



**FROM XL TO S – REDUCE
YOUR POWER BI MODEL
SIZE BY 90%!**

Nikola Ilic

Data Mozart, Microsoft Data Platform MVP

A big thank you to our partners

 plainwater
de kracht van heldere data

 webdashboard
Powered by In Summa

 sogeti
Part of Capgemini

 dashData
power to your people

 BEILO BI
ZEELAND

 infotopics
apps for power bi

 Motio

 Kimura

 FabriCode

 SIGNON
ICT TRAININGEN

 valcon

 KASPAROV
FINANCE & BI

 ilionx

 creates.

 nine
altitudes

 Tabular Editor
Better Data Models Faster

 raedt-BI

 OKVIZ

 Blauwdruk BI

 THE
DATA
COOKS

 DataShareHub

 easydash

 DTX

 GET
RESPONSIVE

 DataHues.

 Quanto
collective analytics

 ANOTHER
DIMENSION

 Control
Over Reports

 DATAKINGDOM

 ONE
PORTAL

 DataCentral

 dexs

 Barbanson
Data Solutions

 mountdata
guide to impact

 Fellowwind

 DATATAKO

 AXELIO

 MINOVA



Nikola Ilic

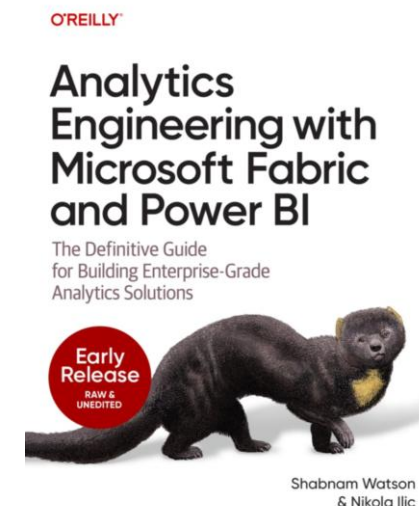
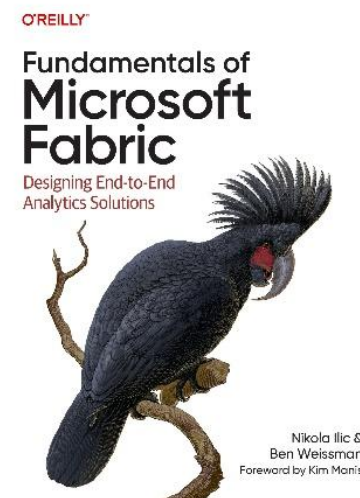
Principal Architect



 data-mozart.com

learn.data-mozart.com

- *I'm making music from the data!*
- Power BI and Fabric addict, blogger, speaker...
- Father of 2, **Barca** & Leo Messi fan...



