

GUIDE YOUR BUSINESS TOWARDS A LOGICAL DATA MODEL

April 2

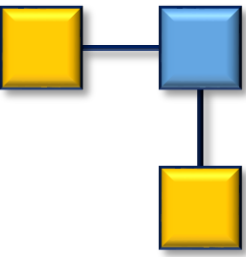
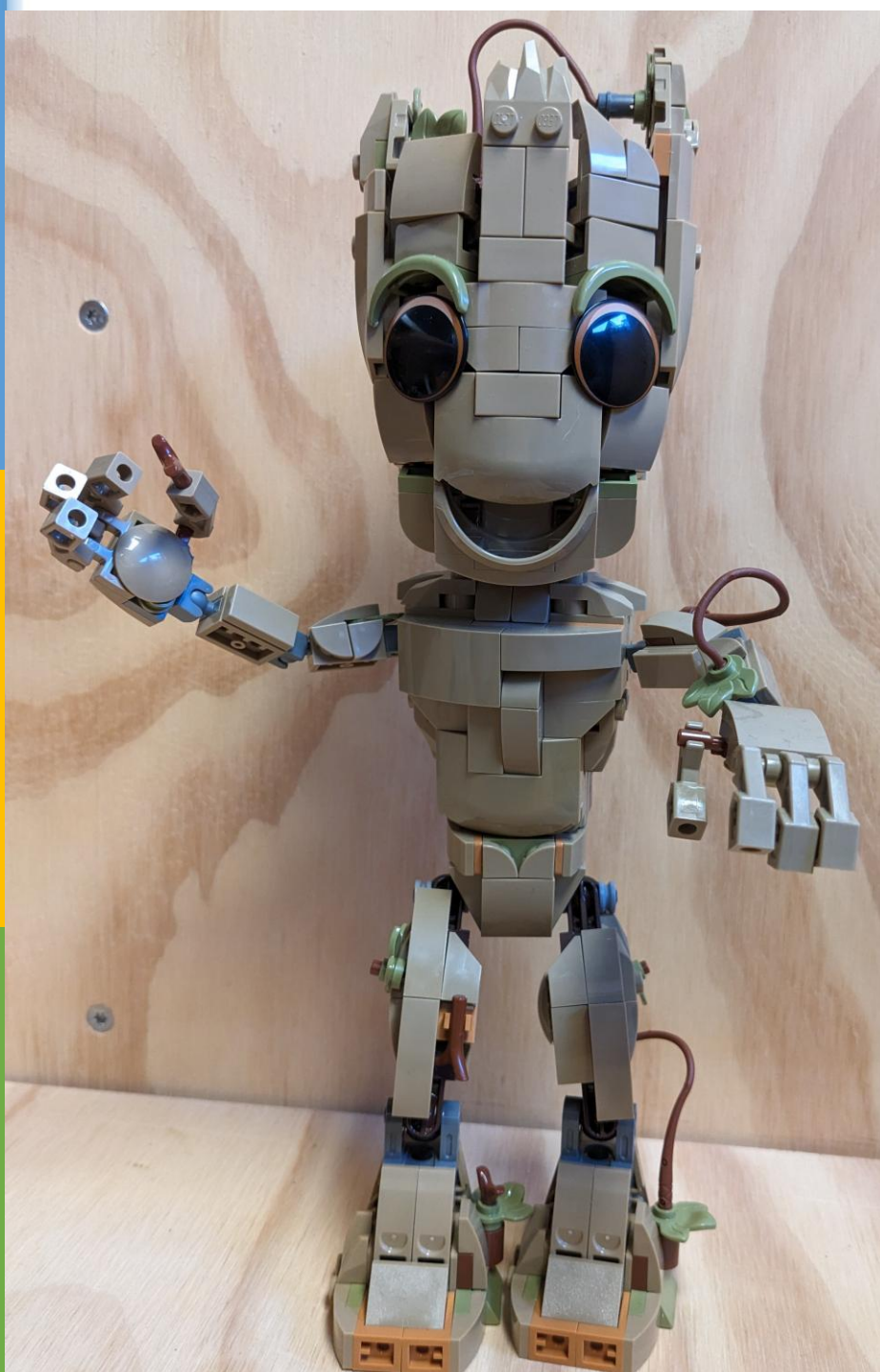
© 2025 Genesee Academy, LLC
25568 Genesee Trail Rd.
Golden, CO 80401 USA +1 303 526 0340

GeneseeAcademy.com

ELMstandards.com

DVstandards.com

GENESEE ACADEMY
World-Class Training

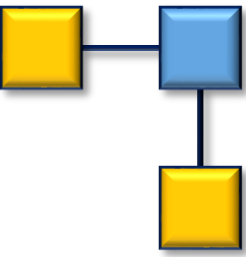


I AM....

Remco Broekmans
Genesee Academy
Trainer
Coach
Data modeler
Dutch
Rock climbing
Walking

.....

Finding the balance

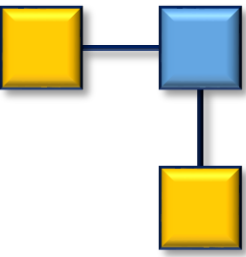


- What is our focus? What is the business talking about?

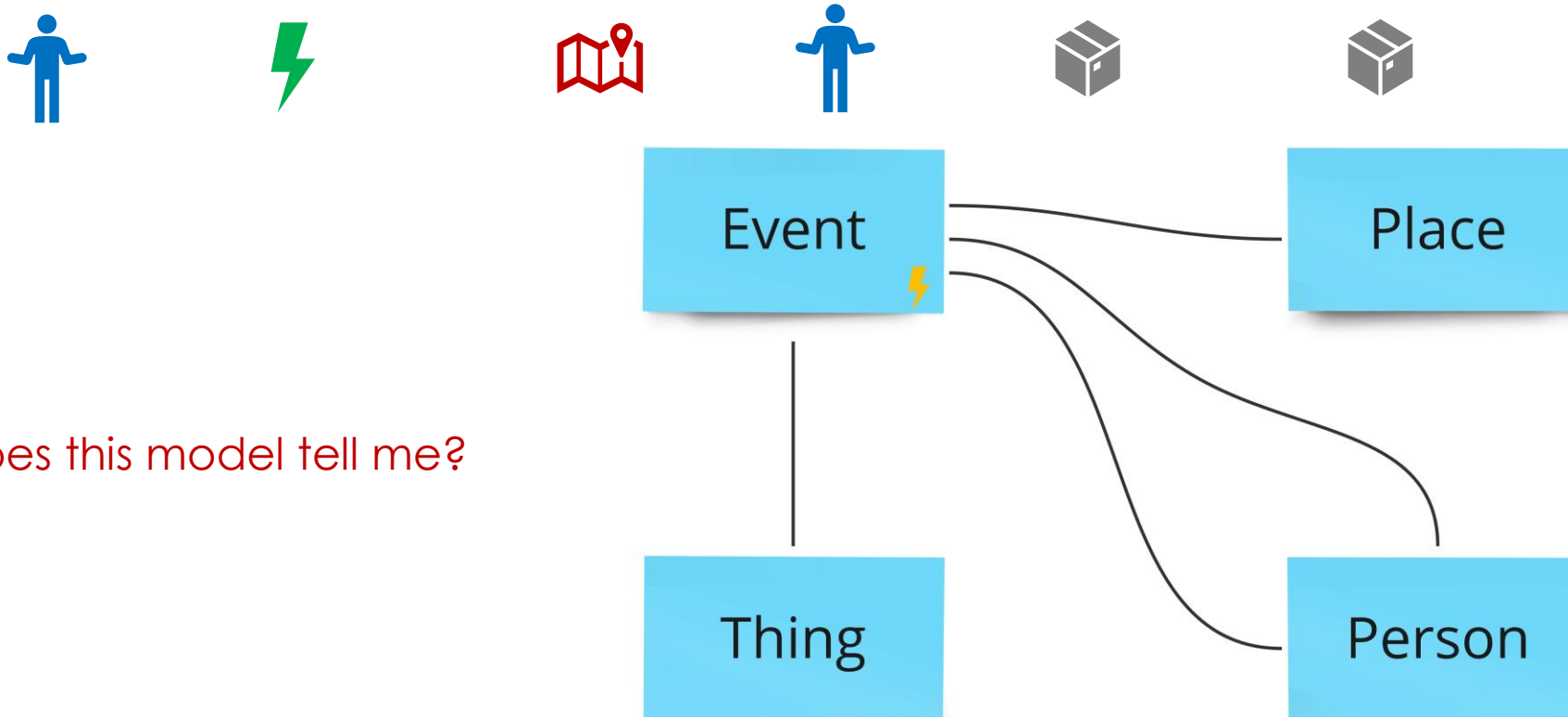
- This is:
 - An organism
- OR
 - A Plant
- OR
 - A tree
- OR
 - A stem
 - A branch
 - A twig
 - A leaf



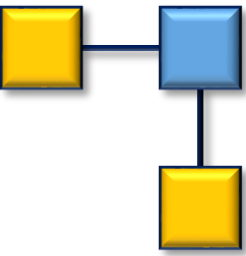
An example



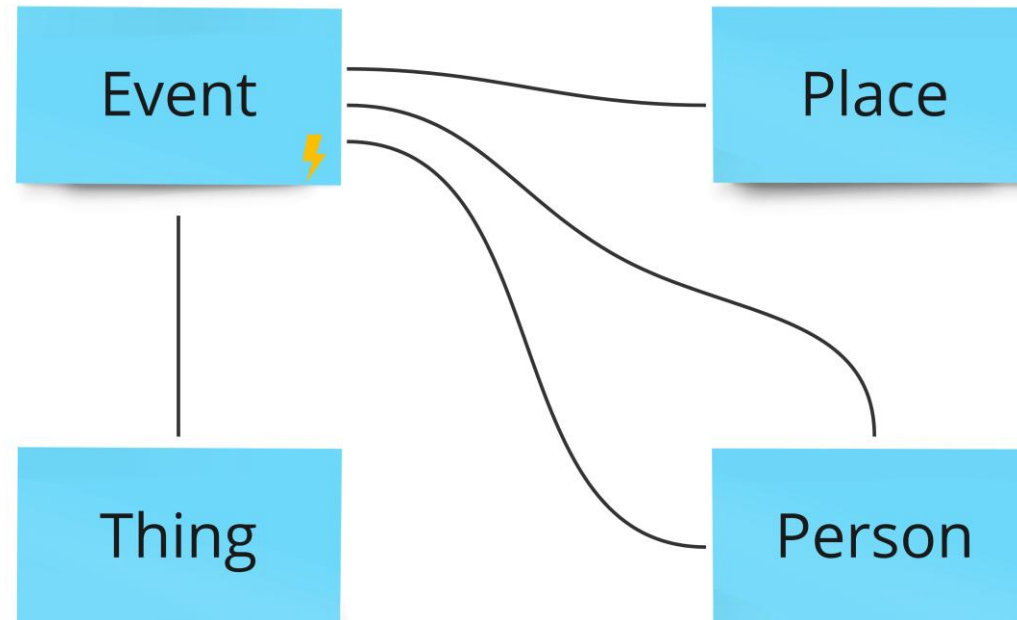
- Remco has an appointment in the hospital with a doctor and got a diagnosis (groin rupture).
- Remco – Appointment – Hospital – Doctor - Diagnosis – groin rupture



An example



- Remco is hospitalized into a bed in a hospital room by a nurse.
- Remco – Hospitalized – Bed – Hospital room - Nurse



What does this model tell me?
And how is this different?



An example

- Remco has an appointment in the hospital with a doctor and got a diagnosis (groin rupture).

Core Business Concept (CBC)

[Remco](#)

[Appointment](#)

[Hospital](#)

[Doctor](#)

[Diagnosis](#)

[Groin Rupture](#)

[Hospitalized](#)

[Bed](#)

[Hospital Room](#)

[Nurse](#)

Core Business Concept (CBC)

[Remco](#)

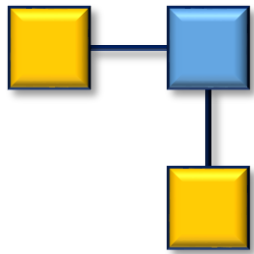
[Appointment](#)

[Hospital](#)

[Doctor](#)

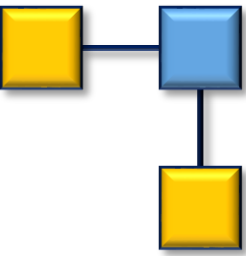
[Diagnosis](#)

[Groin Rupture](#)







- Remco is hospitalized into a bed in a hospital room by a nurse.

