VertiPaq - Engin stojící za Power BI a SSAS Tabular

Roman Lánský

Roman@joyfulbi.com

@rlany

VertiPaq

- Basics and Concepts
  - SSAS Tabular Overview
  - Column Store Principles
- Architecture
  - Data Load
  - Querying
  - Development Process

Basics

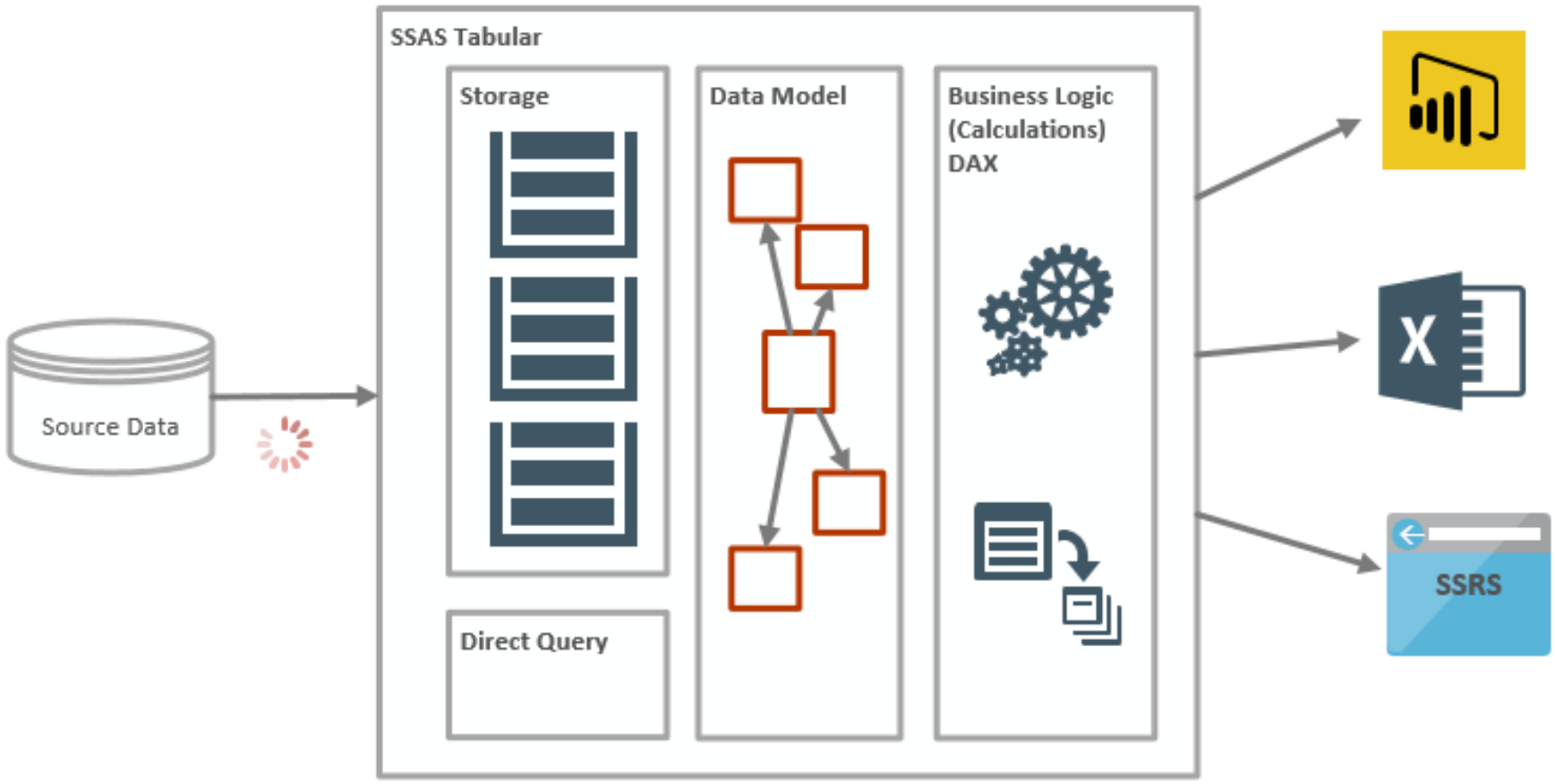
Four Flavors of Same Beast

- SSAS Tabular = Power BI = Excel Power Pivot = AAS (SSAS Azure)
- Power BI a lot wider ecosystem around VertiPaq In memory engine.
  - Cloud + Desktop + Gateway + SSRS Integration …
- Excel Power Pivot
  - VertiPaq extended with Excel features. (Size is very limited in 32b version, Release cycle sticks to Excel so it’s slow)
- SSAS Tabular – adding features from PowerBI