@DataMozart

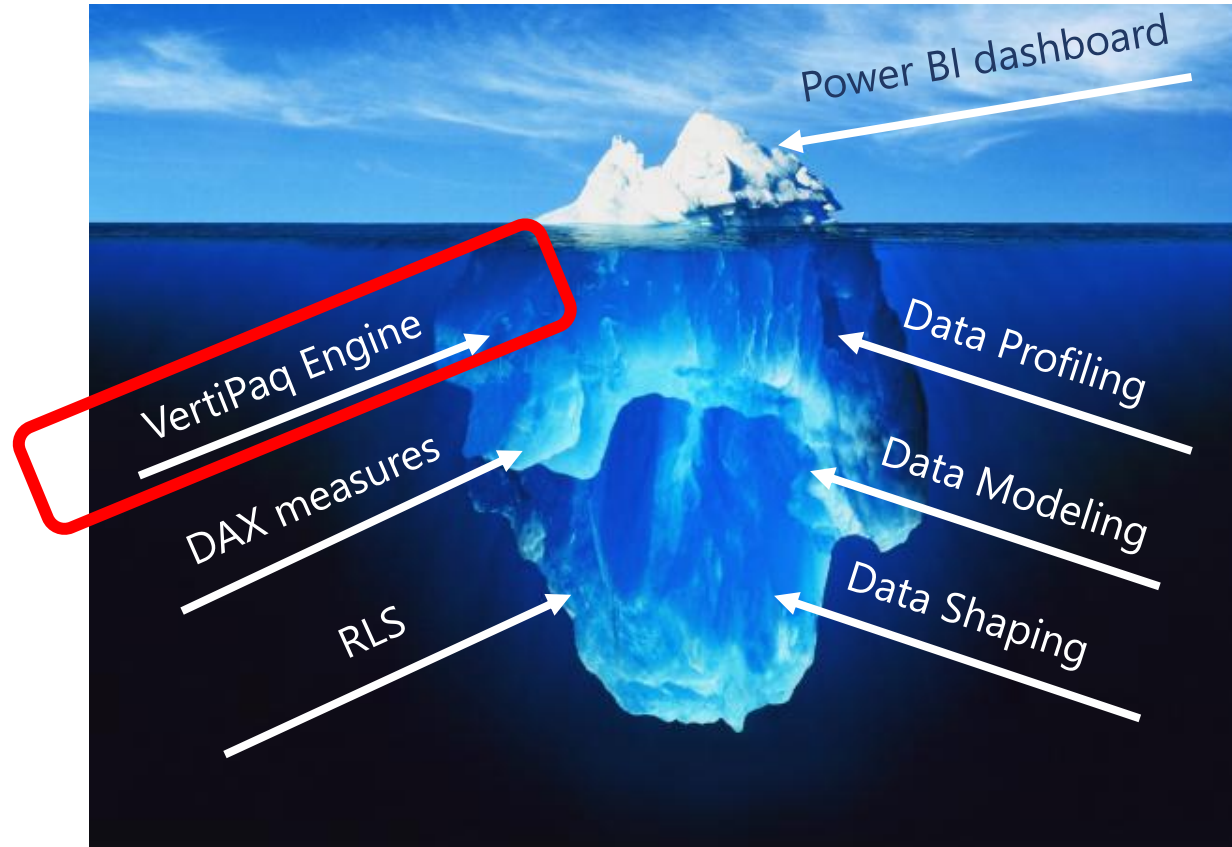


@DataMozart

What to expect today?



Scratching under the surface



Prerequisites

- ✓ Level 300 – 400 Session
- ✓ Knowledge of relational databases
- ✓ Understanding of Power BI and Data Modeling