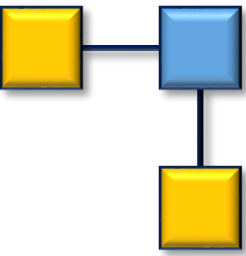
An example







Categorizations NOT Abstraction →

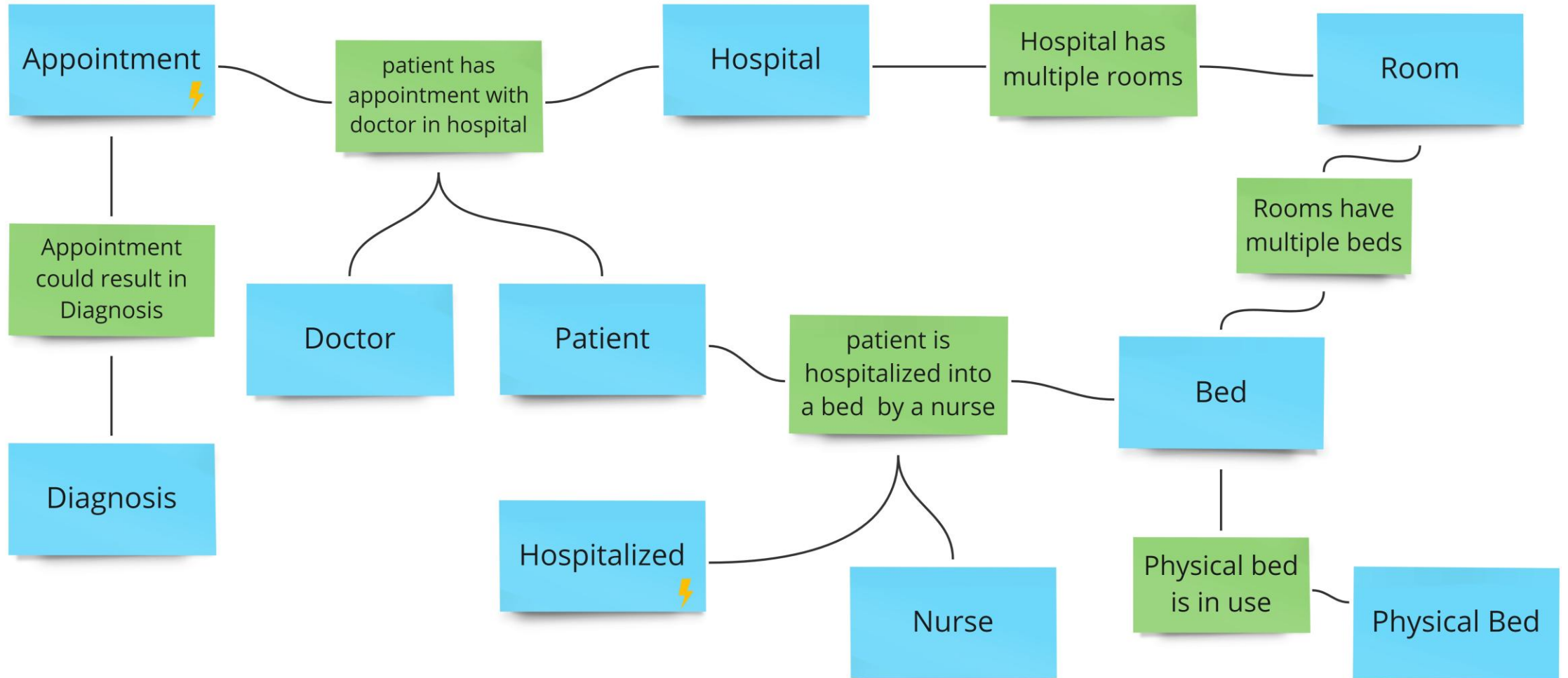
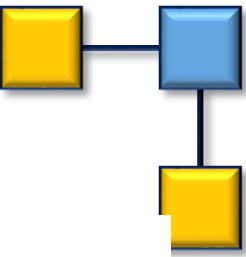
Core Business Concept (CBC)	Event 	Person 	Place 	Thing 	Other Concept
Remco					Instance of Patient
Appointment	x				
Hospital			x		
Doctor		x			Instance of Diagnosis
Diagnosis				x	
Groin Rupture					
Hospitalized	x				Discussion
Bed					
Hospital Room			x		
Nurse		x			

After discussion with Business



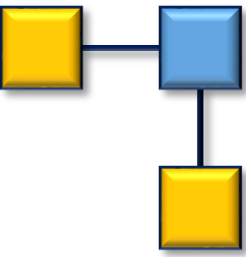
Core Business Concept (CBC)	Event 	Person 	Place 	Thing 	Other Concept
<u>Remco</u>					Instance of Patient
<u>Appointment</u>	x				
<u>Hospital</u>			x		
<u>Doctor</u>		x			
<u>Diagnosis</u>				x	
<u>Groin Rupture</u>					Instance of Diagnosis
<u>Hospitalized</u>	x				
<u>Bed</u>			x		Best Fit
<u>Hospital Room</u>			x		
<u>Nurse</u>		x			
<u>Patient</u>		x			
<u>Physical Bed</u>				x	for maintenance

The model



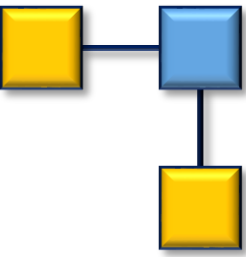
Easy to communicate, clear what is what!

Our road

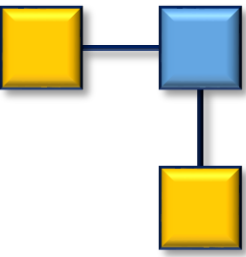


- Use structured & guided method to get to the Ensemble Logical Model.
- Business driven over source driven.
- Walk the path together – business & IT.
- Keep it simple, clear, understandable and useable for all.
- Document while we move forward.
- Be agile and flexible.
- Basically, capture what is important for the business which is not perse what is captured in the sources / applications!
- Find the balance between abstraction/generalization and specification!

What is the Ensemble Logical Model



Data Modeling Forms & Purpose



3rd Normal
Form

Ensemble
Modeling

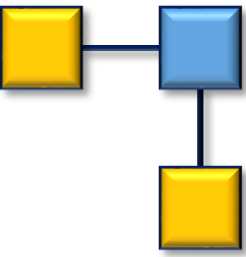
Dimensional

- While **3rd Normal Form** is optimal for **Operational Systems**.
...and **Dimensional** is optimal for **Data Marts**.
...the **Ensemble Modeling** is optimal for the **Data Warehouse (EDW)**.
- **Data Vault** is the leading **Ensemble Modeling** pattern.

REMARK: This is a very short and quick intro without disregarding 3NF & Dimensional as modelling patterns



Deconstructed Burrito...



VS

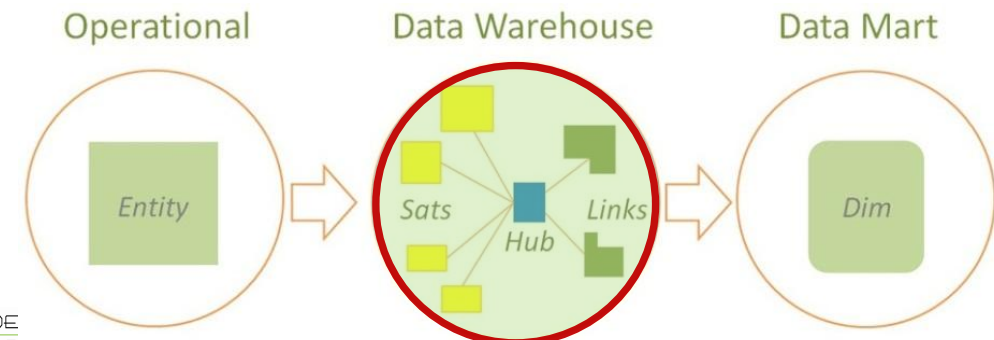
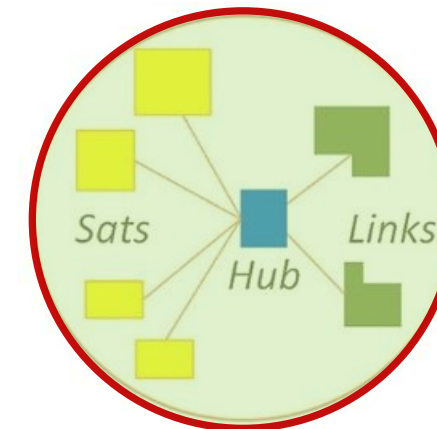
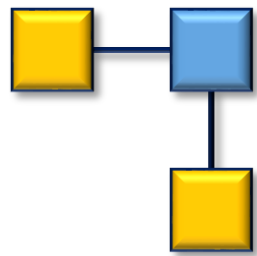
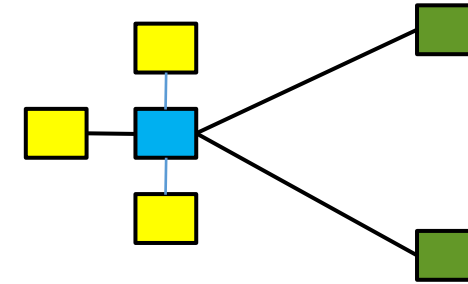


Decomposing while at the same time we keep everything together (unified).
Don't look at the individual parts alone but always see them as a whole.

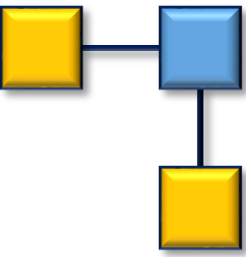
Of course if business is telling a Burrito IS something on its own

Ensemble

- An **Ensemble** is a Core Business Concept
 - Event, Person, Place, Thing...
 - Business-Driven
 - Business-Facing
 - Identifiable by the business
- An **Ensemble** in DV contains only
 - One **Hub**
 - Directly attached **Satellite(s)**
 - All related **Link(s)**

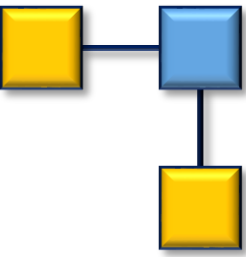


Finding the essence



- Ask questions?
 - What is going on in your organization / department?
 - Explain the process / day-to-day work
- LISTEN
 - Be curious
 - Capture all!
- Find the CBC's – The balance between abstraction or specific!
 - Event/Transactions
 - Person / People
 - Places
 - Things / Objects
 - Other important Concepts/Terms

What is a Core Business Concept?



- A Core Business Concept (CBC);
 - everything that is important for the organization.
 - clear on its own.
 - identifiable by the organization.
 - Can be counted.
- A Core Business Concept is NOT
 - a metric or other descriptive information.
 - Abstracted or generalized (no role playing or typing -see bullit 2&3)
- The right Core Business Concepts are key towards a **Business Focused** and **Understandable** fully integrated data warehouse!