SSAS Tabular

- SSAS two modes
  - SSAS Multidimensional
  - SSAS Tabular (VertiPaq) – adopted from PowerBI
- Model + Data + Query Language + ETL
- In Memory
  - Drive is not your bottleneck anymore, memory is!
  - Data compressed.
- MDX Compatible

SSAS Tabular Basic Architecture



Storage

Columns Store vs Row Store

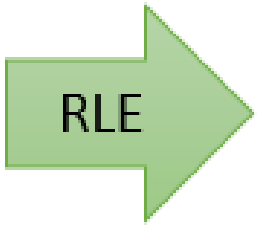
How Many Stock Do We Have in Each Category?

How Is Data Stored?

ID	Product Name	Stock Count	Unit Price USD	Color	Category	Stock Amount USD
48	Beanie	44	20	Red	Accessories	880
58	Belt	5	65	Brown	Accessories	325
60	Cap	66	18	Yellow	Accessories	1188
62	Sunglasses	7	90	Brown	Accessories	630
64	Hoodie with Pocket	7	45	Gray	Hoodies	315
66	Hoodie with Zipper	89	45	Green	Hoodies	4005
68	Long Sleeve Tee	0	25	Green	Tshirts	0
70	Polo	554	20	Blue	Tshirts	11080
76	V-Neck T-Shirt	32	20	Red	Tshirts	640

Columns Store Compression - RLE

Row	Color
1	Blue
2	Blue
3	Blue
4	Blue
5	Green
6	Green
7	Red
8	Red
9	Red



Color	Start	Count
Blue	1	4
Green	5	2
Red	7	3

Columns Store Compression - RLE

Row	Color
1	Blue
2	Blue
3	Blue
...	...
400	Blue
401	Green
...	...
666	Green
667	Red
...	...
1020	Red

Color	Start	Count
Blue	1	400
Green	401	266
Red	667	354

Columns Store Compression - RLE

Row	Color
1	Blue
2	Blue
3	Blue
...	...
400	Blue
401	Green
...	...
666	Green
667	Red
...	...
1020	Red

HASH

Color	ColorID
Blue	1
Green	2
Red	3

Color
1
1
1
...
1
2
...
2
3
...
3

RLE

Dictionary

ColorID	Color
1	Blue
2	Green
3	Red

Data

ColorID	Count
1	400
2	266
3	354

Columns Store Compression - RLE

Sort By Color and Category

Row	Color	Category
1	Blue	Accessories
2	Blue	Accessories
3	Blue	Hoodies
4	Blue	Hoodies
5	Green	Accessories
6	Green	Accessories
7	Red	Hoodies
8	Red	Tshirts
9	Red	Tshirts

Color	Start	Count
Blue	1	4
Green	5	2
Red	7	3

Category	Start	Count
Accessories	1	2
Hoodies	3	2
Accessories	5	2
Hoodies	7	1
Tshirts	8	2

Columns Store Compression - RLE

Ordered by Category and Color

Row	Color	Category
1	Blue	Accessories
2	Blue	Accessories
3	Green	Accessories
4	Green	Accessories
5	Blue	Hoodies
6	Blue	Hoodies
7	Red	Hoodies
8	Red	Tshirts
9	Red	Tshirts

