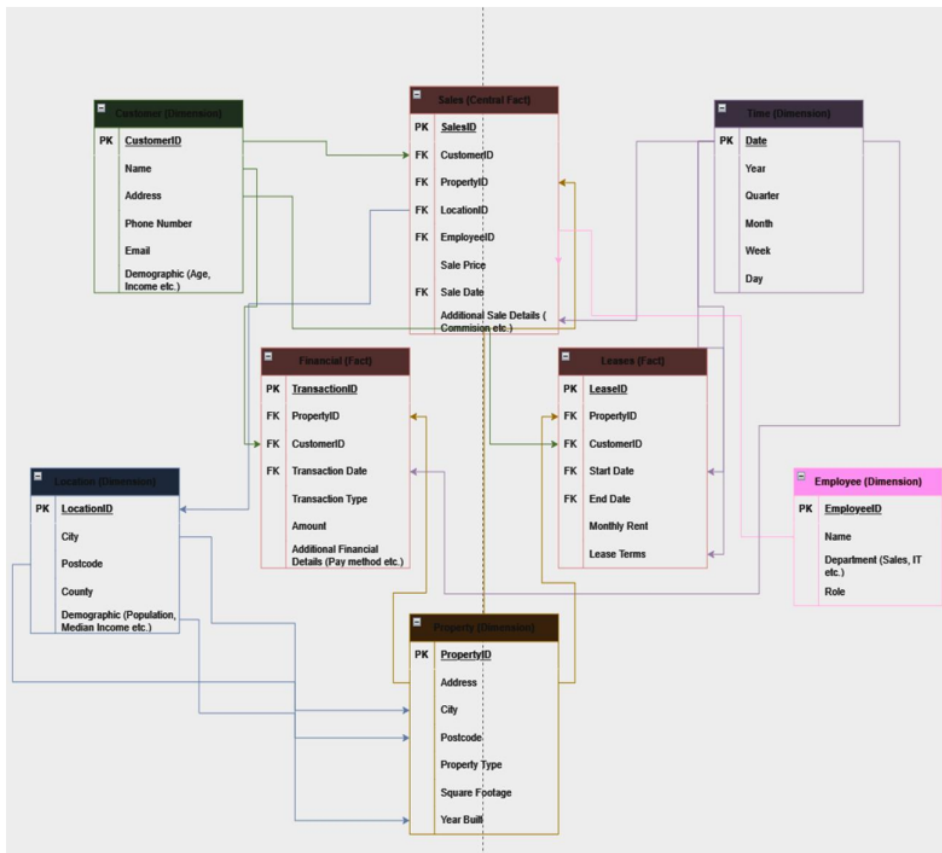


Figure 9. ETL plan

Figure 9 illustrates a rudimentary ETL strategy that may be implemented by the Cambrian organisation. I shall explicate the nature of the fact tables and the sources from which they derive their data.

The Finance table is the initial table that is visible. This table will receive its data from the Time, Property, and Customer tables in order to function as a fact table. Despite having a one-to-many relationship, they will collect information including CustomerID and PropertyID. Additional financial information and the TransactionID, Transaction Type, and Amount will be generated by the table.

The Leases table is the second table that is visible. This table will receive its data from the Property, Customer, and Time tables in order to function as a fact table. Although their relationship will be one-to-many, they will collect information including PropertyID and CustomerID. Lessee terms and monthly rent will be generated by the table itself.

The Sales table is the third table that is visible. The data for this primary fact table will be derived from the following tables: Customer, Finance, Property, and Time. Despite having a one-to-many relationship, they will collect information including CustomerID, PropertyID, LocationID, and EmployeeID. Additional Sale Details, SalesID, and Sales Price are generated by the table itself.

The Customer table is the fourth table that is accessible. The data for this dimensions table will be derived from the financial, sales, and leases tables.

The Time table is the fifth table that is visible. The data for this Dimensions table will be derived from the Financial, Sales, and Leases tables.

The Employee table is identified as the sixth table that is visible. The Sales table will provide the data for this table, which will function as a dimensions table.