What we Model

Core Business Concepts (CBC)

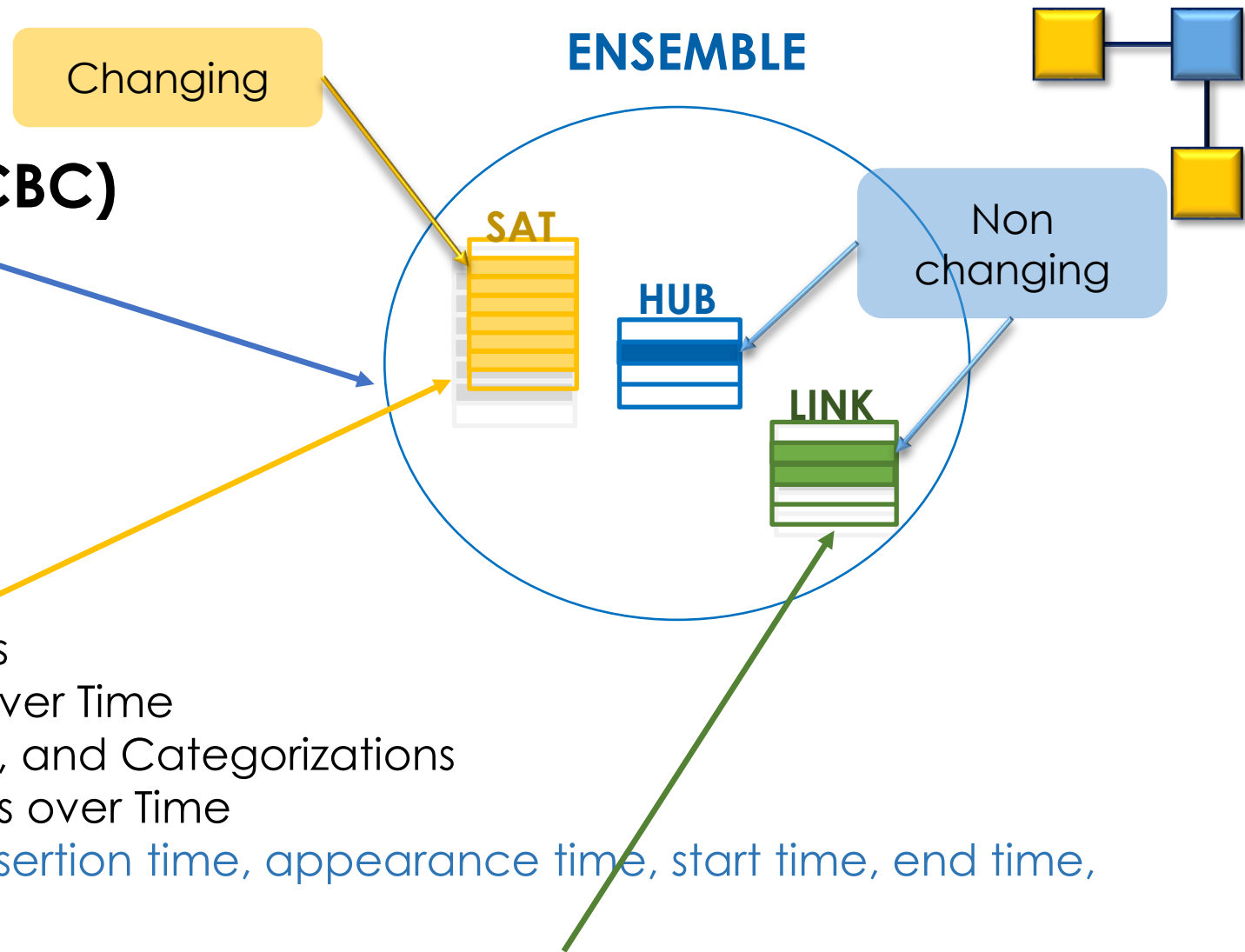
- Event
- Person
- Thing
- Place
- Other Concept
- Organization (optional)

Context & History

- Descriptive Attributes
- Changing Context over Time
- Types, Classifications, and Categorizations
- Status/State changes over Time
- All relevant times (assertion time, appearance time, start time, end time, etc.)

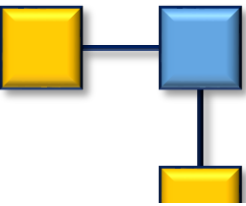
Natural Business Relationship (NBR)

- Unique, Specific NBR
- Based on Naturally Correlated Concepts





First 3 artifacts are all on the Core Business Concepts



Genesee Academy - Proprietary

Core Business Concept (CBC)	Event	Person	Place	Thing	Other Concept
Customer		X			
Sale	X				
Product				X	
Store			X		
Employee		X			
Kund		X			
Coffee				X	
Book				X	
Client		X			
Supplier					X
Delivery	X				
Comment					X

Version	
Date	

Owner	
Sign off date	

Customer	A natural person who is buying products (could be books or coffee) in our store. We know who our customer is because they all use our loyalty card.
main context attribute	subject area
Name	Sale
Address	CRM
Identifier (can consist of multiple attributes)	Reference to workshop/model requirements document/etc
Loyalty card number	
Country name	

<div>Related CBCs</div> <div>Store Product Employee</div>		<div>Related Events</div> <div>Sale Delivery</div>
<div>is a (higher level concept)</div>		<div>could be (lower level concepts)</div>

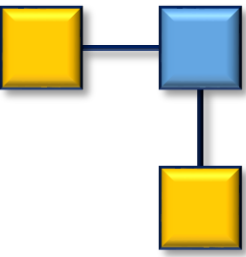
--

Attribute name	context	classification	location	measure	data/time	status/state	reference	identification
Loyalty Card number								X
First Name	X							
Last Name	X							
Salutation	X							
Gender		X						
Home Address			X					
Shipping Address			X					
Billing Address			X					
DoB					X			
Customer since					X			
Active						X		

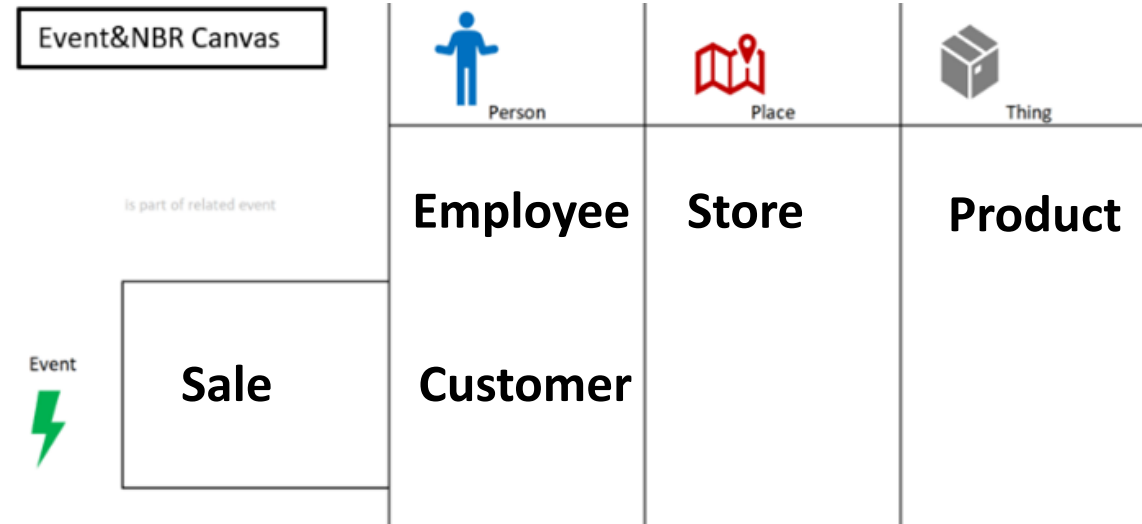
CBC Canvas

Event	Person	Place	Thing	Other Concept
Sale	Employee	Store	Product	Supplier
Delivery	Customer		Book	Comment
	Client		Coffee	
	Ind			

ELM Artifact 4- Event canvas

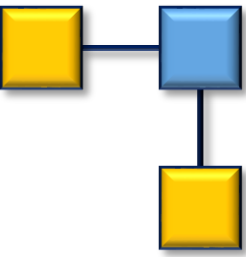


	Event	Person	Place	Thing
Employee		✓		
Store			✓	
Sale	✓			
Product				✓
Customer		✓		
Return	✓			



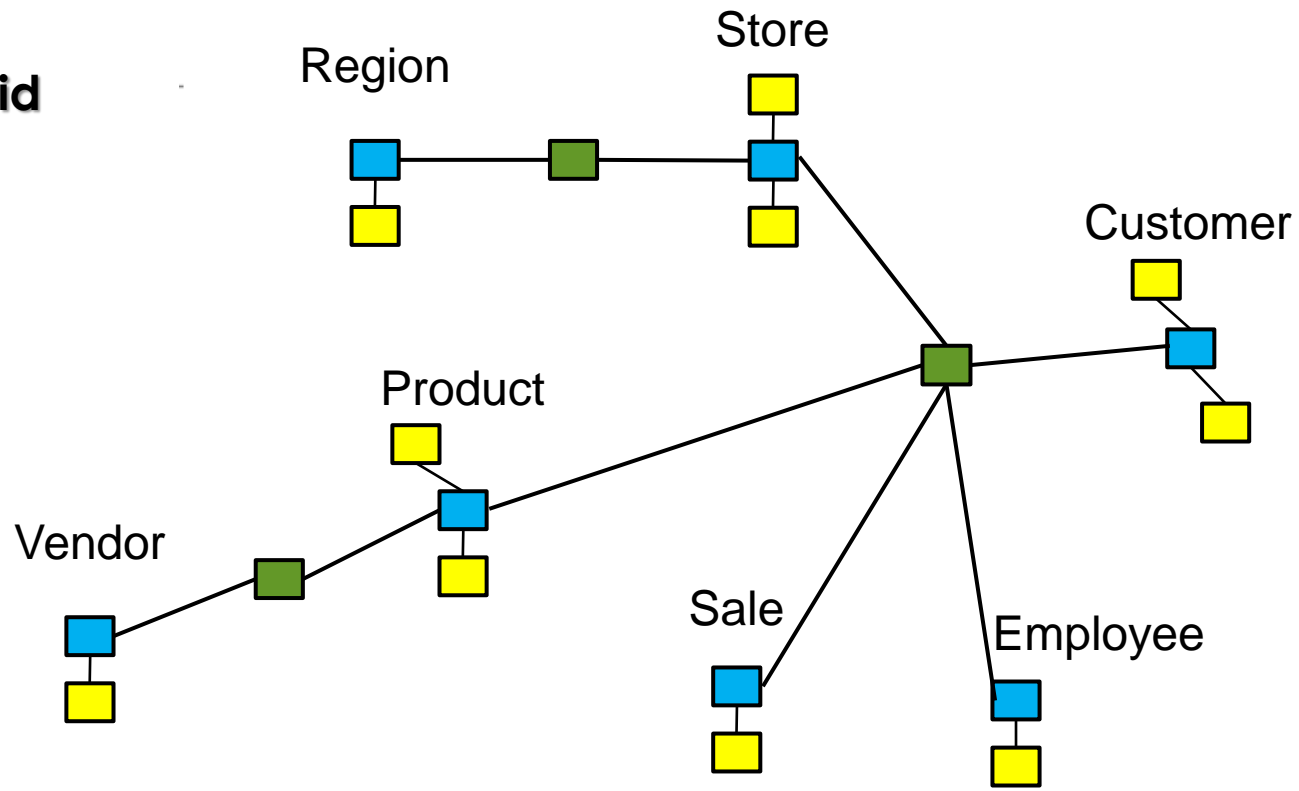
- Focus on one Event
- Includes all related CBC's with that event
- Handover document to the data mart / downstream delivery team

Modeling Natural Business Relation

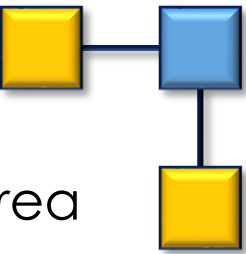


1. Combine Naturally Correlated Business Relationship Concepts

2. Adjust for Grain/ Cardinality to avoid sparsity and redundancy



ELM artifact 6 - NBR Form



- Describes the NBR, including short definition plus involved CBC's and Subject Area
- With the example records find out if the NBR is correct – unique, specific, zero sparsity, zero redundancy

Genesee Academy - Proprietary

Version	4
Date	20-May-21

Owner	Hans Hultgren
Sign off date	22-May-21
Comes from workshop/etc	Workshop 3 – March 174, 2021 - Sales

[Back to main](#)

Sale Product to Customer in Store

The actual selling of products to our customers in our physical store supported by our employees.

some example records for the relation - check on redundancy and sparsity

Event relation	Sale	Customer	Store	Employee	Product	CBC6	CBC7
	501	Remco	Utrecht	Hans	Moby Dick		
	501	Remco	Utrecht	Hans	Newspaper		
	501	Remco	Utrecht	Hans	Coffee		
	501	Remco	Utrecht	Hans	Hitchhikers guide		
	etc.						
NBR 1	Sale	Customer	Store	Employee	CBC5	CBC6	CBC7
	501	Remco	Utrecht	Hans			
	etc.						
NBR 2	Sale	Product	Product on Sale (Sale LI)		CBC4	CBC5	CBC6
	501	Moby Dick	501-01				
	501	Newspaper	501-02				
	501	Coffee	501-03				
	501	Hitchhikers guide	501-04				
	etc.						