Category	Start	Count
Accessorie	1	4
Hoodies	5	3
Tshirts	8	2

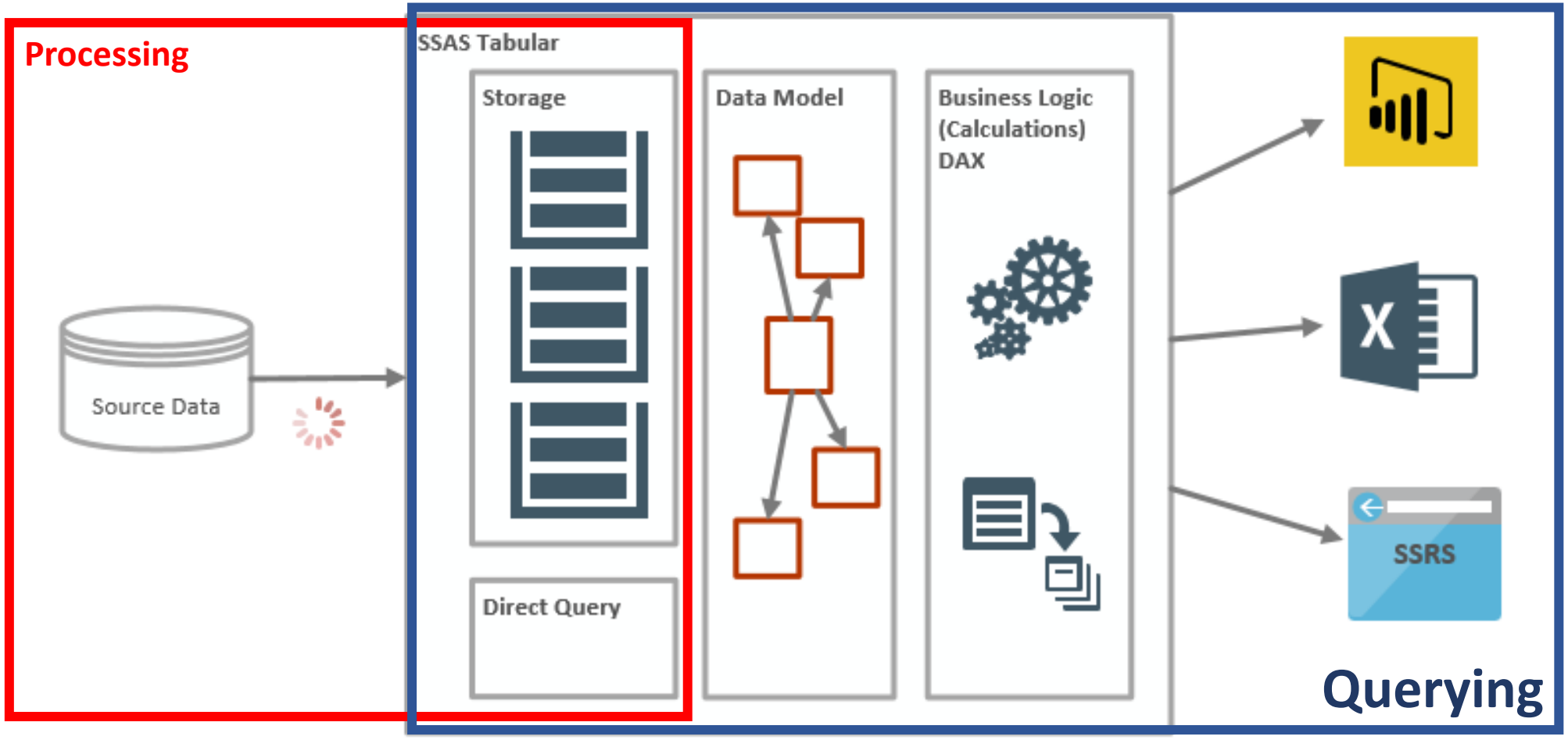
Color	Start	Count
Blue	1	2
Green	3	2
Blue	5	2
Red	7	3

Columns Store Compression - Summary

- **Data Type Size of the Column** – impact only on size of dictionary
- **Distribution of Values in the Column** – repeated values will be more compressed.
- **Combination of the values across columns matters**
- **Sort order might have a big impact**
- **Happens in Segments**

