29/09/2025

DIMENSIONAL MODEL FOR HANDLEBARHAVEN

Coert, Miss Madeline

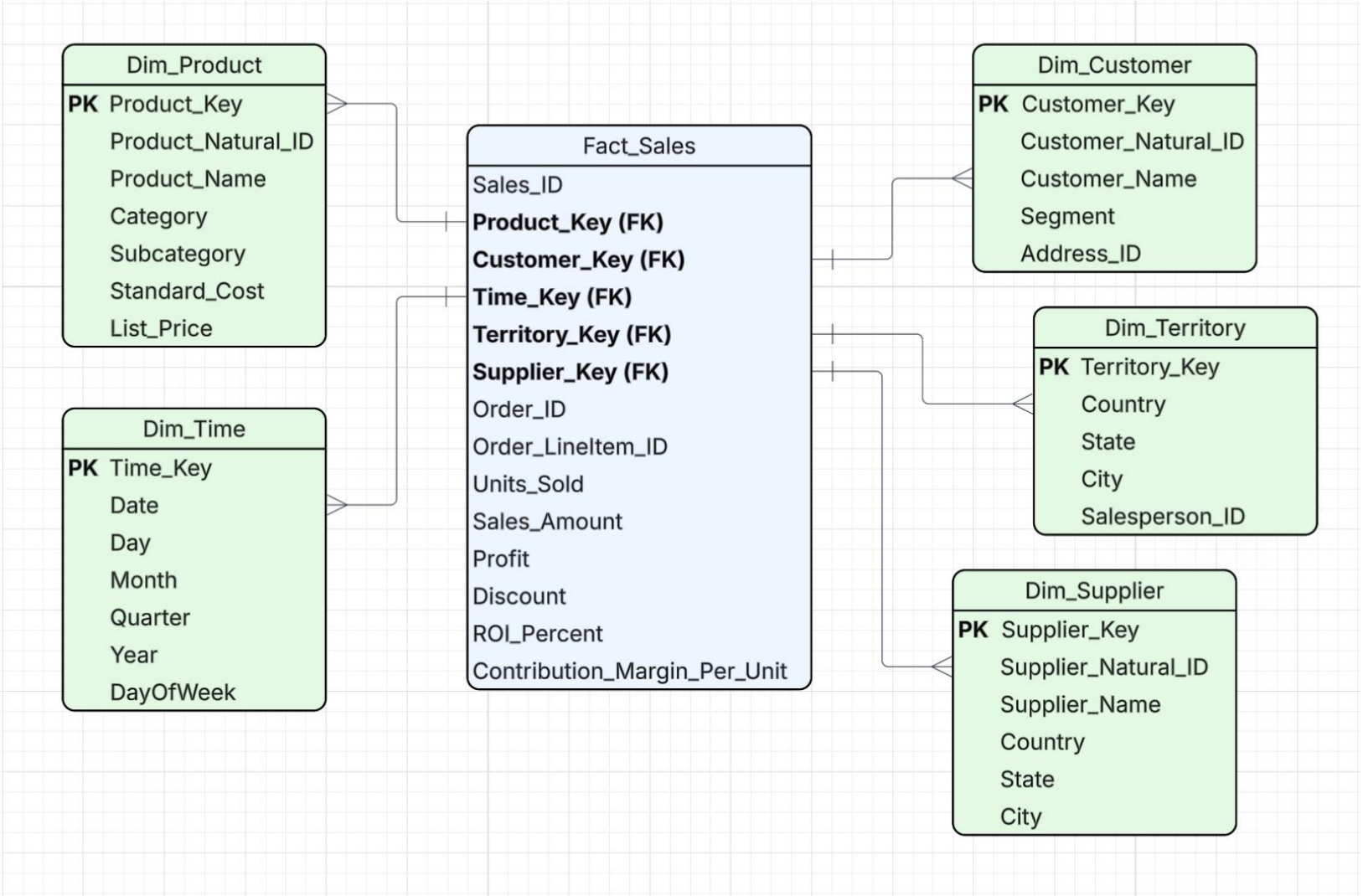
Table of Contents

Executive Summary

In today's business environment, companies generate vast amounts of information every day. While this data is valuable, it can quickly become overwhelming without the right tools to turn it into meaningful insight. Business leaders need to be able to see clearly which customers, products, regions and time periods drive performance so they can make confident, data-driven decisions. Dimensional modelling provides a structured way to organise data so that insights are easy to access, understand and act upon. It transforms raw information into a business ready format that managers and executives can use without needing technical expertise (Kimball & Ross, 2013).