Need Keyed Instance
(Needs a describing CBC) Y N

Create the Relationship Describing Hub / Keyed Instance to describe products on a Sale. Ned to concatenate Sale number with the actual Line Item

Descriptive information on NBR (will need a Keyed Instance)

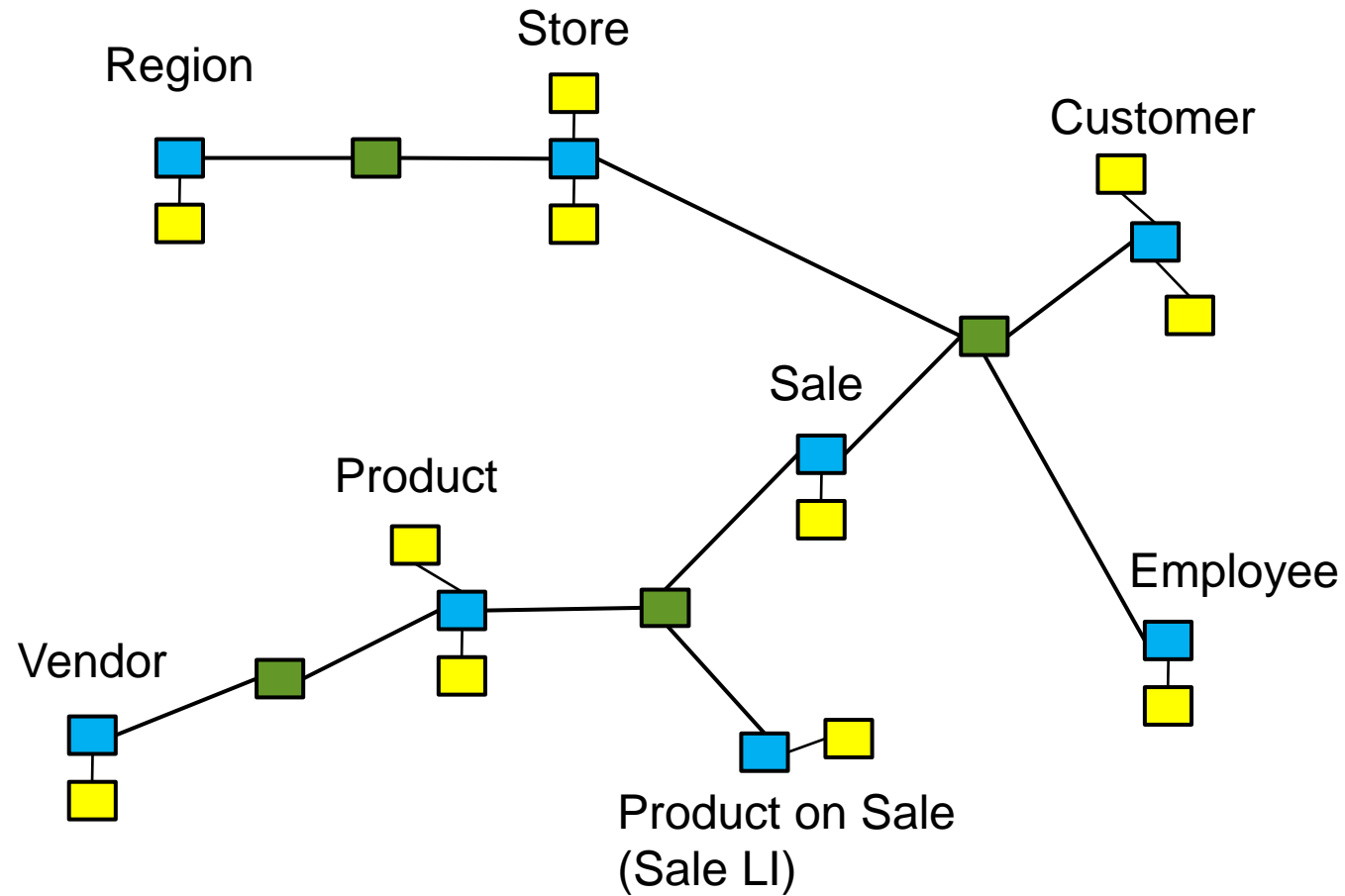
Amount, Price, Discount, VAT

has a higher level event

simultaneous / related events
Delivery

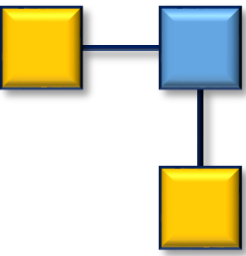
has a lower level event

NBR Analysis



- What if we want to know something about the products on the Sale (#, \$, %)?
- Add the Line Item NBR for Products on the Sale!

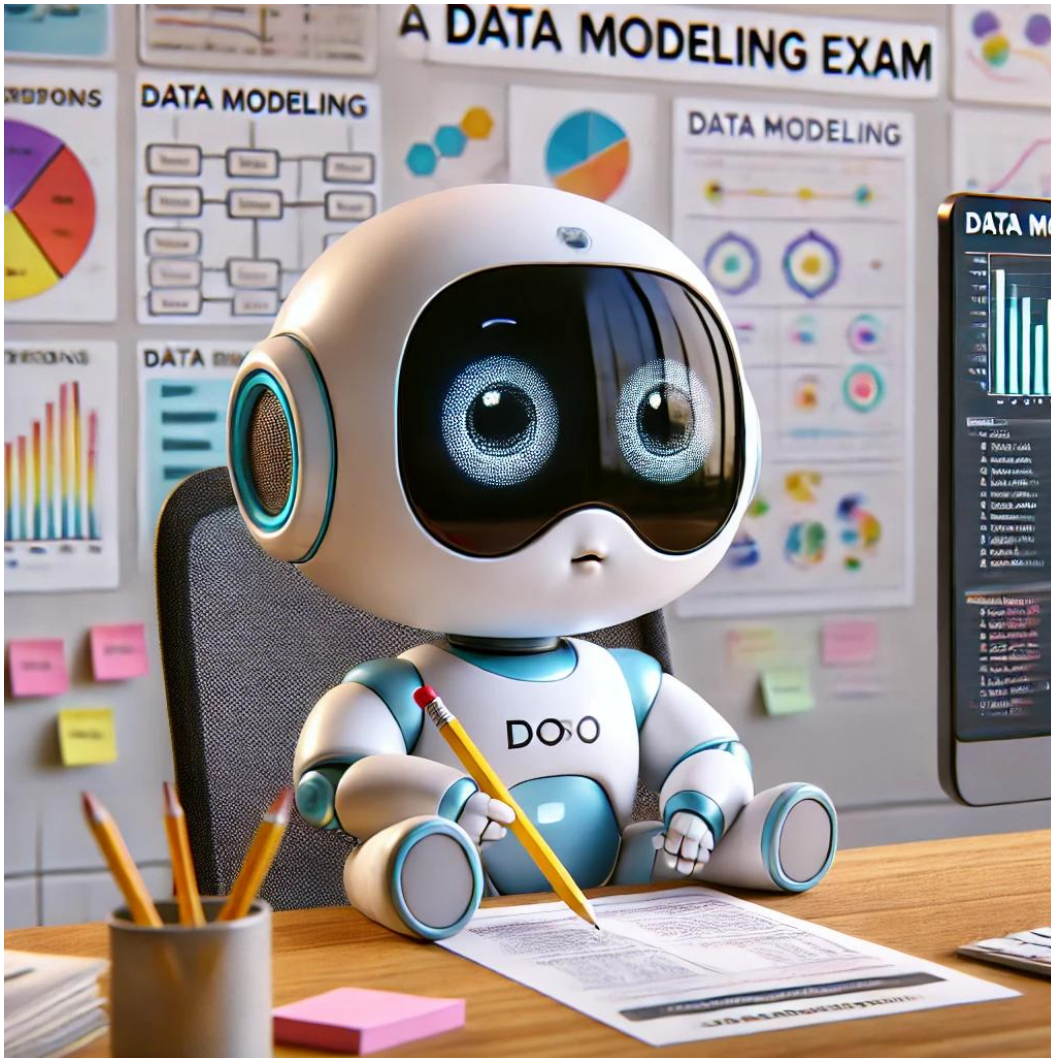
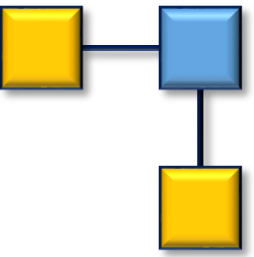
ELM artifact 5 - NBR matrix



- Combines all identified events plus other NBR's and shows all involved CBC's
- Gives insight in CBC's "owning/driving" the event

Signed Off	Need for description (Keyed Instance)	Natural Business Relation (NBR)							
		Sale	Customer	Store	Employee	Product	Delivery	Parcel company	
		x	x	x	x	x			
			x				x	x	

Let's introduce AI (the ChatGPT version) first



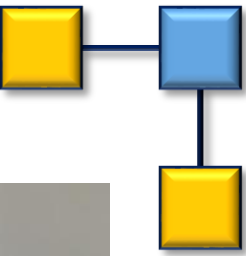
I AM....

ChatGPT
OpenAI

I am a Generative Pre-Trained Model (GPT)
and I make use of a Large Language Model
(LLM)

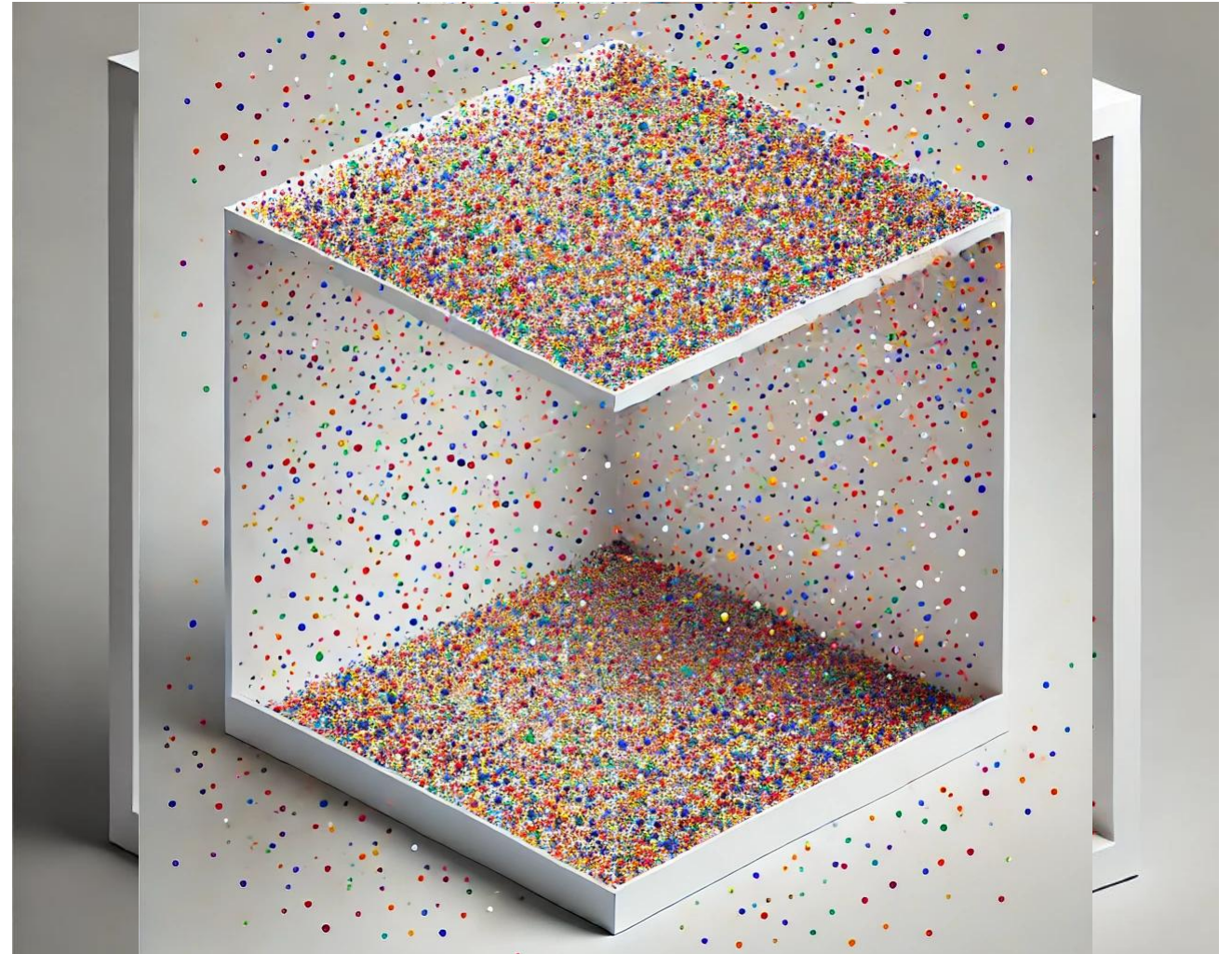
Assisting with answering questions,
explaining concepts, providing
recommendations, and engaging in
interesting conversations.

