CMPG 321

GROUP GP10

Request for Proposal: ClearVue Ltd

Sponsor: Dr. J. Pretorius

Executive Sponsor: Mrs. J. Thandi

Group Members:

Jacques van Heerden (Student Number: 35317906)

Rohann Venter (Student Number: 25130757)

Francois Verster (Student Number: 40723380)

Christo Prinsloo (Student Number: 21052239)

Erika Haasbroek (Student Number: 37673149)

Stefan Venter (Student Number: 39066894)

Table of Contents

Table of Tables

Table of Figures

Executive Summary

ClearVue Ltd. requires a Business Intelligence (BI) system that enhances its reporting capabilities and facilitates informed decision-making through the analysis of sales data. The new BI system must produce time-based reports (daily, weekly, monthly, quarterly, annually), be adaptable to future supplier-centred analytics and support hierarchical data.

MongoDB will be used as the NoSQL database system that can address the company's financial structure. This database will integrate with Apache Kafka, a streaming tool for real-time payment transactions. The visualisation platform Power BI will also be used to create interactive dashboards that will assist the management and analysts.

The new NoSQL-based BI system will deliver more flexible, efficient, and scalable performance compared to traditional systems, such as Microsoft Access. All this will be achieved by keeping ClearVue's analytical needs and data requirements in mind.

Requirements Definition Approach

Structured processes, including internal data analysis, stakeholder identification, and secondary research, will be employed to determine the requirements for the Business Intelligence system. The data files provided: Sales Line.xlsx, Payment Lines.xlsx, Products.xlsx, and Purchases Headers.xlsx will be revised to get an understanding of ClearVue's existing data structures and reporting requirements. This analysis will identify key operational entities, including suppliers, sales representatives, customers, and products.

Additionally, various Internet sources will be utilised to determine and select the best-suited database technologies for ClearVue's day-to-day operations and additional requirements, specifically supporting a non-standard financial year, flexibility, and time-based reporting. Sources used also include extra information on NoSQL from MongoDB (n.d.) and comparative analyses between NoSQL and SQL from GeeksforGeeks (2025). These sources also highlighted precisely why a NoSQL database solution will be ideally suited for ClearVue.

AI assistant tools, namely ChatGPT, Grok, and Claud AI, were utilised to provide me with practical schema design examples illustrating how NoSQL can meet ClearVue's requirements and some of the potential shortcomings that may be encountered during system design.

Business Requirements

Analysis and reporting of sales data

ClearVue's sales data files need to be analysed to generate reports for daily, weekly, monthly, quarterly, and annual sales performances. The reporting method(s) must align with ClearVue's unique financial year format (last Saturday of preceding month to last Friday of current month format). The sales data reporting and analysis must enable them to make impactful strategic decisions that will lead to a boost in sales performance.

Supplier-focused analytics using supplier data

ClearVue's plans to analyse supplier data require a flexible data solution. The solution must be flexible and scalable to accommodate data needs.

Evolving data needs

A flexible data model that supports hierarchical and semi-structured data. Scalability for growth, agility for diverse data and support for real-time analytics.

Capability to stream data

The proposed system must be able to process and update data, such as sales data, in real-time. This will enable real-time insights, benefit decision makers, and organise data in a BI-friendly structure (star schema) for analytical queries.

Incorporate a NoSQL-based BI system.

A user-friendly BI dashboard will be implemented and used for querying and visualisations. The NoSQL database will assist with scalability, flexibility, and the ability to manage diverse data structures.

Performance & usability

The proposed system must enable fast query response times and provide easy and intuitive navigation for its end users. Data accuracy must also be maintained at a high standard, as it has a significant impact on business performance and decision-making. To maintain data accuracy, it must be consistent and validated across multiple datasets.

High-Level Business Objectives

ClearVue Ltd's strategic objectives must align with those of an advanced decision support system.

Data Centralisation

ClearVue's data is currently divided into multiple Excel documents. This could lead to data inaccuracies and inefficiencies. By centralising data into a single NoSQL database, the company can have a single point of access to all company data. When company data is centralised in a single repository, it can be easily manipulated to produce meaningful information through comparison, cross-referencing, and other methods.

Operational Efficiency

ClearVue requires a BI system that addresses the inefficiencies of its current reporting workflow, which currently involves significant time and effort and increases the risk of human error. By utilising a NoSQL database integrated with Apache Kafka, the need to run manual queries will be largely eliminated. Daily, weekly, monthly, and other reports can be automated, saving the company time and resources. With the above, management will have more time to focus on strategic decision-making.

Accessibility and Usability

The proposed BI system can make statistics derived from company data available to relevant staff members and stakeholders. Data can be presented visually, by filtering on desired data and displaying meaningful results via graphs and interactive dashboards.

Usability is essential, as reports must be as easily understandable as possible, without requiring technical expertise or extensive staff training.

Enhanced Decision Making

The overall high-level business objectives are to identify and analyse the

1. Trend components, i.e. long-term upward or downward movements,
2. Seasonal components, i.e. predictable patterns within the financial year,
3. Cyclical components, i.e. economic or market-driven fluctuations, and
4. Irregular components, i.e. unexpected variations. (*Components of Time Series | Forecasting Class Notes | Fiveable*, n.d., p. 1). This is especially important to visualise when making strategic decisions and will ultimately **enhance the decision-making process**.

Data Driven Operations

Data-driven operations are another high-level business requirement that involves the effective implementation of Apache Kafka, integrating NoSQL, allowing decision-makers to act on real-time data and achieve operational excellence through advanced anomaly detection techniques (Falkonry, 2024).