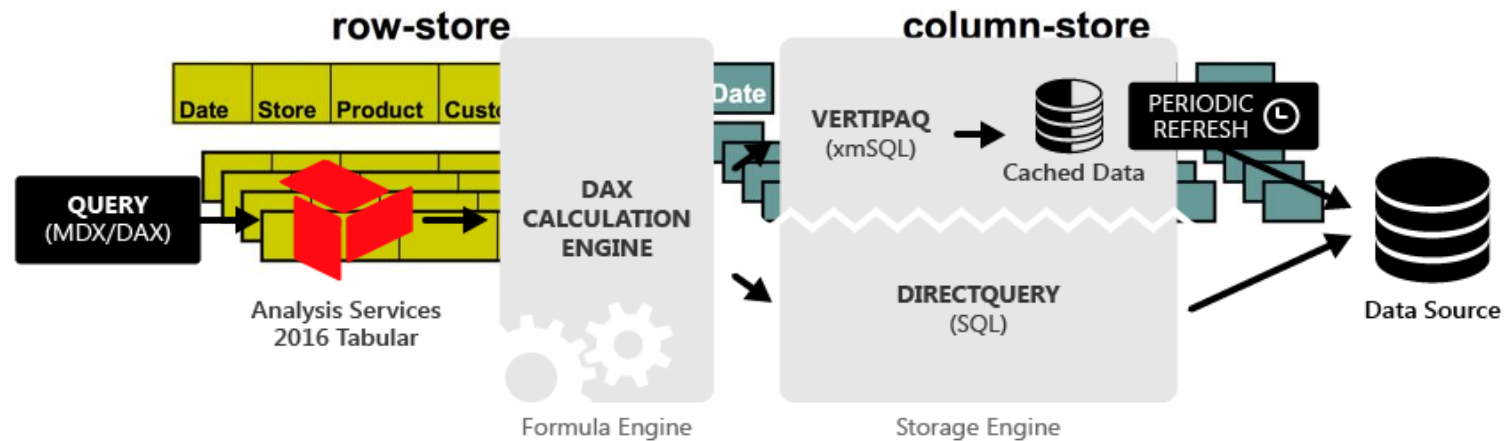
Contents at a glance

1. What is VertiPaq?
2. How does VertiPaq store the data?
3. How can we help VertiPaq to get the optimal data model?
4. Demo



What is a VertiPaq?

- ✓ Row-store vs columnar databases
- ✓ Formula engine vs storage engine

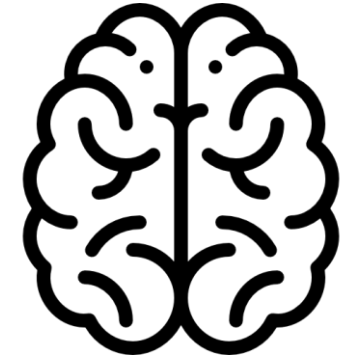


Source: The Definitive Guide to DAX, 2nd Edition (M.Russo & A.Ferrari)



Formula Engine

- ✓ „Brain“ of Power BI!!!
- ✓ Understands DAX and translates it
- ✓ Works in a single thread!



1. ACCEPTS THE REQUEST
2. PROCESS THE REQUEST
3. GENERATES QUERY PLAN
4. EXECUTES THE QUERY



Storage Engine

- ✓ „Muscles“ of Power BI !!!
- ✓ 3 possible options to choose from *
- ✓ Works in multi-thread !



*

1. Import mode

2. Direct Query

3. Dual



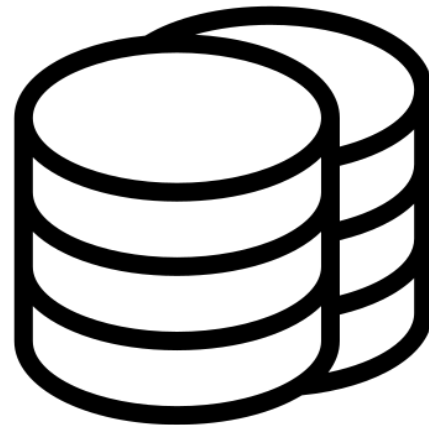
Import Mode

- ✓ Based on VertiPaaS
- ✓ Data stored in-memory as a snapshot
- ✓ Data is being refreshed periodically



Direct Query

- ✓ Data is retrieved from the data source at the query time
- ✓ Data stored in original source before, during & after the query execution



Dual Mode

- ✓ Two versions of the same table
- ✓ Data is in-memory, but can also be retrieved directly from the source



Import Mode in-depth

- ✓ Reads data source
- ✓ Encodes & compresses data
- ✓ Establish dictionary
- ✓ Build relationships
- ✓ Calculated columns & tables



How VertiPaq stores the data

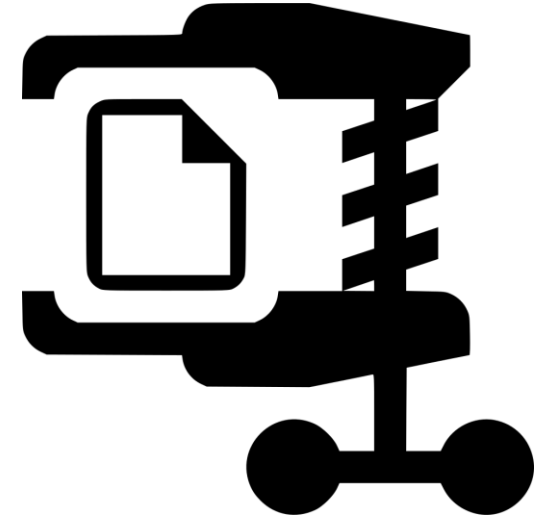
- ✓ VertiPaq is a columnar in-memory database
- ✓ VertiPaq chooses optimal compression algorithms for each column!