Architecture

SSAS Tabular Basic Architecture

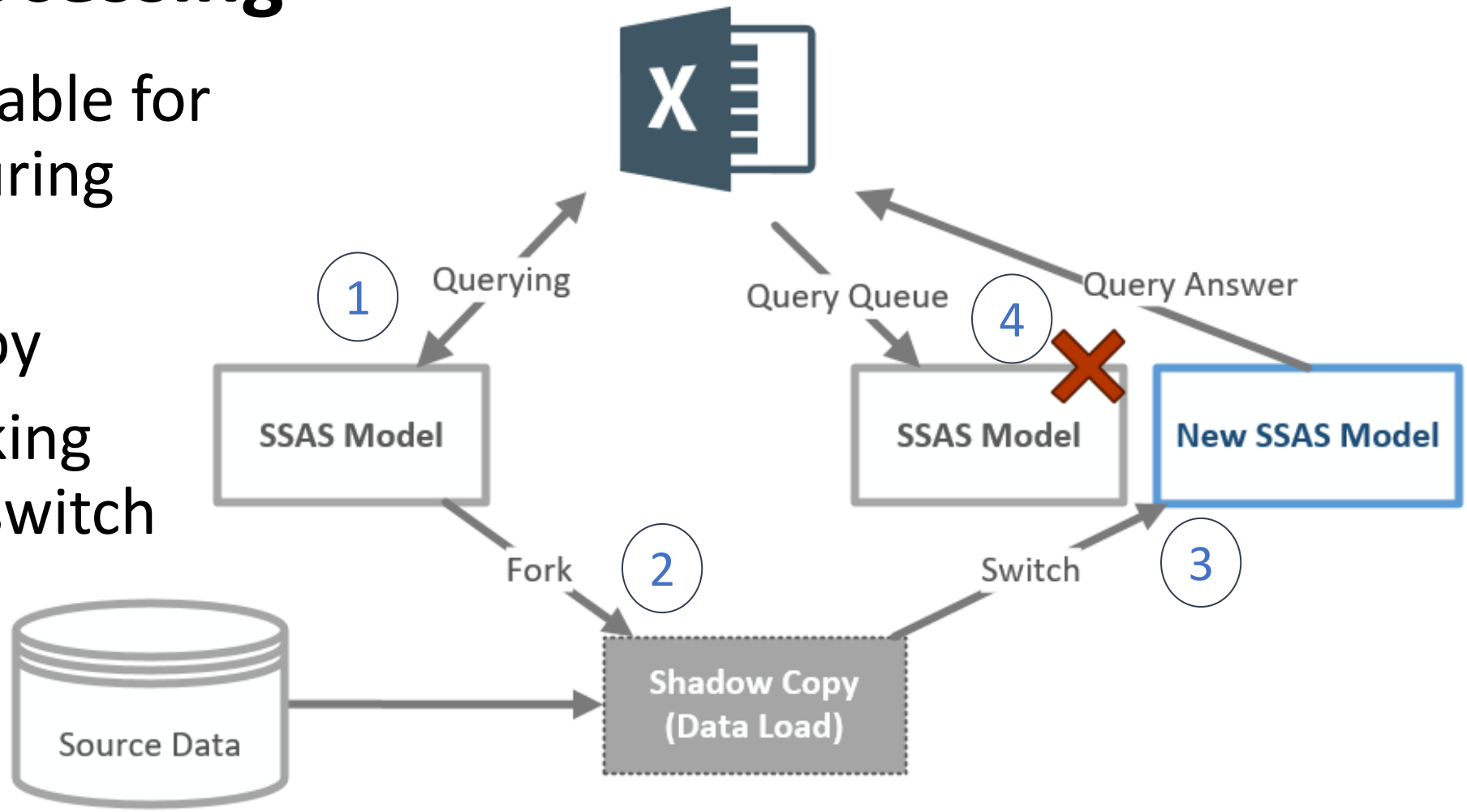


Loading of the data

Model size
is not
memory consumption!!

Data Processing

- Model available for querying during data load
- Shadow copy
- Query blocking during the switch