Data-driven explanation:

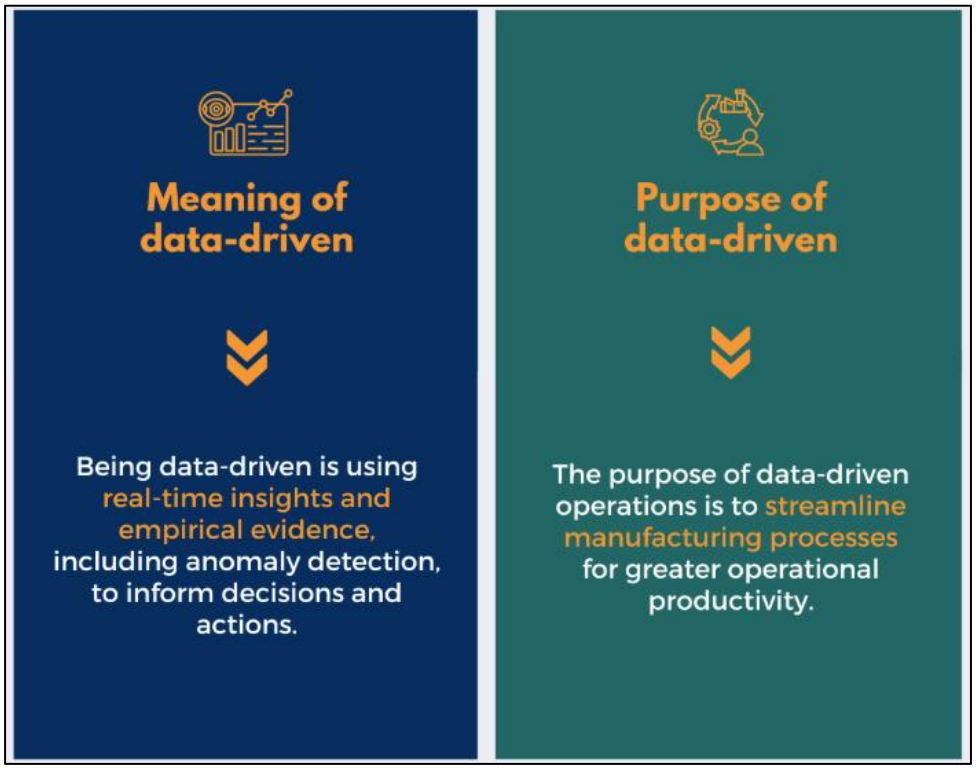

Figure 1: Data-Driven Explanation (Data-Driven-Operations-Image-1.png, 2024)

Scalable Supplier-Centred Analytics

Develop BI capabilities that can be easily adjusted to meet future requirements for supplier-focused reporting and analysis, enabling a more comprehensive assessment of lead times, cost impacts, and supplier reliability. At a later stage, the refinement of the Power BI system will be automated, allowing it to adapt to new data, with minimal interaction required, making it ideal for scalability.

The Advantages of a Scalable Power BI System

Firstly, scalability will thus allow for flexibility, which means the models need to be able to accommodate change (BI Consulting Pro, 2024).
Secondly, scalability will enable data growth, allowing models to handle increased data volumes with accurate and acceptable reporting performance (BI Consulting Pro, 2024).
Ultimately, scalability will lead to a reduction in complexity, making the management of this advanced decision support system significantly easier (BI Consulting Pro, 2024).

Table 1: Advantages of Scalable BI Systems

Analytical and Information Requirements

In the high-level business objectives, the components of a time-series analysis were briefly mentioned. In this section, we will dive deeper into what this means with expected examples.

The requirements of this decision support system or Power BI system are to support both descriptive and predictive analytics through the following capabilities:

Time Series Analysis

Component	Description	Example
Trend Component	Identify long-term sales movement patterns (Pickering, 2019)	12-month moving average of total sales.
Seasonal Component	Detect predictable fluctuations within ClearVue's non-standard financial year. (SINGH et al., n.d.)	Seasonal Index per product category.
Cyclical Component	Analyse business cycles driven by economic or market conditions. (SINGH et al., n.d.)	Sales growth rate compared to changes in GDP.
Irregular Component	Isolate anomalies or unexpected events affecting sales. (SINGH et al., n.d.)	Sales Variance Analysis During Supply Chain Disruptions.

Table 2: Time-series Analysis Components

Casual Modelling

When the variable to be predicted is impacted by or connected with other variables in the model, the forecasting model is classified as a causal model (RENDER, 2018). An example would be daily sales, which may depend on factors such as average temperature and humidity (RENDER, 2018).

Real-Time Analytics

Incorporate real-time sales and payment transaction monitoring via Apache Kafka integration, for example, the Live sales per hour dashboard.

