



ADVANCING ANALYTICS

Data Modelling Showdown

Unravelling the mysteries of Different
Data Modelling Approaches



ABOUT ME



Ust Oldfield (he/him)

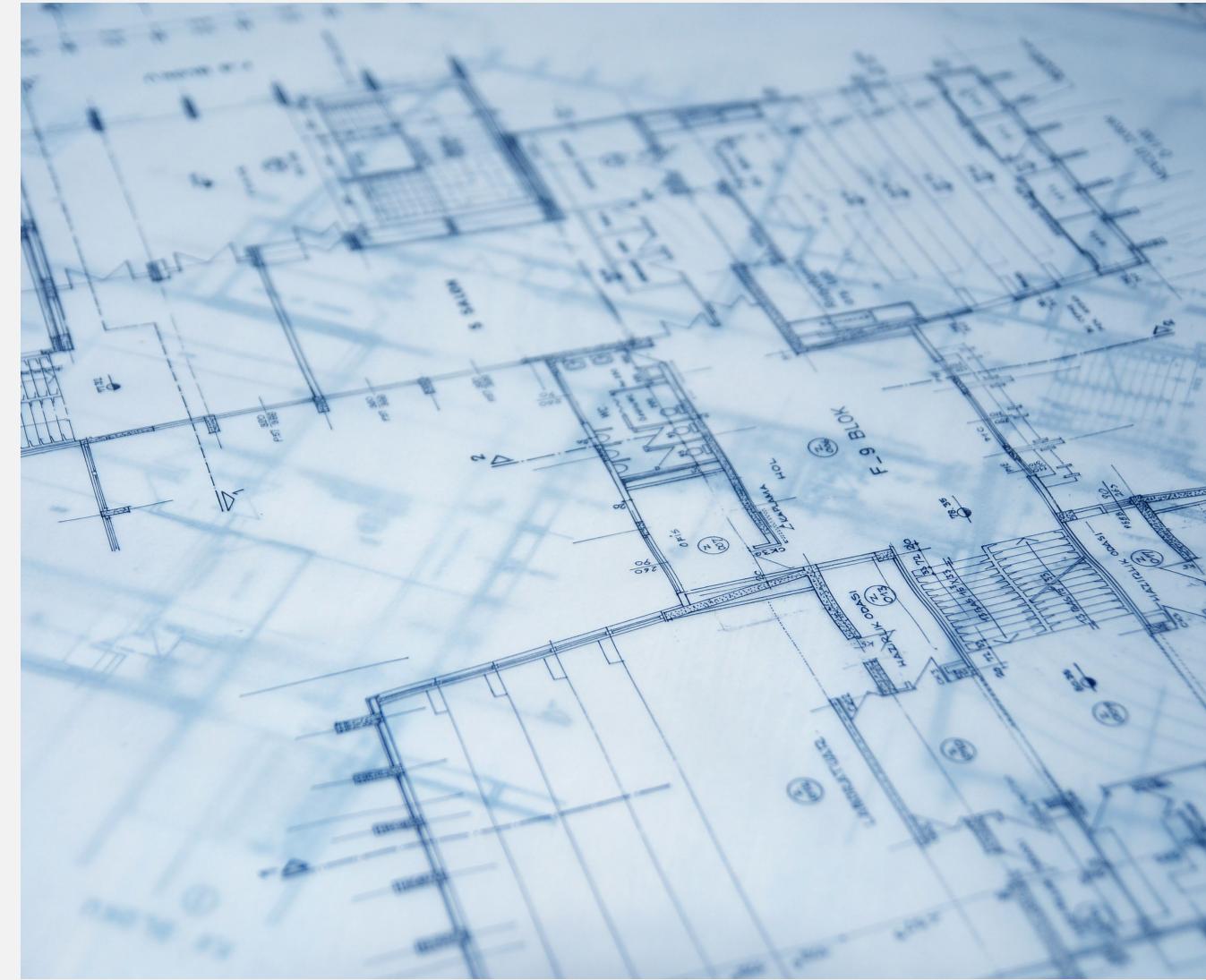
Head of Analytics



EVERYTHING, EVERYWHERE, ALL AT ONCE



WHAT IS A DATA MODEL?



NOUNS AND VERBS



Customer **Buys** a Product

Employee **Orders** Stock

Company **Manufactures** a Product

ARE ALL MODELS EQUAL?



ARE ALL MODELS EQUAL?



ARE ALL MODELS EQUAL?



FOUR PERSPECTIVES TO MODEL

User, Conceptual, Logical & Physical



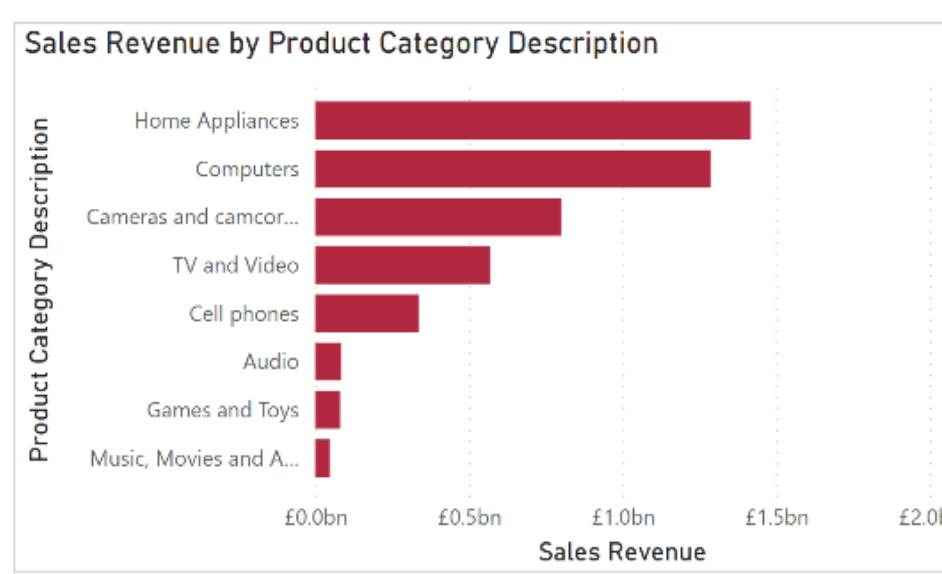
USER MODEL

Traditional

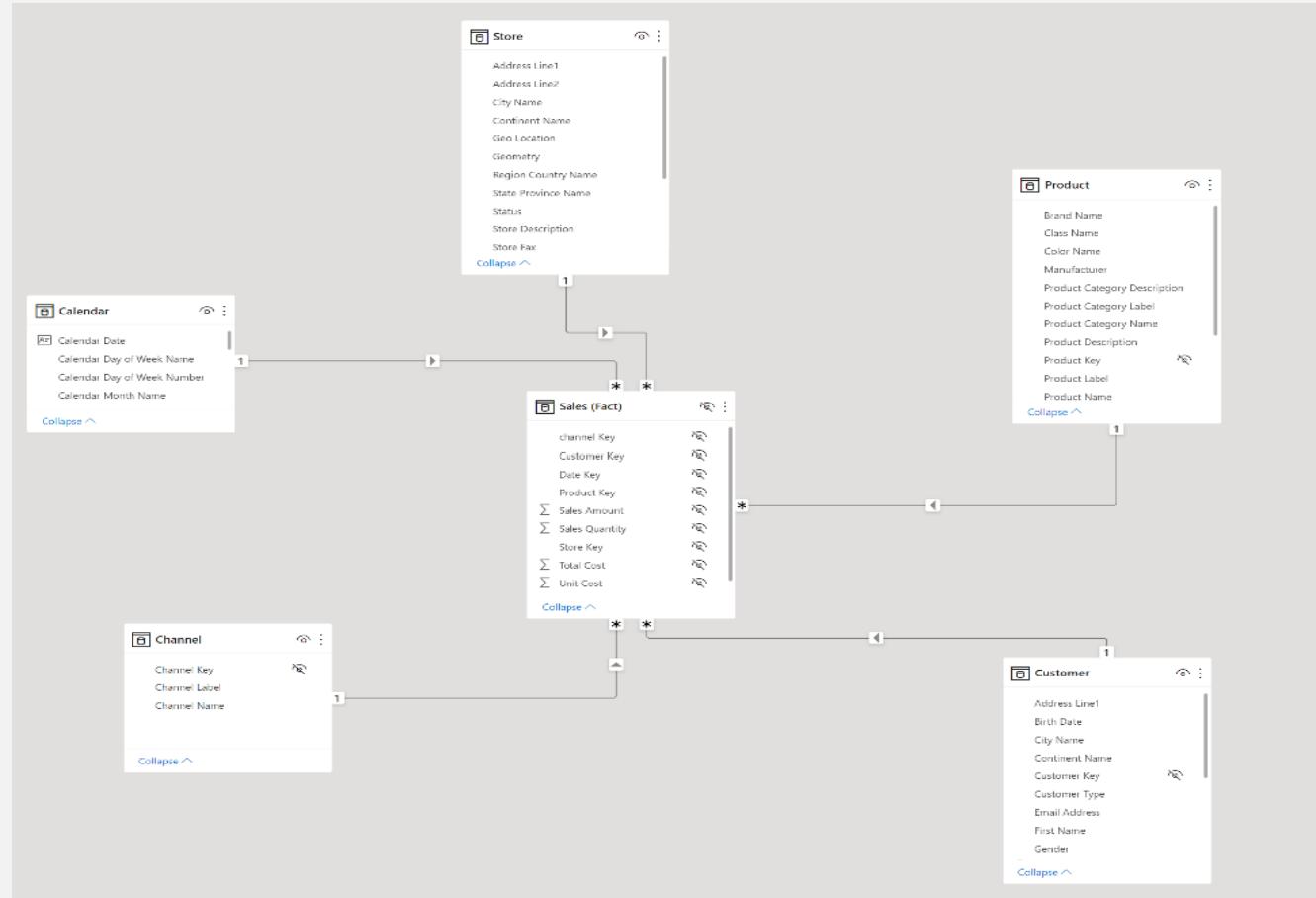
Brand Name	Product Category Description	Sales Revenue	Sales Unit Quantity
Fabrikam	Cameras and camcorders	£487,526,342	883,067
Contoso	Home Appliances	£409,170,649	882,783
Litware	Home Appliances	£385,483,883	1,573,761
Wide World Importers	Computers	£330,211,537	695,956
Adventure Works	Computers	£320,481,334	702,509
Proseware	Computers	£309,664,690	1,187,117
Fabrikam	Home Appliances	£269,066,882	696,947
The Phone Company	Cell phones	£215,784,039	821,932
A. Datum	Cameras and camcorders	£200,888,455	824,994
Adventure Works	TV and Video	£166,734,817	322,470
Contoso	Computers	£153,596,659	1,384,550
Litware	TV and Video	£151,118,320	292,627
Fabrikam	Computers	£145,913,491	262,156
Southridge Video	TV and Video	£126,161,992	494,359
Total		£4,623,947,002	24,648,387