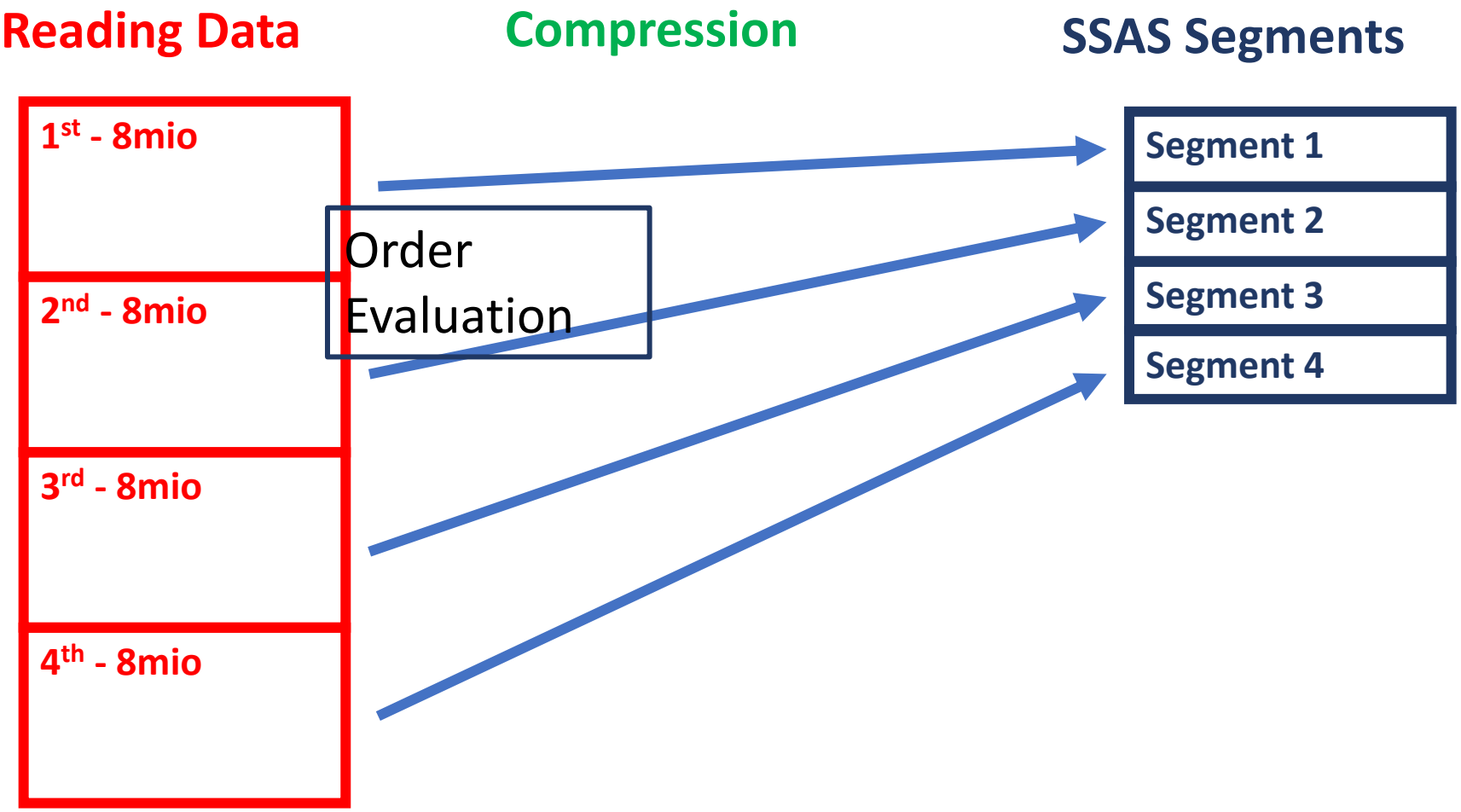
SSAS Tabular Loading Data

- Data Size

Rows	3,000,000
Int Columns	3
Float Columns	3
Nvarchar ()	50
Data size Bytes	9,108,000,000