Claud AI were used to create a sample Power BI dashboard to illustrate the Analytical and Informational requirements (Business Performance Dashboard, 2025)

Figure 2: Key Performance Indicators

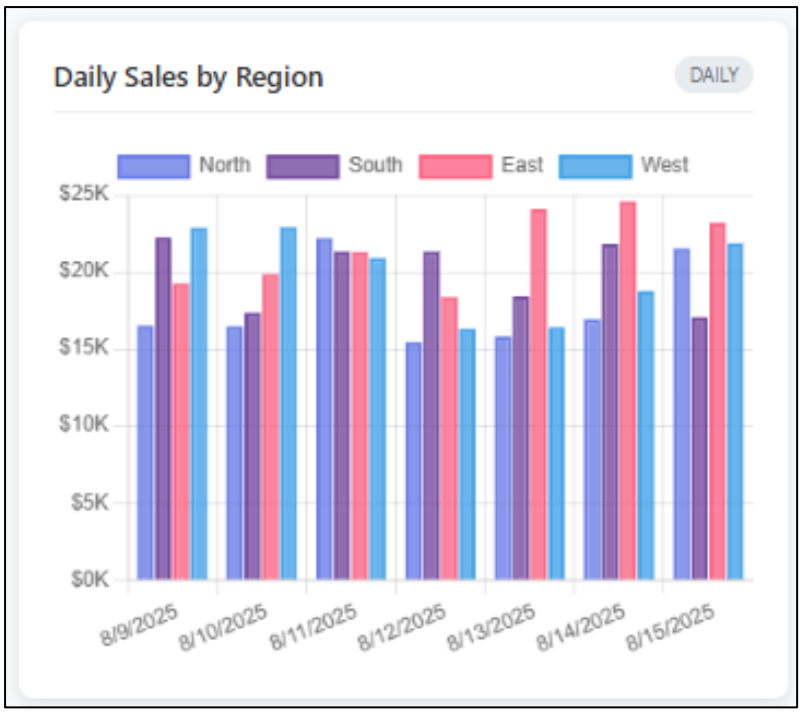

Figure 3: Daily Sales by Region

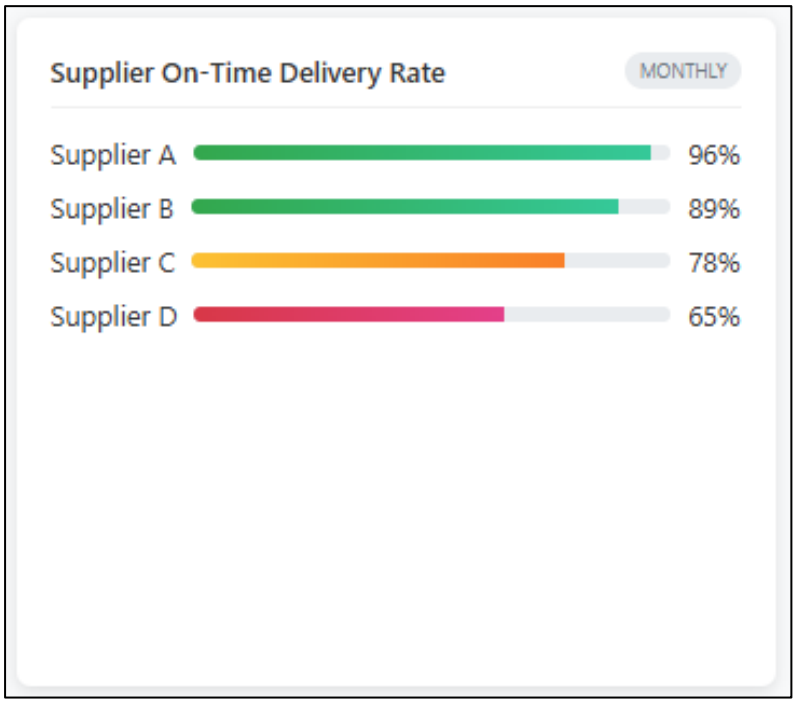

Figure 4: Supplier On-Time Delivery Rate

Figure 5: Quarterly Product Trends

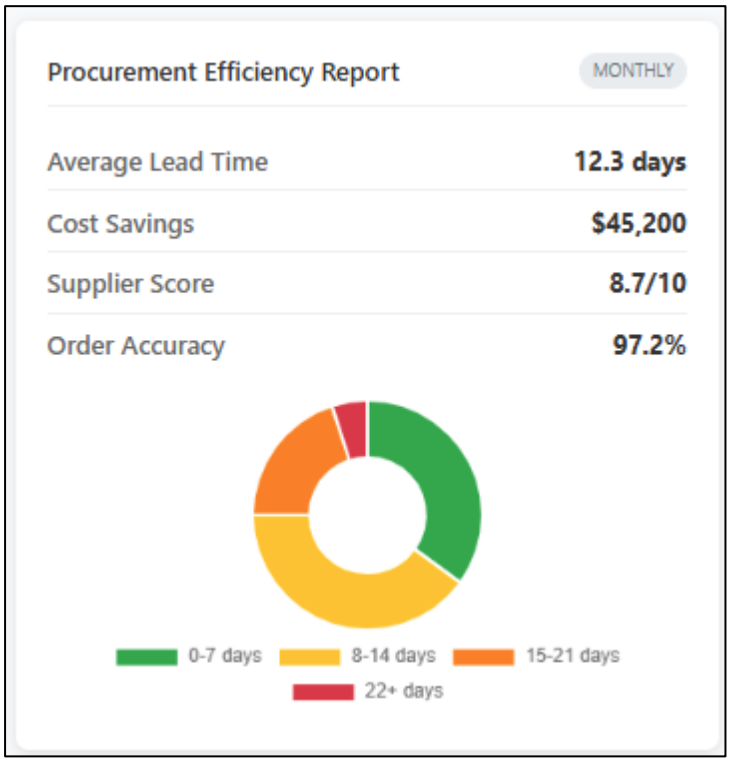

Figure 6: Procurement Efficiency Report