USER MODEL

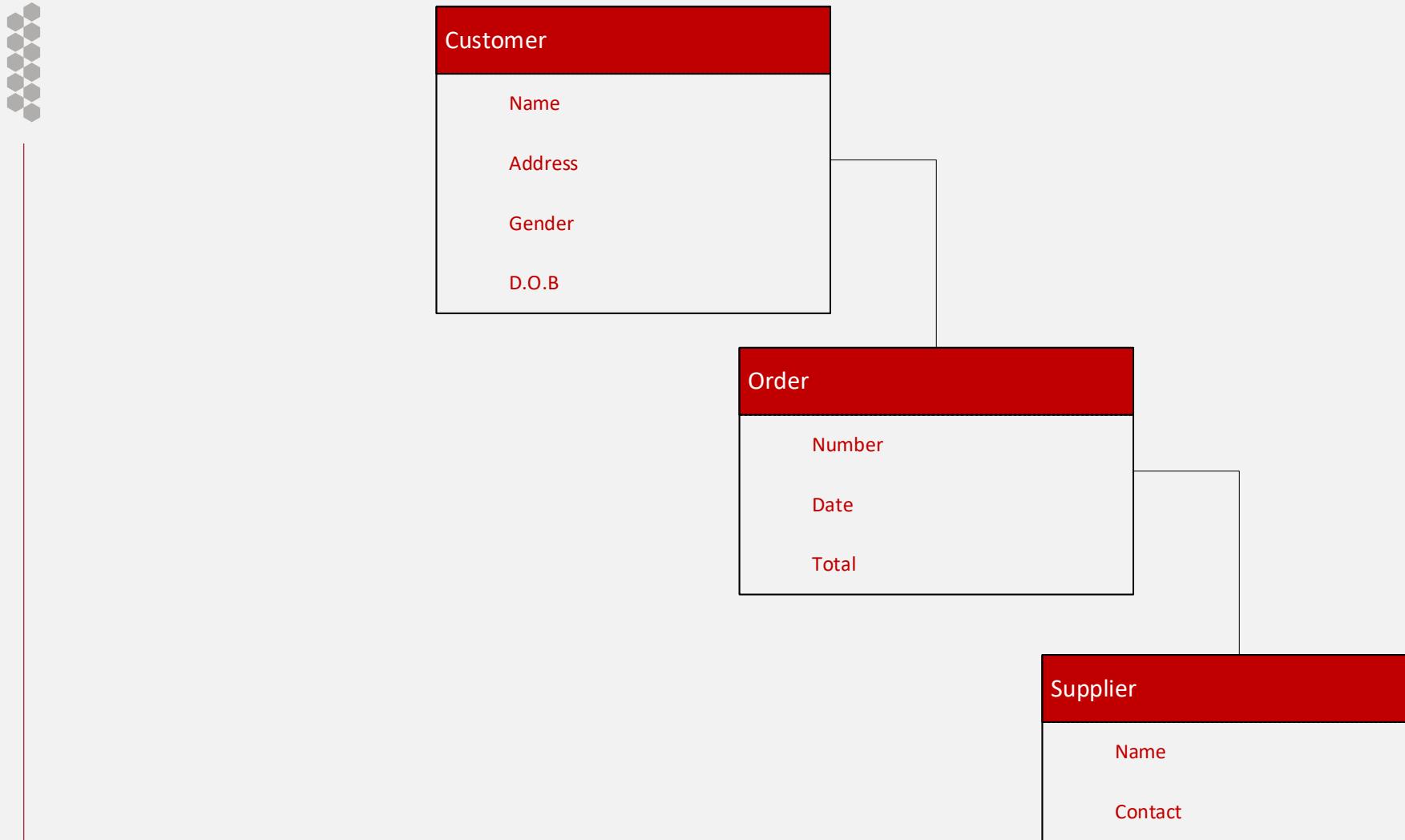
Enhanced



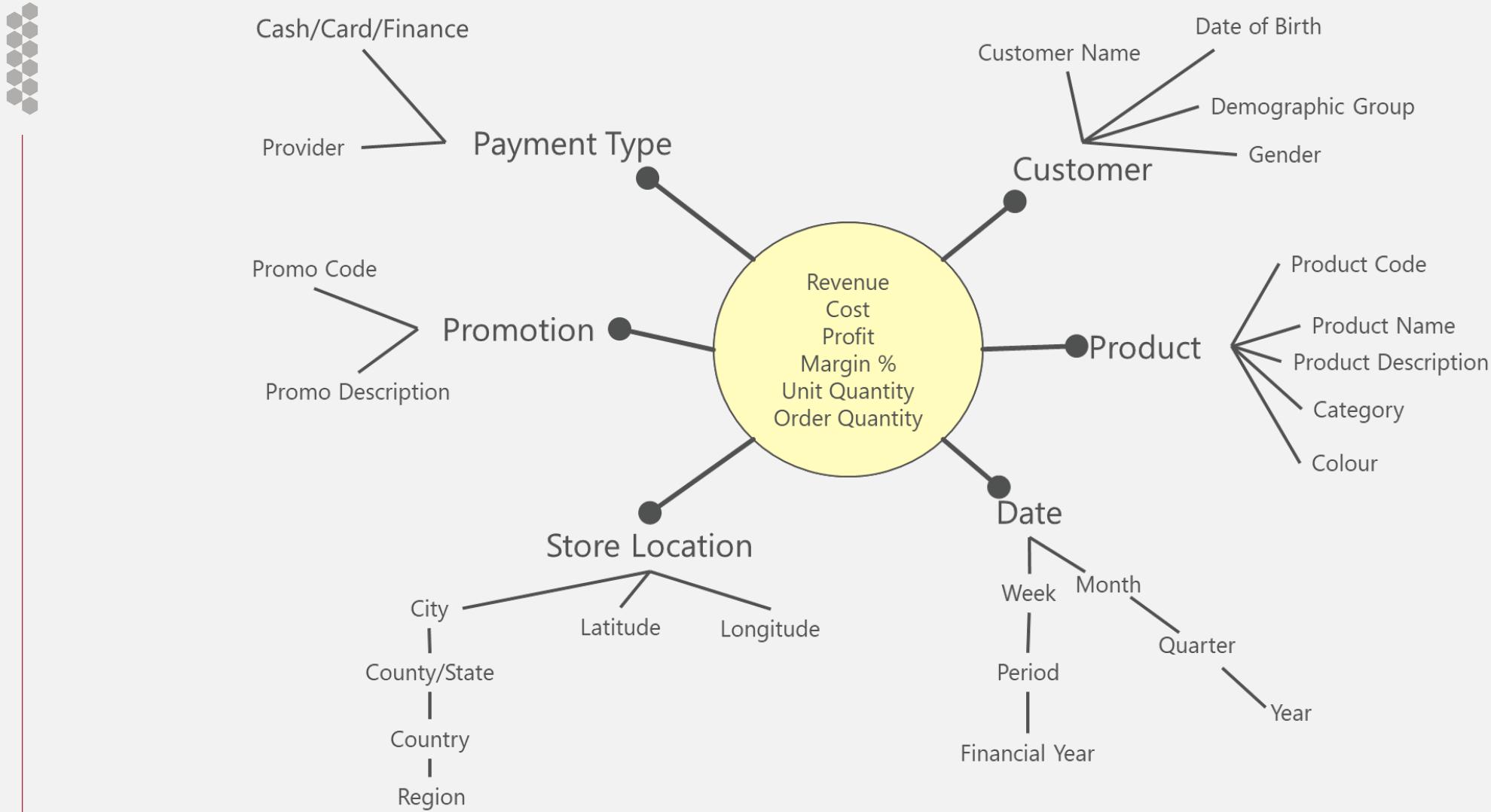
PHYSICAL MODEL



LOGICAL MODEL



CONCEPTUAL MODEL



WHAT IS A TRAIN?