Data Size GB	8.48

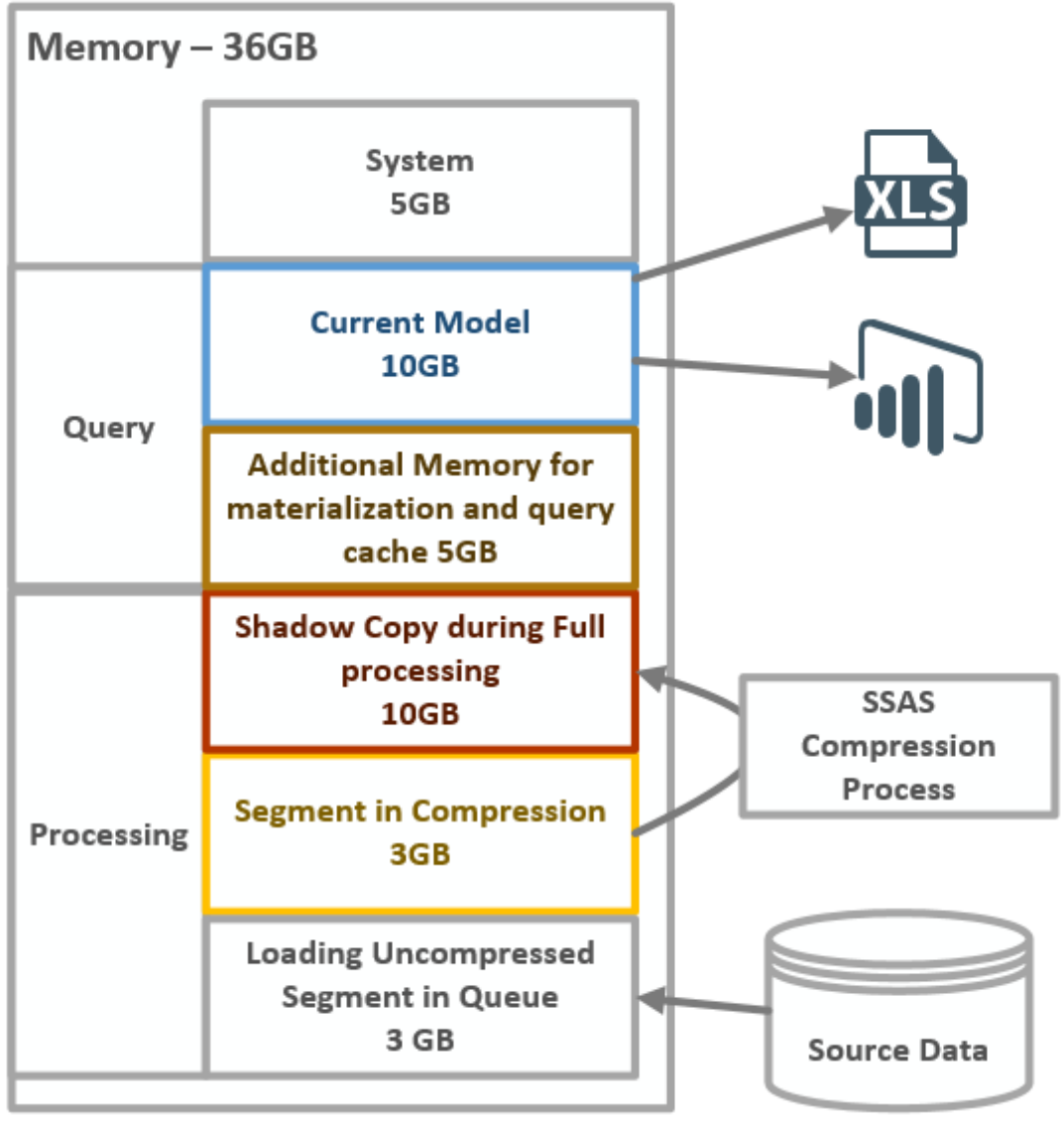
DataType	Bytes	Info
Int	4	4 or 8 Bytes
float	8	Bytes
varchar	1	Byte per character
nvarchar	2	Bytes per character

