tableau Day

BNP PARIBAS

# Advanced Workshop:
# Designing Efficient Workbooks

Chrystelle| Sales Consultant

17/10/2019


My workbook
is slow!
Why ???

# Best Practices for Designing Efficient Tableau Workbooks

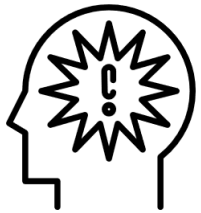
Tableau 10 Edition

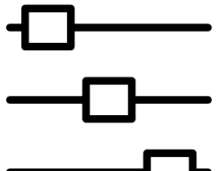
Alan Eldridge
Tableau Software

https://www.tableau.com/learn/whitepapers/designing-efficient-workbooks

# In a Nutshell…

Performance is *not* an after-thought





There is no one “Silver Bullet” …

If it isn’t fast in database, it won’t be fast in Tableau

If it isn’t fast in desktop, it won’t be fast in Server

Design Choices correlate to Performance

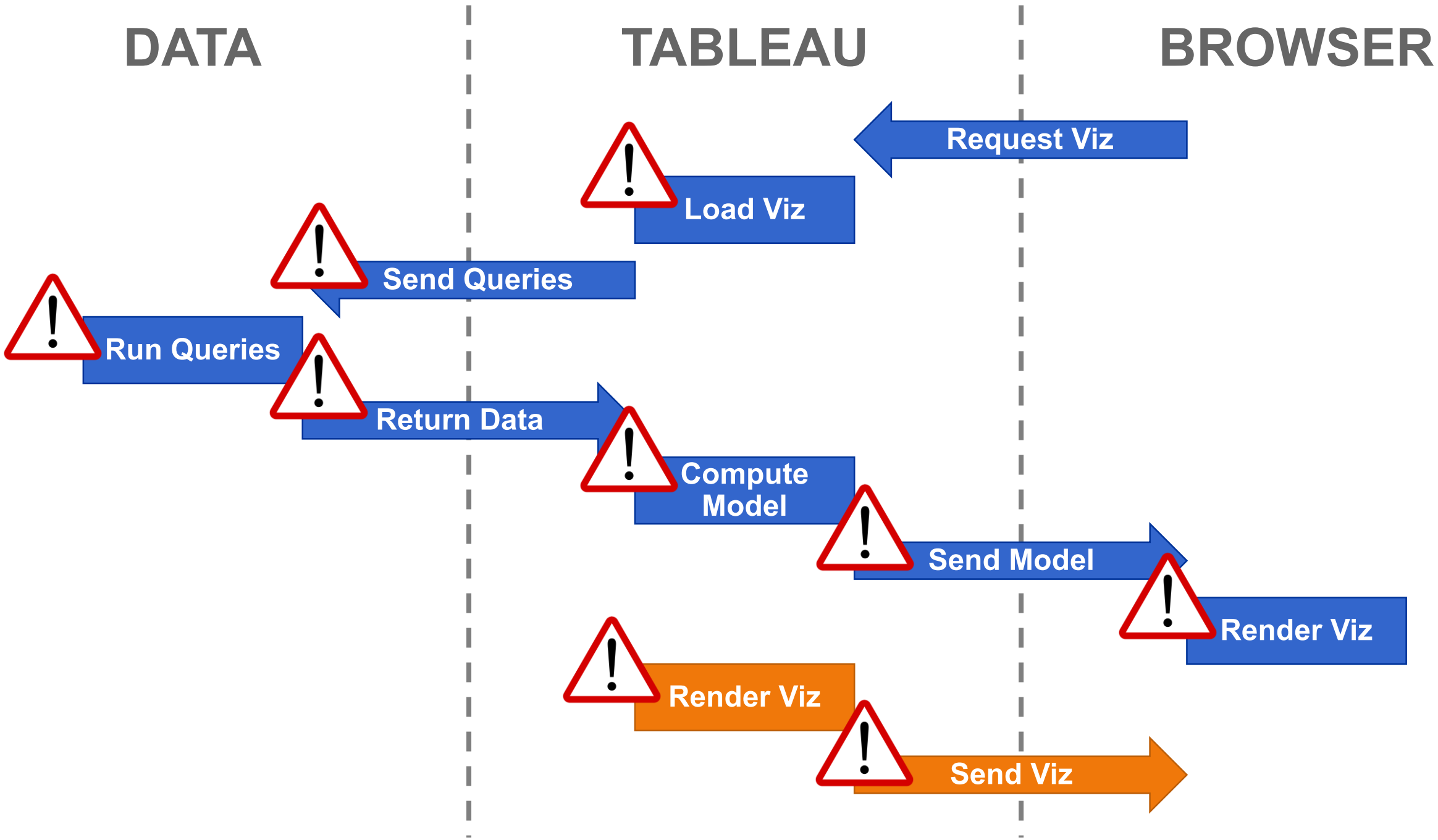
**Interpretation:**

[*“Avoid”*, *“Minimize”*] ≠ *“Do Not Use”*

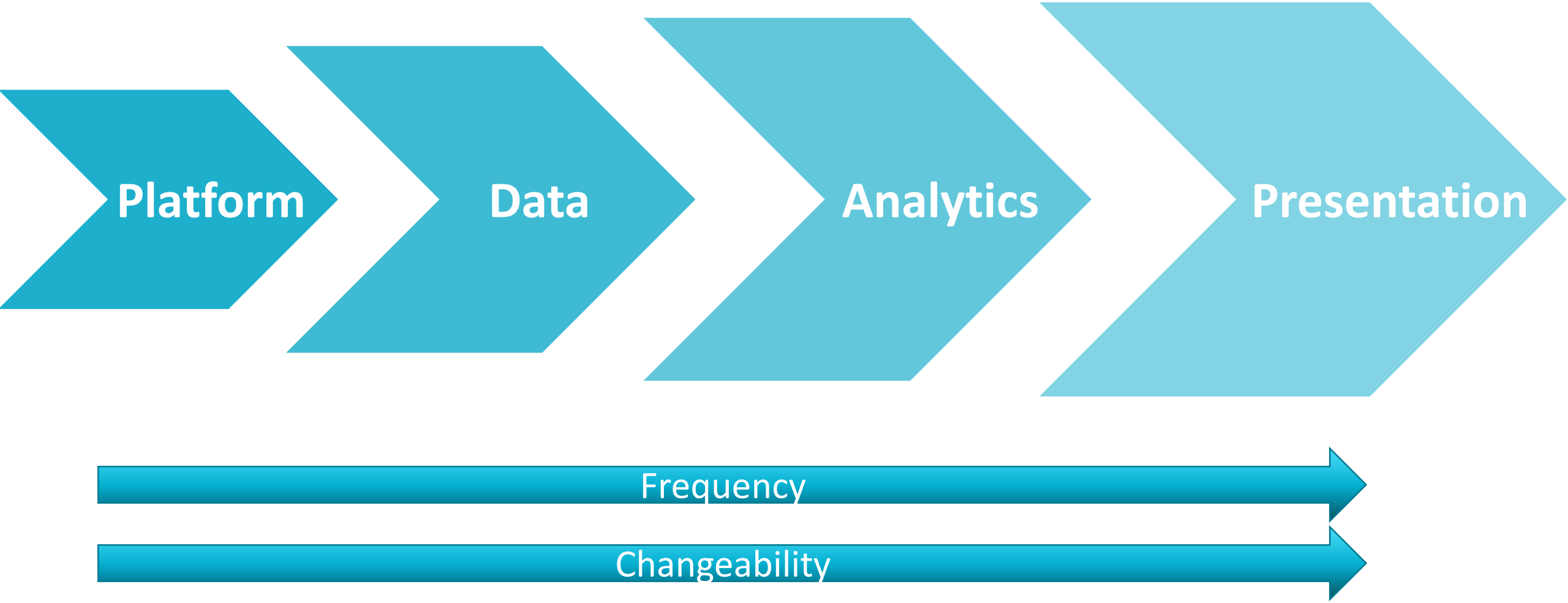
**Key points – Look for this symbol!**

**>> ‘everything in moderation’**

# Why Is My Workbook Slow?


DATA
TABLEAU
BROWSER
Request Viz
Load Viz
Send Queries
Run Queries
Return Data
Compute Model
Send Model
Render Viz
Render Viz
Send Viz

