



June 13-15, 2022 · Data platforms · Data engineering · Big data

# BUDAPEST DATA FORUM

Adattárházak, data engineering  
és big data konferencia

Június 13.  
Tutorial nap  
személyes

Június 14.  
Konferencianap  
személyes + streaming

Június 15.  
Konferencianap  
csak online

## Data Modeling with the Unified Star Schema

*Francesco Puppini*

June 15th, 2022 - Budapest (virtual conference)

# About myself



Consultant in Business Intelligence since 2001

- Business Objects, SQL
- But then.. Deeper into the source: → DWH, Teradata
- Qlik, Tableau
- Now Astrato



LinkedIn:

Francesco Puppini

# About Astrato

[Product](#)[Solutions](#)[Pricing](#)[Company](#)[Resources](#)[Book a demo](#)[Sign up free](#)

# Play the data where it lives.

Next generation cloud BI for modern data teams, with live data insights and dashboards on top of Snowflake

[Sign up now](#)[WATCH VIDEO](#)

<https://astrato.io/>

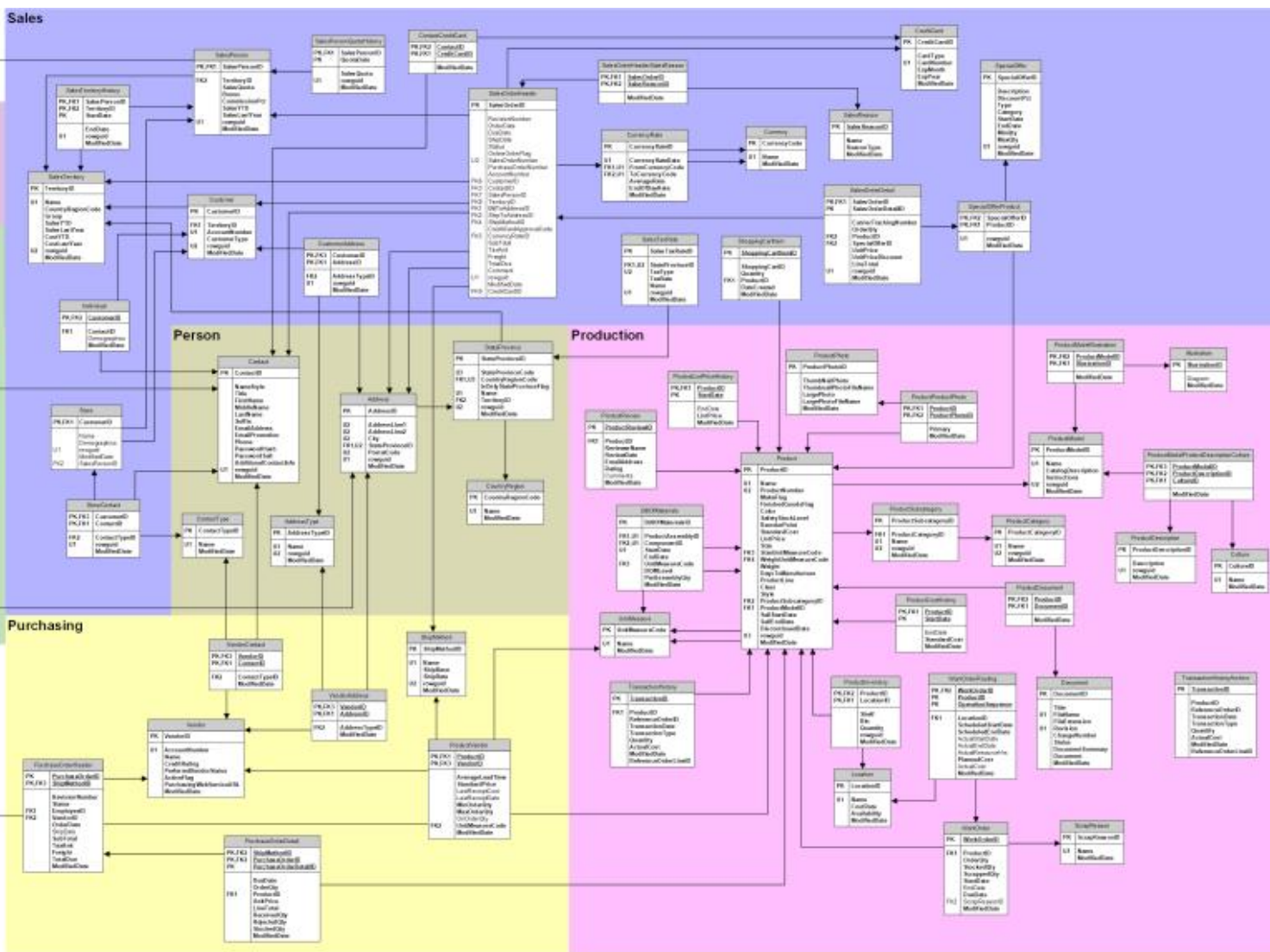
# Business Intelligence: delivering information to the business



How do we achieve this?