The purpose of this report is to design a star schema for Handlebar Haven that will support an accurate analysis of sales, customers, suppliers and territories. At this point of planned expansion, it is essential that the business fully understands where its strengths and weaknesses lie. The proposed model will allow Handlebar Haven to quickly identify its most profitable customers, products and markets, while also highlighting underperforming areas that may require change. By providing a clear and accessible view of business performance, the model will equip decision makers with the insights they need to minimise risk and maximise growth opportunities.

Star Schema Diagram



Dimension Model Design

Introduction

This dimensional model has been developed to transform Handlebar Haven's operational sales and purchasing data into an analytical platform that directly answers the five executive questions raised by the CEO. The company is at a critical stage of planned expansion, and the leadership team requires a system that provides transparent and accurate insights into which customers, products, territories, timeframes and categories are driving profitability. The model follows the Kimball style star schema principles, which emphasise clarity, performance and flexibility (Kimball & Ross, 2013). This ensures that executives and analysts can query the dataset and arrive at consistent answers without ambiguity. By placing sales at the centre of the model and surrounding it with conformed dimensions, Handlebar Haven gains a platform that not only addresses the current questions but is extensible for future analytics, supporting long-term decision-making in a competitive and fast paced market.

Fact Table Design & Grain

At the centre of the schema lies the Fact_Sales table. This fact table was designed at the finest possible level of detail with one row per order line item. Each row captures the sale of a single product to a customer within an order, including the associated measures and links to relevant dimensions. The decision to adopt this transaction level grain was driven by the CEO's explicit need to understand profitability not just at the customer or category level but down to the individual products. For example, without line level detail, the company could identify that a customer generated a certain amount of revenue in a quarter, but it would be unable to see which products contributed disproportionately to that revenue or which ones dragged profitability down. By adopting the lowest grain, executives can identify high performing SKUs, pinpoint low value items and make informed decisions about which subcategories to retain or replace.

Furthermore, the fact table supports aggregation at multiple levels, such as monthly sales by state, quarterly profit by product category and annual contribution margin by customer segment. This flexibility addresses four of the five CEO questions directly, since customer, product, territory and time analyses all begin with the ability to aggregate transaction level data in multiple ways. The transaction grain also protections the company's future analytical needs. As data volumes continue to grow, Handlebar Haven will require less changes to its star schema when answering more granular questions. Instead, executives will always have the option to drill down into the lowest level of granularity. This bottom-up modelling choice ensures that no insights are lost and future use cases such as customer level churn modelling or promotion effectiveness analysis, can be accommodated without rebuilding the warehouse.

Fact Table Measures & Calculation Logic

The facts selected for inclusion in the model are directly aligned with the CEO's questions about profitability and performance. Each row of the fact table contains numeric measures that allow profitability to be evaluated by customer, product, territory and time. These measures include units sold, sales amount, profit, discount and return on investment percentage. In addition, contribution margin per unit has been calculated by dividing profit by units sold at the line-item level. This measure is particularly powerful, as it highlights which products, customers, or regions generate the highest profitability efficiency, rather than simply the highest revenue.

The natural composite keys Order_ID and Order_LineItem_ID were preserved within the fact table for traceability. This ensures that every analytical result can be reconciled back to the operational source if needed for auditing or validation. However, the surrogate key of Sales_ID, was also introduced as the primary key. This surrogate simplifies joins, improves query performance and supports integration with BI tools that rely on simple integer-based keys.

By modelling profit as revenue minus cost and ensuring that cost is sourced from the appropriate record, the fact table allows executives to understand profitability in a way that aligns with real-world business conditions. Discounts are stored explicitly, ensuring visibility into how pricing strategies affect margins. ROI% allows for comparisons between products with different cost structures, giving the CEO a strong basis for comparing investments across categories. Collectively, these facts form the quantitative foundation of the model, ensuring that each of the five CEO questions can be answered with confidence.

Dimension Design & Hierarchies

Surrounding the fact table are five key dimensions: Product, Customer, Territory, Time and Supplier. Each dimension has been carefully modelled to support natural business hierarchies and analytical paths. The Product dimension provides a Category to Subcategory to Product hierarchy, capturing list price, standard cost and product identifiers. This dimension enables executives to analyse profitability both at the SKU level and at higher levels of aggregation, such as comparing subcategories of road bikes against mountain bikes. The ability to build from product to category is central to answering the CEO's fifth question about which categories or subcategories should be replaced. The Customer dimension stores customer attributes and classifies customers into retail stores or individual consumers. This allows segmentation analysis, helping identify not only the most profitable customers overall but also whether Handlebar Haven's growth is being driven more by bulk retail relationships or direct to consumer e-commerce sales.