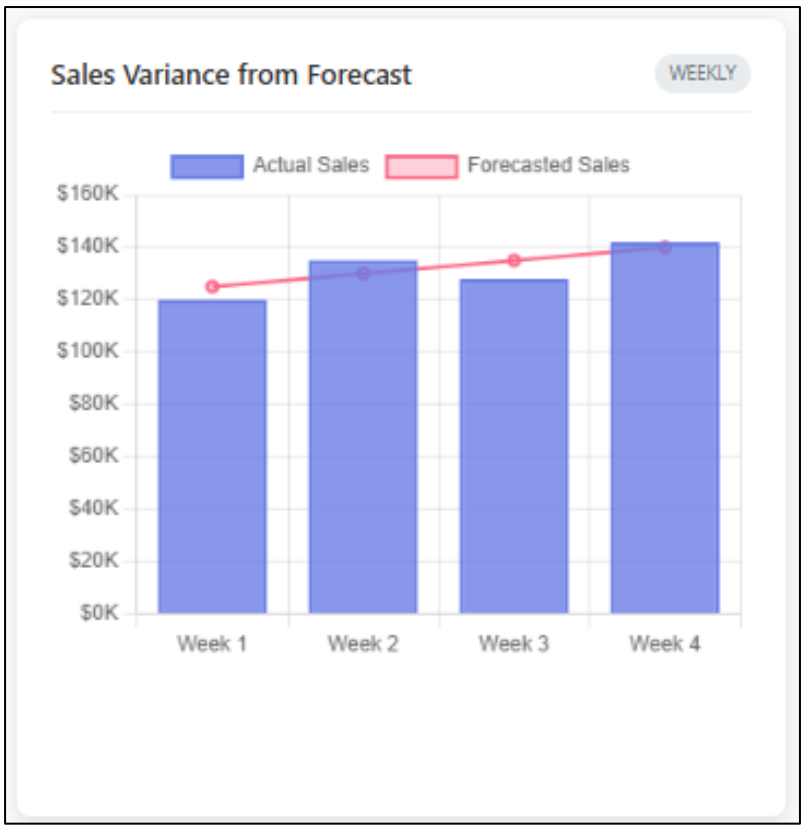

Figure 7: Sales Variance from Forecast

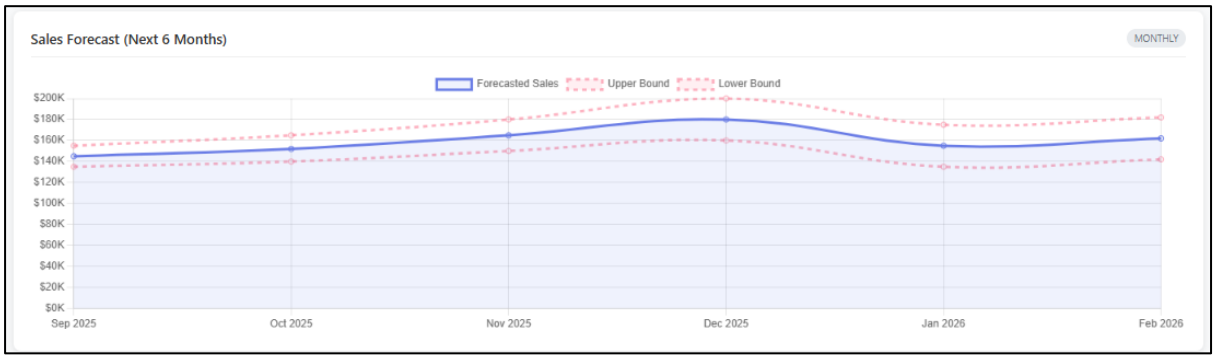

Figure 8: Sales Forecast (Next 6 Months)

Figure 9: Seasonality Analysis

In summary, the BI system must provide descriptive and predictive analytics using time series components (trend, seasonal, cyclical, irregular) and causal models to understand and forecast sales, supplier, and operational performance.