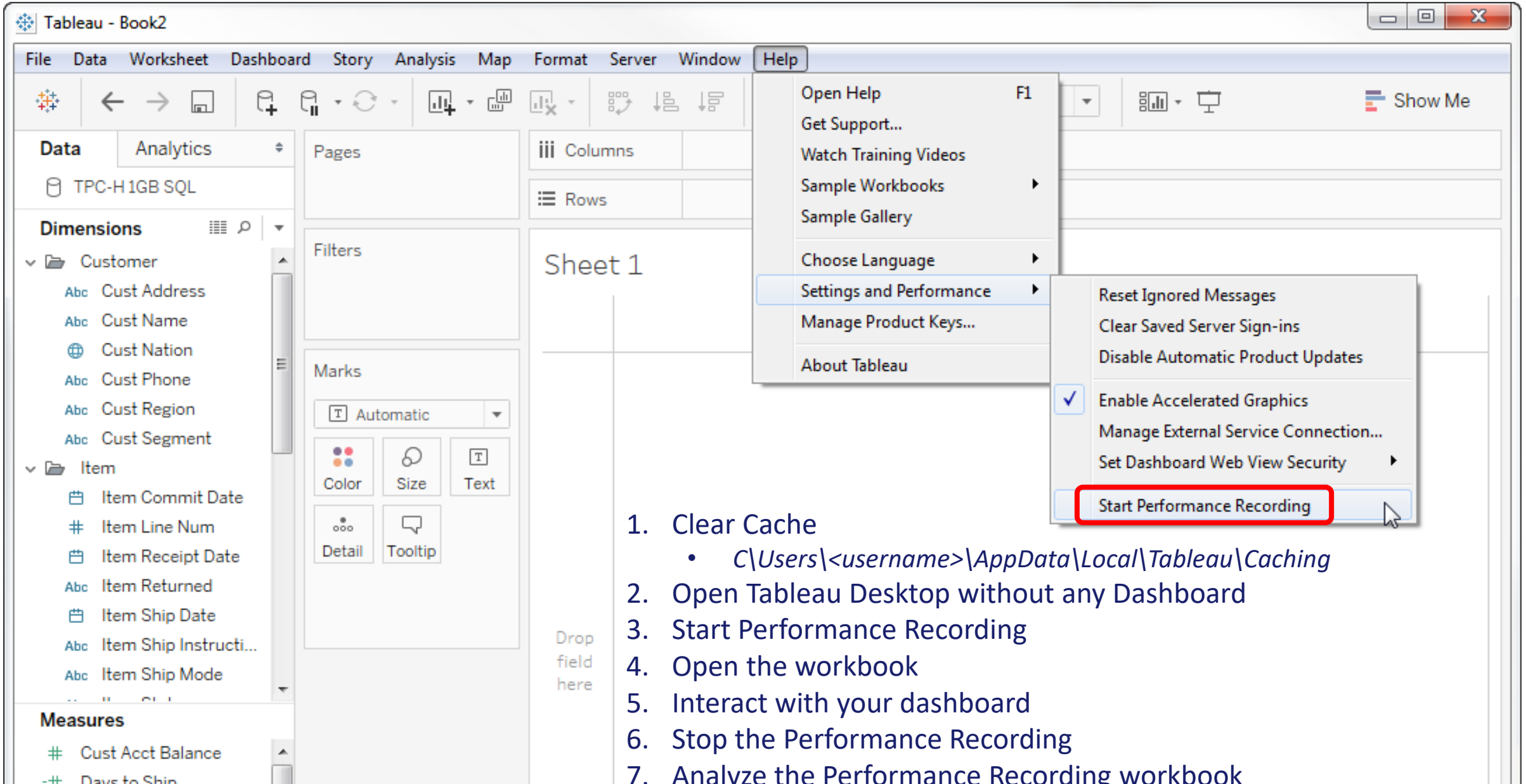
# Concept: Visual Pipeline



# Finding the Problem

# When possible troubleshoot using Tableau desktop even if the complaint is about Server ;-)

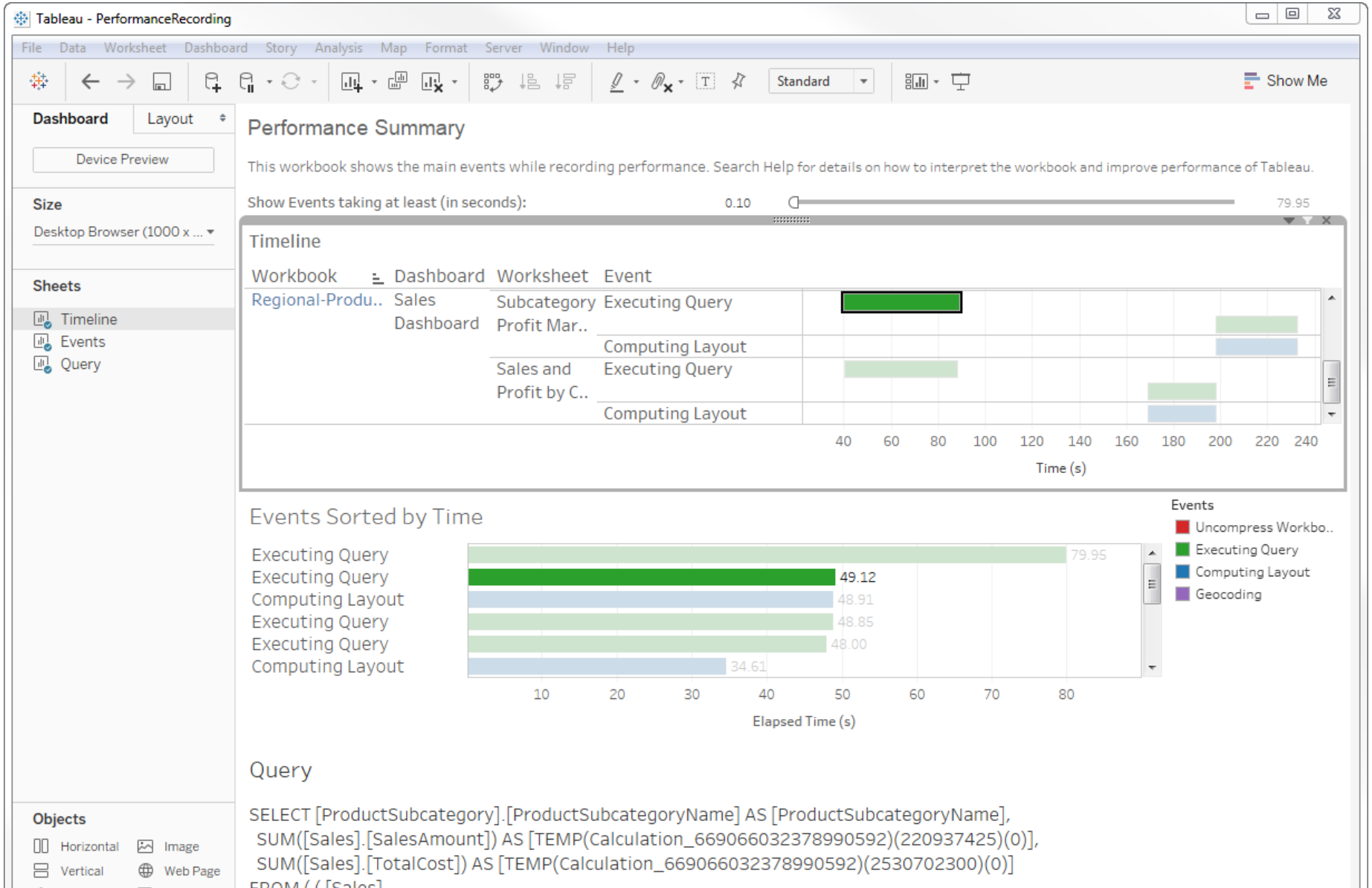
# Performance Recorder



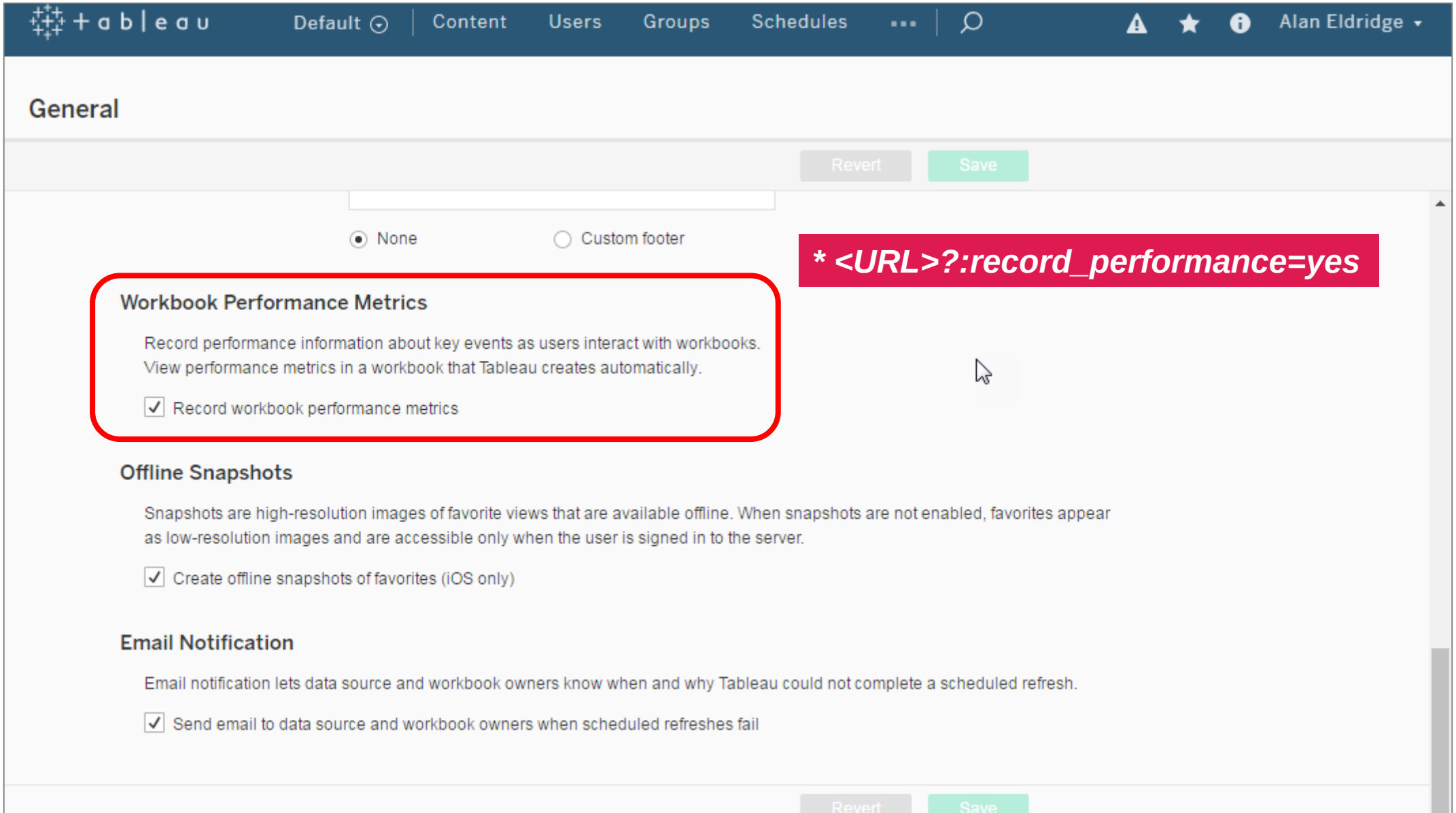

1. Clear Cache
   - *C\Users\<username>\AppData\Local\Tableau\Caching*
2. Open Tableau Desktop without any Dashboard
3. Start Performance Recording
4. Open the workbook
5. Interact with your dashboard
6. Stop the Performance Recording
7. Analyze the Performance Recording workbook