ID	Name	Color	Unit Price
1	Camcorder	Red	112.25
2	Camera	Red	97.50
3	Smartphone	White	100.00
4	Console	Black	112.25
5	TV	Blue	1,240.85
6	CD	Red	39.99
7	Touch screen	Blue	45.12
8	PDA	Black	120.25
9	Keyboard	Black	120.50



Encoding Types

1. Value Encoding
2. Hash Encoding (Dictionary)
3. RLE (Run-Length-Encoding)



Value Encoding

- ✓ The most desirable
- ✓ Works exclusively with integers

Phone Calls	Difference
4100	100
4500	500
4200	200
4500	500
4000	0
5000	1000
4700	700
4500	500
4700	700
5000	1000
4000	0
4100	100

Max Value = 5000 (13 bits)

Max Value = 1000 (10 bits)

123



Hash Encoding

- ✓ The most frequent
- ✓ VertiPaq creates a dictionary!

Subjects	Index	Index	Subjects
Complaint	0	0	Complaint
Bonus Minutes	1	1	Bonus Minutes
Complaint	0	2	New Contract
Complaint	0	3	Termination
New Contract	2		
Complaint	0		
New Contract	2		
Termination	3		
Bonus Minutes	1		
Termination	3		
Complaint	0		
Complaint	0		



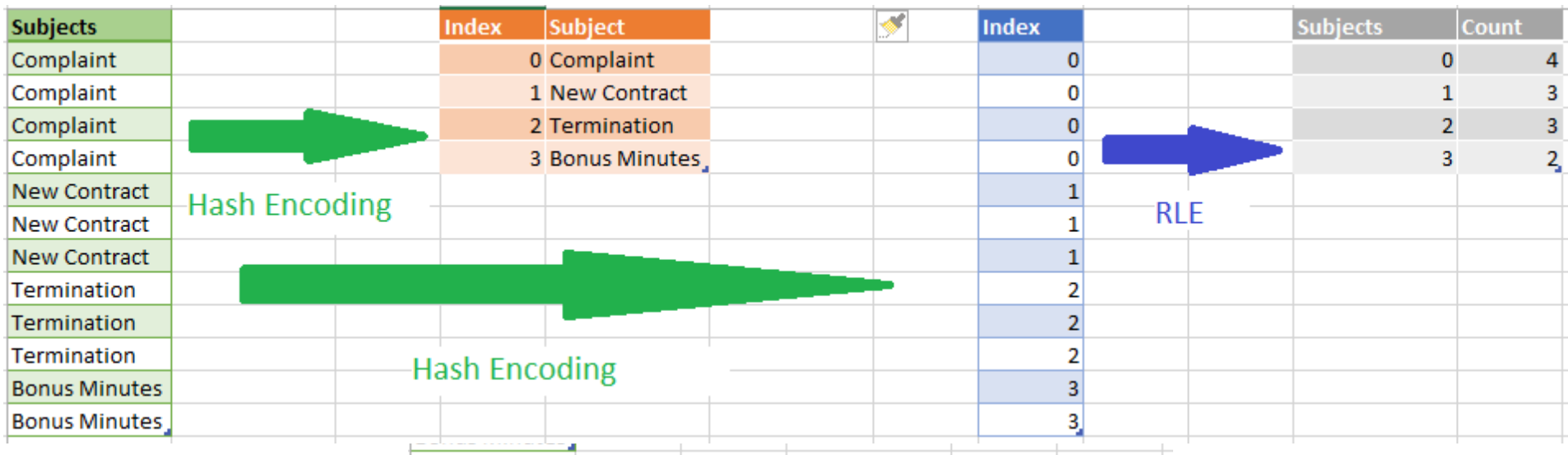
Hash Encoding

- ✓ VertiPaq is practically **data type independent!**
- ✓ Data type affects column size – **it's a myth!!!**
- ✓ **Cardinality** is the main factor