- Made up of rolling stock – carriages and an engine
- 12:09 from Gare du Nord to St Pancras
- Today's 12:09 from Gare du Nord to St Pancras
- Marketing concept e.g. Orient Express

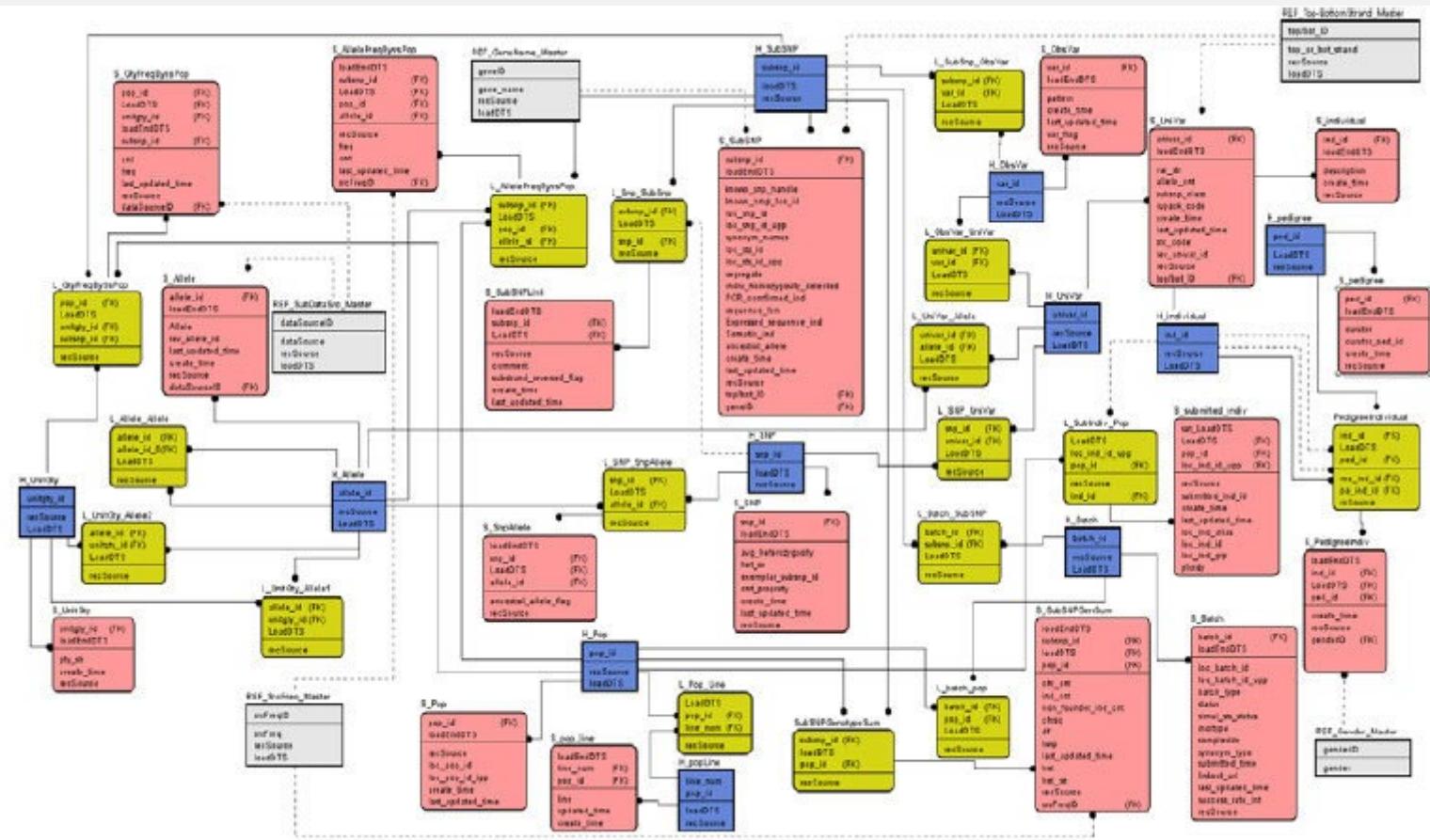


The Showdown



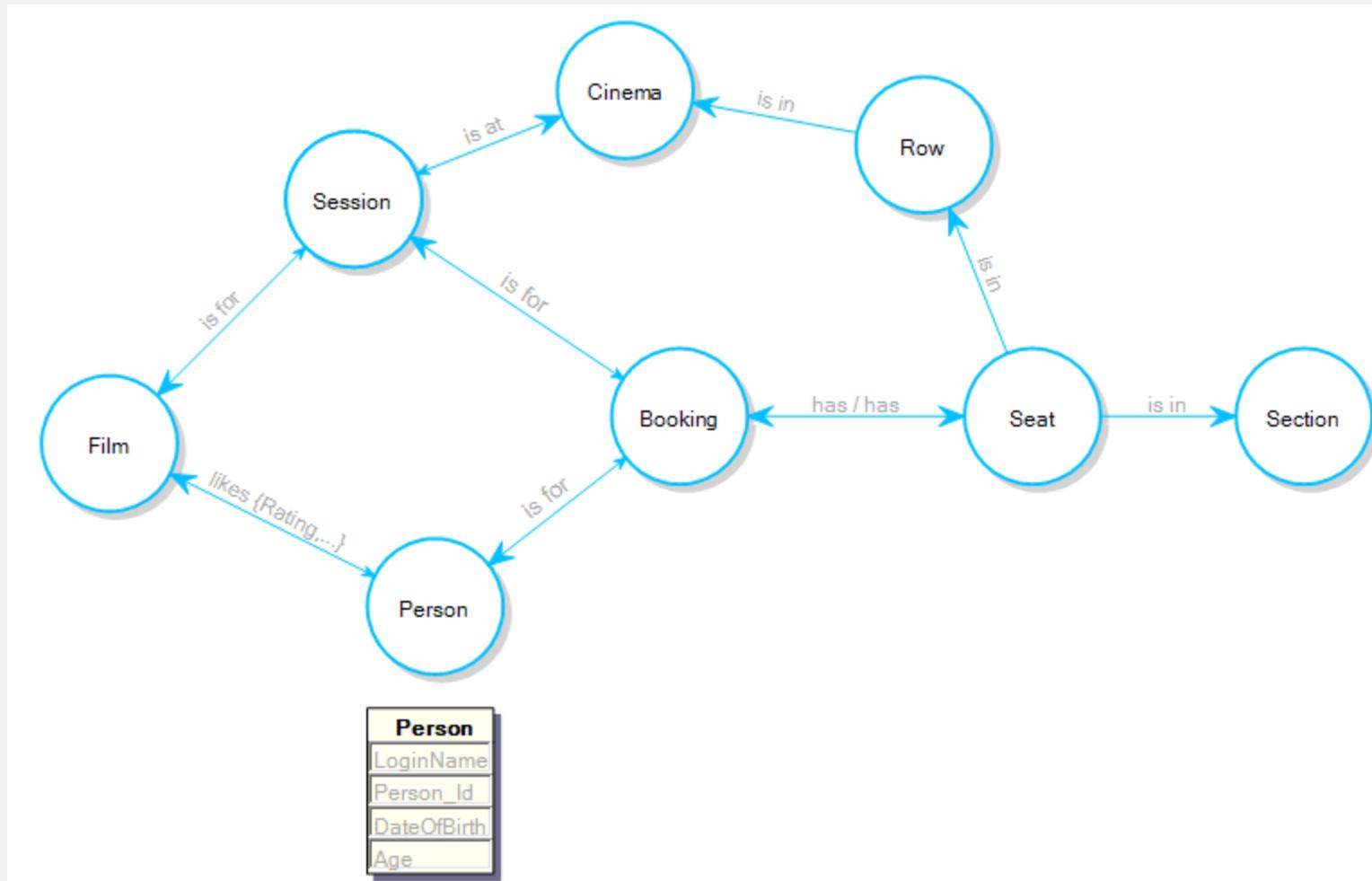
DATA VAULT





GRAPH







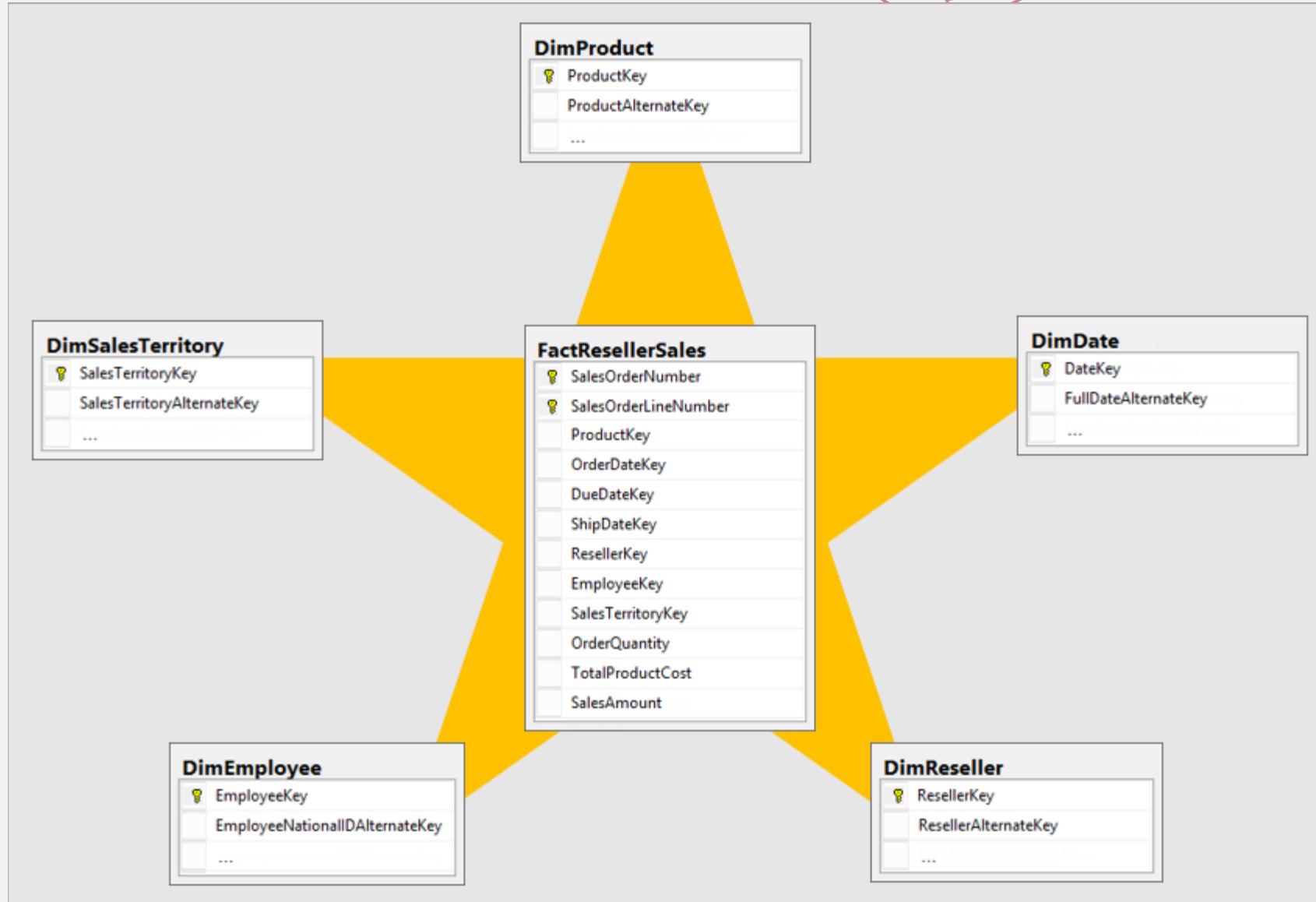
DIMENSIONAL



Star Schema

Snowflake







ENTITY CENTRIC



Channel_Name	Category	Customer_Count	Total_Revenue	CLV
Organic Search	Accessories	1	80	80
Organic Search	Clothing	1	190	190
Paid Search	Clothing	1	70	70
Social Media	Accessories	1	40	40
Social Media	Clothing	1	150	150
Social Media	Footwear	1	160	160

ONE BIG TABLE





Customers

A 3x4 grid of 12 empty squares, used for drawing or writing practice.