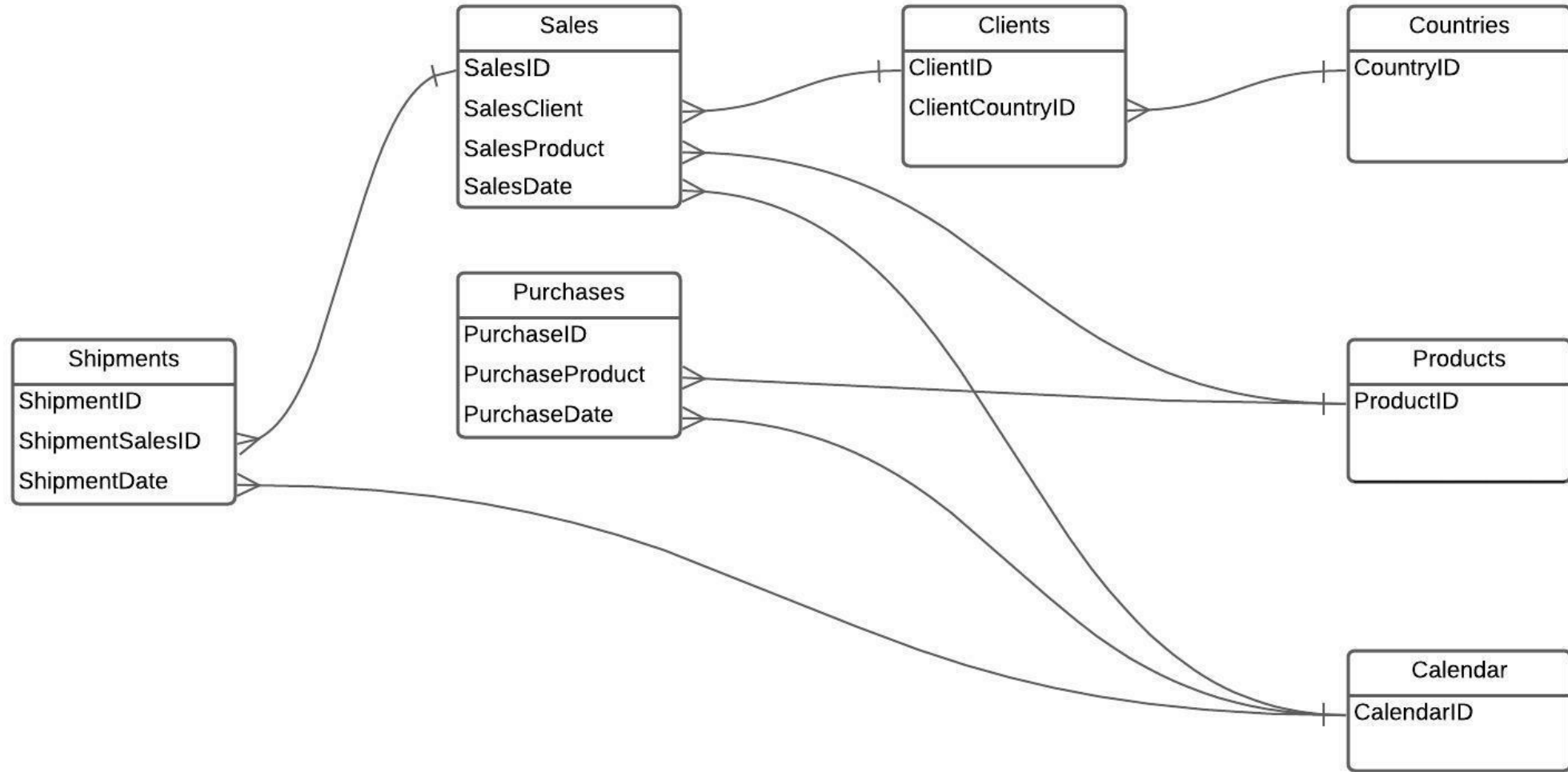


### AdventureWorks OLTP Schema



But instead, BI is slow, expensive, and very often it produces incorrect numbers (not matching the data source). Why is this happening?

# Let's use a very simple example based on Excel



With a simple data set like this, what can possibly go wrong?!

# Sales and Shipments

SalesID	Date	YM	SalesClient	Product	SalesPaymentType	SalesQuantity	SalesAmount
SA01	01-Jan	2020-01	Bill	PR01	PayPal	1	100
SA02	02-Jan	2020-01	Bill	PR02	Card	1	100
SA03	02-Jan	2020-01	Francesco	PR02	PayPal	2	200
SA04	03-Jan	2020-01	Francesco	PR03	Revolut	1	100
SA05	04-Jan	2020-01	Francesco	PR01	Card	4	400

ShipmentID	ShipmentSalesID	ShipmentDate	ShipmentYM	ShipmentProduct	ShipmentType	ShipmentQuantity	ShipmentAmount
SH01	SA01	01-Jan	2020-01	PR01	Plane	1	100
SH02	SA02	02-Jan	2020-01	PR02	Truck	1	100
SH03	SA03	02-Jan	2020-01	PR02	Truck	2	200
SH04	SA04	03-Jan	2020-01	PR03	Truck	1	100
SH05	SA05	04-Jan	2020-01	PR01	Plane	3	300
SH06	SA05	03-Feb	2020-02	PR01	Plane	1	100

What happens if we join these two tables together?

## Let's load them into Tableau

### Sales and Shipments

Sales ID	Shipment ID	Shipment Amount	Sales Amount
SA01	SH01	100.0	100.0
SA02	SH02	100.0	100.0
SA03	SH03	200.0	200.0
SA04	SH04	100.0	100.0
SA05	SH05	300.0	400.0
	SH06	100.0	400.0
Grand Total		900.0	???

What will be the Grand Total of Sales Amount displayed by Tableau?  
900 or 1,300?

LIVE DEMO



If you load it by “Relationship” (aka In-Memory Association):




## Sales and Shipments

Sales ID	Shipment ID	Shipment Amount	Sales Amount
SA01	SH01	100.0	100.0
SA02	SH02	100.0	100.0
SA03	SH03	200.0	200.0
SA04	SH04	100.0	100.0
SA05	SH05	300.0	400.0
	SH06	100.0	400.0
Grand Total		900.0	900.0

By default, the displayed result is 900. This number is correct!

But if you load it by “Join”:



A Venn diagram with two overlapping circles. The left circle is labeled 'Sales' and the right circle is labeled 'Shipments'. The intersection of the two circles is shaded blue.

Sales ID	Shipment ID	Shipment Amount	Sales Amount
SA01	SH01	100	100
SA02	SH02	100	100
SA03	SH03	200	200
SA04	SH04	100	100
SA05	SH05	300	400
	SH06	100	400
Grand Total		900	1,300

By default, the displayed result is 1,300. This number is clearly wrong.

## Key concept

***Some particular combinations of tables cannot be joined together***

# In-Memory Association vs JOIN

With the In-Memory Association, the displayed numbers are always correct.

But this operation only exists in a few BI tools (Qlik, Tableau, Power BI, Good Data,...). Unfortunately, it does not exist in SQL. The SQL language uses the JOIN instead.