The Location table is the eighth table that is visible. The data for this dimensions table will be derived from the Sales and Property tables.

The Property table is identified as the eighth table that is visible. This table will receive its data from the Sales, Financial, and Leases tables in order to function as a Dimensions table.

Data Quality and Information Governance issues

Data warehouse systems must have good data quality and information governance during ETL and normal operation. Consider these crucial issues:

Data accuracy is crucial while extracting data from source systems. Incorrect data can affect analysis and decision-making. Data profiling and validation checks should be used during ETL to find and fix errors.

Data Completeness: The data warehouse must have all necessary data. Missing data can bias analysis and insights. ETL should include data completeness checks to find missing data and fill gaps.

Data consistency ensures data is uniform and coordinated across sources and systems. Different naming standards, data formats, and definitions might cause inconsistencies. To ensure consistency, ETL should use data standardisation methods including cleansing and transformation. [31].

Timely data is essential for informed decision-making. Delays in data extraction, transformation, or loading might affect data warehouse insights. To ensure timely data availability, monitor data latency and deploy effective ETL operations.

Data integrity: Data integrity provides data accuracy and consistency throughout its lifecycle. Keep referential integrity, enforce data constraints, and prohibit unauthorised changes. Data integrity constraints and access controls protect warehouse data.

Data Governance: Data governance includes data asset management rules, procedures, and controls. Data ownership, governance, and quality requirements are defined. Data quality and integrity require continuous monitoring and enforcement of data governance principles.

Metadata Management: Metadata contextualises warehouse data. Documenting data lineage, definitions, and linkages is essential to metadata management. Comprehensive metadata management improves data discovery, interpretation, and quality. [32].

Security: Data must be protected from unauthorised access, disclosure, and tampering. To protect sensitive data in the data warehouse, utilise encryption, access controls, and

authentication. Data security is maintained by security audits and compliance checks.

Data Privacy: GDPR, CCPA, and HIPAA compliance is essential for privacy protection. To comply with privacy laws, sensitive data should be anonymized, pseudonymized, and masked. Implementing privacy-by-design principles embeds privacy into the data warehouse system.

Data Retention and Archiving: Data lifecycle management requires data retention policies. Archiving and removing old data optimises storage and keeps data relevant. Ensure legal and regulatory data retention compliance.

Overall, data quality and information governance issues must be addressed throughout the data lifecycle, from extraction to storage and analysis. Organisations may assure data quality, integrity, and governance in their data warehouse system by employing rigorous procedures, controls, and technology. [33].

Bibliography

- [1] S. REDDY, R. SRINIVASU, P. CHANDER RAO and S. R. RIKKULA, "Data warehousing, data mining, olap and oltp technologies are essential elements to support decision-making process in industries," (*IJCSE*) *International Journal on Computer Science and Engineering*, vol. 2, no. 9, pp. 2865-2871, 2010.
- [2] S. Harizopoulos, D. J. Abadi, S. Madden and M. Stonebraker, "OLTP through the looking glass, and what we found there," in *Making Databases Work: the Pragmatic Wisdom of Michael Stonebraker*, Association for Computing Machinery and Morgan & Claypool, 2018, p. 409–43.
- [3] U. Dayal and S. Chaudhuri, "An overview of data warehousing and OLAP technology," *ACM SIGMOD Record*, vol. 26, no. 1, pp. 65-74, 1997.
- [4] P. PONNIAH, "The Compelling Need for Data Warehousing," in *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*, New York / Chichester / Weinheim / Brisbane / Singapore / Toronto, JOHN WILEY & SONS, INC., 2001, pp. 9-11.
- [5] P. Ponniah, "Planning and Project Management," in *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*, John Wiley & Sons, Inc., 2001, p. 70.
- [6] S. S. Baboo and P. R. Kumar, "Next Generation Data Warehouse Design with OLTP and," *International Journal of Computer Applications*, vol. 72, no. 13, pp. 45-49, 2013.
- [7] "Big Data Processing Architectures," in *Data Warehousing in the Age of Big Data*, 2013, pp. 35-36.
- [8] C. Longbottom and R. Bamforth, "Optimising the data warehouse," Quocirca, 2013.
- [9] M. VELICANU and G. MATEI, "Database Vs Data Warehouse," *Revista Informatica Economică*, vol. 3, no. 43, pp. 91-94, 2007.
- [10] P. Tiwari, S. Kumar and A. C. Mishra, "Improved performance of data warehouse," *International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 94-101, 2017.
- [11] A. Vaisman and E. Zimányi, "Data Analytics: Exploiting the Data Warehouse," in *Data Warehouse Systems*, 2014, pp. 329-383.
- [12] T. Das and A. Mohapatro, "A Study on Big Data Integration with Data Warehouse," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 9, no. 4, pp. 188-191, 2014.
- [13] P. Ponniah, "Infrastructure as the foundation for data warehousing," in *Data Warehousing Fundamentals for IT Professionals*, John Wiley & Sons, 2011, p. 190.