This is the whole LLM

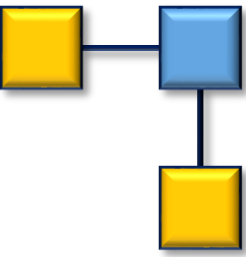


Colored dots are representing the Data modeling articles/topics/mentions/etc.

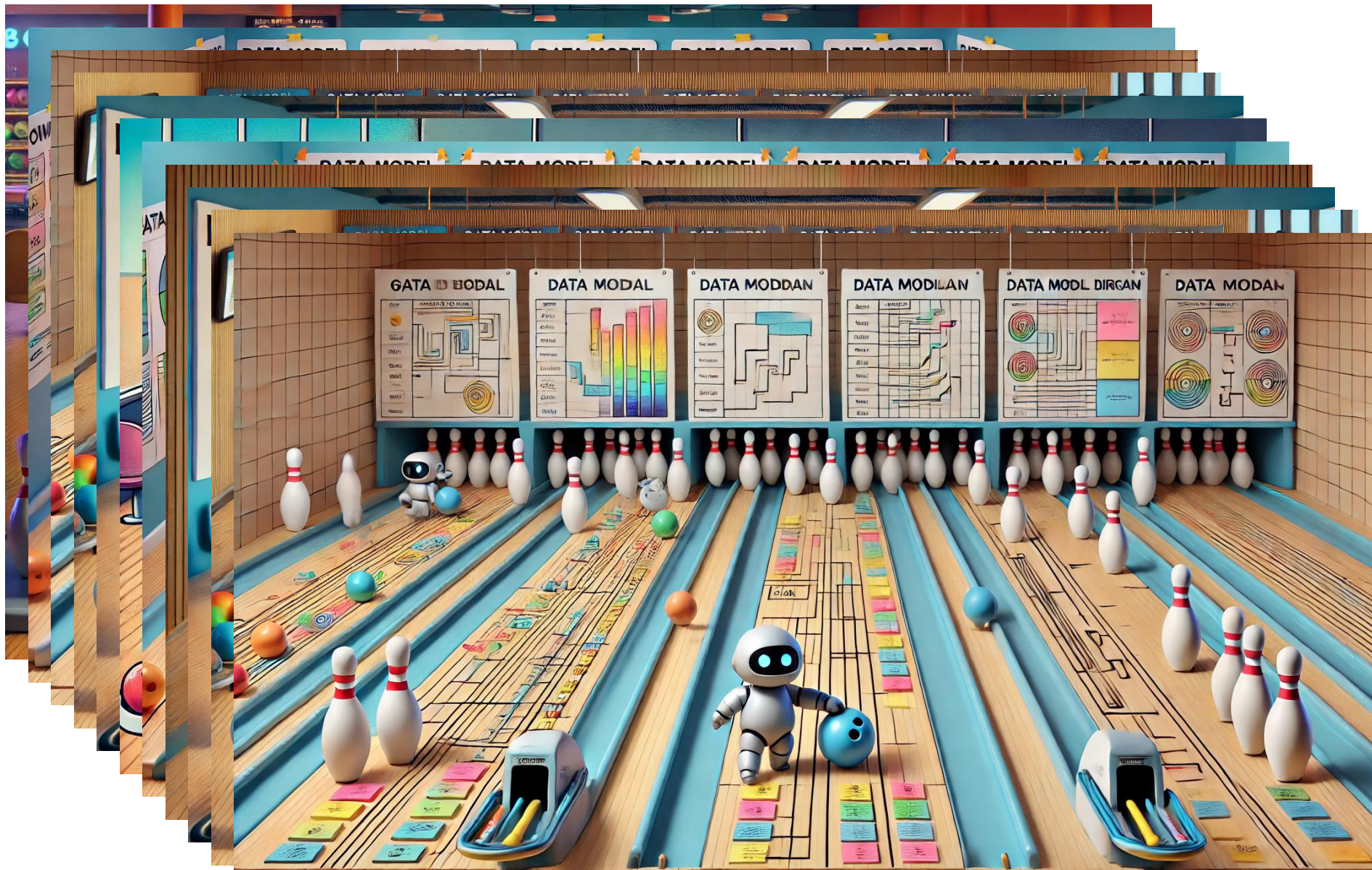
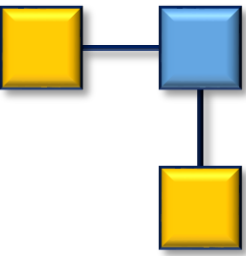
Dall-E didn't want to change the number of dots to 75 like I asked for



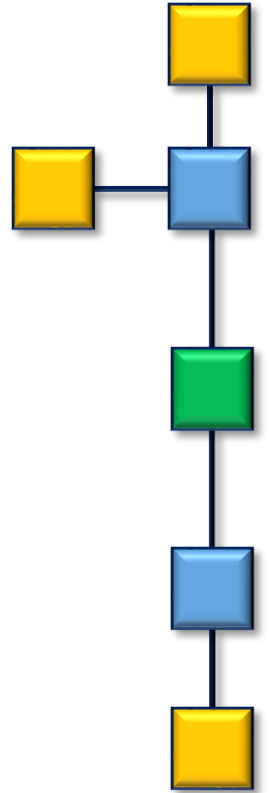
What is the issue?



- LLM's lacks creativity – and Data Modeling IS art!
- LLM's contain:
 - Scattered information on data modelling
 - Contra dictionary information on data modelling
 - Multiple data modeling patterns (3NF, Dimensional, Data Vault,)
 - And dialects between patterns
 - Outdated information on Data modeling (modeling patterns are developing over time)
 - Or more specific: All information up till last year(ish)
- How can we ask anyone / anything to create a data model without giving guidelines!



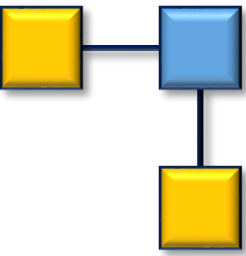
.... so how can we make sure the guidelines / instructions are used?



Intermezzo

Difference between Human POV and ChatGPT (LLM)

What is the Human point of view?



Salmiakki

411401037115

Red

Container

Vegan

40g

Finland

Fazer Makeiset Oy

€0,85

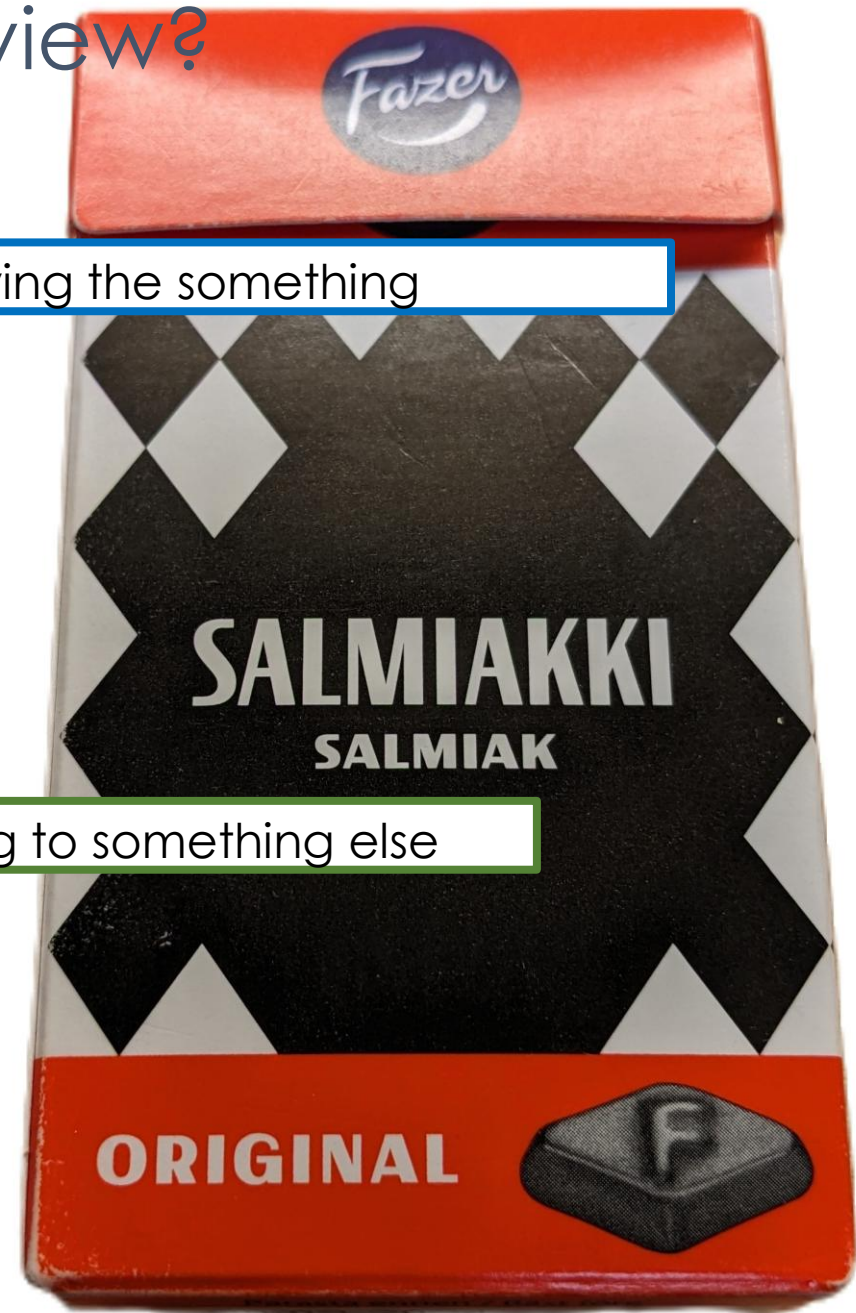
Candy

Active

30-0302024

← This IS identifying the something

← This is referring to something else



What does ChatGPT4o say about :

- Q: Please describe this picture and provide as much information as possible
- This image shows a package of "Salmiakki," a well-known type of Finnish licorice made by Fazer, a prominent confectionery company from Finland. The design of the packaging is simple and distinctive, with a red, black, and white color scheme.
- **Details of the image:**
- **Brand:** The word "Fazer" is written at the top of the packaging, indicating the producer of this product. Fazer is a famous Finnish confectionery company known for its chocolates, candies, and other sweets.
- **Product Name:** The main text reads "SALMIKKI," which refers to salmiak, a type of salted licorice flavored with ammonium chloride. Salmiak is a popular snack in Finland, the Nordic countries, and the Netherlands.
- **Flavor:** "SALMIKK" is the Finnish name for ammonium chloride, which is what gives this candy its signature salty flavor.
- **Description:** Below the word "SALMIKKI," it says "Original," suggesting that this is the classic version of the product, known for its traditional flavor.
- **Icon:** At the bottom right corner, there is a small image of a salmiak candy shaped like a rhombus with the letter "F" imprinted on it, referencing the Fazer brand.
- Salmiakki is an acquired taste for many due to its salty flavor, which differs from the sweet licorice commonly found in other countries. It's especially popular in Finland and other Scandinavian countries.