**But then, how do we obtain the correct numbers in SQL?**

**In SQL the solution for 30 years was always the same: AD HOC.**

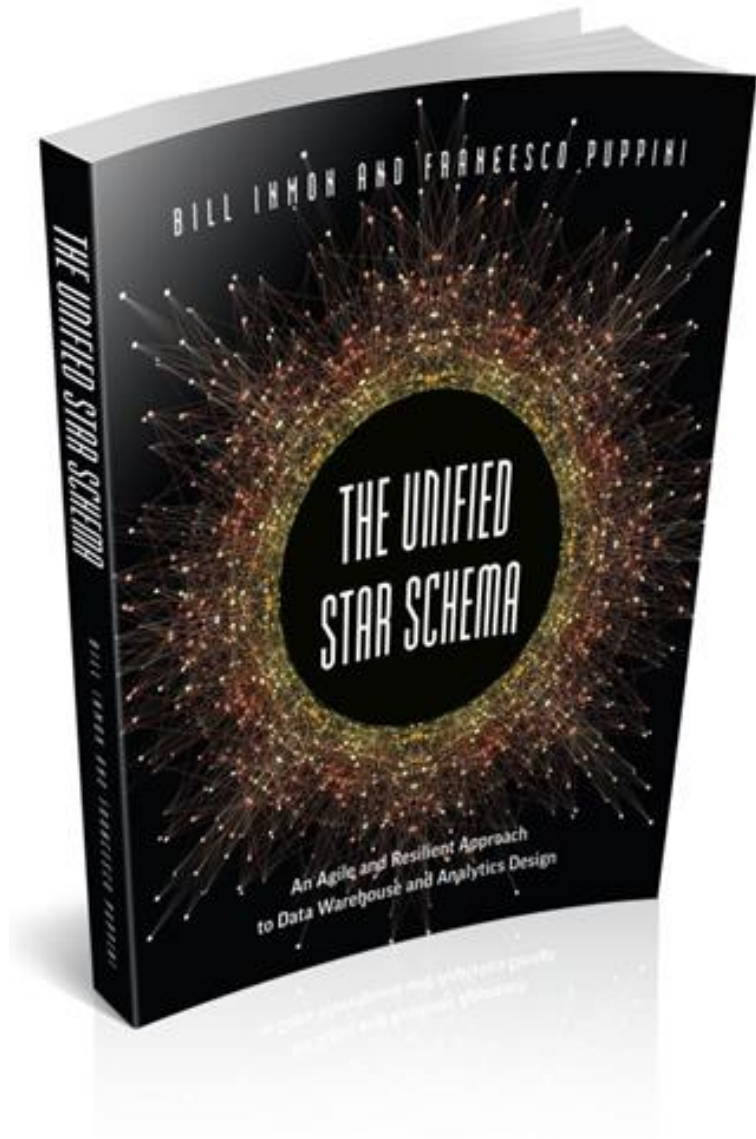
Today, for each new question, we need the help of an OPERATOR



**But.. Is it possible to make the end users  
INDEPENDENT from the OPERATOR?!**



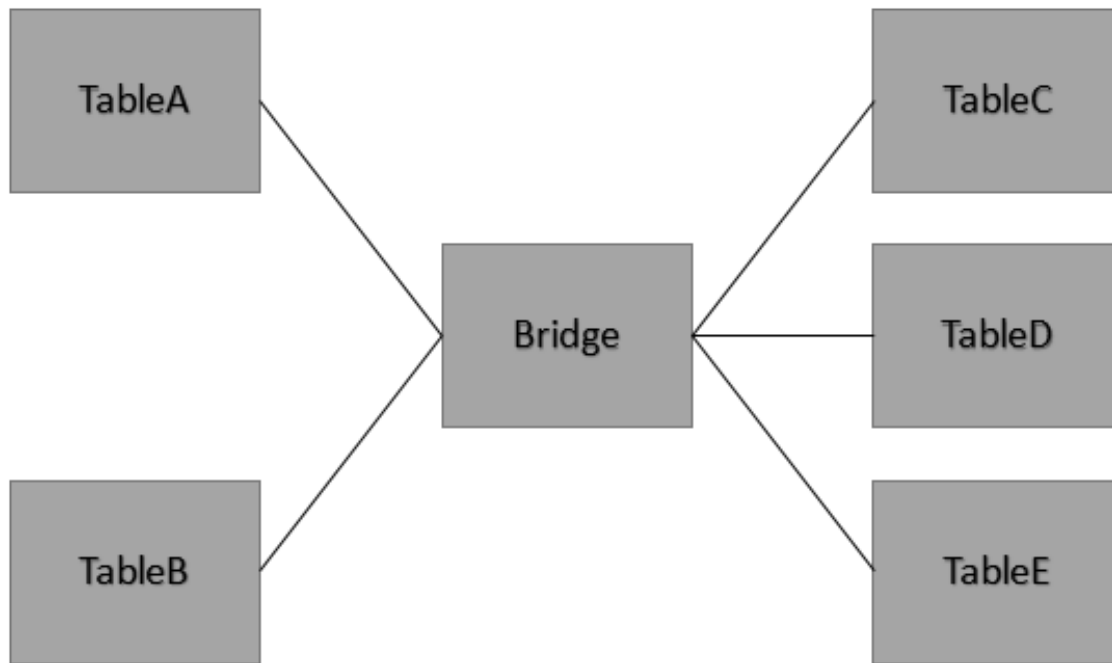
# The Unified Star Schema (USS Approach)



Bill Inmon, Francesco Puppini

# The Unified Star Schema (USS Approach)

All the tables are connected to a central table, called the “Bridge”.  
The Bridge is a UNION of “Stages” containing keys and measures



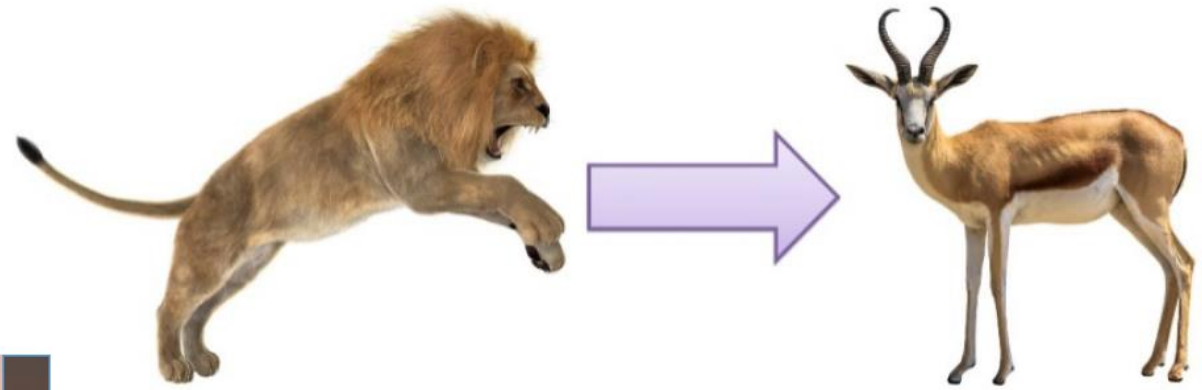
Potentially, all the tables of all your data marts can fit into one single star schema.

But... How does it work?!?!?

And... why no one thought about it before?

# Intuition n.1: ORIENTATION

There is always a hunter and a prey

The screenshot shows the American Airlines website. At the top, there's a banner with the text "Fly to the U.S. and beyond" and "Book your trip". Below this is a navigation bar with "Log in", "Join AAdvantage", "Book", "Manage trips / Check-in", and "Flight status". The "Book" section is active, showing options for "Flight", "Hotel", and "Car". There are radio buttons for "Return" (selected) and "One way", and a checkbox for "Redeem miles". The "From" field is set to "VCE" and the "To" field is "City or airport". The "Adults (12+)" field is set to "1" and the "Children (2-11)" field is set to "0". There are "Depart" and "Return" date fields, both showing "dd/mm/yyyy". A "Search" button is at the bottom right. On the left, there's a "Log in" section with fields for "AAdvantage # or username", "Need AAdvantage number?", "Last name", and "Password", along with a "Remember me" checkbox and a "Log in" button.