Products

A 3x3 grid of nine empty boxes, intended for drawing or writing practice.



Orders

A 3x3 grid of nine empty boxes, intended for drawing or writing practice.



One Big Table

A blank 10x10 grid for drawing or plotting.

WHEN TO CONSIDER WHICH MODELLING APPROACH



Scenario	Dimensional	Data Vault	Graph	One Big Table	Entity Centric
Change Tracking	✓	✓	✓		
Detailed Change Tracking		✓			
Simplified Queries	✓			✓	✓
Complex Relationships		✓	✓		
Complex Hierarchies	✓	✓	✓		
Performance & Scalability	✓	✓		✓	✓
Ad-hoc Analysis & Reporting	✓				
Data Analytics	✓			✓	✓
Consistency	✓	✓			
Ease of Maintenance	✓				✓
Separation of Business Rules		✓			
Geospatial Analysis	✓		✓		

THE VICTOR



MIXED MODEL ARTS

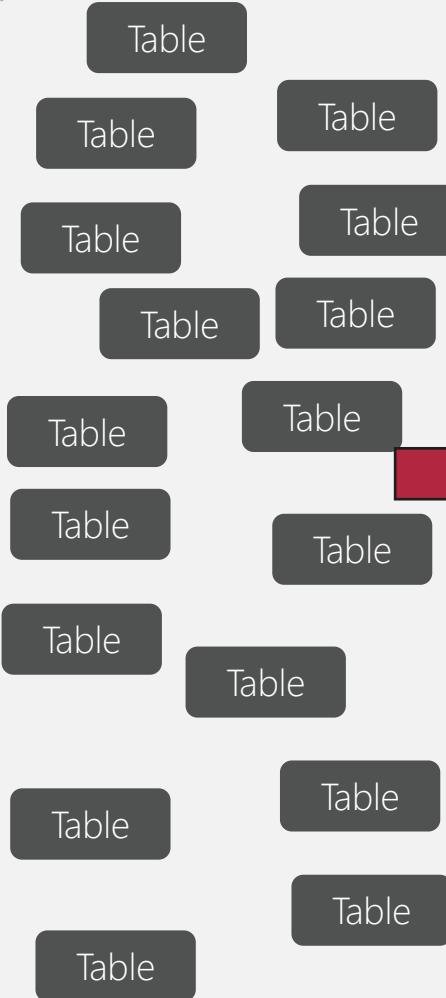
For when one approach doesn't satisfy all use cases



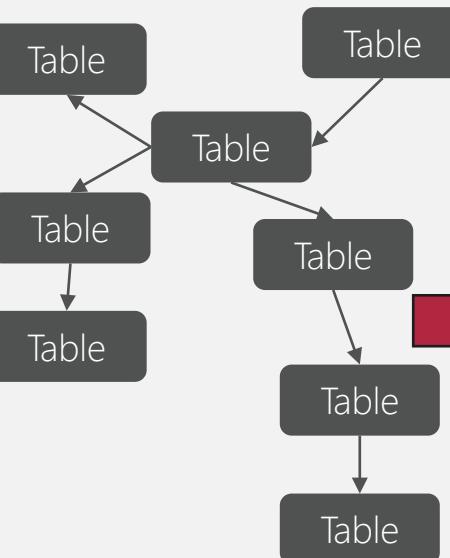
PRACTICAL MIX & MATCH



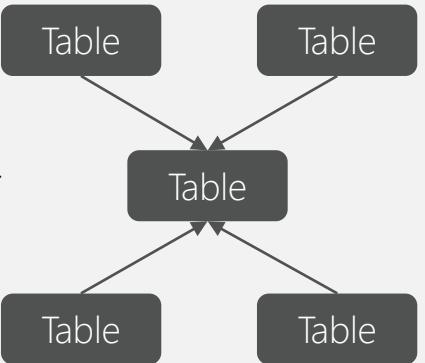
Incoming from Source



Data Vault Modelled



Dimensional Modelled

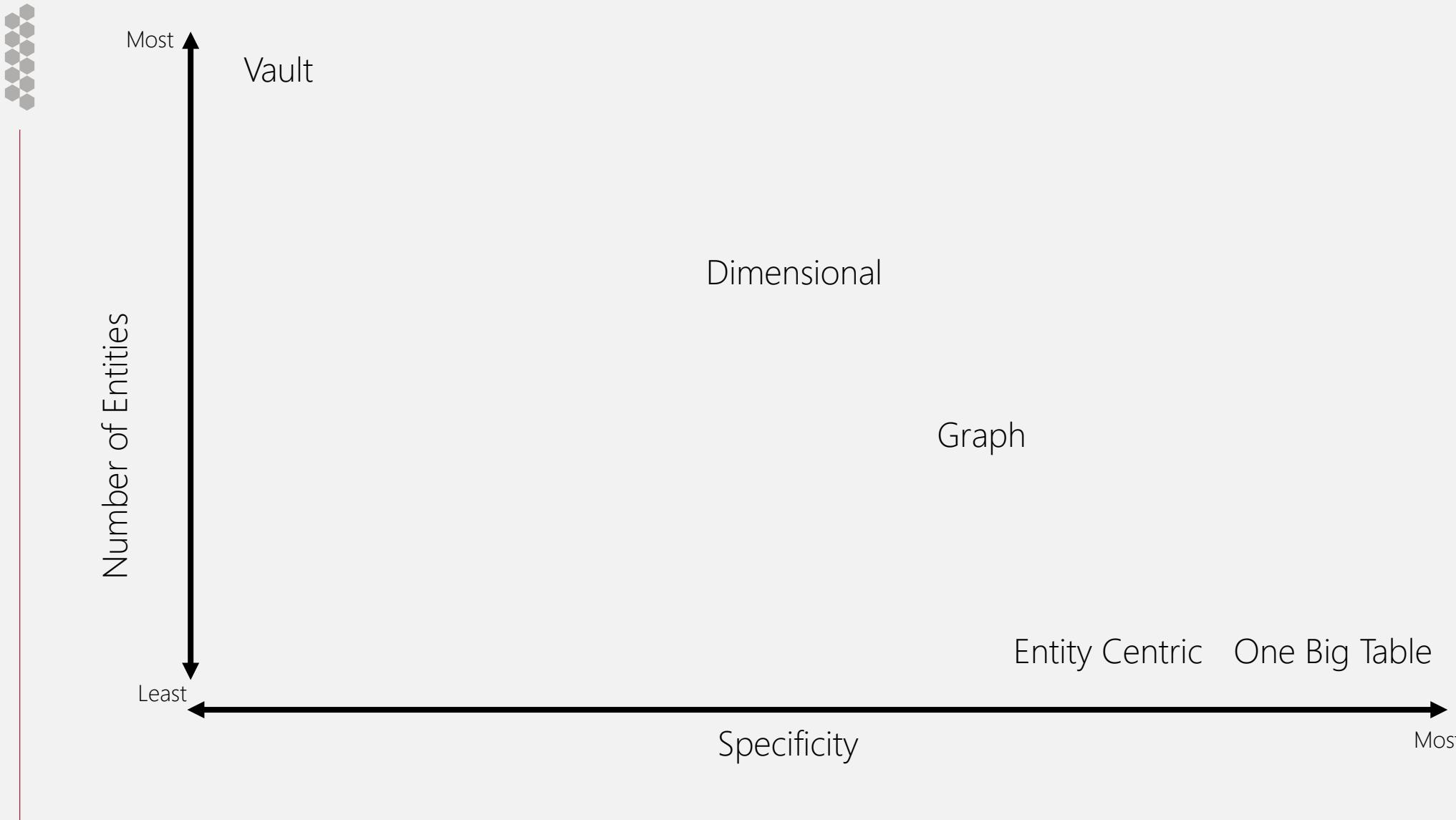


Niche Modelled



Entity

GENERIC TO SPECIFIC



DESIGNING A MODEL

In an Agile manner



WHAT IS BEAM*?



Agile Data Warehouse Design

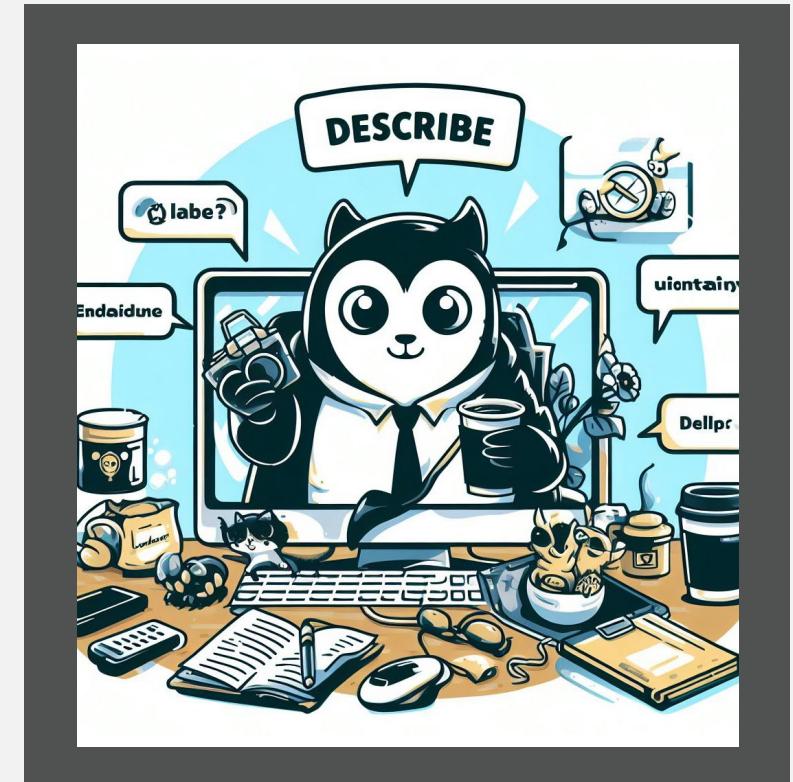
*Collaborative Dimensional Modeling,
from Whiteboard to Star Schema*