The Territory dimension follows a geographic hierarchy of Country to State to City, with salespeople linked to regions. This structure supports analysis of profitability by region and salesperson, directly answering the CEO's third question about which territories are most profitable. The Time dimension captures dates at the day level, with hierarchies that work up to month, quarter and year. Attributes such as day of week enable more detailed seasonality analyses, including the CEO's interest in profitability by day of week. Finally, the Supplier dimension provides visibility into vendors, distinguishing between internal manufacturing and external suppliers.

This allows Handlebar Haven to analyse not just which products are profitable, but whether in house or externally sourced products contribute more to margins. Product and Time are implemented as conformed dimensions, ensuring consistency across both sales and purchasing facts. This means that profitability calculations are synchronised across all business processes, eliminating the risk of misaligned metrics.

Keys, Traceability & ETL Considerations

The schema balances the use of natural keys for traceability and surrogate keys for performance and maintenance. Natural keys such as Order_ID and Product_ID were preserved in the fact and dimension tables to ensure that all analytics can be reconciled with operational source systems. These keys are critical for auditing, as they provide a direct link back to the original transactions. However, surrogate keys were introduced as the primary keys in each dimension and in the fact table. Surrogates provide technical advantages, including reduced storage, faster joins and the ability to manage slowly changing dimensions (Dragos-Paul, 2011). For example, if a customer moves to a new address or a salesperson is moved to a different region, the system keeps track of the changes so that past sales are still connected to the right customer or salesperson information.

ETL processes are responsible for assigning surrogate keys through lookup logic during load routines. They must also manage slowly changing dimensions with appropriate business rules, Type 1 for correcting errors and Type 2 for capturing historical changes (Simitsis, 2005). Additionally, fact load routines must reference the correct dimension version, particularly for products where historical prices and costs change over time. This ensures that profitability metrics are consistent and historically accurate. By combining surrogate keys for usability with natural keys for traceability, the schema ensures that analysts have both performance and accountability. Executives can rely on aggregated insights, while auditors can confidently drill back to individual transactions when needed.

Product History, Discounts, ROI & Assumptions

Handlebar Haven's operational system records product costs and list prices in the Products_History table, which is updated every six months. This introduces an important nuance where profitability cannot be calculated using today's costs but must be aligned with the costs that were effective when a sale was made. To address this, the ETL process matches each sales transaction with the relevant record in Products_History based on the order date. Profit is then calculated as revenue minus the historical cost, ensuring that profitability metrics are time appropriate.

Discounts are also incorporated into the model. When a special offer or bulk discount applies, the actual unit price on the sales line is treated as authoritative. A discount value is derived by comparing the historical list price with the actual sales price, providing visibility into the financial impact of discounting strategies. Furthermore, ROI is calculated by dividing profit by historical cost, giving executives a consistent way to compare how effectively money is converted into returns across different products and time periods. The model assumes that Standard_Cost in the product history is a reasonable replacement for true company cost, with flexibility to substitute actual invoice costs if those become available. These considerations help ensure that executives are not misled by oversimplified metrics and that profitability analysis reflects the real conditions under which sales occurred.

Conclusion

This star schema directly addresses each of the CEO's strategic questions. It identifies key customers by linking profitability to the Customer dimension. It determines the most and least profitable products and categories through the Product hierarchy. It evaluates profitability across territories and salespeople through the Territory dimension. It analyses seasonal performance through the Time dimension. Finally, it distinguishes profitable versus underperforming categories for replacement, using the Contribution Margin and ROI measures.

Beyond answering current questions, the model establishes a strong foundation for further company growth. By extending the schema to incorporate external data sources such as competitor pricing or market demand trends the organisation can move beyond internal reporting to true market benchmarking. The structure also enables advanced use cases like predictive analytics for demand forecasting, churn modelling or discount effectiveness, preparing executives with forward looking insights rather than purely historical reporting. To protect the reliability of these insights, Handlebar Haven should implement ongoing data governance practices, ensuring that historical costs, discounts and dimension changes are consistently validated and auditable. Over time, the schema can also support scenario based modelling, enabling leadership to evaluate the financial impact of strategic decisions before committing resources.

In summary, these recommendations position Handlebar Haven not only to answer its immediate strategic questions but also to anticipate future opportunities, reduce risk and embed data-driven decision-making across the organisation. This ensures the company's leadership can expand confidently, backed by analytics that are transparent, accurate and scalable.