Reservations

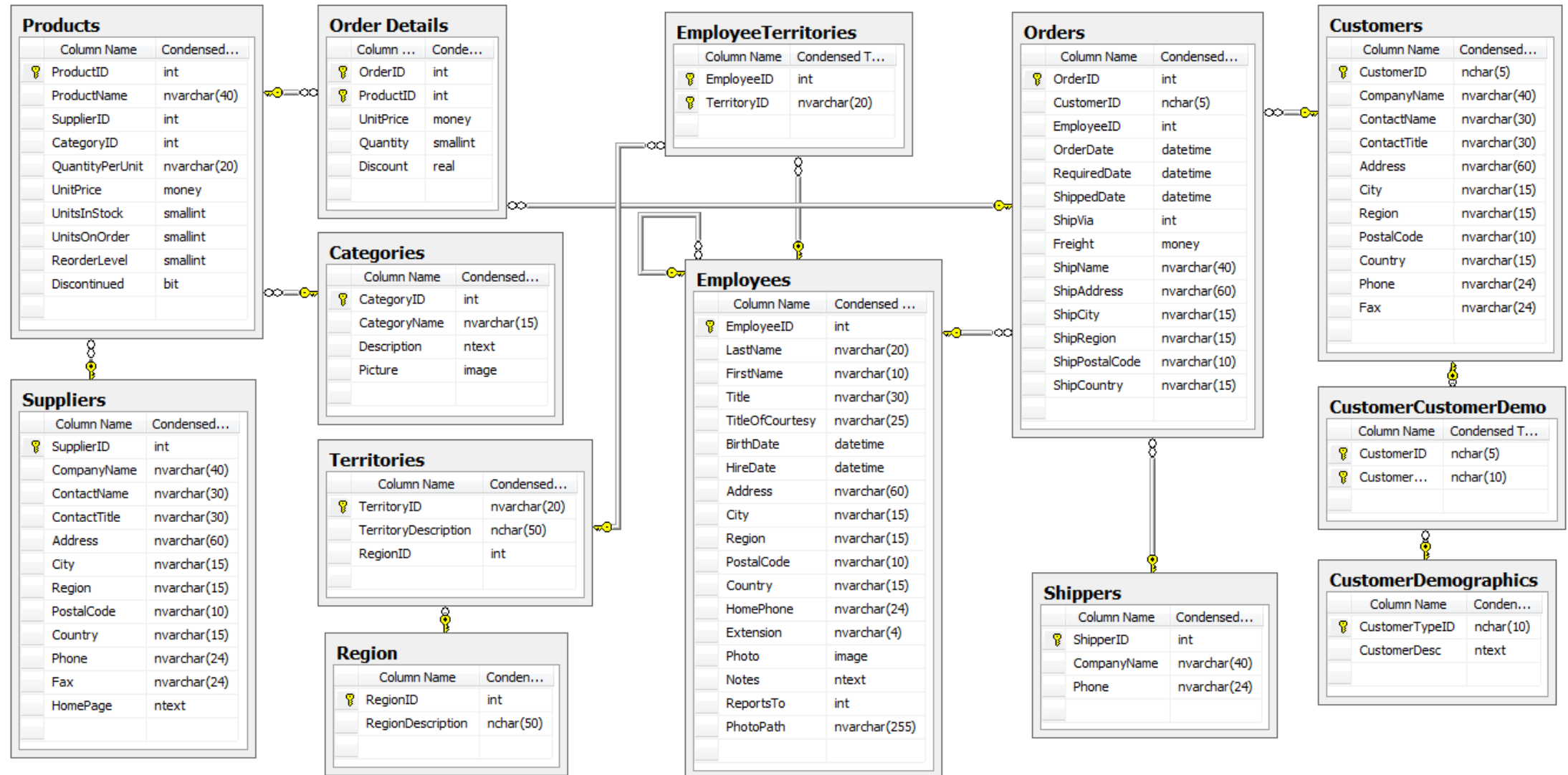
A rectangular box representing the "Reservations" entity. Inside the box, the attribute "PassengerID" is listed.

Passengers

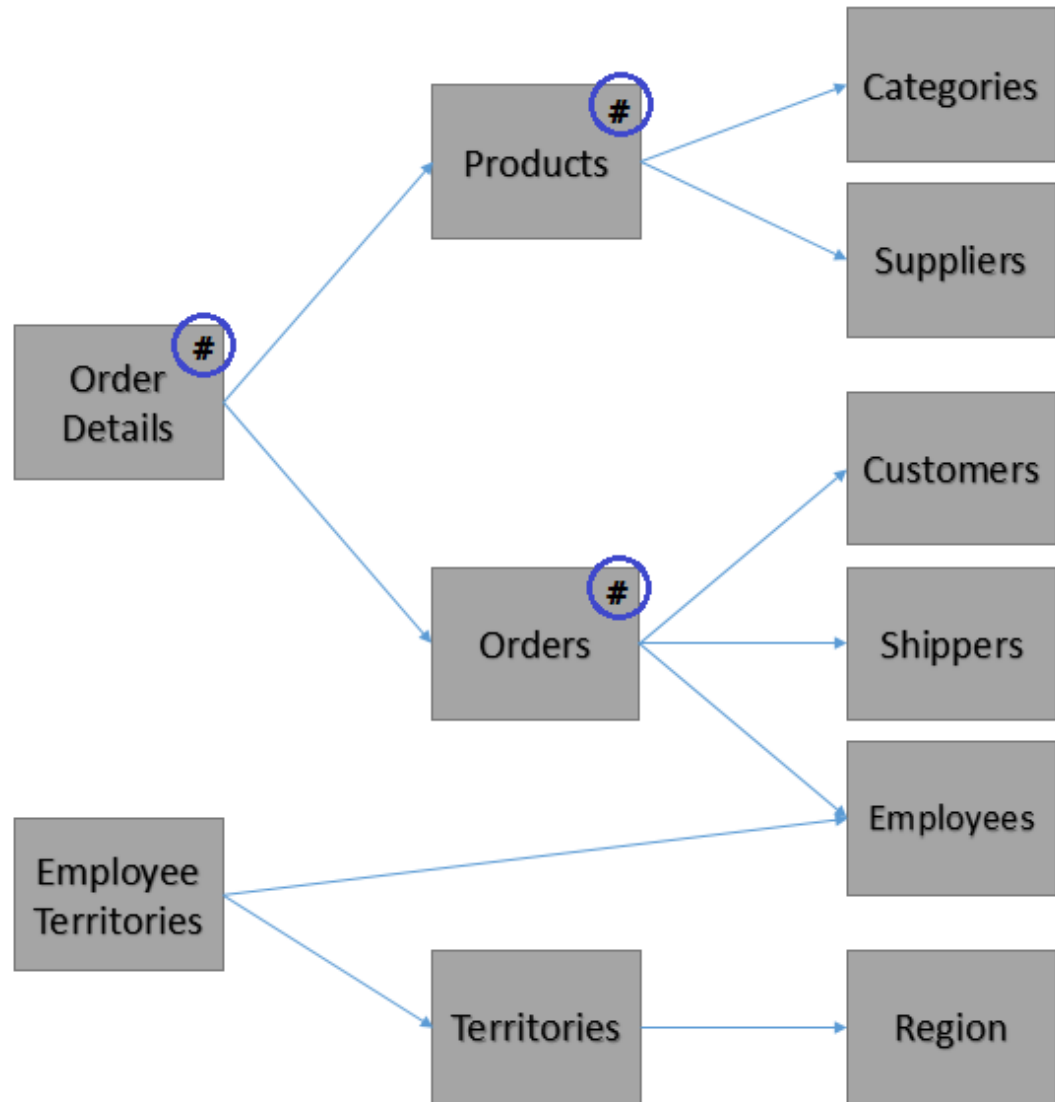
A rectangular box representing the "Passengers" entity. Inside the box, the attribute "PassengerID" is listed.