# Performance Recorder

https://help.tableau.com/current/server/en-us/perf_record_interpret_server.htm



# Performance Recorder

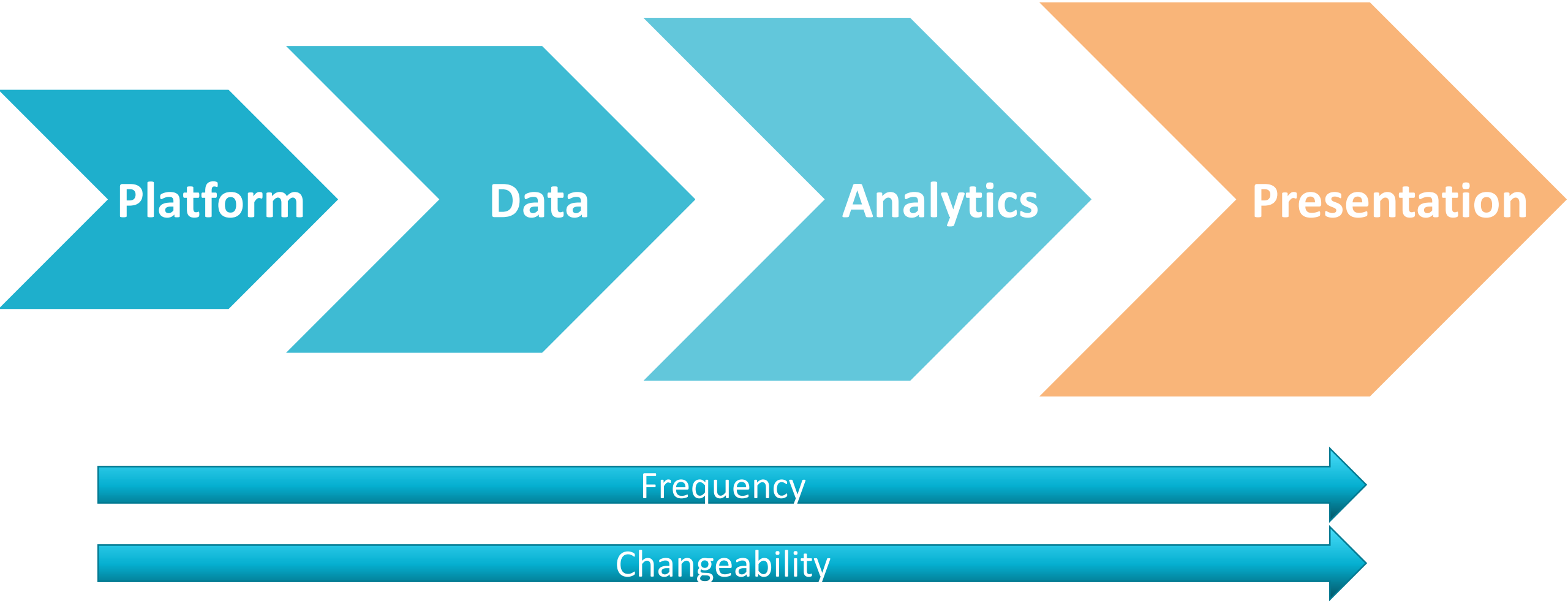


# Finding the problem … after Performance Recorder

- Log files — Errors
  - Desktop Analysis
  - Tableau Log Viewer
  - Log Shark
- Server Performance Views — Usage

- Browser Tools / Network Tracers — Platform
- TabMon
- ….. and …..
- Tableau Server Management Add-On ☺ — 360° View for Perf