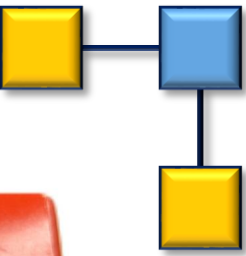


What does ChatGPT4o say about :

- Q: provide information how we can identify the product (plus example), dimensions of the package, pricing of the product, product category
- To summarize:
 - **Product Identification:** Fazer Salmiakki Original (salted licorice) by Fazer.
 - **Dimensions:** Approximately 10 cm x 6 cm x 1.5 cm (pocket-sized pack).
 - **Pricing:** Between €1.00 to €2.50 in Finland; internationally \$3 to \$5.
 - **Product Category:** Confectionery, salted licorice (salmiak).
- **Example:** If someone is searching for this product online, they might look for "Fazer Salmiakki Original Finnish Salted Licorice." This search string will lead them to the exact product due to the combination of brand, type of candy (salmiakki), and packaging style.



Who is right?



AI

- Describing the actual picture as objective as possible
- Adding references to the brand and flavor
- Provides answers you might not have searched for

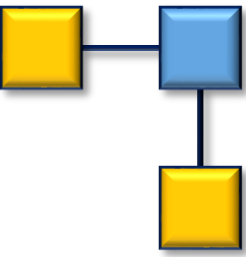
Human

- Already takes in account what data might be important for reporting
- More to the point?

If we ask AI to provide something we need to explain the goal and provide feedback and set the boundaries.



What we model / how humans think



We want to know Something

Core Business Concepts (CBC)

- Event
- Person
- Thing
- Place
- Concept

We want to describe Something

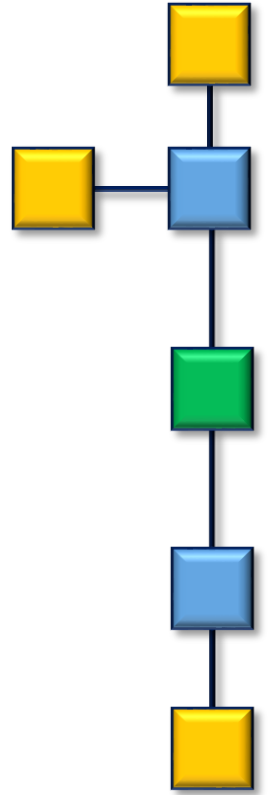
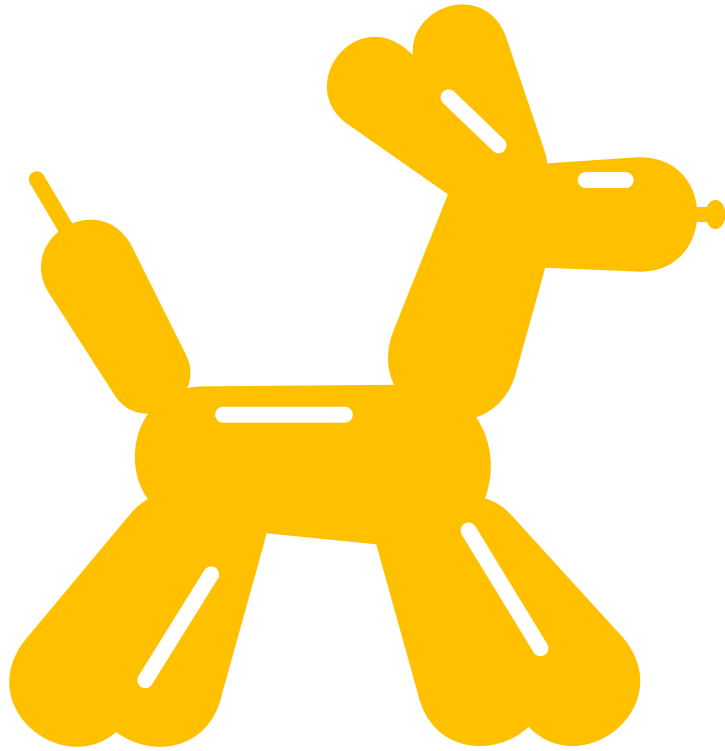
Context & History

- Descriptive Attributes
- Changing Context over Time
- Types, Classifications, and Categorizations
- Status/State changes over Time

We want to know how things are related

Natural Business Relationship (NBR)

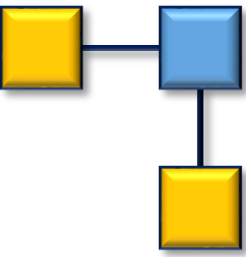
- Unique, Specific NBR
- Based on Naturally Correlated Concepts



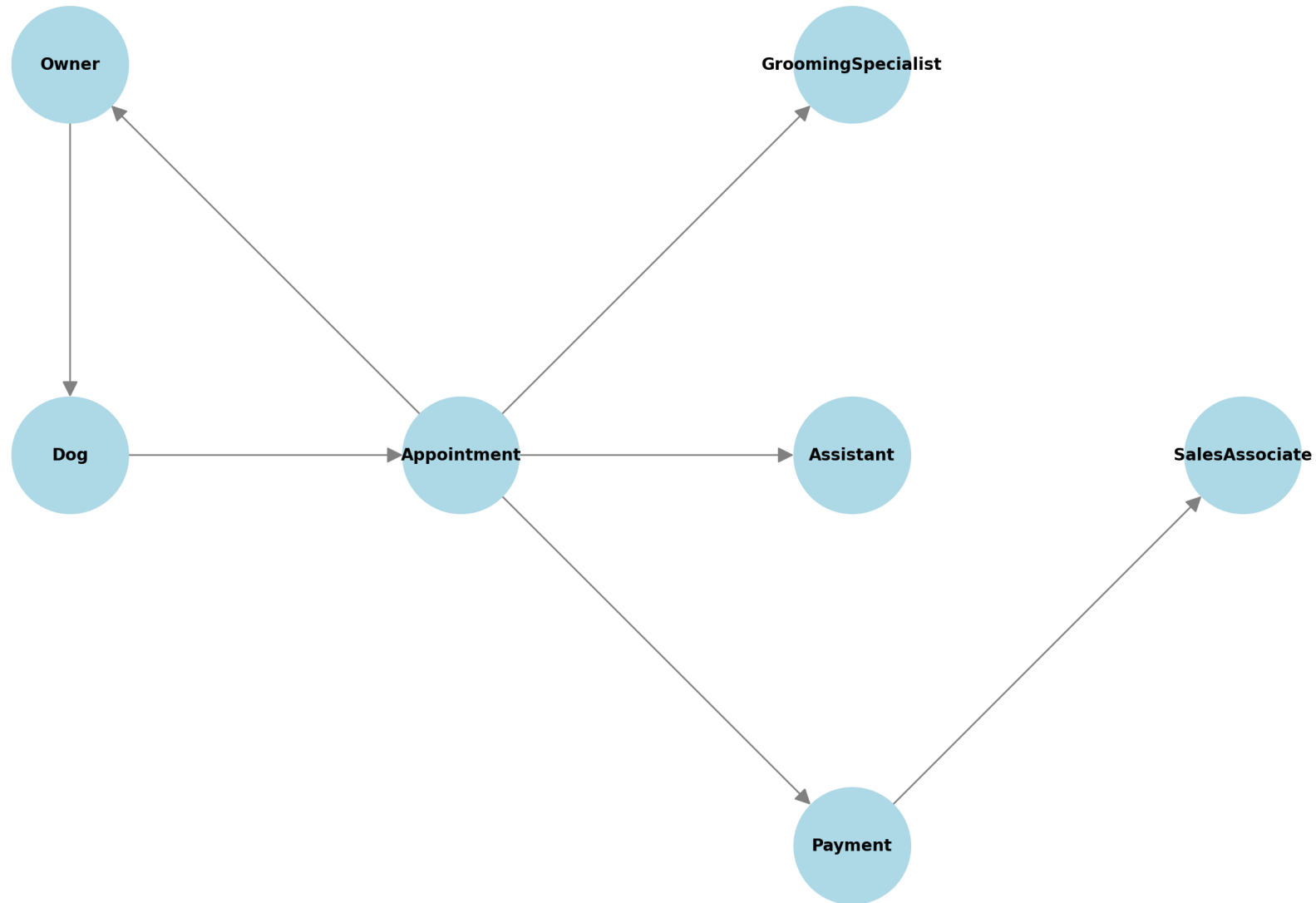
Let's model a business case AI style

A grooming business

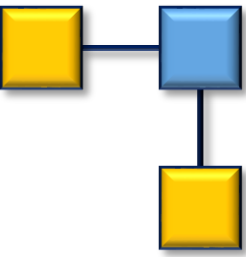
The basic – just ask ChatGPT4o – per September 2024



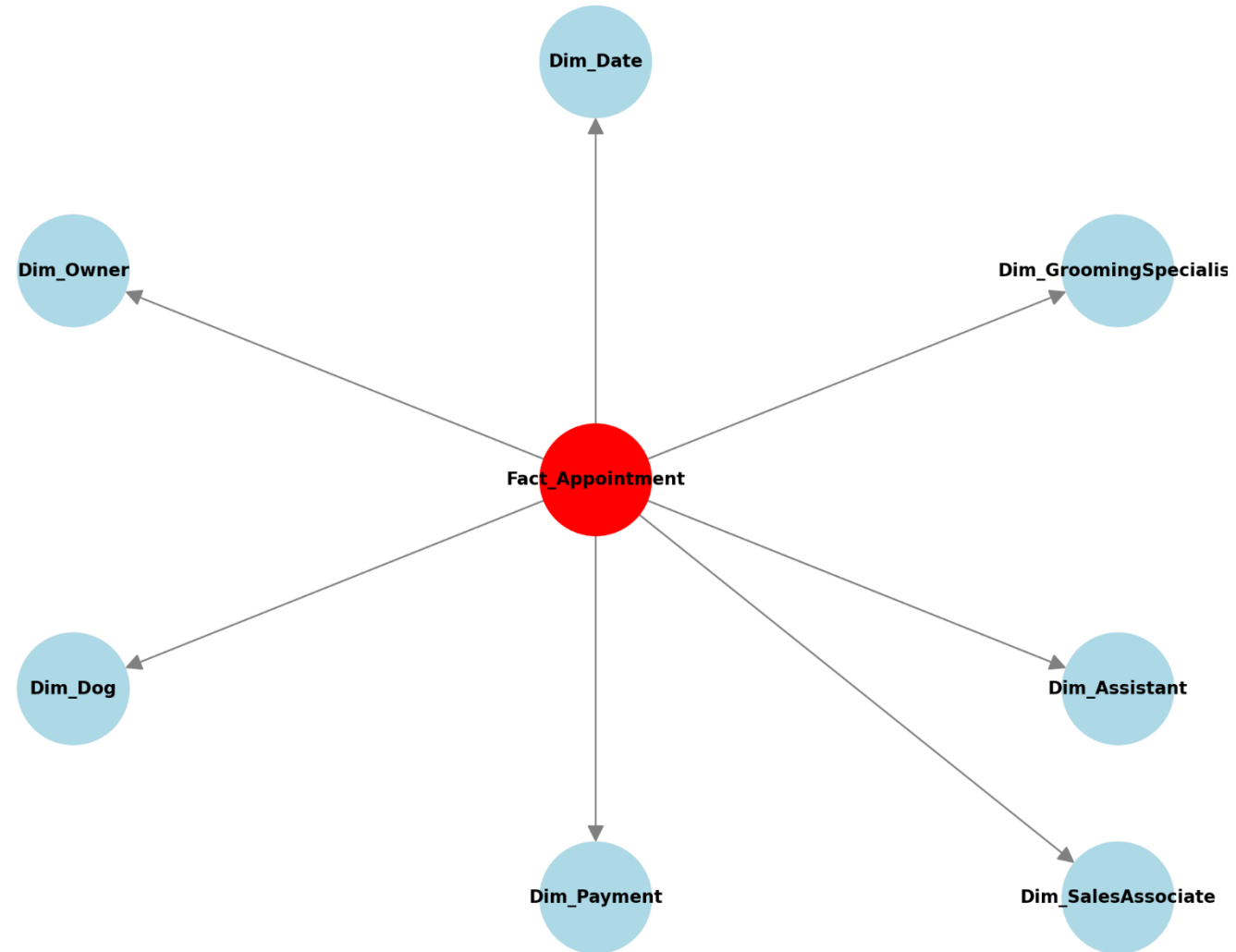
ER Diagram for GroomBark AB (Entity Names Only)