Preliminary Source System Analysis

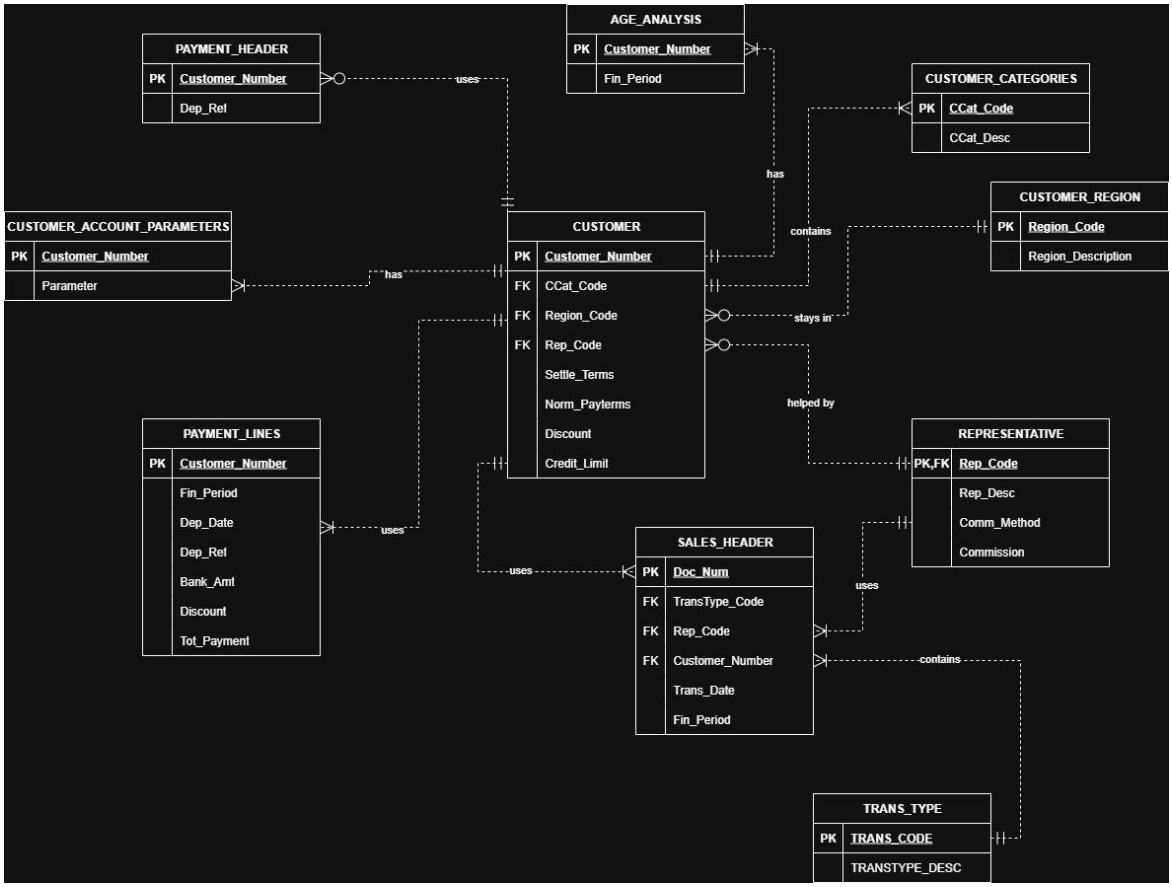

Figure 10: ClearVue Ltd ERD (part 1)

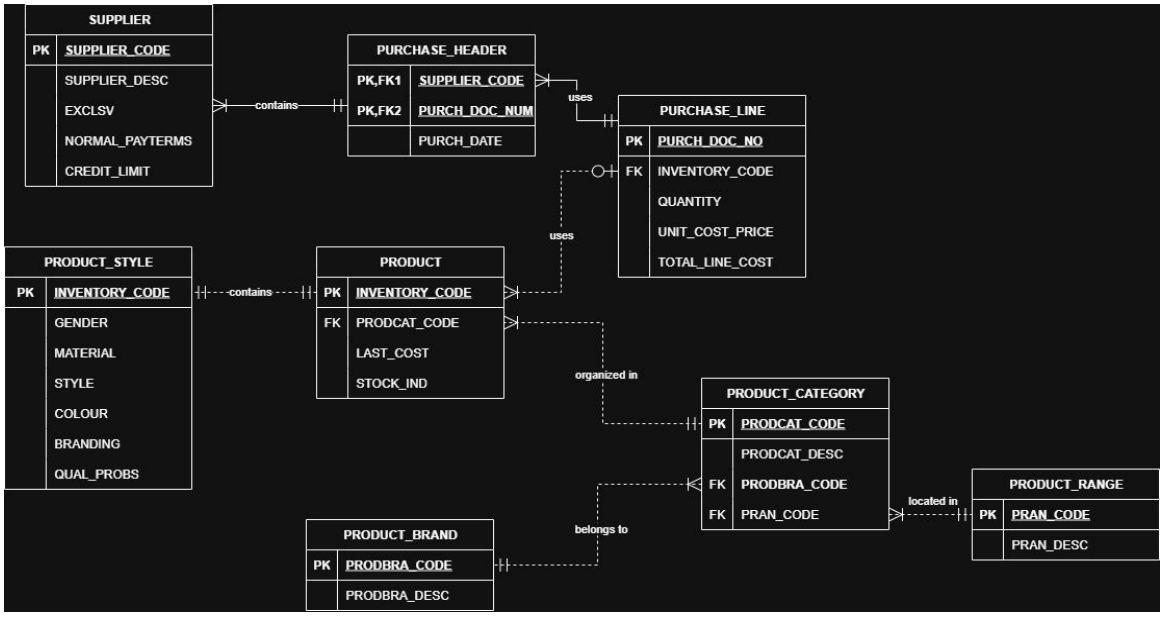

Figure 11: ClearVue Ltd ERD (part 2)

ClearVue Ltd Data Structures

Relationships

- Each payment header links with one age analysis.
- Each customer account parameter connects with one customer.
- Each customer links with one customer category.
- Each customer lives in one customer region.
- Each customer connects with one representative.
- Each sales header contains one transaction type.
- Each sales header engages one representative.
- Each sales header connects with one customer.
- Each payment header engages one payment line.
- Each sales header contains one payment line.

Hierarchies

- The customer hierarchy starts with customers, each belonging to customer categories, within a customer region, with representatives, selling to customers.
- The sales hierarchy starts with sales headers, has transaction types and payment headers with payment lines.

Data Structures

- Payment Header contains the customer number and the deposit reference.
- Age Analysis contains the customer number and the financial period.
- Customer Account Parameters contain the customer number and parameter.
- Customer contains customer number; customer category code; region code; representative code; terms of settlement; normal payment terms; discount; credit limit.
- Customer Categories contains the customer category code and the customer category description.
- Customer Region contains region code and region description.
- Representative contains representative code; representative description; commission method; commission.
- Payment Lines contain customer number; financial period; deposit date; deposit reference; bank amount; discount; total payment.
- Sales Header contains document number; transaction type code; representative code; customer number; transaction date; financial period.
- Transaction Type: Consists of transaction type code; transaction type description.