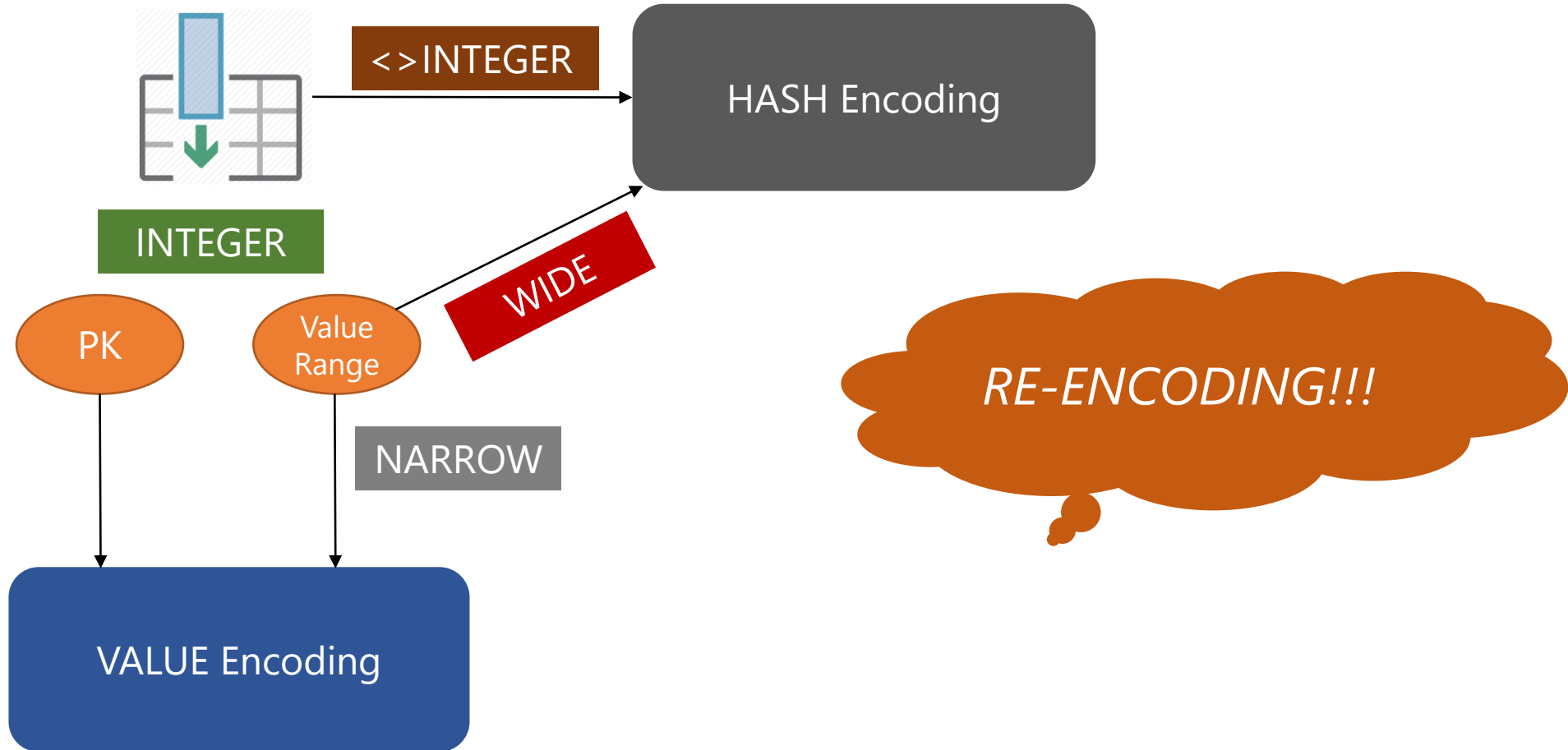


RLE (Run-Length-Encoding)

- ✓ Mapping table with ranges
- ✓ Highly dependent on the ordering within column
- ✓ RLE always comes after Hash encoding!