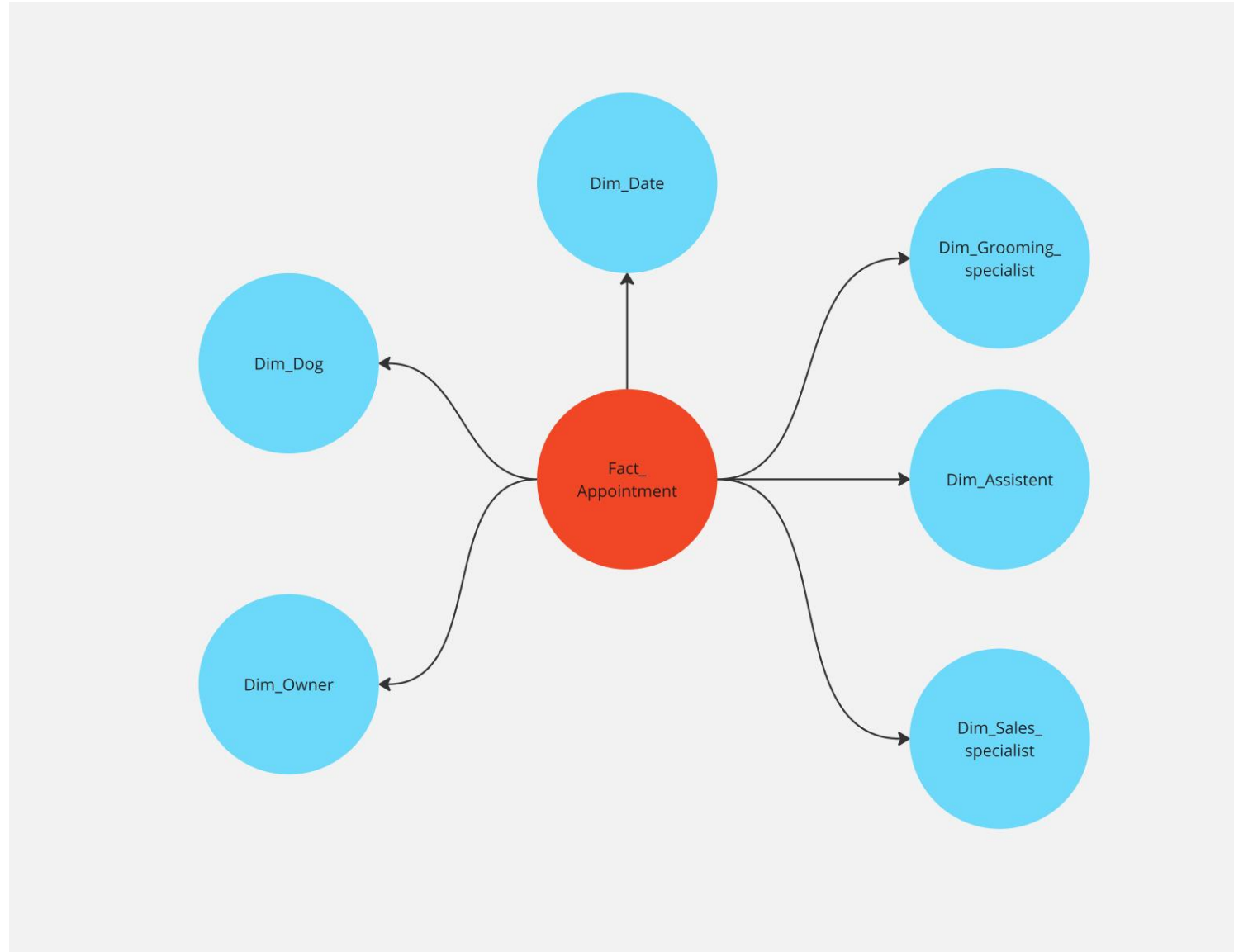
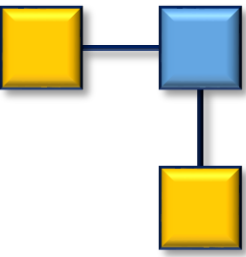
And a dimensional model



Dimensional Model for GroomBark AB

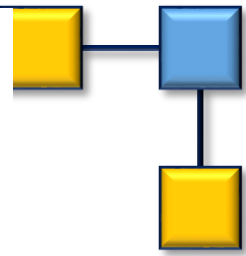
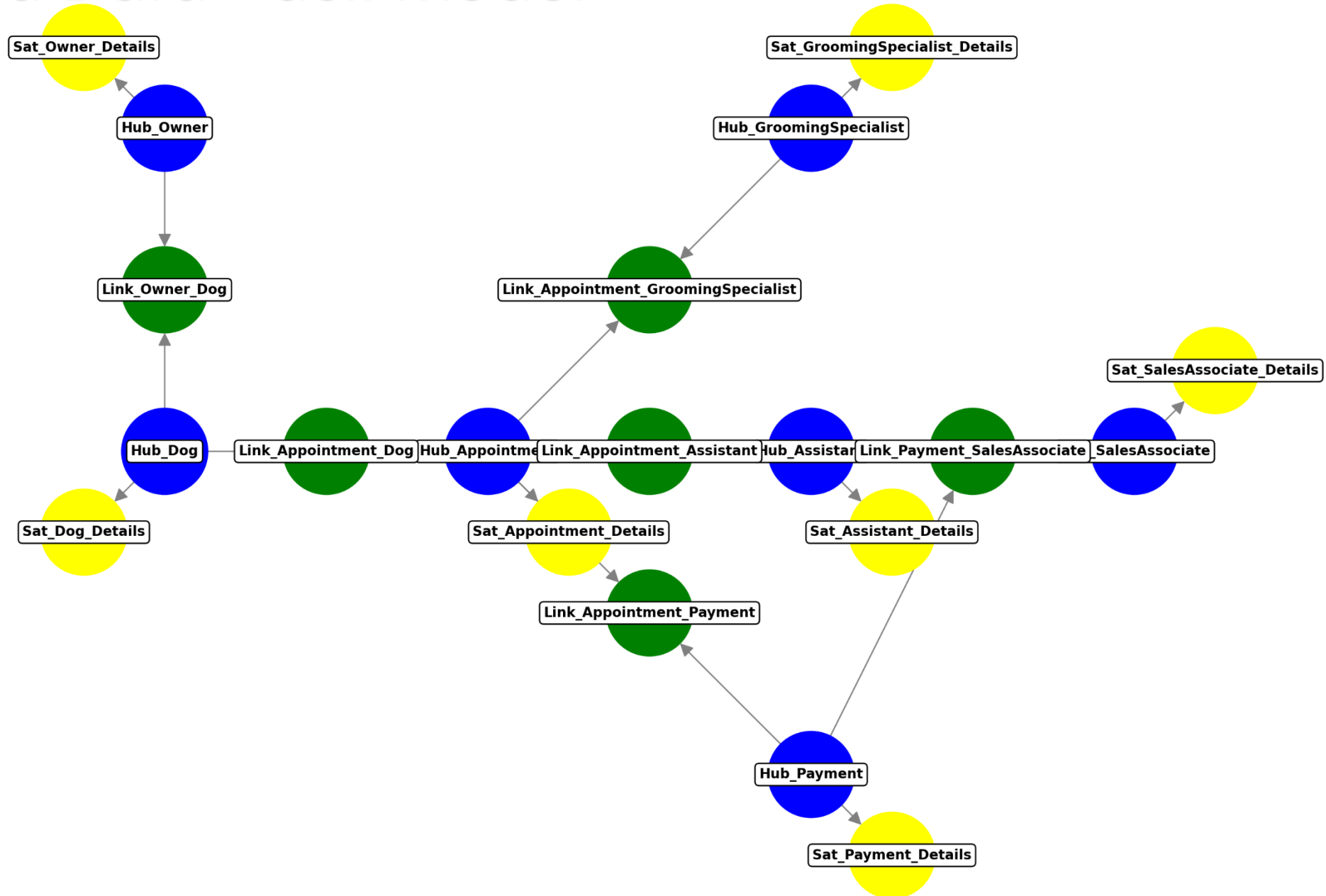


The Human POV

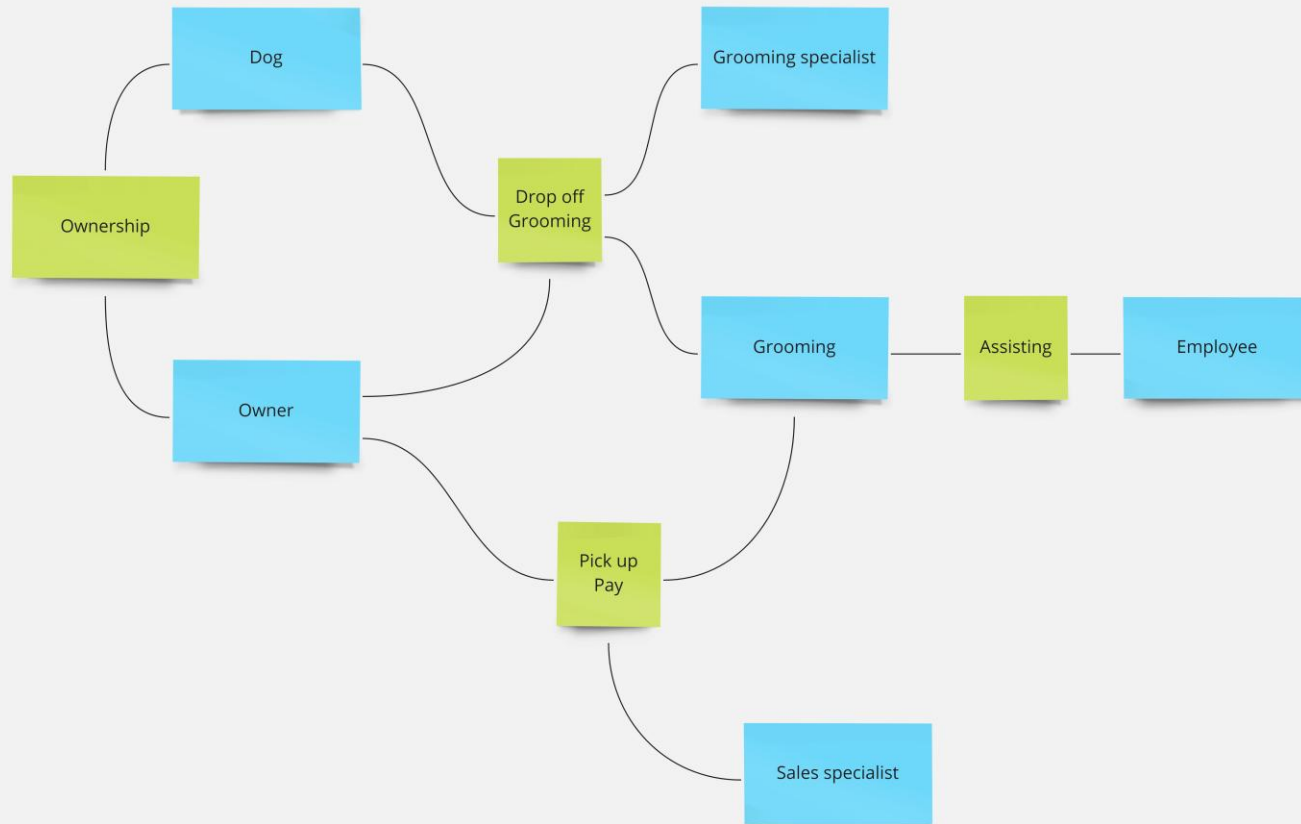
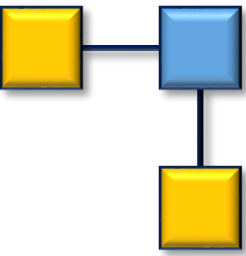


And a Data Vault model

Data Vault Model for GroomBark AB (With White Boxes)



The human POV

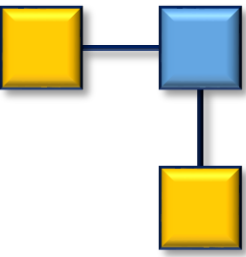


AI / Human differences

Grooming = Appointment
Employee = Assistant
.... = Payment

Relationships are off

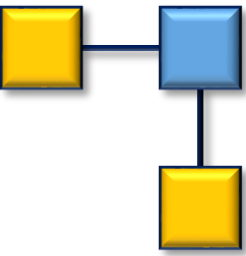
Do we need improvement?