- [14] P. Ponniah, "Distinguishing Characteristics," in *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*, John Wiley and Sons, Inc., 2004, pp. 16-18.
- [15] P. Ponniah, "The Architectural Components," in *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*, John Wiley and Sons, Inc., 2001, pp. 134-141.
- [16] "Data Extraction, Transformation, And Loading," in *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*, John Wiley and Sons, Inc., 2001, pp. 279-288.
- [17] R. Kimball, "Responsibilities," in *The Data Warehouse Toolkit*, 2004, pp. 394-399.
- [18] "Data Marts and Data Warehouses," in *Introduction to Data Mining and its Applications*, pp. 75-150.
- [19] A. Simon, "Building an Enterprise Data Management," in *Modern Enterprise Business Intelligence and Data Management*, pp. 63-75.
- [20] "Developing scalable management information system with big financial data using data mart and mining architecture," *Information Processing & Management*, vol. 60, no. 3, 2023.
- [21] S. Williams, "Data Mart Strategy: Creating Intelligence Across the Supply Chain," *Communications of the Association for Information Systems*, vol. 5, no. 9, 2001.
- [22] "Dimensional Modeling using Star Schema for Data Creation," *Oriental Journal of Computer Science and Technology*, vol. 10, no. 4, pp. 745-754, 2019.
- [23] "How to Model a Data Warehouse," in *Agile Data Warehouse Design*, 2011, pp. 11-13.
- [24] R. Kimball, "A First Course on Dimensional Modeling," in *The Data Warehouse Lifecycle Toolkit: Tools and Techniques for Designing, Developing and Deploying*, 1998, pp. 76-112.
- [25] R. Thierauf, *Effective Business Intelligence Systems*, Quorum Books, 2001.
- [26] T. Löfgren, D. Gravem and G. Haraldsen, "A glimpse into the businesses' use of internal and external data sources in decision-making processes," *BLUE-ETS Conference on Burden and Motivation in Official Business Surveys*, 2011.
- [27] P. O'Neil, B. O'Neil and X. Chen, "The Star Schema Benchmark (SSB)," 2007.
- [28] R. Kimball, "Real-Time ETL Systems," in *The Data Warehouse ETL Toolkit*, 2004, pp. 426-427.
- [29] R. Kimball, "Extracting," in *The Data Warehouse ETL Toolkit*, 2004, p. 104.
- [30] R. Kimball, "Metadata," in *The Data Warehouse ETL Toolkit*, 2004, pp. 351-380.
- [31] R. Mahati, *Data Quality: Dimensions, Measurement, Strategy, Management, and Governance*, 2018.

- [32] M. Souibgui, "Data quality in ETL process: A preliminary study," *Procedia Computer Science*, vol. 159, pp. 678-687, 2019.
- [33] H. N. BENKHALE and D. BERRABAH, in *Data Quality Management For Data Warehouse*, pp. 1-8.
- [34] ChatGPT, <https://chat.openai.com/>
- [35] Draw.io, <https://app.diagrams.net/>

DW and DM Assignment 1

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

57 /100

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