The ORIENTATION tells us about the Many-To-One relationship

# Northwind from Microsoft



# Northwind ORIENTED



With oriented relationships we can detect the problems.

OK. But how to we PREVENT these problems?



## Intuition n.2: using the UNION

If these two cans  
are tables:



This is a JOIN

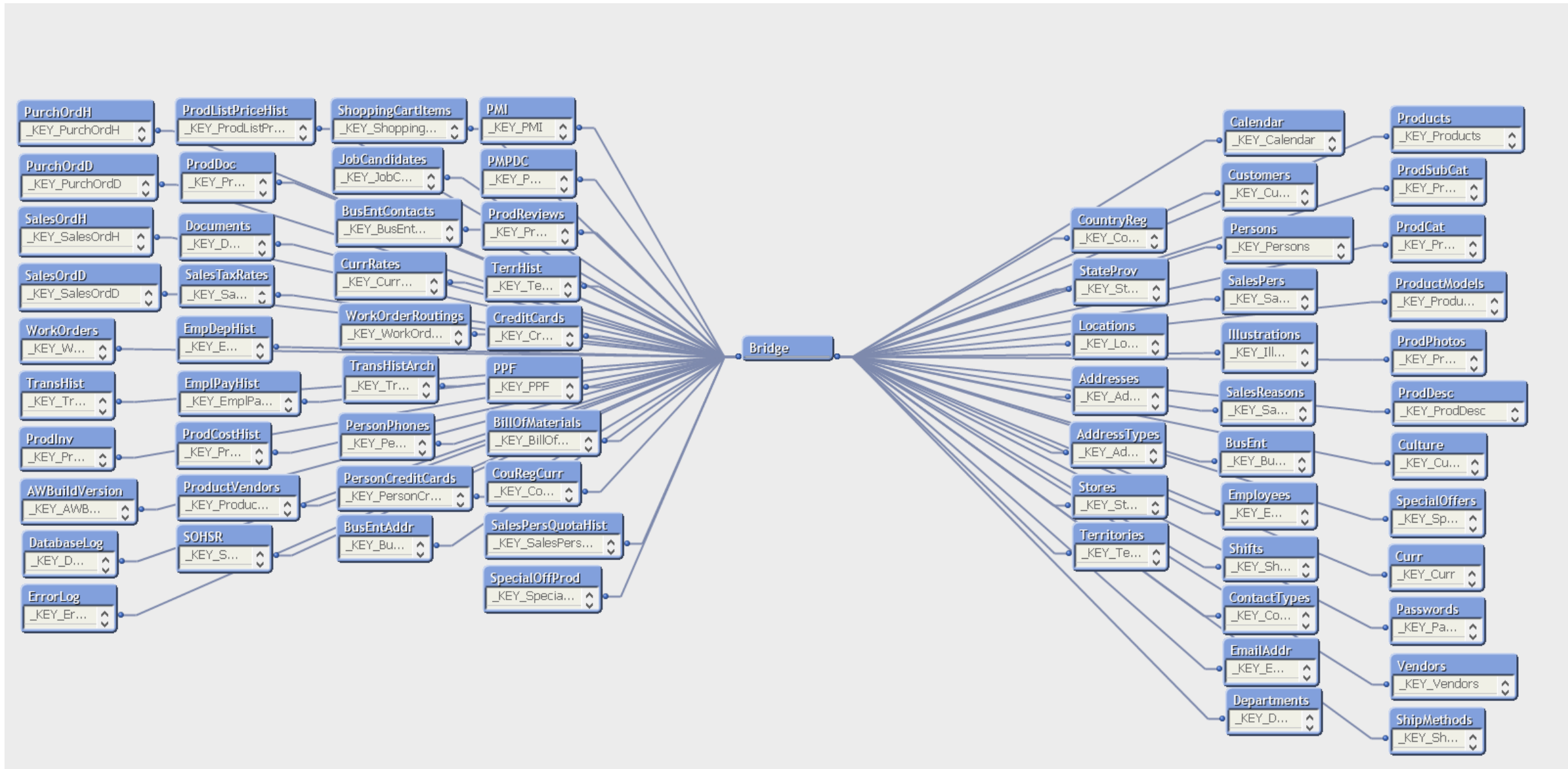


... and this is a  
UNION



LIVE DEMO

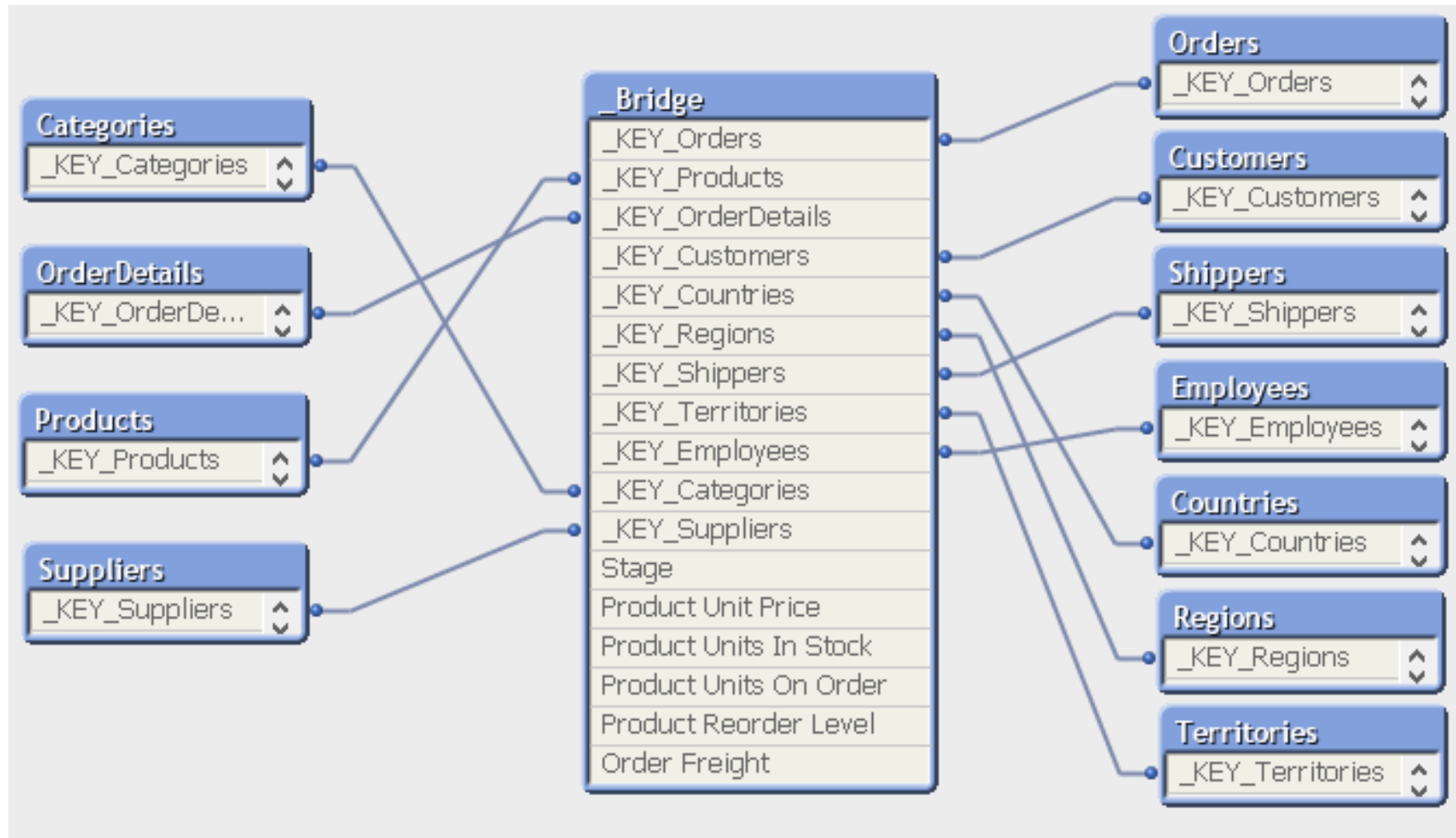
# This is AdventureWorks organized as a USS