Lawrence Corr
with Jim Stagnitto



THE 3 D's

Discover, Describe, Document



THE 7 W's

WHO

WHAT

WHERE

WHEN

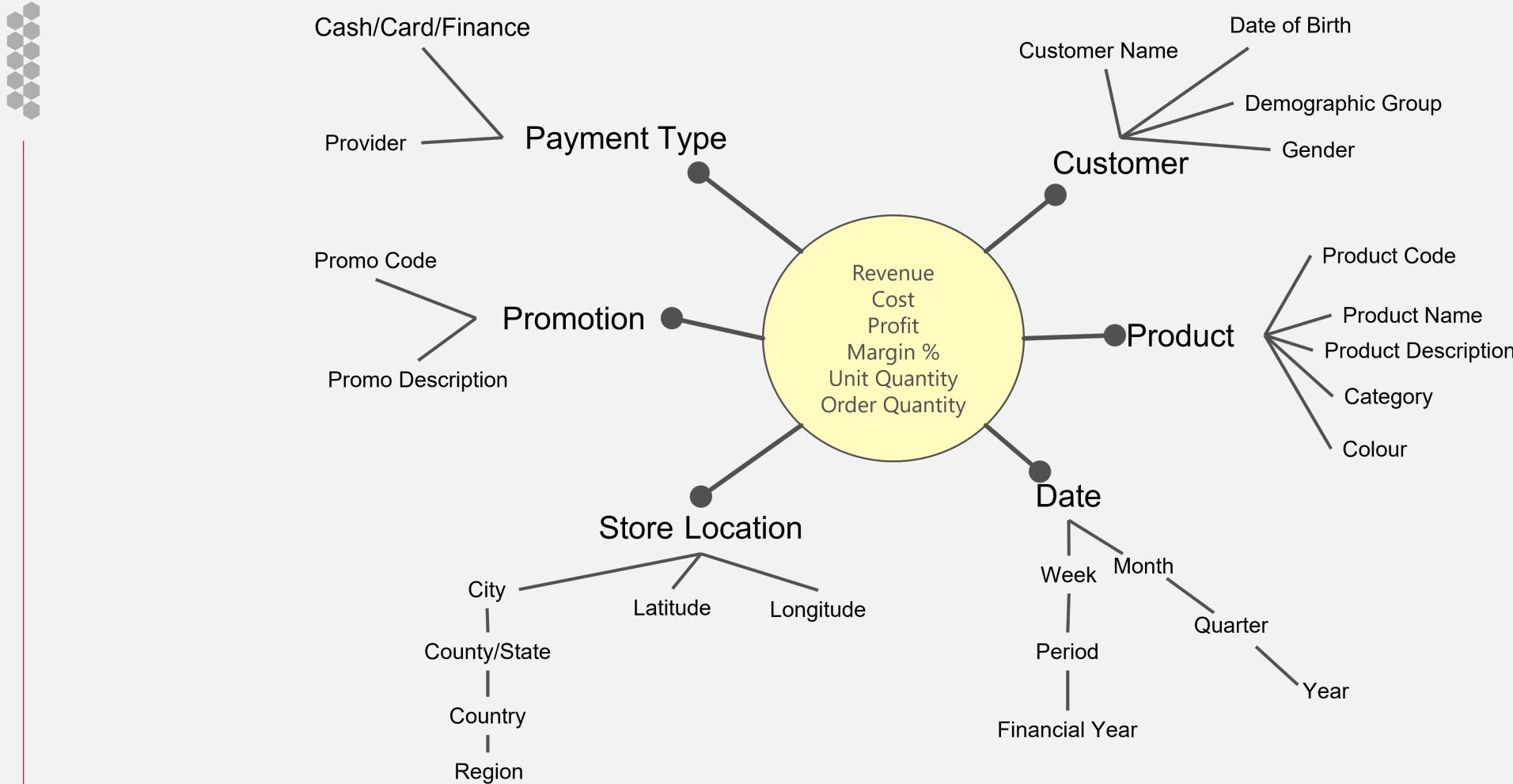
WHY

HOW

HOW MANY



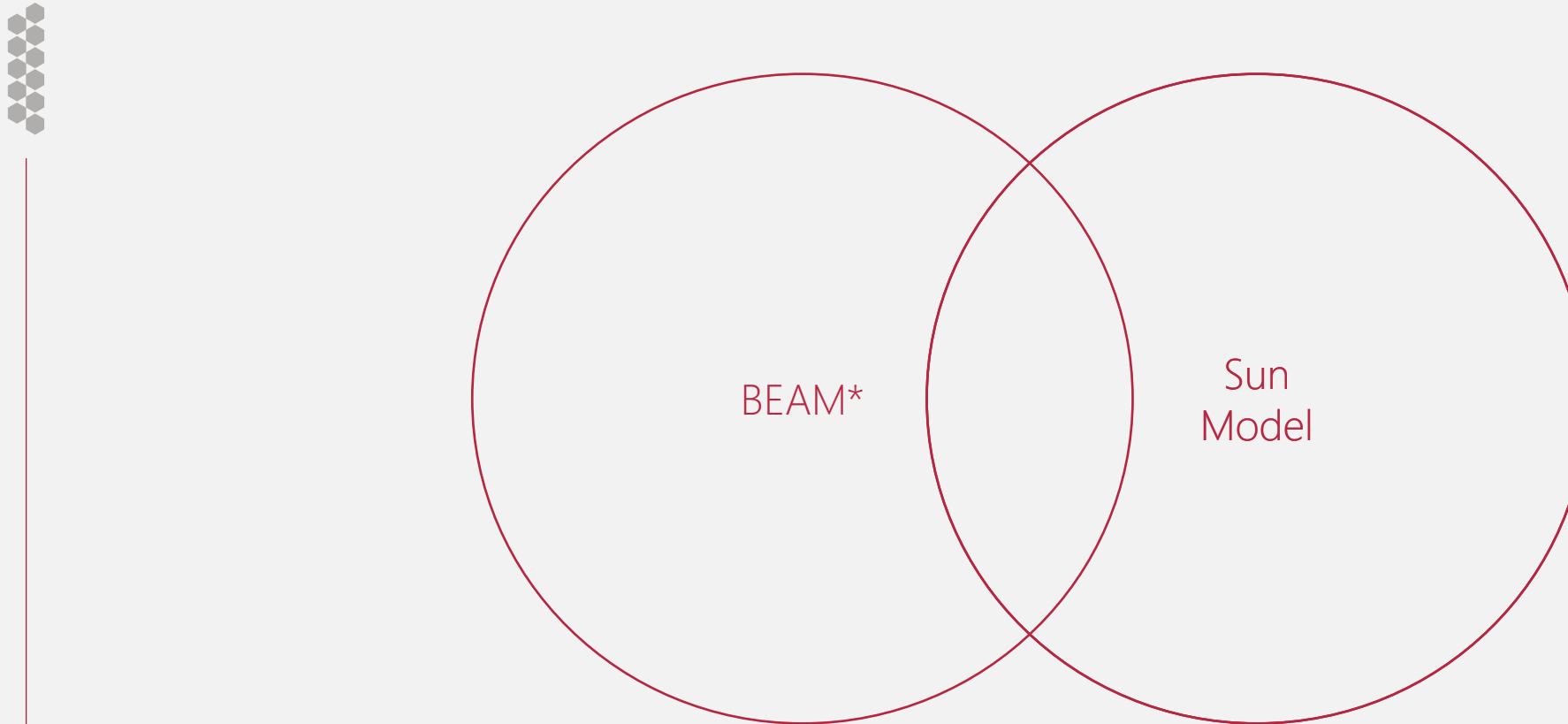
WHAT IS SUN MODELLING?