Preliminary Success Criteria

The proposed NoSQL-based BI system will be evaluated against the following preliminary criteria, which align with ClearVue Ltd's business requirements and high-level objectives. These criteria ensure that the new system outperforms the old Excel-based processes while meeting the needs of both users and the business.

Category	Measure of Success	Target	Measurement Method
User satisfaction	The overall user experience and ease of use for the new Power BI dashboards.	Satisfaction score of at least 4/5 from the end-users.	Surveys were conducted post-implementation. Feedback meetings after rollout.
Data accuracy	Data consistency and validation across several reports.	98%+ accuracy across validated datasets.	Manual and automated scripts validating data to the original Excel sources.
Query performance	The average response time for analytical queries.	Less than 3 seconds for 95% of queries.	Using systems logs and several performance testing tools, like MongoDB profiler.
Scalability	The ability to handle the growth of data and business needs (adding supplier analytics).	The system must be able to handle 2x the current data volume without performance degradation. Degradation. New data types must be supported within 1 week.	Load testing using simulated data.
Business impact	Enhance the outcomes of decision-making, particularly in	Forecast variances must be reduced by at least 10% - 20%.	By comparing pre- and post-systems forecast accuracies.

	terms of the accuracy of sales forecasts.		
Usability and adoption	How users adapt and frequently use the system.	Over 80% of users utilise the system dashboards every week.	Using different analytics of Power BI, like login frequency and report views.
Real-time analytics	The latency in processing streamed data.	Less than 5 seconds to do processing and update the dashboard.	Utilising monitoring tools provided by Power BI to track end-to-end latency.

Table 3: Preliminary Success Criteria

The above criteria will be refined during the design phase using stakeholder feedback and testing. Achieving these targets will demonstrate that the new system meets ClearVue's business needs.

Conceptual NoSQL Design

A document-oriented NoSQL database model will be used for this system. Document databases naturally support hierarchical data through embedding, eliminating the need for complex NoSQL joins typically used in relational databases. Document models scale horizontally by sharding collections, which suits the growth in sales and payment data over time.

The data the application will manage is discussed extensively in the following:

1. Sales data: This encompasses the core transactional data that is related to all the sales activities of ClearVue Ltd. Some of the attributes used include sales amounts, the dates of every transaction, and the associated performance metrics of the company. The data will be structured to align with ClearVue's unique financial year, which is defined as a financial month starting on the last Saturday of the preceding month and ending on the last Friday of the current month. This structure will ensure that the application is compatible with the company's financial reporting requirements.
2. Customer data: The information related to the customers includes identifiers (customer number), demographic details (customer region), and purchase history.
3. Payment headers: This will include information regarding the payment transactions, such as payment dates and amounts.

4. Product brands: This will include product categories and their subcategories. ClearVue facilitates analytics of sales, inventory management, and potential supplier-focused reporting.
5. Hierarchical data: Data for product categories and other nested relationships that is semi-structured. A NoSQL database's flexibility enables complex, non-relational data structures that facilitate efficient queries and analysis.
6. Streaming data: Real-time data feeds will be incorporated to provide live insights, such as transaction monitoring. This flexible compatibility will support dynamic decision-making and operational responsiveness.

The application will support queries to enable robust sales performance analysis across multiple timeframes, aligned with ClearVue Ltd.'s financial year structure. The specific queries include:

Sales Reports:

Daily sales reports: queries to summarise and analyse daily sales data to provide daily metrics (e.g., sales, number of sales, best-selling products).

Weekly sales reports: queries to summarise sales data for a week, based on the reporting week (e.g., Monday to Sunday), to follow short-term trends and sales performance.

Monthly sales reports: queries to summarise sales for a month (e.g. January 31 to February 27 for February 20xx), to provide metrics such as total sales, customer activity, or product performance.

Quarterly sales reports: queries to analyse sales for a quarter, providing the ability to assess mid-term sales performance and trends.

Annual sales reports: queries to provide a holistic sales analysis for a financial year, as part of long-term planning and strategy.

Hierarchical product queries: using hierarchical data to analyse product work by categories or brands (e.g. sales by category or subcategory).

Real-time analytics queries: queries to manage streaming data with a real-time reporting ability. For example, streaming data to monitor sales in real-time, or reporting the number of transactions in real-time for alerts.

Supplier analysis queries (for the future): queries designed to provide insights and analysis on supplier data, including supplier performance, product supply trends, and other relevant information. As the system matures, we will look to support supplier analytics.

Custom analysis queries: Queries based on a star schema (if possible, in the NoSQL design) will provide support for more extensive analysis or complex analytical queries,

such as identifying trends in sales, customer segments, and cross-referencing sales with payment data or product data.

JSON examples for the different collections that will be imported to MongoDB

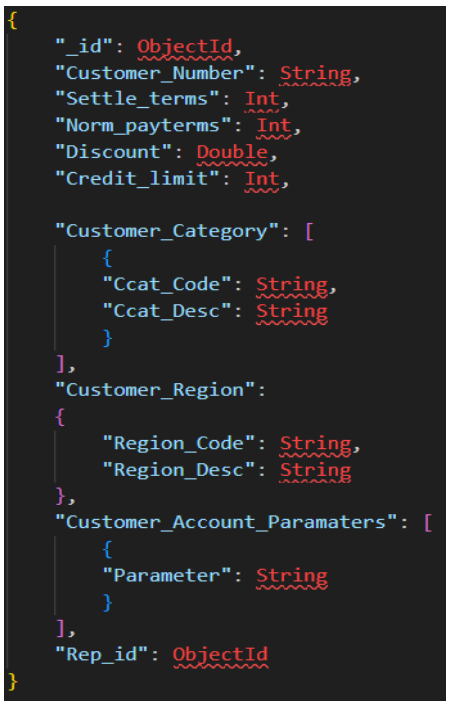

Figure 12: Customer Collection JSON

Stores data that supports sales analysis, credit management and customer-based decision making.