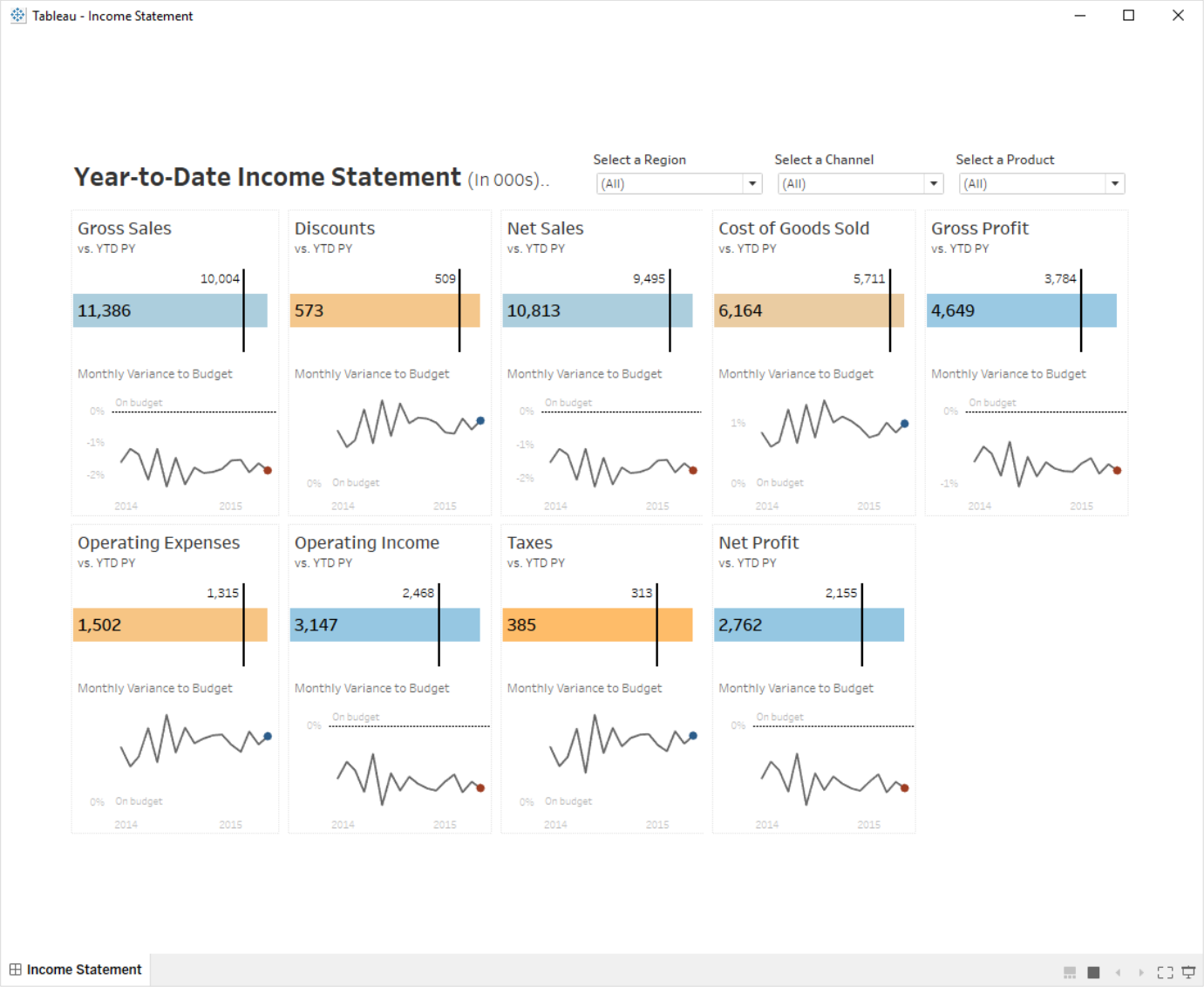
# Presentation



# Presentation Layer

## Concerned with

- Dashboards
- Worksheets



# Dashboards

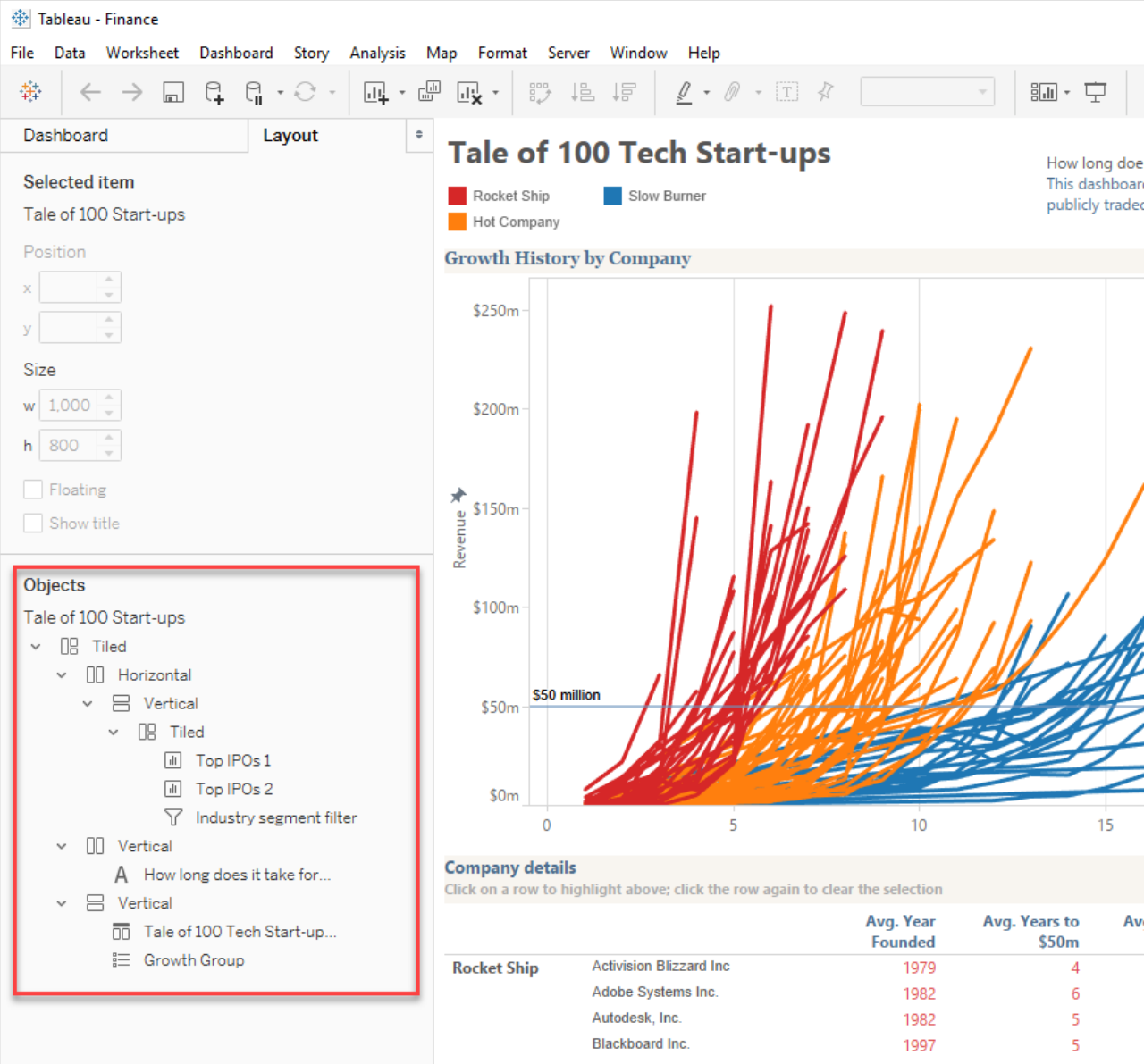
## Made up of zones

## Can be data driven

Worksheets, filters, parameters, page controls, legends

## Or non-data driven

Text, images, web content, blanks, layout containers

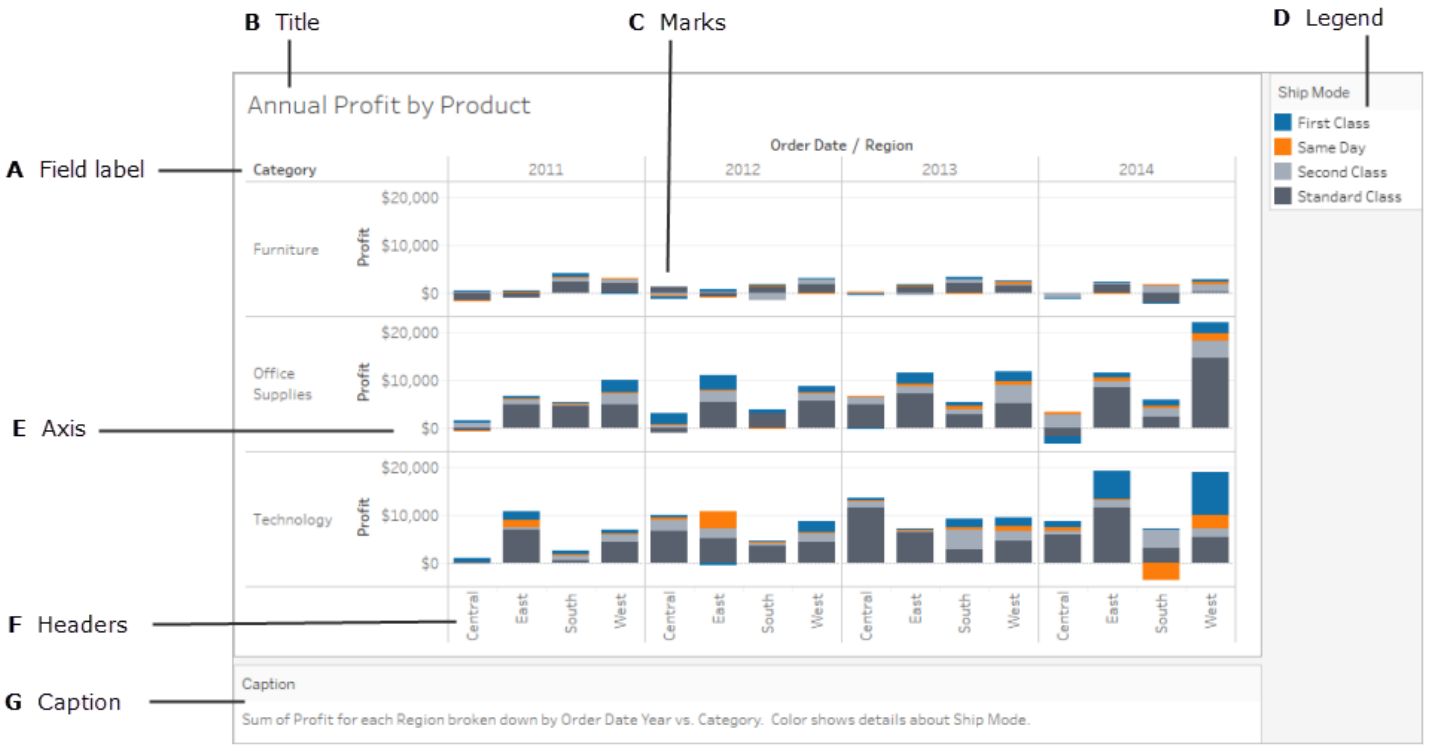


# Worksheets

## The main data driven zone is the worksheet

## It’s made up of

Headers, axes, panes, cells, marks, titles, captions, field labels, legends

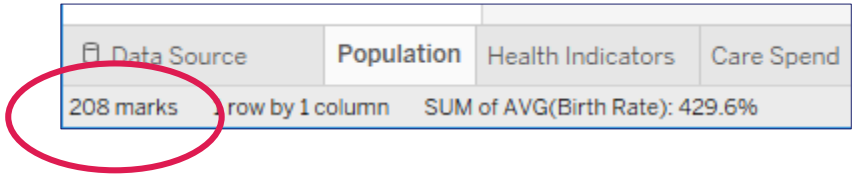


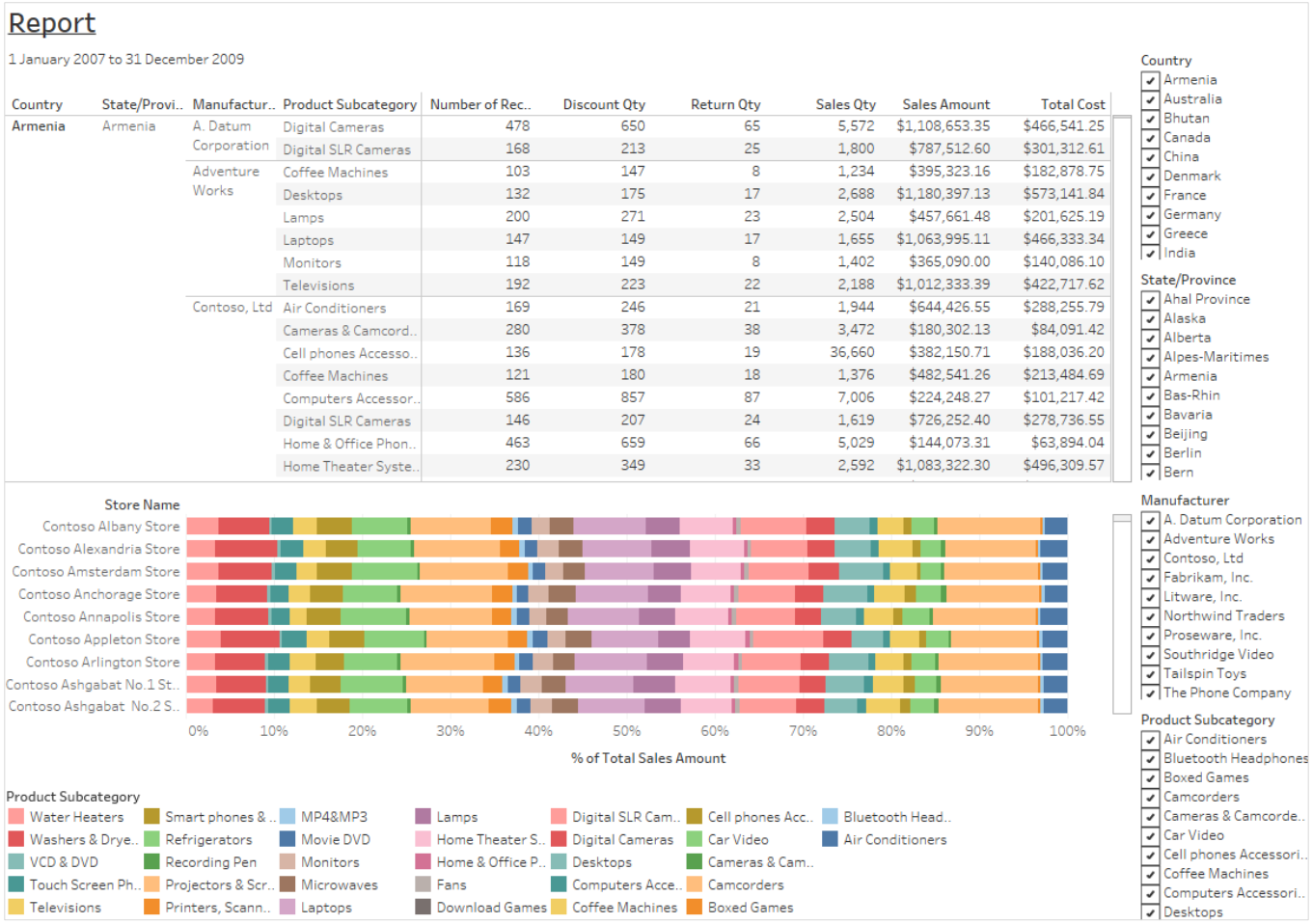
# Worksheets—Best Practice

## Keep worksheets simple

## Minimise the number of marks

- Don’t retrieve data you don’t need
- Don’t render data you can’t see



Report

1 January 2007 to 31 December 2009

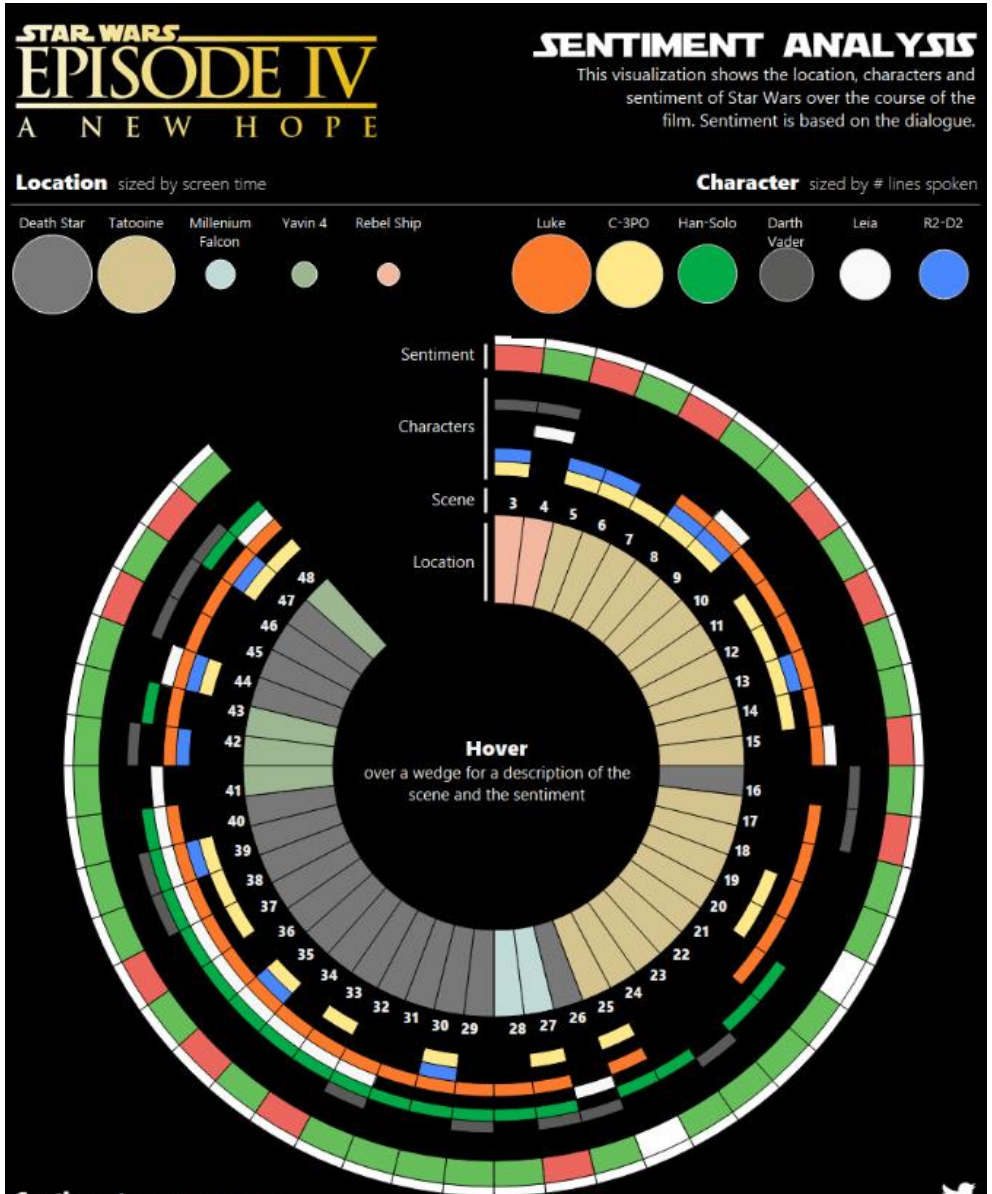