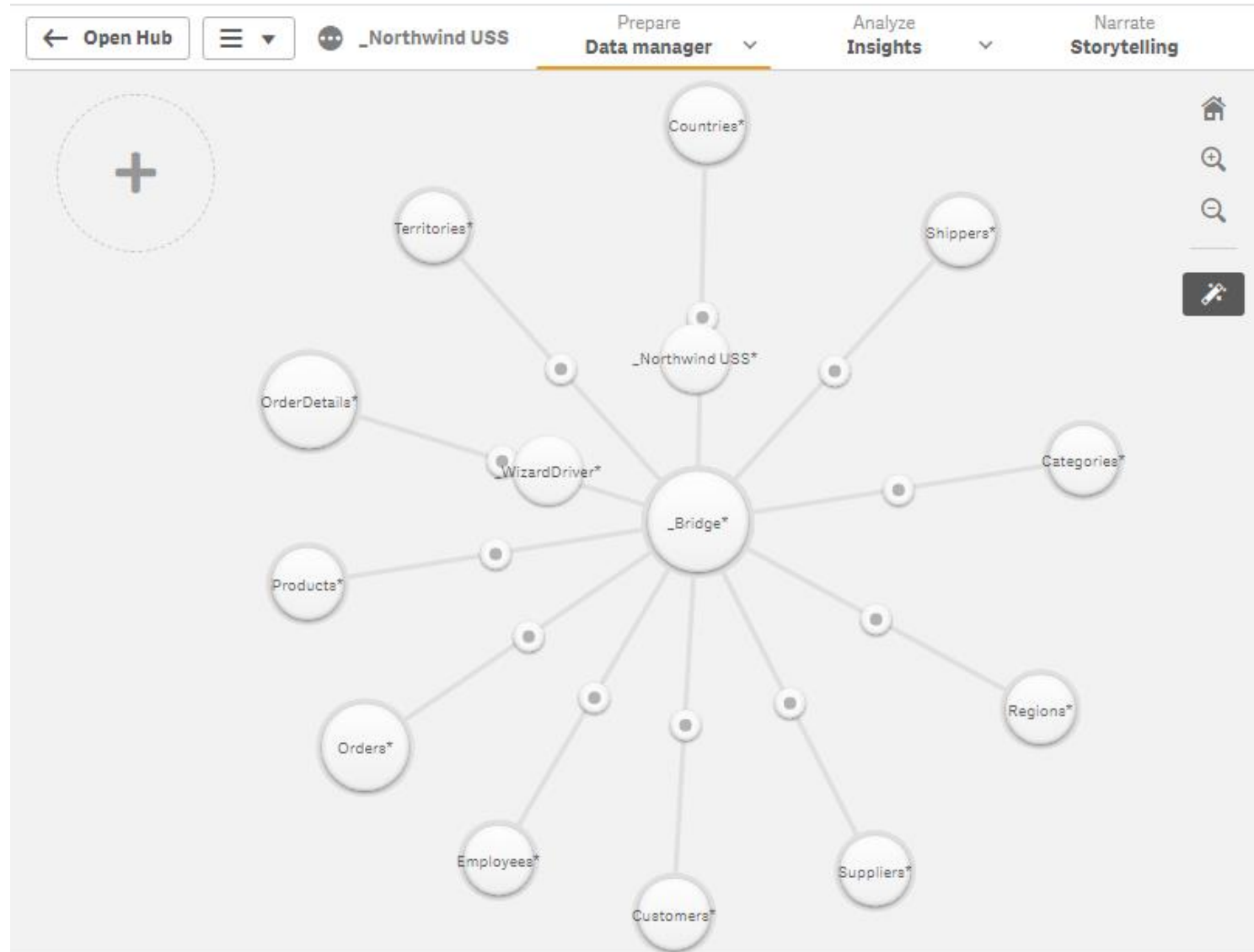
# The USS with Tableau



# The USS with QlikView

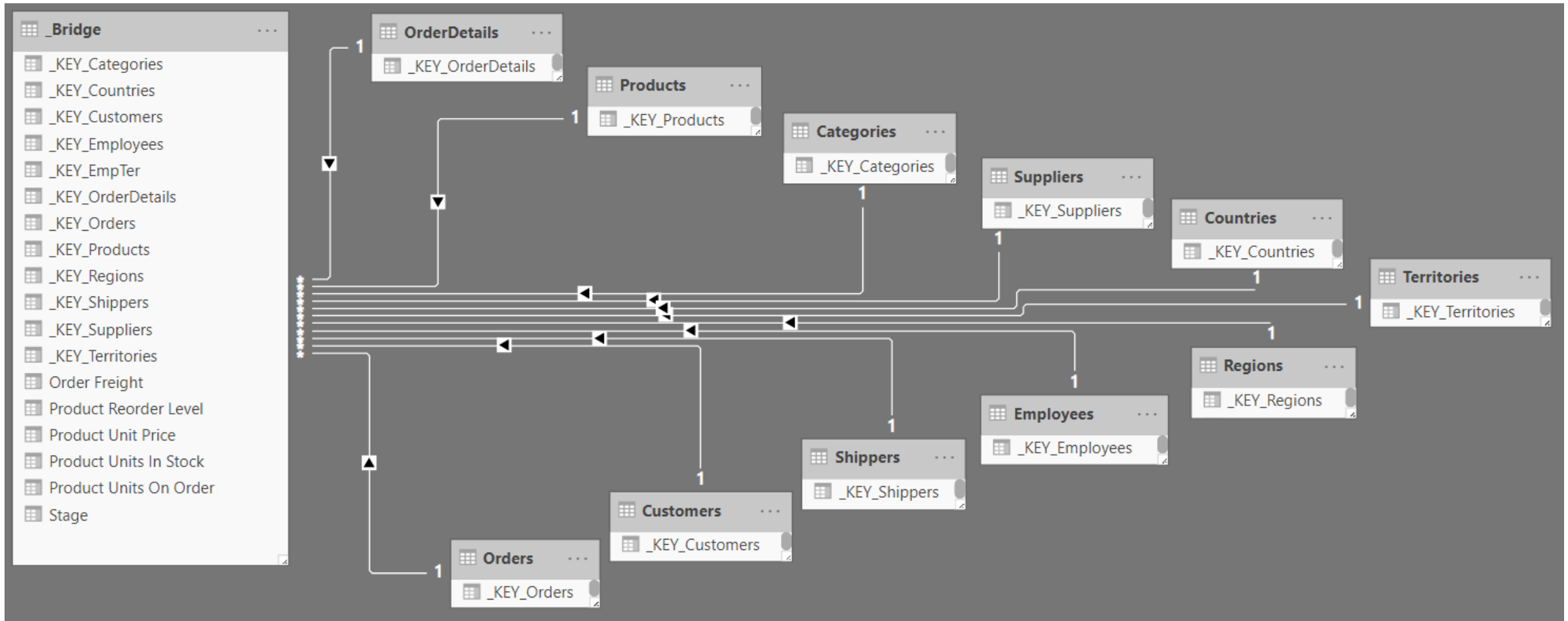


# The USS with Qlik Sense

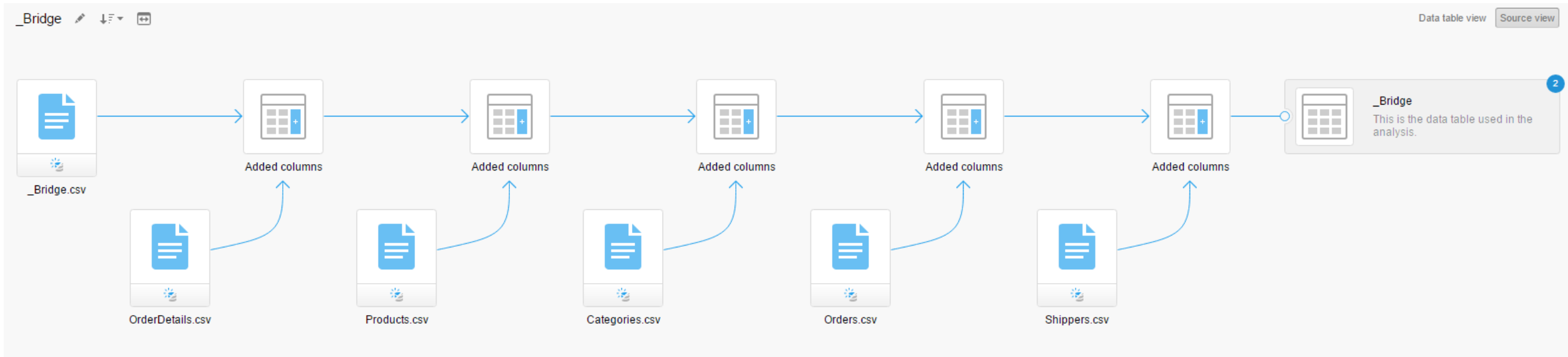




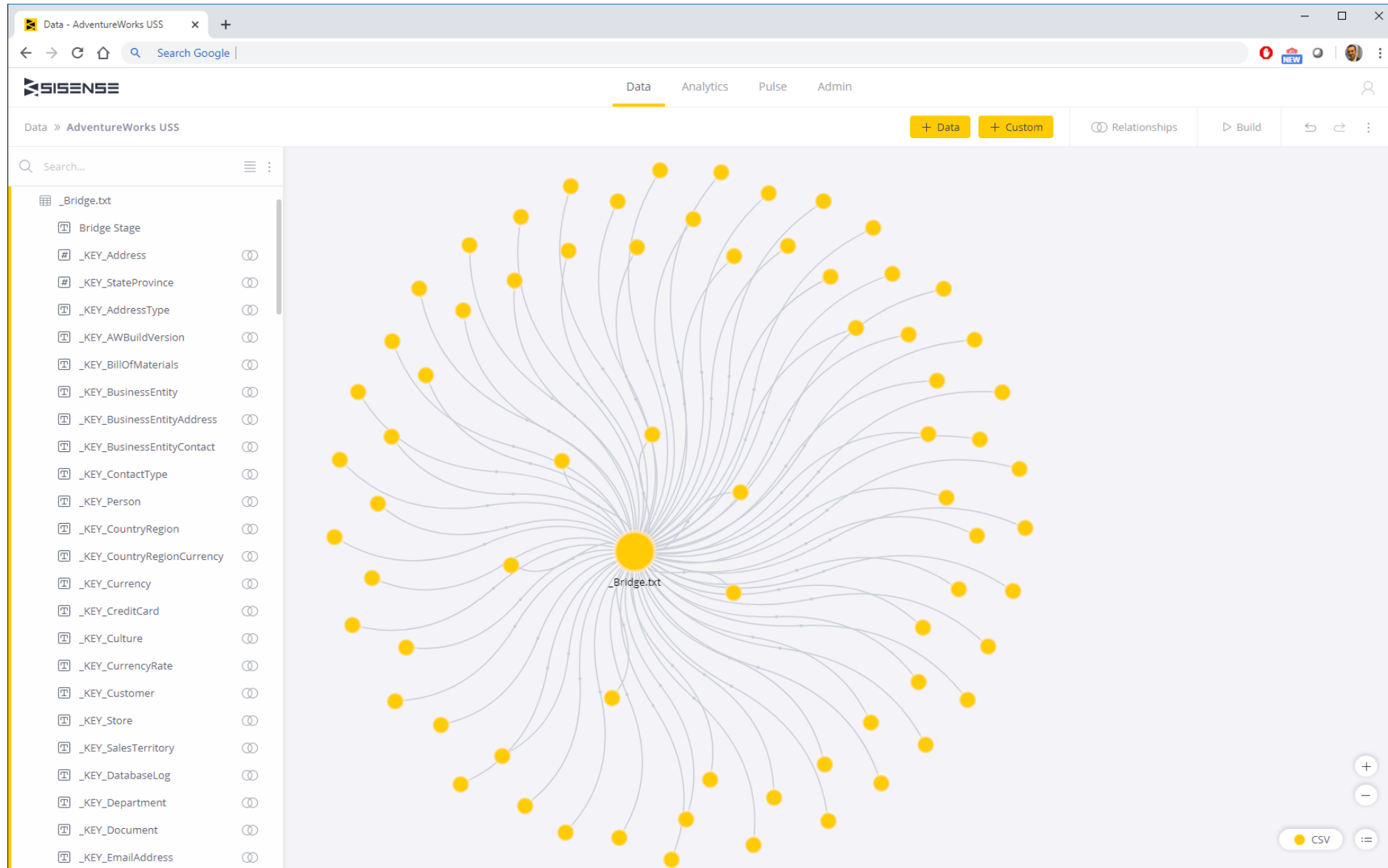
# The USS with Power BI



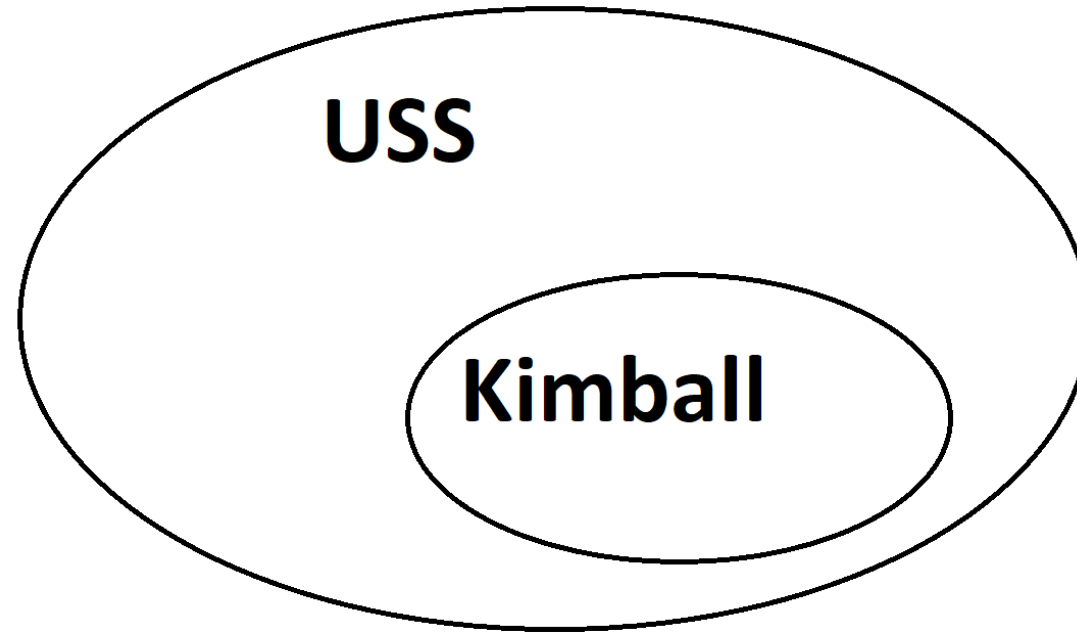
# The USS with Tibco Spotfire



# The USS with Sisense



The Unified Star Schema is an EXTENSION of Kimball's approach



Whatever can be done with the Dimensional Modeling can be done also with the USS. But the USS goes far beyond!