SSAS Tabular Loading Data

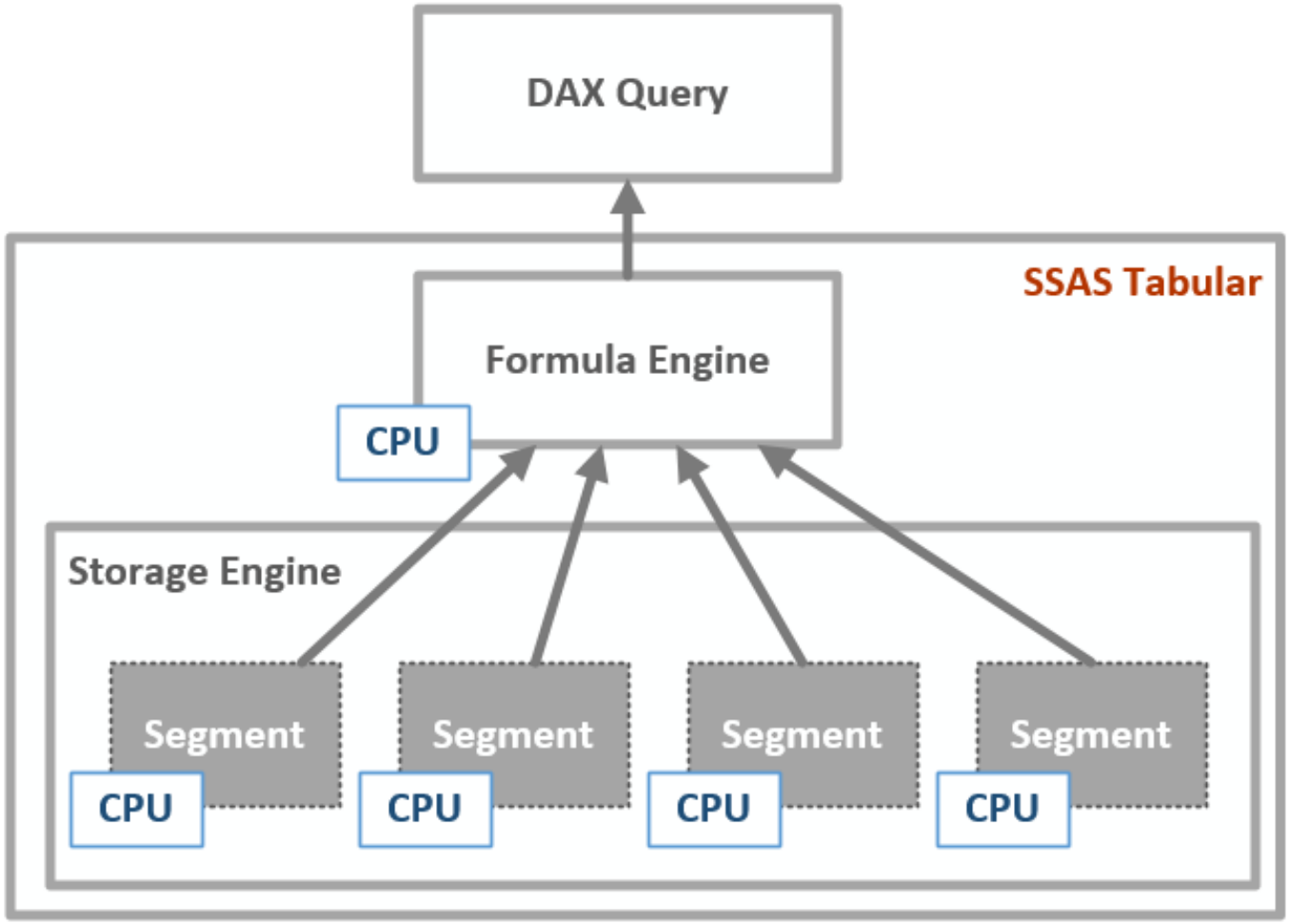


Data Processing Memory

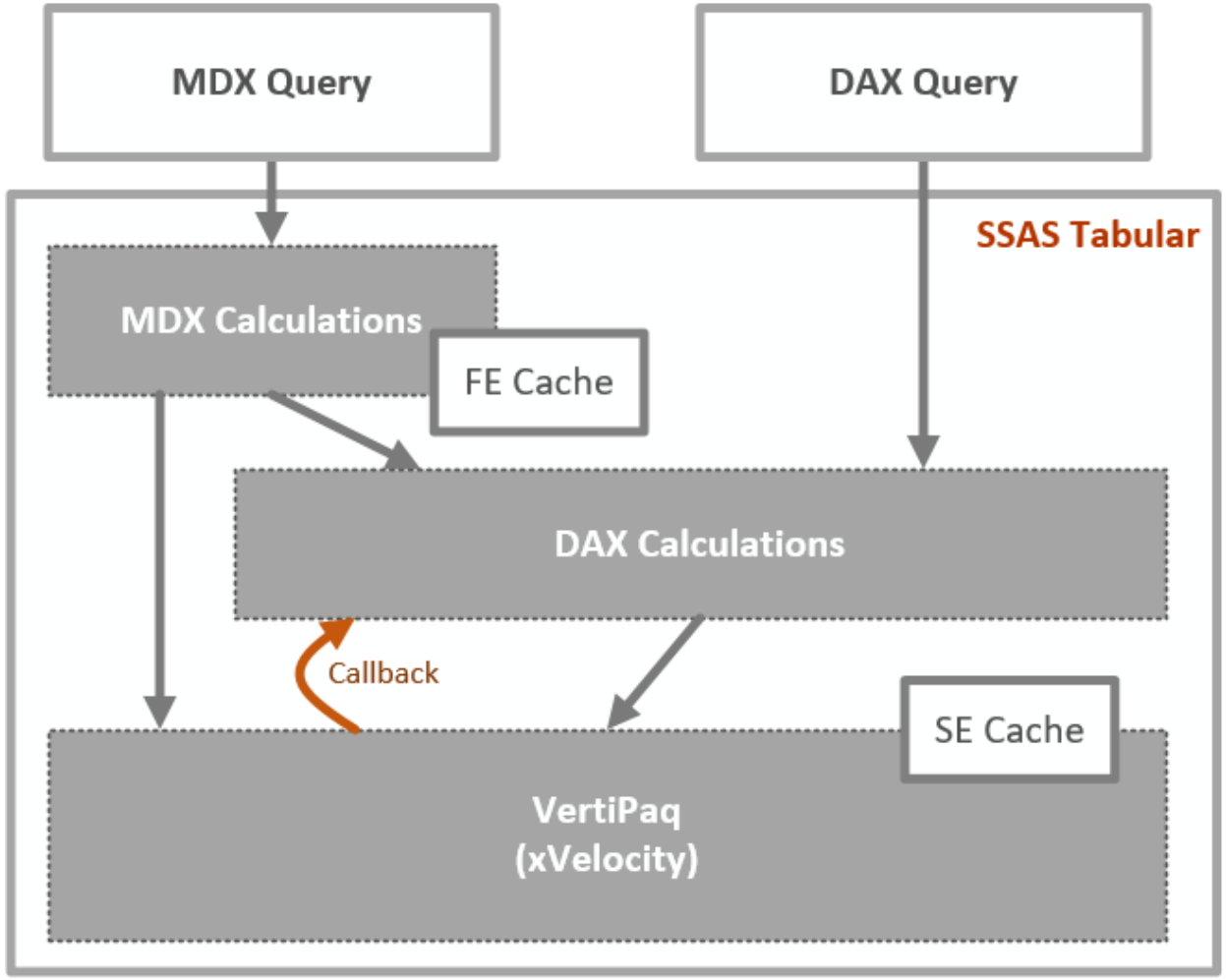
- Process for one partition, If you have more you have in parallel more segments in compression
- Starts with twice size of the segment
- One processing Thread = Two uncompressed segments in memory
- Query memory for materialization – can be at the end bigger than model itself