| Country | State/Provi.. | Manufactur.. | Product Subcategory | Number of Rec.. | Discount Qty | Return Qty | Sales Qty | Sales Amount | Total Cost |
|---|---|---|---|---|---|---|---|---|---|
| Armenia | Armenia | A. Datum Corporation | Digital Cameras | 478 | 650 | 65 | 5,572 | $1,108,653.35 | $466,541.25 |
| | | | Digital SLR Cameras | 168 | 213 | 25 | 1,800 | $787,512.60 | $301,312.61 |
| | | Adventure Works | Coffee Machines | 103 | 147 | 8 | 1,234 | $395,323.16 | $182,878.75 |
| | | | Desktops | 132 | 175 | 17 | 2,688 | $1,180,397.13 | $573,141.84 |
| | | | Lamps | 200 | 271 | 23 | 2,504 | $457,661.48 | $201,625.19 |
| | | | Laptops | 147 | 149 | 17 | 1,655 | $1,063,995.11 | $466,333.34 |
| | | | Monitors | 118 | 149 | 8 | 1,402 | $365,090.00 | $140,086.10 |
| | | | Televisions | 192 | 223 | 22 | 2,188 | $1,012,333.39 | $422,717.62 |
| | | Contoso, Ltd | Air Conditioners | 169 | 246 | 21 | 1,944 | $644,426.55 | $288,255.79 |
| | | | Cameras & Camcord.. | 280 | 378 | 38 | 3,472 | $180,302.13 | $84,091.42 |
| | | | Cell phones Accesso.. | 136 | 178 | 19 | 36,660 | $382,150.71 | $188,036.20 |
| | | | Coffee Machines | 121 | 180 | 18 | 1,376 | $482,541.26 | $213,484.69 |
| | | | Computers Accessor.. | 586 | 857 | 87 | 7,006 | $224,248.27 | $101,217.42 |
| | | | Digital SLR Cameras | 146 | 207 | 24 | 1,619 | $726,252.40 | $278,736.55 |
| | | | Home & Office Phon.. | 463 | 659 | 66 | 5,029 | $144,073.31 | $63,894.04 |
| | | | Home Theater Syste.. | 230 | 349 | 33 | 2,592 | $1,083,322.30 | $496,309.57 |

# Worksheets—Best Practice

**Not all marks are the same**

## Avoid:

- Polygon marks
- Manufactured chart styles like donut, Sankey, sigmoid curved lines, etc.
- Maps, if you don’t really need them—tiles can take time to load