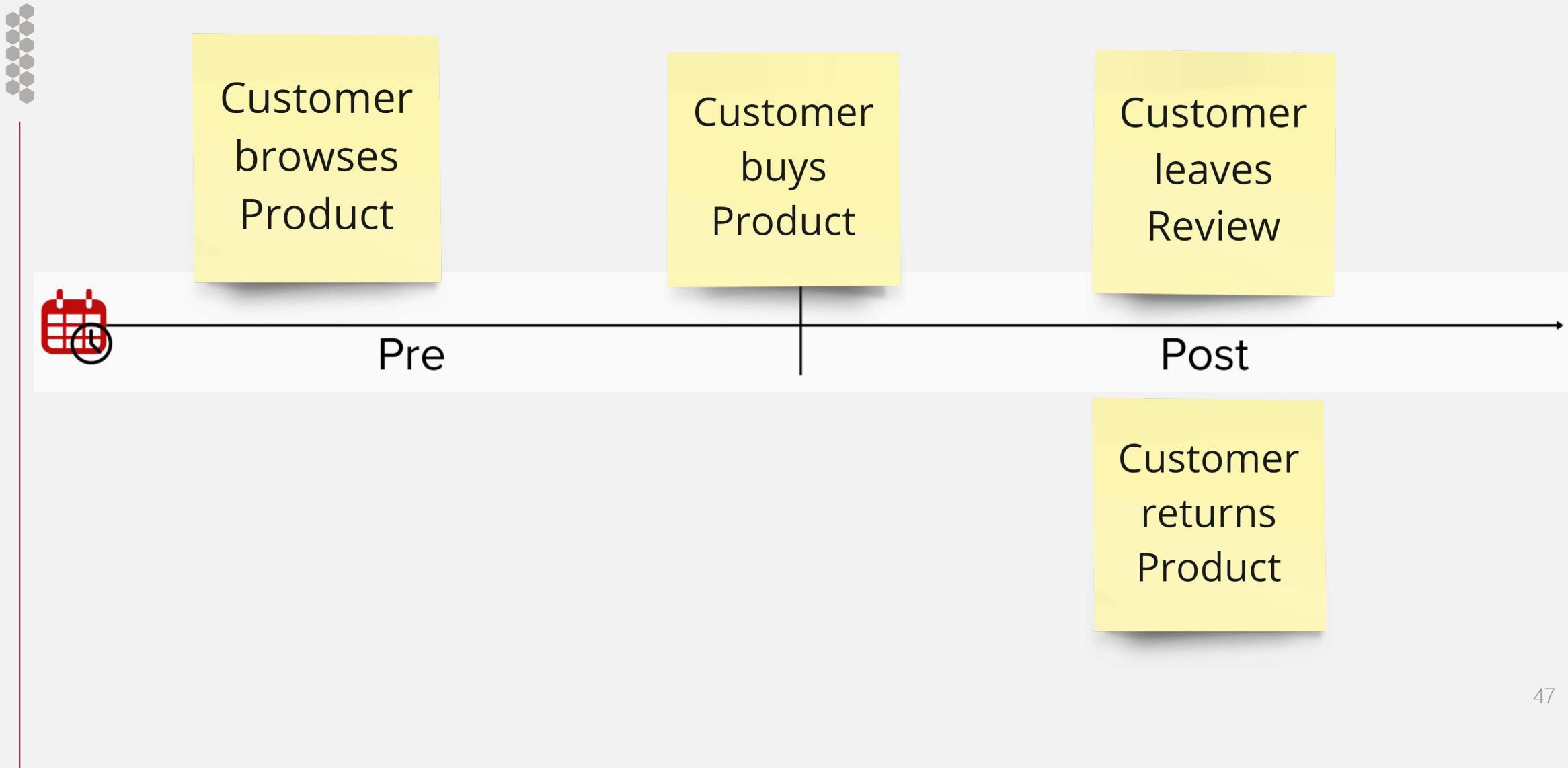
SUNBEAM



WHAT IS SUNBEAM?



EVENT IDENTIFICATION



CUSTOMER BUYS PRODUCT

HOW MANY

WHO

WHAT

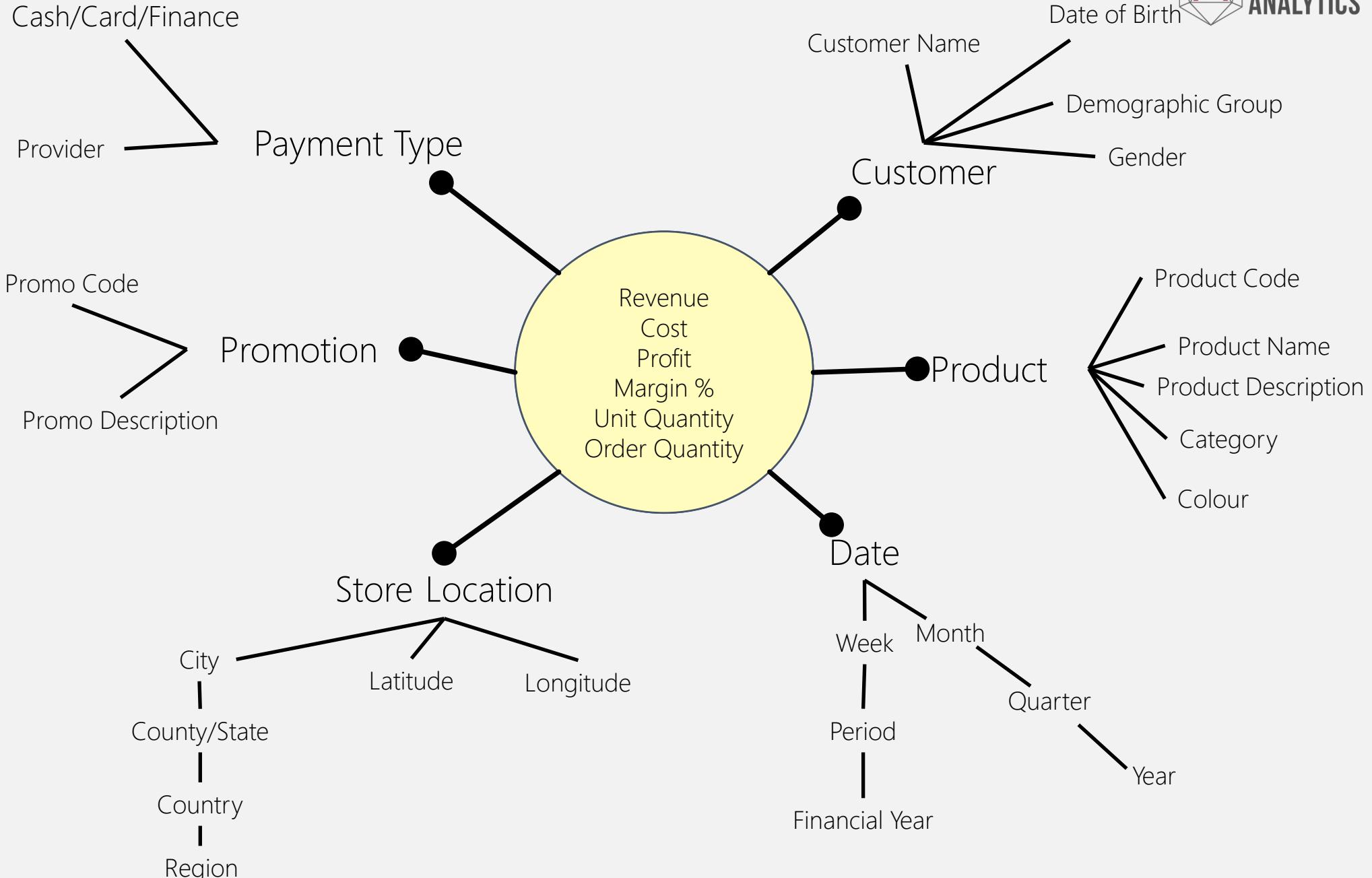
WHERE

WHEN

WHY

HOW

HOW OFTEN



TIPS & TRICKS



THE 3 D's

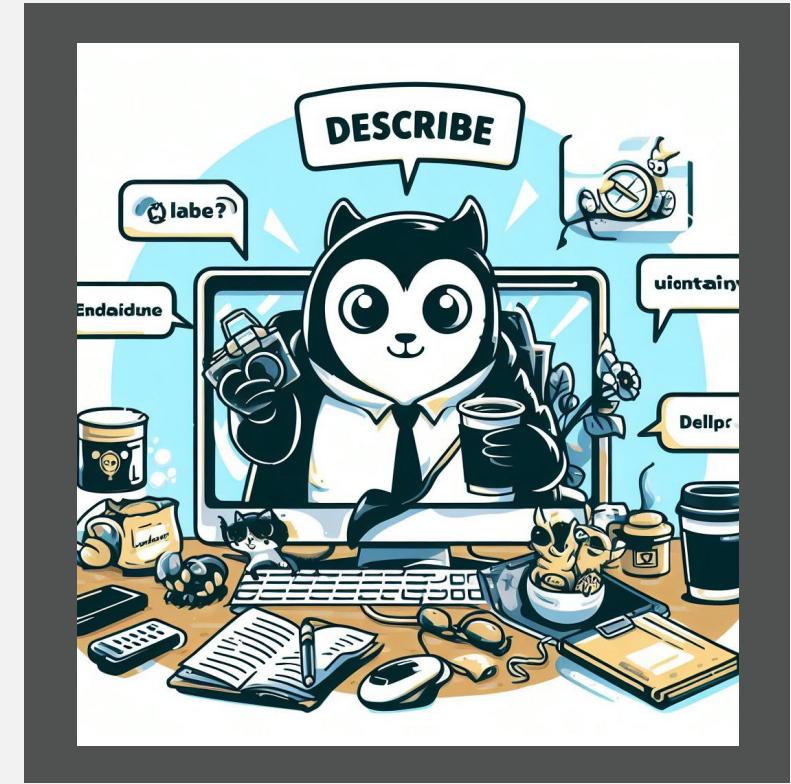


Discover, Describe, Document

Who Does What?

Use the 8 Ws

Sun Modelling



REDUCE, REUSE, RECYCLE

50%

Of a data model will follow generic patterns



- 25% will follow industry patterns
- 25% will be unique to an org

ANALYTICAL DATA PRODUCTS

And data models



PRODUCT ATTRIBUTES



DEFINING ANALYTICAL DATA PRODUCTS



A domain-bound set of data services, including the underlying infrastructure & code to produce/maintain