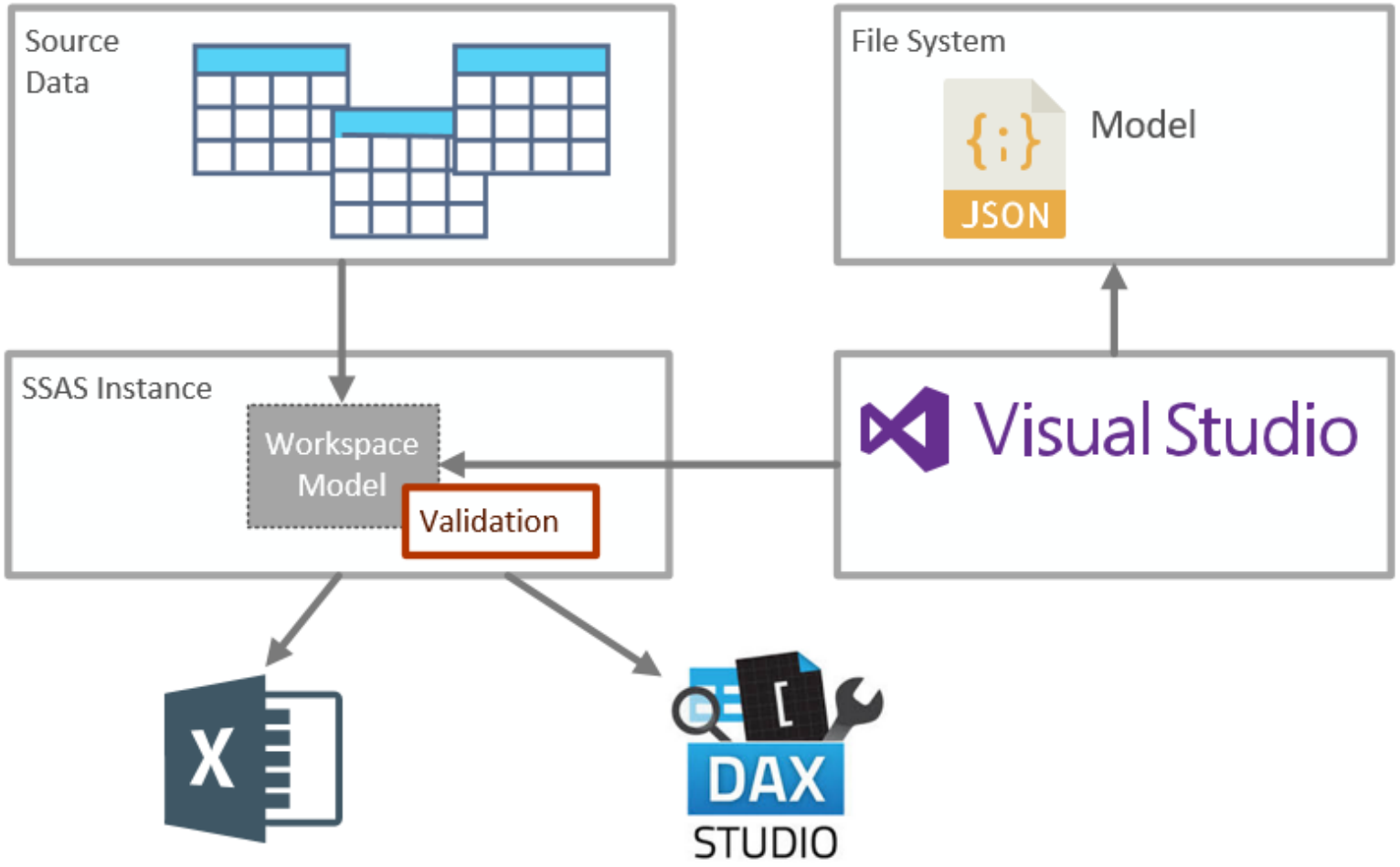
Query Processing



Query Processing



Development Process for SSAS



SSAS Best Practices for Development

- Use Tabular Editor for PB as well now!
- View layer
  - Name of columns = Business names
  - Fact table – TOP X for Dev purpose
  - Clear Contract
- No Calculated columns if not necessary (lookup to smaller table)

SSAS Best Practices for Development

- Know your data, model and users – better knowledge better optimization
- Star model is your target
  - Simple = better performance
- Be careful about big dimensions
- Be careful with Bi-directional filters (can have side effects)
- Size matters

Development

- Dax Studio
  - Model Information
  - Measures development
  - Performance tuning
- Tabular Editor
  - Basic functions
  - Scripting
- Dax formatter

Questions?

Presenter

- MSBI Consultant & SSAS Enthusiast & Blogger
- Email: Roman.Lansky@JoyfulCraftsmen.com
- Twitter: @rlany
- Blog: http://joyfulcraftsmen.com/blog/