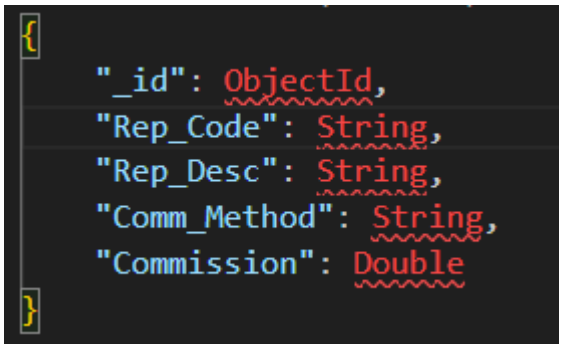

Figure 13: Representative Collection

Captures information on sales representatives and their commission structures. This will help track sales performance by rep and increase accountability in sales reports.

Figure 14: Payment Collection

Records customer payments and payment lines, supporting discount tracking, financial reporting and data consistency.

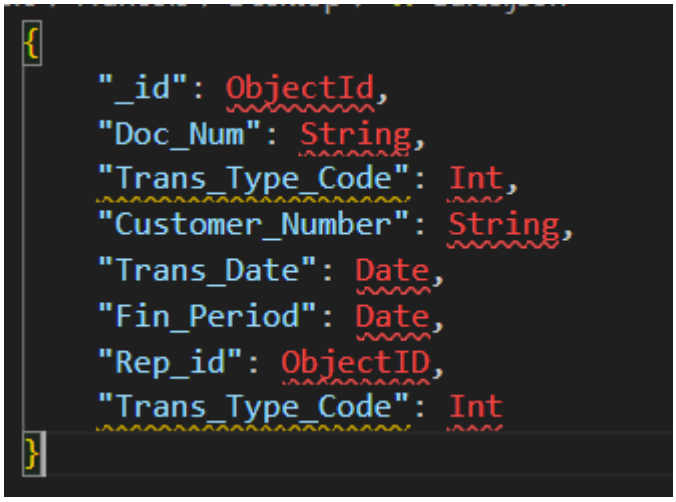

Figure 15: Sales Collection

Captures transaction-level data that will be used for daily, weekly, monthly and financial-year reports. This will be one of the core datasets used for reporting.

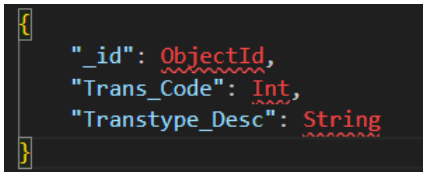

Figure 16: Transaction Types Collection

A central lookup of transaction codes and descriptions, which is referenced by Sales. This is kept separate because it is a business definition that changes slowly. Keeping this separate decouples it from high-volume sales data, allowing transaction types to be added/renamed without rewriting historical records.

Figure 17: Age Collection

Stores outstanding balances by time bucket, enabling real-time credit assessments and informed cash flow decision-making.

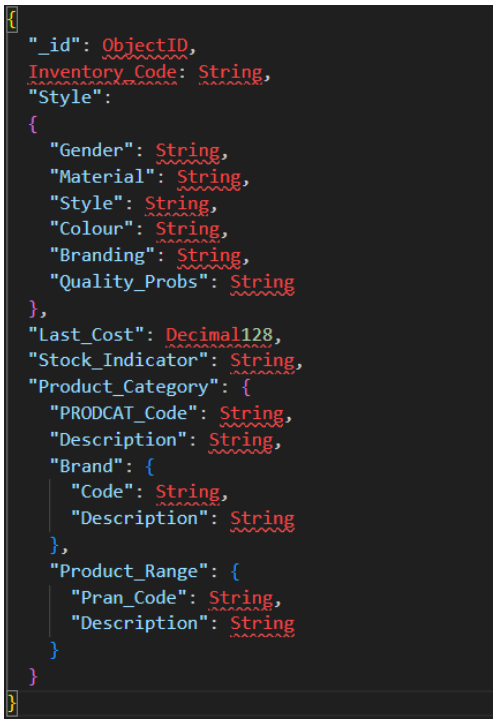

Figure 18: Product Collection

Captures information on ClearVue's products, enabling reporting and informed decision-making regarding products being sold.

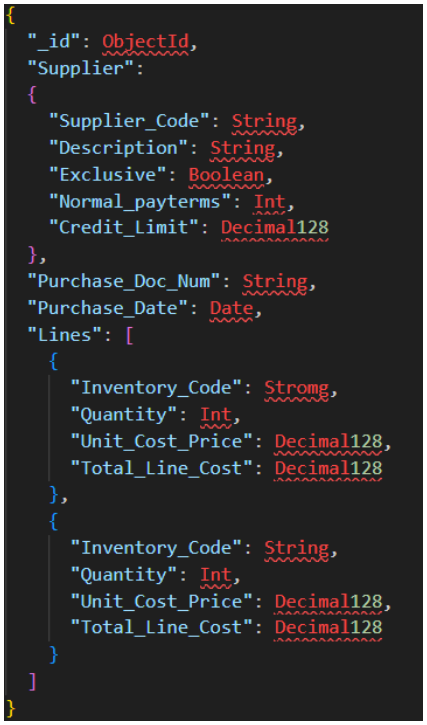

Figure 19: Product Purchase Collection

Stores product and supplier attributes, which enable ClearVue to create supplier-focused analytics, evaluate their supplier performance, and assess their contribution to sales.