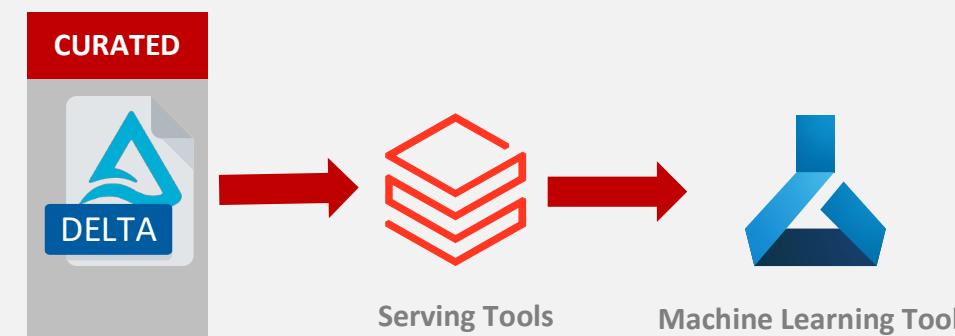
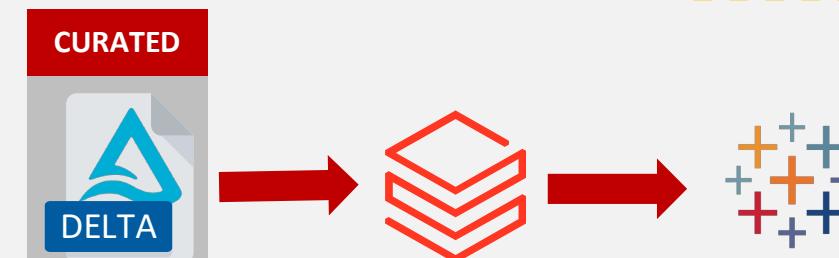
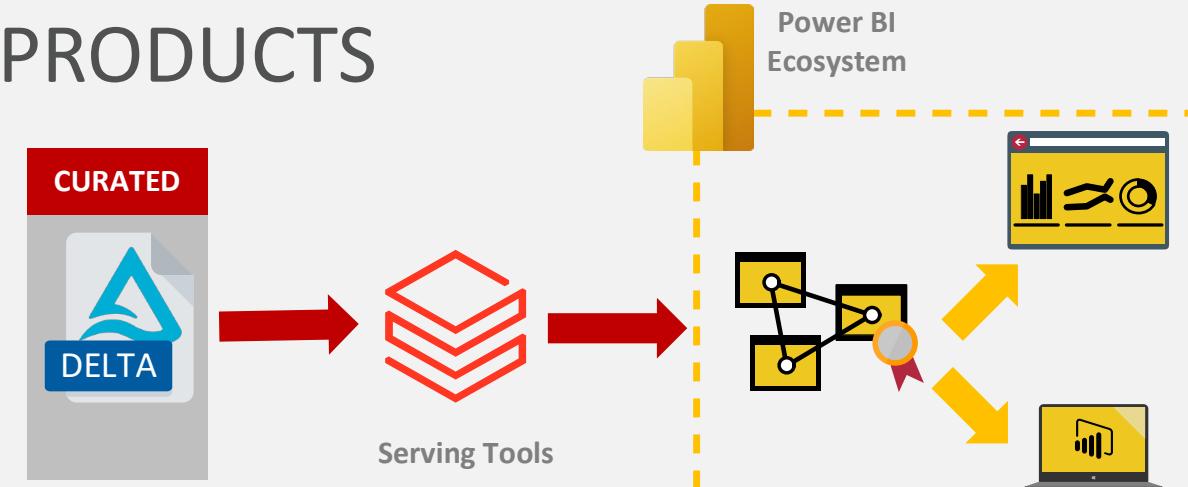
It Is:

- A Data Model
- A Reporting Table
- An ML Model

It is NOT:

- A Dashboard
- An Application
- An Entire Data Warehouse / Lake

DELIVERING DATA PRODUCTS



COMPANY STRATEGY



Increase Revenue



Reduce Cost



Maximise Profit

THE PURPLE CHOCOLATE COMPANY



Raising the price of the product to customers, so that farmers are better paid



Make the bars smaller, so that I can maintain manufacturing costs



Invest in a sustainable manufacturing plant



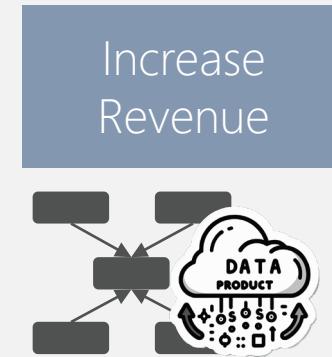
PURPLE CHOCOLATE & DATA PRODUCTS



Sales



Marketing



Direct to Customer

Wholesale

Retail Partners



Suppliers



Stock

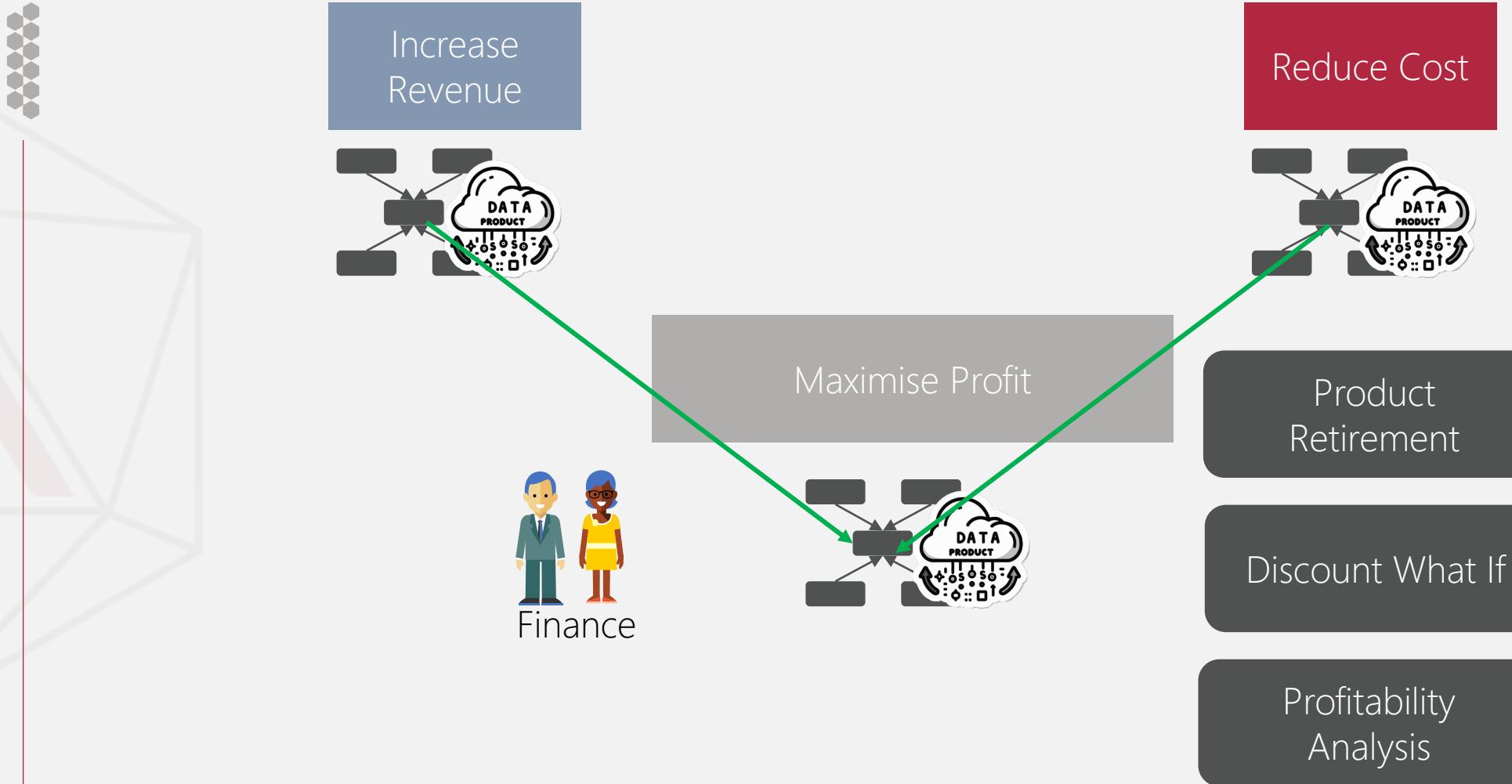


Farms

Manufacturing

Warehouse

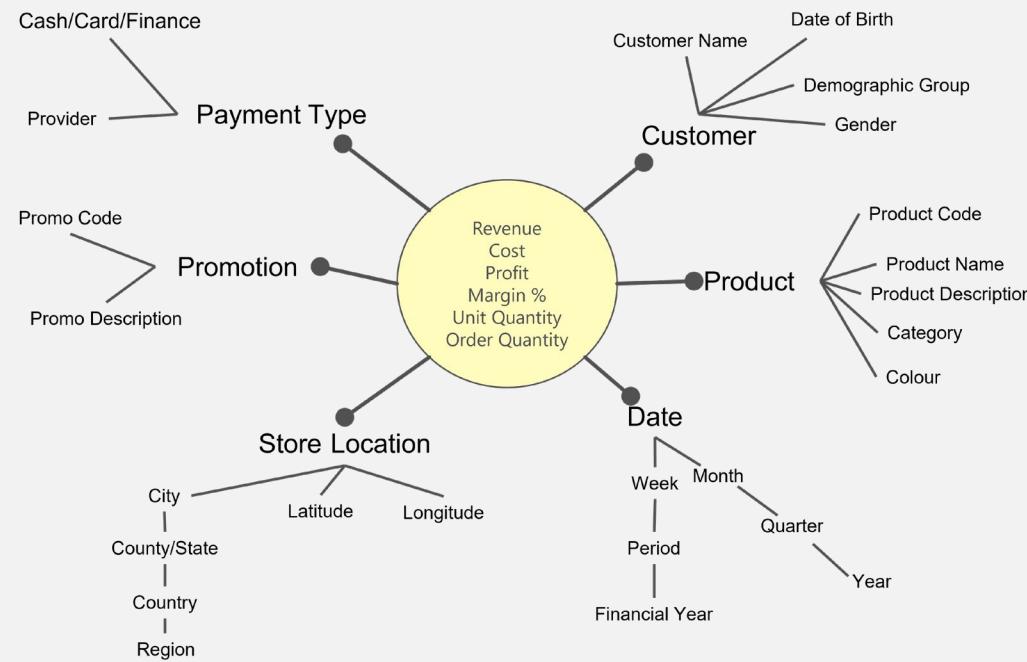
PURPLE CHOCOLATE & COMPOUND VALUE



SunBeam and Analytical Data Products



SunBeam



Analytical Data Products



ASKING QUESTIONS

Of the Conceptual Data Model



What products makes us the most profit?



Do our stores have enough inventory for the next month of forecasted sales?



What are our net revenue sales for last month, and how does that compare to the same period in the previous year?