Data Dictionary

Attribute Name	Brief Description	Source Table/ Field	Transformation
Sales_ID	Surrogate key for fact table		
Product_Key	Foreign key to Product dimension	Products_History.Product_ID	
Customer_Key	Foreign key to Customer dimension	Customers.Customer_ID	
Time_Key	Foreign key to Time dimension	Sales_Orders.Order_Date	
Territory_Key	Foreign key to Territory dimension	Sales_Orders.Territory_ID	Join Sales_Orders → Sales_Persons → Territory_ID
Supplier_Key	Foreign key to Supplier dimension	Supplier.Supplier_ID	
Order_ID	Sales order number	Sales_Order_Details.Sales_Order_Number	
Order_Lineitem_ID	Sales order line ID	Sales_Order_Details.Order_Line_ID	
Units_Sold	Units sold	Sales_Order_Details.Order_Qty	
Sales_Amount	Total amount for line item	Sales_Order_Details.Line_Total	
Profit	Profit on the line item	Sales_Order_Details.Unit_Price, Sales_Order_Details.Order_Qty, Products_History.Standard_Cost	(Unit_Price × Order_Qty) – (Standard_Cost × Order_Qty
Discount	Difference between list price and unit price. (assumed negotiated price).	Products_History.List_Price, Sales_Order_Details.Unit_Price	List_Price – Unit_Price
ROI_Percent	Return on investment %	Sales_Order_Details.Unit_Price, Sales_Order_Details.Order_Qty, Products_History.Standard_Cost	(Profit ÷ (Standard_Cost × Order_Qty)) × 100
Contribution_Margin_Per_Unit	Margin per unit sold	Sales_Order_Details.Unit_Price, Sales_Order_Details.Order_Qty, Products_History.Standard_Cost	Profit ÷ Units_Sold

Dim_Product			
Product_Key	Surrogate key	Surrogate key (ETL)	
Product_Natural_ID	Natural key	Products.Product_ID	
Product_Name	Product name	Products.Product_Name	
Category	Product category ID	Product_Categories.Product_Cat_ID	
Subcategory	Product subcategory ID	Product_Sub_Category.Product_Subcategory_ID	
Standard_Cost	Standard Cost	Products_History.Standard_Cost	
List_Price	List price	Products_History.List_Price	

Dim_Time			
Time_Key	Surrogate Key	Surrogate key (ETL)	
Date	Order date	Sales_Orders.Order_Date	
Day	Day from Date	Derived (from Date)	
Month	Month from Date	Derived (from Date)	
Quarter	Quarter from Date	Derived (from Date)	
Year	Year from Date	Derived (from Date)	
DayOfWeek	Day of week from Date	Derived (from Date)	
Dim_Customer			
Customer_Key	Surrogate key	Surrogate key (ETL)	
Customer_Natural_ID	Natural key	Customers.Customer_ID	
Customer_Name	Full name	Customers.First_Name + Last_Name	
Segment	Customer segment	Customers.Person_Type (e.g., Retail/Consumer)	
Address_ID	Address foreign key	Customer_Address.Address_ID	

Dim_Territory			
Territory_Key	Surrogate key	Surrogate key (ETL)	
Country	Country name	Country_Region_Code.Country_Name	
State	State Name	State.State_Name	
City	Customer City	Customer_Address.Client_City	
Salesperson_ID	Sales person ID	Sales_Persons.Sales_Person_ID	

Dim_Supplier			
Supplier_Key	Surrogate key	Surrogate key (ETL)	
Supplier_Natural_ID	Natural key	Supplier.Supplier_ID	
Supplier_Name	Supplier name	Supplier.Supplier_Name	
Country	Supplier country	Supplier.Country	
State	Supplier address	Supplier.Address	
City	Supplier city	Supplier.City	

References

Dragos-Paul, P. O. P. (2011). Natural versus Surrogate Keys. Performance and Usability. *Database Systems*, 55. Retrieved from: http://dbjournal.ro/archive/4/6_Dragos_Pop.pdf

Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling*. John Wiley & Sons. Retrieved from: https://books.google.com.au/books?hl=en&lr=&id=4rFXzk8wAB8C&oi=fnd&pg=PR27&dq=importance+dimensional+modelling+data+analytics+&ots=3q7TRfZ5QK&sig=PIUpnkD9DPiLLY71AzTxjxkfgRY#v=onepage&q=importance%20dimensional%20modelling%20data%20analytics&f=false

Simitsis, A., Vassiliadis, P., & Sellis, T. (2005). Optimizing ETL processes in data warehouses. In *21st International Conference on Data Engineering (ICDE'05)* (pp. 564-575). Ieee. Retrieved from: https://ieeexplore.ieee.org/abstract/document/1410172

AI Usage Statement:

AI assistance (ChatGPT) was used throughout the project to edit and polish written paragraphs, ensuring readability, clarity and alignment with professional tone. It was also applied to refine formatting and descriptions in the data dictionary. All analytical design, modelling and strategic insights were developed independently.