AI Usage Log

AI Tool	Prompt	Reason	Output
ChatGPT	Can you design the NoSQL collections for this data? [I provided an image of our first ERD.] Currently, I have split my Customer, Sales, and Payments sections. The reason for this is that there is a high volume of business sales and purchases.	I wanted to compare ChatGPT's designs of the Customer, Sales, and Payment collections to my own designs. I also verified my designs and fixed any errors or missing features.	https://chatgpt.com/share/68b
ClaudAI	Create a sample Power BI Dashboard for the following requirements	I visualised the Informational and Analytical Requirements section, as it is much easier to convey the connotation when	https://claude.ai/share/a1c917

		presented with a visual representation of how the requirements will be used to achieve an end-product.	
Grok	Using the information provided, can you provide me with ideas for Preliminary Success Criteria in a table format? Please provide reasons for it	I used Grok to generate ideas for the Preliminary Success Criteria. I used the output ideas to create the Preliminary Success Criteria for the proposed solution.	https://grok.com/share/c2hhcn
Grok	Using the executive summary and Requirements definition approach, as well as the other business background information I have provided, do you agree with our business requirements	I wanted to verify that my business requirements match the background information of the business.	https://grok.com/share/c2hhcn
Grok	Can you explain NoSQL collections and when I should split up collections or keep them as one	I struggled to understand how NoSQL collections work, so I used the provided Grok output to gain a better understanding. I used this knowledge to help design the document and proposed system.	https://grok.com/share/c2hhcn
Grok	Could you please provide me with two different NoSQL approaches to this? Key-Value and Document database (with added attachment of the ERDs)	I wanted to know which NoSQL approach would be best suited for ClearVue Ltd.'s new solution	https://grok.com/chat/6fc90a6

Table : AI Usage Log

References

BI Consulting Pro. (2024, May 24). What is scalability in Microsoft Power BI | DP-600 Ep15 #microsoftfabric #dp600 #powerbi [Video]. YouTube. https://www.youtube.com/watch?v=2VymWcpoLGo

Business Performance Dashboard. (2025, August 15). Claud AI. Retrieved August 15, 2025, from https://claude.ai/public/artifacts/a05b6487-18bc-4203-954c-e959b21412d5

ChatGPT - NoSQL collections design. (n.d.). ChatGPT. https://chatgpt.com/share/68b45651-dea0-8000-ae07-53e7f217600f

Components of Time Series | Forecasting Class Notes | Fiveable. (n.d.). Fiveable. Retrieved August 14, 2025, from https://library.fiveable.me/forecasting/unit-2/components-time-series/study-guide/Yuqvby3KgGXmk5IT

Data-Driven-Operations-Image-1.png. (2024, January). https://falkonry.com/. https://falkonry.com/wp-content/uploads/2024/01/Data-Driven-Operations-Image-1.png

Evaluation of Business Requirements. (2025). Grok.com. https://grok.com/share/c2hhcmQtNA%3D%3D_6f5e7978-be06-4983-a5d2-9e267047ce5f

Falkonry. (2024, January 30). Data Driven Operations | Falkonry. Retrieved August 14, 2025, from https://falkonry.com/glossary/data-driven-operations/

GeeksforGeeks. (2025, July 11). SQL vs NoSQL: Which One is Better to Use. GeeksforGeeks. Retrieved August 13, 2025, from https://www.geeksforgeeks.org/sql/sql-vs-nosql-which-one-is-better-to-use/

Guidelines for Deciding. (2025). Grok.com. https://grok.com/share/c2hhcmQtNA%3D%3D_f5e646bd-96d4-467a-9244-3b7d3a7b76b2

MongoDB. (n.d.). What is NoSQL? NoSQL databases explained. Retrieved August 13, 2025, from https://www.mongodb.com/resources/basics/databases/nosql-explained

Pickering, M. (2019, March 15). Trend analysis - Level 4 study tips - AAT Comment. AAT Comment. Retrieved August 15, 2025, from https://www.aatcomment.org.uk/learning/study-tips/trend-analysis-level-4-study-tips/

Preliminary Success Criteria Table. (2025). Grok.com. https://grok.com/share/c2hhcmQtNA%3D%3D_8916b757-7f92-4a8b-bfe3-4a462be65b43

RENDER, B. (2018). Quantitative Analysis for Management: THIRTEENTH EDITION. In R.

STAIR, JR. M., M. HANNA E., & T. HALE S. (Eds.), Quantitative Analysis for Management (13th ed., pp. 165–192). Pearson Education Limited.

SINGH, J., SINGHAL, S., & Ya-Lun-Chau. (n.d.). TIME SERIES ANALYSIS.

https://moirabaricollegeonline.co.in/attendence/classnotes/files/1608735484.pdf

What Are NoSQL Collections? (2025). Grok.com.

https://grok.com/share/c2hhcmQtNA%3D%3D_f5e646bd-96d4-467a-9244-3b7d3a7b76b2