DRIVING ACTION

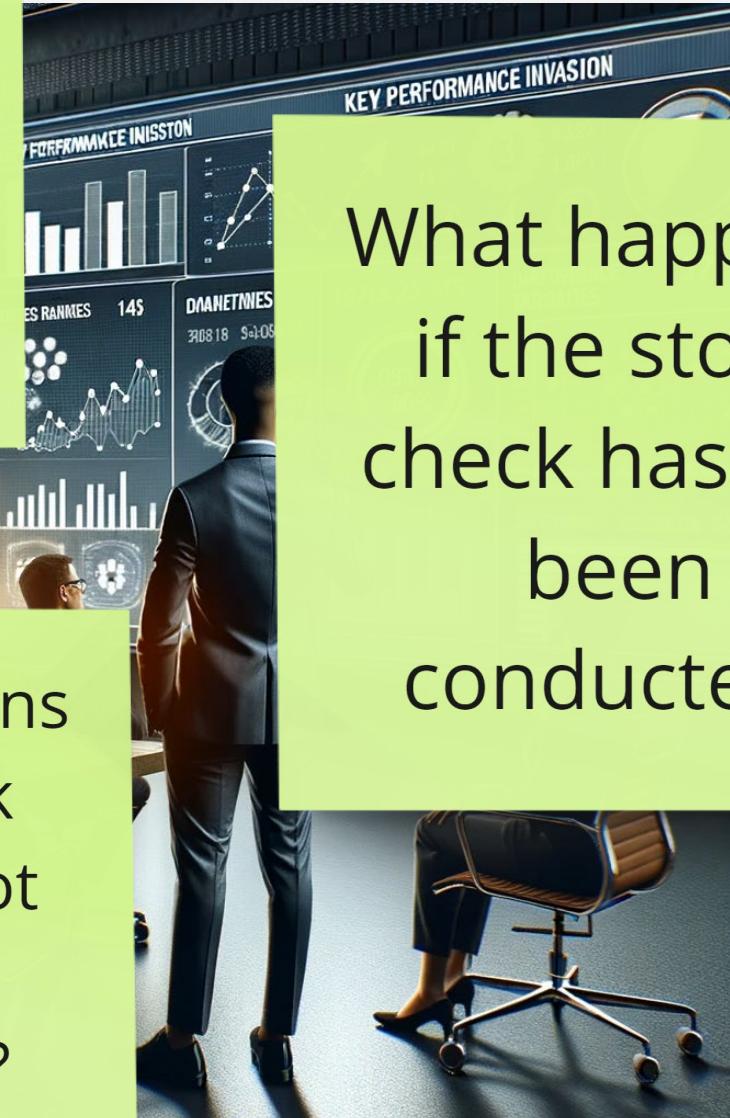
From User Questions

Do our stores have enough inventory for the next month of forecasted sales?

What happens if there's not enough inventory?

What happens if the stock data has not been refreshed?

What happens if the stock check has not been conducted?



CONCEPTUALLY REAL

Turning Concepts into Reality



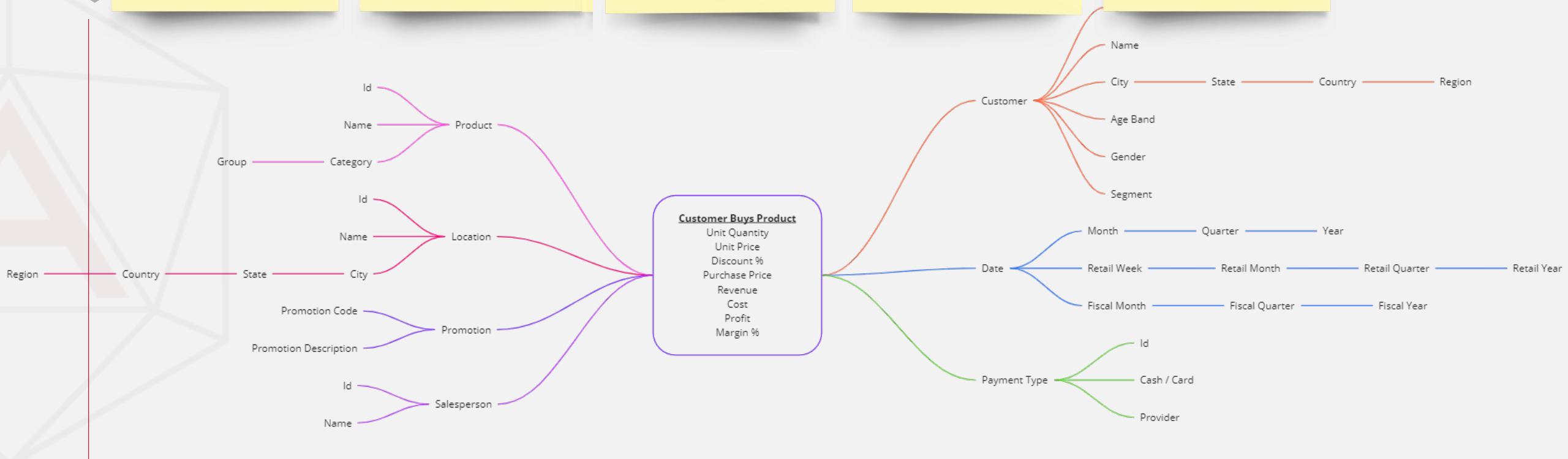
What products makes us the most profit?

What are our net revenue sales for last month, and how does that compare to the same period in the previous year?

What's the sales forecast for the next month?

Which products perform best? Are there any which I should consider retiring?

Which locations perform best? Which states should I expand in or contract presence?



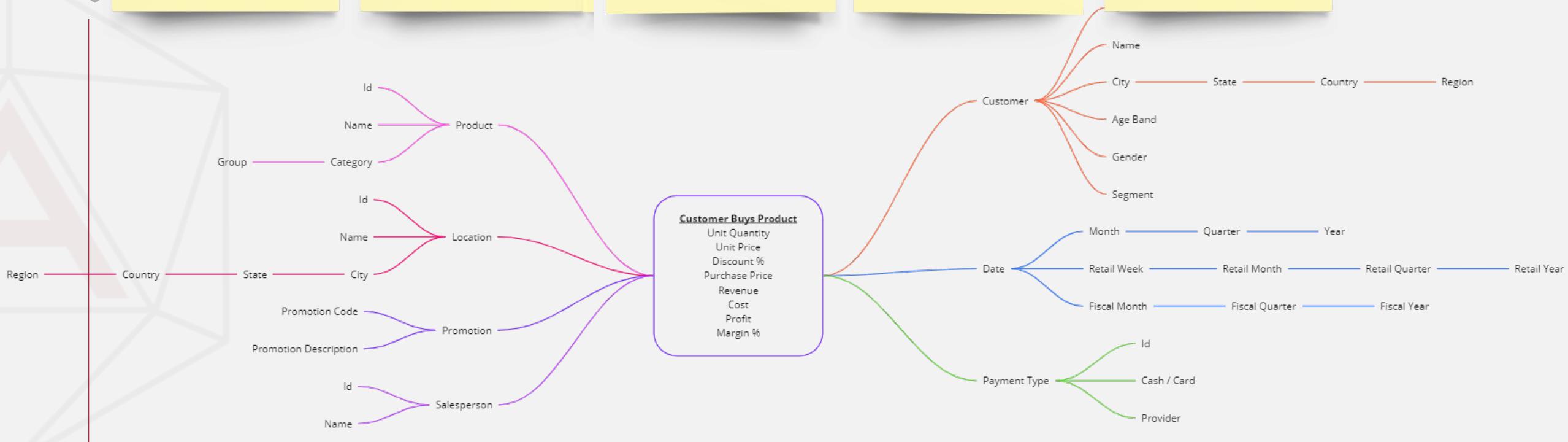
What products makes us the most profit?

What are our net revenue sales for last month, and how does that compare to the same period in the previous year?

What's the sales forecast for the next month?

Which products perform best? Are there any which I should consider retiring?

Which locations perform best? Which states should I expand in or contract presence?



What's the sales forecast for the next month?

Which locations perform best? Which states should I expand in or contract presence?

What are our net revenue sales for last month, and how does that compare to the same period in the previous year?

Which products perform best? Are there any which I should consider retiring?

What products makes us the most profit?



DATA PRODUCT BACKLOG

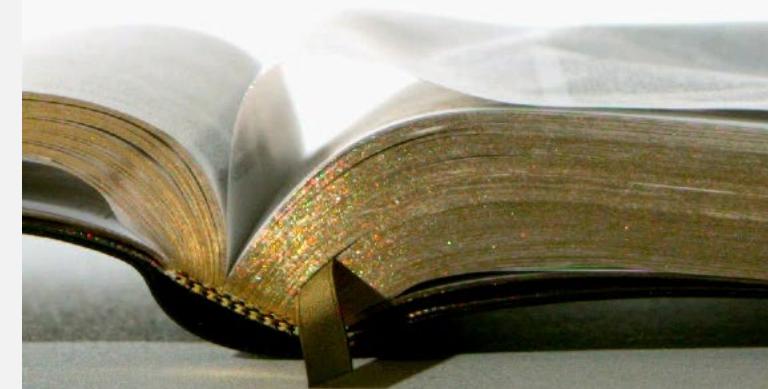
User Story	Details	Iteration
As a Financial Planner, I want to forecast sales so that I can predict demand	Thin SalesOrder fact to focus on Revenue and Forecast Revenue. Date Dimension	1
As a Store Planner, I want to know which Locations are performing well so that I can make expansion decisions	Location Dimension Expand grain of fact to include Location	2
As a Financial Planner, I want to know month on month comparison, so that I can decide if performance is improving	New measures: - Revenue Last Month - Revenue Month on Month	3
As a Product Analyst, I want to know the performance of products, so that I can choose which ones to retire	Product Dimension Expand grain of fact to include Product	4
As a Financial Planner, I want to know which products are the most profitable, so that I can make strategic investments in their production	Add remaining metrics to fact table: - Cost - Profit - Etc	5

FURTHER READING



To get your brain juices flowing

- [Agile Data Warehouse Design - Corr & Stagnitto \(2011\)](#)
- [The Data Warehouse Toolkit – Kimball & Ross \(2013\)](#)
- [The Nimble Elephant – Giles \(2012\)](#)
- [Building a Scalable Data Warehouse with Data Vault – Linstedt & Olschimke \(2015\)](#)
- [Introduction to SunBeam – Bloye \(2023\)](#)
- [Data Vaulting: From a Bad Idea to Inefficient Implementations – TIMi \(2022\)](#)
- [Data Products For Dummies – Mohan, Adams & Mullen \(2023\)](#)
- [Entity Centric Modeling for Optimising KPIs – Scherding \(2023\)](#)





THANK YOU