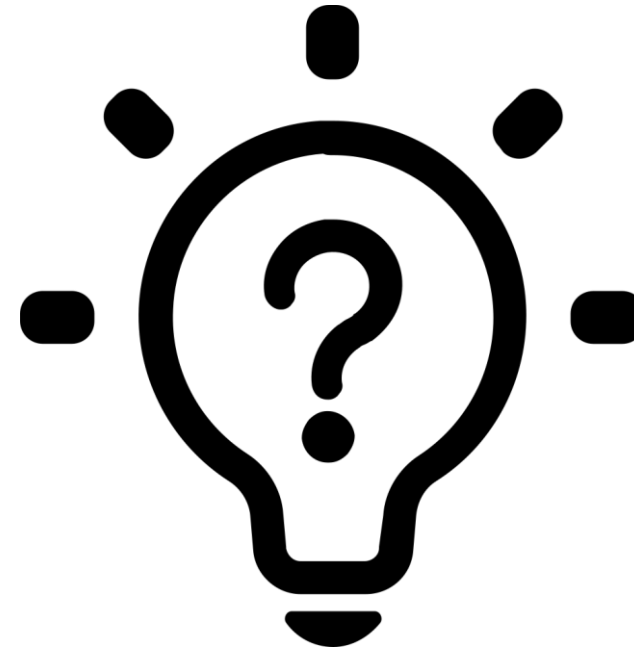


How VertiPaq "thinks"



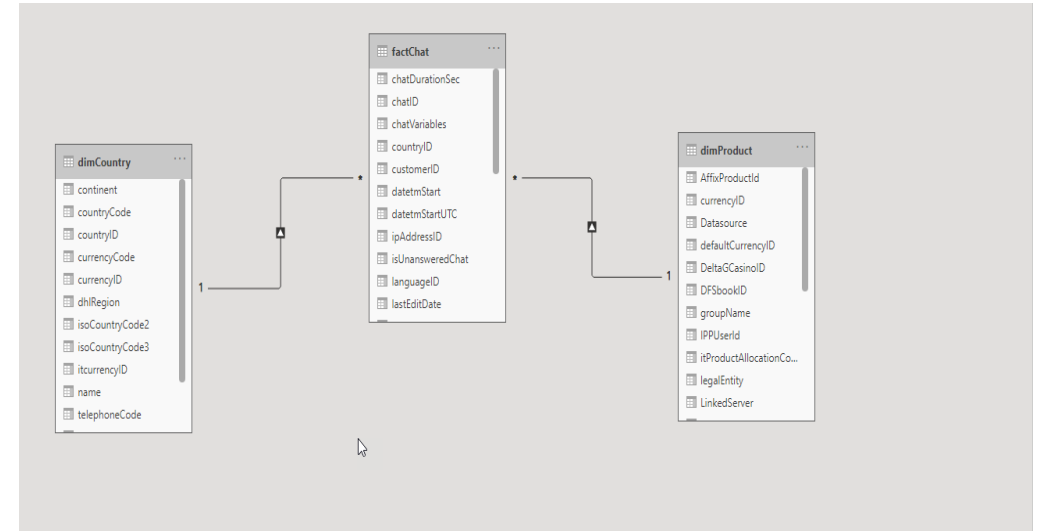
Choosing optimal algorithm

- ✓ Cardinality
- ✓ Data distribution
- ✓ Number of rows
- ✓ Column data type



Understanding relationships

- ✓ Fast transfer of filters between the tables
- ✓ Costs depend on cardinality!
- ✓ Aggregations can help



Understanding materialization

- ✓ Related to columnar databases
- ✓ Datacache temporary storage
- ✓ Late vs Early materialization



Understanding aggregations

- ✓ Help the Storage Engine to work faster
- ✓ Reducing the number of columns and rows
- ✓ NOT for optimizing DAX calculations!!!