# Chapters of the book

<b>Part I: Architecture</b>	<b>11</b>
<b>Chapter 1: Data Marts and the Dimensional Model</b>	<b>13</b>
<b>Chapter 2: Dimensional Modeling Concepts</b>	<b>23</b>
<b>Chapter 3: Data Mart Evolution</b>	<b>29</b>
<b>Chapter 4: Transformations</b>	<b>35</b>
<b>Chapter 5: The Integrated Data Mart Approach</b>	<b>43</b>
<b>Chapter 6: Monitoring the Data Mart Environment</b>	<b>49</b>
<b>Chapter 7: Metadata and Documentation in the Data Mart Environment</b>	<b>55</b>
<b>Chapter 8: The Evolution toward Integrated Data Marts</b>	<b>61</b>



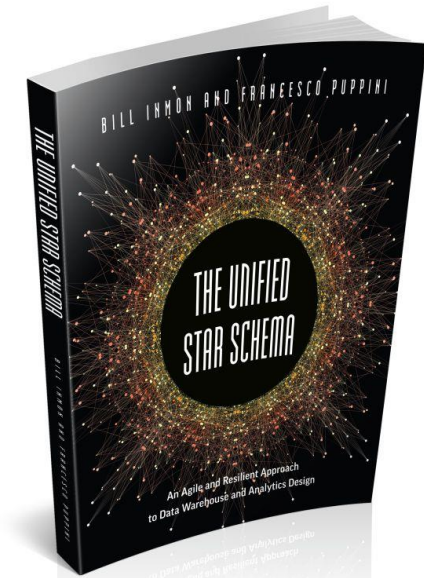
# Chapters of the book

<b>Part II: The Unified Star Schema</b>	<b>67</b>
<b>Chapter 9: Introduction to the Unified Star Schema</b>	<b>69</b>
<b>Chapter 10: Loss of Data</b>	<b>91</b>
<b>Chapter 11: The Fan Trap</b>	<b>115</b>
<b>Chapter 12: The Chasm Trap</b>	<b>143</b>
<b>Chapter 13: Multi-Fact Queries</b>	<b>173</b>
<b>Chapter 14: Loops</b>	<b>199</b>
<b>Chapter 15: Non-Conformed Granularities</b>	<b>219</b>
<b>Chapter 16: Northwind Case Study</b>	<b>239</b>



# THANK YOU!

## Q & A



<https://technicspub.com>