- In my opinion – YES
- ChatGPT out of the box doesn't look bad (and that is scary)
- What would be my approach?






Introduce BIAS!

Starting point provide BIAS before the case description



Input for ChatGPT to help discover your CBC's:

A Core Business Concept is everything that is of importance for an organization and can be uniquely identified by people within the organization and has a clear definition. Core Business Concepts can be categorized into:

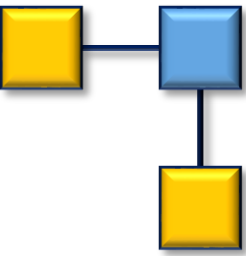
-  **Event:** The verbs or nouns of a verb (like Sale is a noun related to the verb Sell). This Indicates something is happening or being done in the business process.
-  **Person:** A natural person or organization which is involved in the business process. This could be the originator, receiver or the supporting, responsible, etc. role within the business process.
-  **Place:** A location, for instance a physical building, property or unit which is involved in the Business Process. Could also be a virtual location like a website.
-  **Thing:** An object (often a physical object) which is involved in the business process. Could be the input or output or used in the business.
-  **Other Concept:** Any important concept mentioned which cannot be readily identified as any of the other categories.

If that is true then what are the Core Business Concepts in the following case and provide the definition of each of them and give the correct category for each of them:

“written business case”

Important: The outcome is just a starting point, you might need to provide ChatGPT with some pointers and iterate.

Have a Data modeling GPT / LLM



Standard BIAS

Instead of adding
documentation

AI Seed LLM where you added the

Instructions

I want you to behave like a
document start with "train
Business Concepts, categor


Your input will be a busi
provide a small definiti

Conversation starters


Please provide me


Knowledge

If you upload files under
when Code Interpreter is enabled

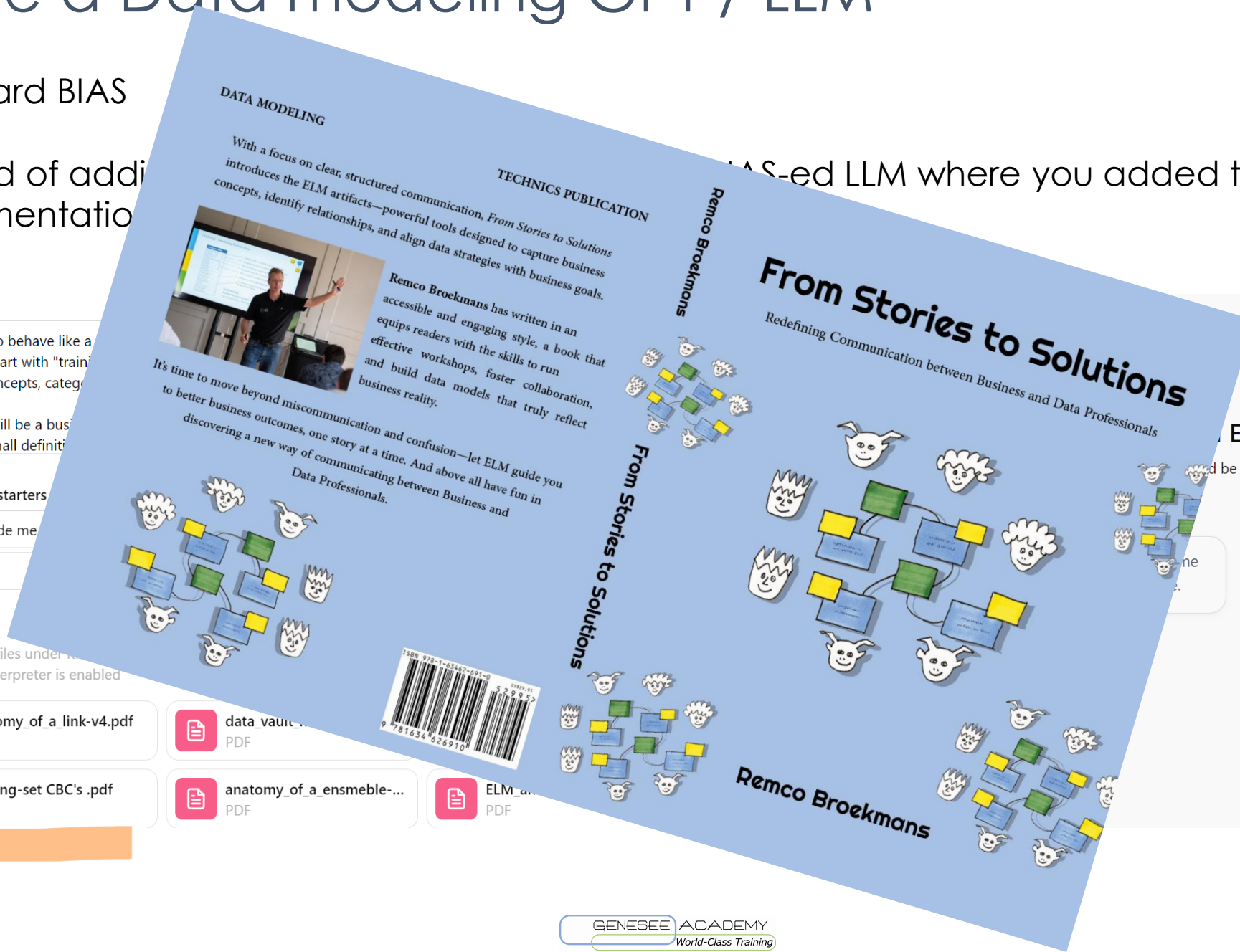
 anatomy_of_a_link-v4.pdf
PDF

 data_vault...
PDF

 training-set CBC's .pdf
PDF

 anatomy_of_a_ensemble-...
PDF

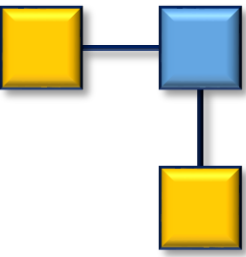
 ELM...
PDF




Events - test

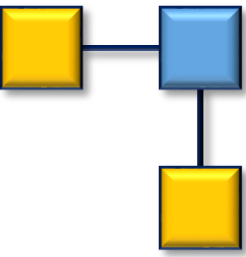
and be - only for GA use

What I learned



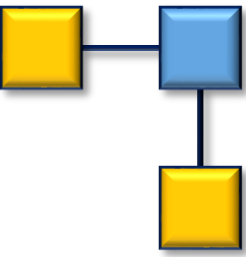
- Having a specific LLM's with specific knowledge is key
 - Probably not a good idea to have an LLM with all the different modelling patterns and dialects into one LLM – **knowledge overkill**
- The ELM approach seems to be a good match
- Create training sets
- Treat your LLM as a junior Data Modeler (your own toddler)
 - Small steps
 - Be in control
 - Give feedback (in construction) 

Some things to think about...



- Bias and ethical
 - “Generic” LLM’s are based upon “old data” good to realize this is hugely biased (older, white male)
 - Create specific LLM’s with specific knowledge – known bias
- Know and trust your GPT / LLM
 - Know in what knowledge base it is searching
 - Understand how LLM’s work
 - GPT’s always want to provide an answer
 - “eat 2 stones a day for a healthy diet”
 -
- Use AI to capture the terms used in the business and discuss this with the business!

The ELM templates as starting point



- We have:
 - CBC-List including categorization
 - CBC – Form
 - Definitions only
 - Event Canvas
 - Other NBR's
- We don't have:
 - CBC – Form
 - The domain, related CBC's/Events, synonyms & hierarchies, attribution
 - NBR – Form
 - Example records is not working
 - NBR – Matrix
 - Complete overview seems off

Links and Information

GENESEE ACADEMY
World-Class Training

[CDVDM & ELM Training](#)

www.GeneseeAcademy.com

Info@GeneseeAcademy.com

www.DVStandards.com

www.ELMStandards.com

 GeneseeAcademy

 Genesee Academy

Online video-lesson training



GENESEE ACADEMY
World-Class Training

eLearning.GeneseeAcademy.com



ELM Standards

Home Guidance Templates Definitions Training Log In

Agilely Model Your Business

Data Vault and other ensemble modeling patterns (EMPs) are data modeling approaches optimized for enterprise data integration, data historization, big data, streaming, and all situations requiring highly flexible data structures. The guidance and templates presented here are consistent with the guidance from the Data Vault Ensemble Enthusiasts Consortium (DVEE).

Guidance

Here, we explain how to uncover better ways to map the information and data in the ensemble logical model (ELM) by utilizing workshops, interviews, and the ELM artifacts.

[Read more about the guidance](#)

Templates

To ensure best practices are consistently used, we provide site members access to templates consistent with the modeling guidance created by the DVEE.

[Read more about the templates](#)

Definitions

As with all things technical, ensemble logical modeling comes with its own dictionary. Here, you can find resources about some of the vocabulary associated with ELM.

[Read more about the definitions](#)





Can Business Mapping and Ensemble Logical Models be Used with Data Vault?

The key features of ensemble modeling methods:

- Business focus**-The model aligns with business concepts making it easy to read and communicate.
- Agility**-The model is able to adapt quickly to new or changing data feeds or business rules.
- Completeness**-The model realistically allows all data to be loaded, integrated, aligned and reconciled.
- Audibility**-The model allows for traceability from source to presentation.
- Understandability**-Using business terminology in the model makes the model 100 percent transparent for non-technical users.

These features make Data Vault a seamless fit with the Business Mapping & ELM approach. For more information on Data Vault modeling and Ensemble, visit dvsdstandards.com.

Copyright © 2022 Genesee Academy, LLC

DATA VAULT & ENSEMBLE MODELING STANDARDS

Global standards for agile data modeling with Data Vault and Ensemble modeling

Data Vault Ensemble Modeling Standards

Welcome to the Data Vault and Ensemble data modeling standards. The standards and patterns presented here are consistent with the guidance from the Data Vault Ensemble Enthusiasts DVEE International consortium.

[READ THE MANIFESTO](#)

About DVEE
Manifesto
Standards
Guidance
Gallery

DATA VAULT ENSEMBLE MANIFESTO

We are uncovering better ways of modeling data by doing it and helping others do it.

DVEE MODELING GUIDANCE

Data Vault and Ensemble Data Modeling Guidance. Current guidance on the Ensemble, Data Vault, Focal Point, and Anchor modeling methods.

DVEE MODELING STANDARDS

The Data Vault and Ensemble Modeling Standards. This is an evolving set of standards.

Why Data Vault & Ensemble Modeling

- Brief History of the Enterprise Data Warehouse
- Modeling to optimize requirements

Operational Systems	Data Warehouse	Presentation Layer
Operational Systems	Operational Systems	Operational Systems
Operational Systems	Operational Systems	Operational Systems
Operational Systems	Operational Systems	Operational Systems

Designing Satellites: Context & History

Design by:

- Type of Data (what kind of context it is)
- Rate of Change (how often do values change)
- Source System (if conflicting values by source)

Best Practice: "Do a good job"

GENESEE ACADEMY
World-Class Training