7.346 rows

	Date	subject	userID
1	2017-07-10	Abandoned chat	4537
2	2017-07-10	Cashback	4453
3	2017-07-10	Cashback	4392
4	2017-07-10	Account query	4396
5	2017-07-10	Deposit assistance	4493
6	2017-07-10	No deposit bonus	4493
7	2017-07-10	Deposit bonus	3782
8	2017-07-10	Deposit bonus	4582
9	2017-07-10	Cashback	4493
10	2017-07-10	Free bonus	4392
11	2017-07-10	No deposit bonus	4537
12	2017-07-10	Account query	419
13	2017-07-10	Cashback	4453
14	2017-07-10	Account query	4396
15	2017-07-10	Free bonus	4392
16	2017-07-10	Deposit bonus	4582
17	2017-07-10	Game query	4178
18	2017-07-09	No deposit bonus	4514
19	2017-07-09	Cashback	527



	Date	subject	totalChats
1	2017-07-10	Abandoned chat	447
2	2017-07-09	Abandoned chat	286
3	2017-07-10	Cashback	554
4	2017-07-10	Account query	457
5	2017-07-09	Cashback	448
6	2017-07-09	Casino	33
7	2017-07-09	Deposit assistance	151
8	2017-07-09	Deposit bonus	396
9	2017-07-09	Account query	393
10	2017-07-10	Comp Bonus	112
11	2017-07-10	Documents	25
12	2017-07-10	Comp Points	1
13	2017-07-09	Free bonus	556
14	2017-07-10	Deposit bonus	389
15	2017-07-09	Documents	40
16	2017-07-09	Game query	80
17	2017-07-09	No contact	301
18	2017-07-09	No deposit bonus	523
19	2017-07-10	Successful Deposit	5

45 rows



DEMO



Let's make PIZZA...

- Pizza Bread
- Tomato sauce
- Ham
- Cheese



INGREDIENTS





Let's go shopping...



Summary

- ✓ Keep only necessary columns !!!
- ✓ Reduce the column cardinality
- ✓ Keep only necessary rows
- ✓ Aggregate when possible
- ✓ Avoid calculated columns!
- ✓ Use proper data types
- ✓ Disable Auto Date/Time





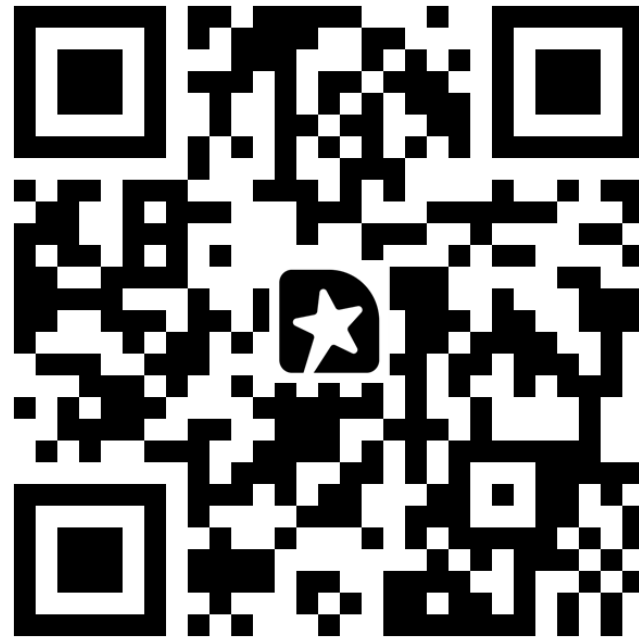
Data-mozart.com



@DataMozart

QUESTIONS

Loved it? Learned something? Tell us!
Share your feedback in just 1 minute



From XL to S - Reduce your Power BI model size by 90%!





THANK YOU!



Data-mozart.com



[@DataMozart](https://twitter.com/DataMozart)