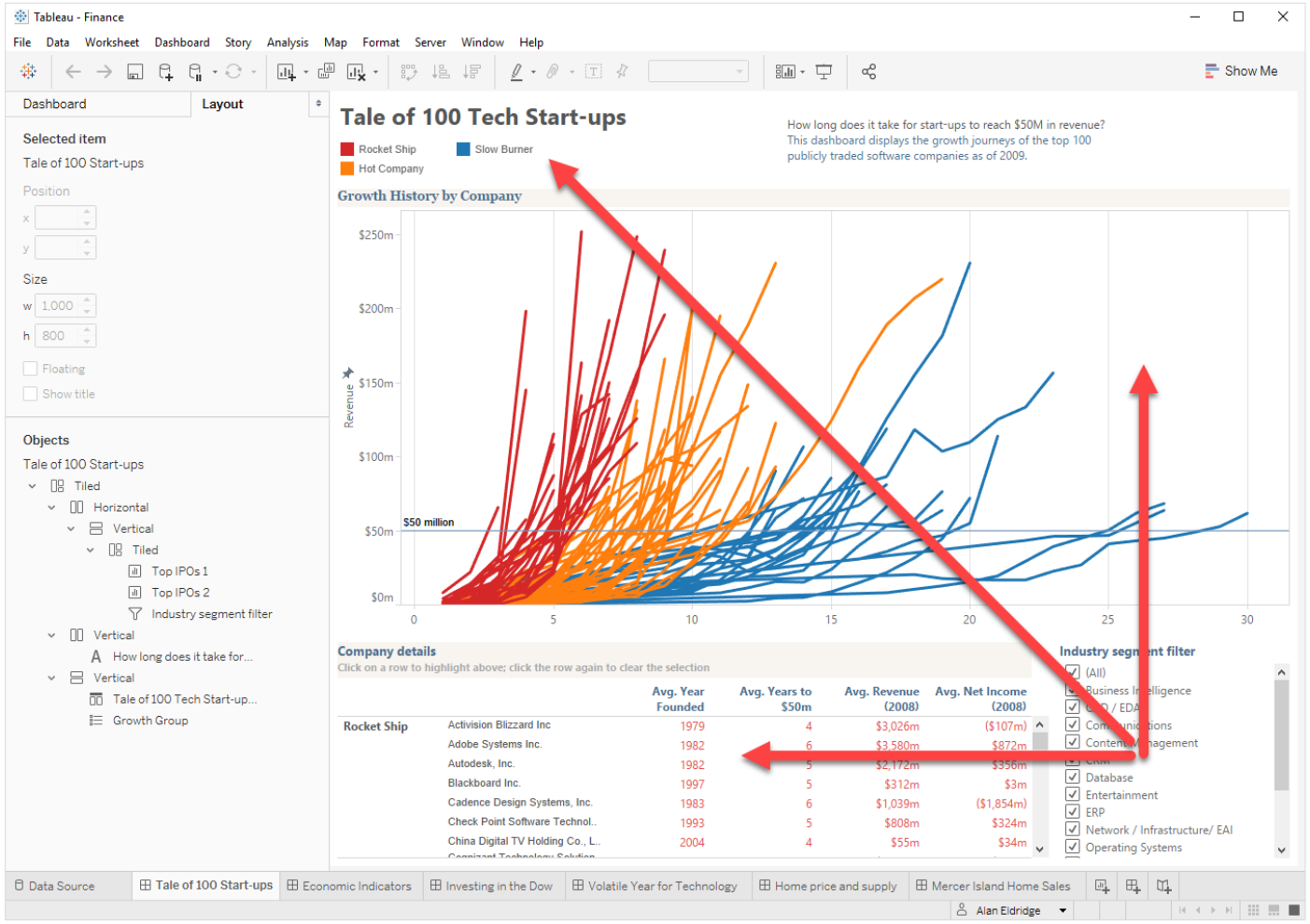


# Dashboards

## Interacting with dashboard elements can trigger events

- Filters
- Parameters
- Highlighting
- Pages
- Actions—filter, highlight, URL

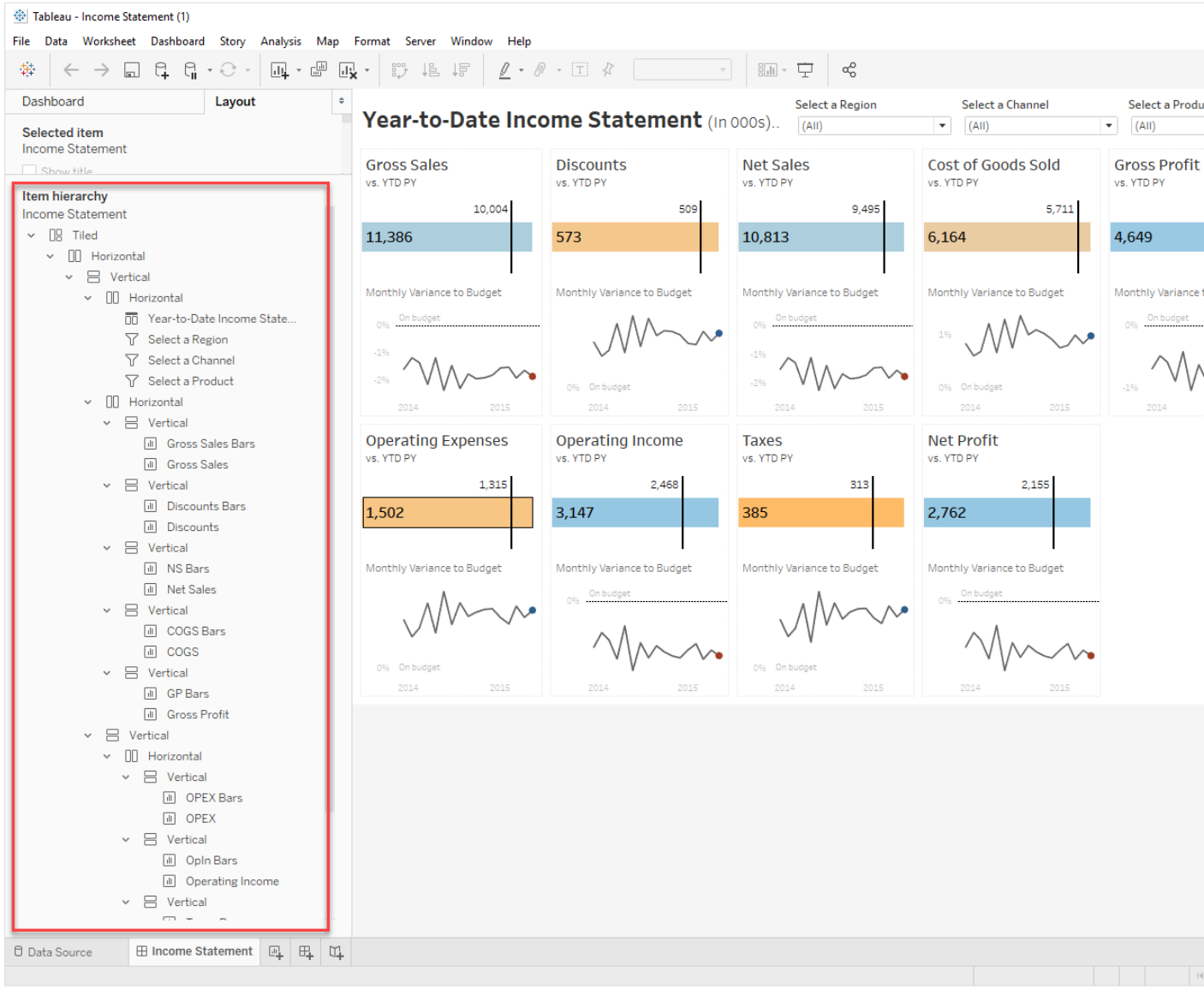
## Can cause other dashboard elements to update



# Dashboards—Best Practices

## Minimise number of zones

## Each data driven zone = 1 query



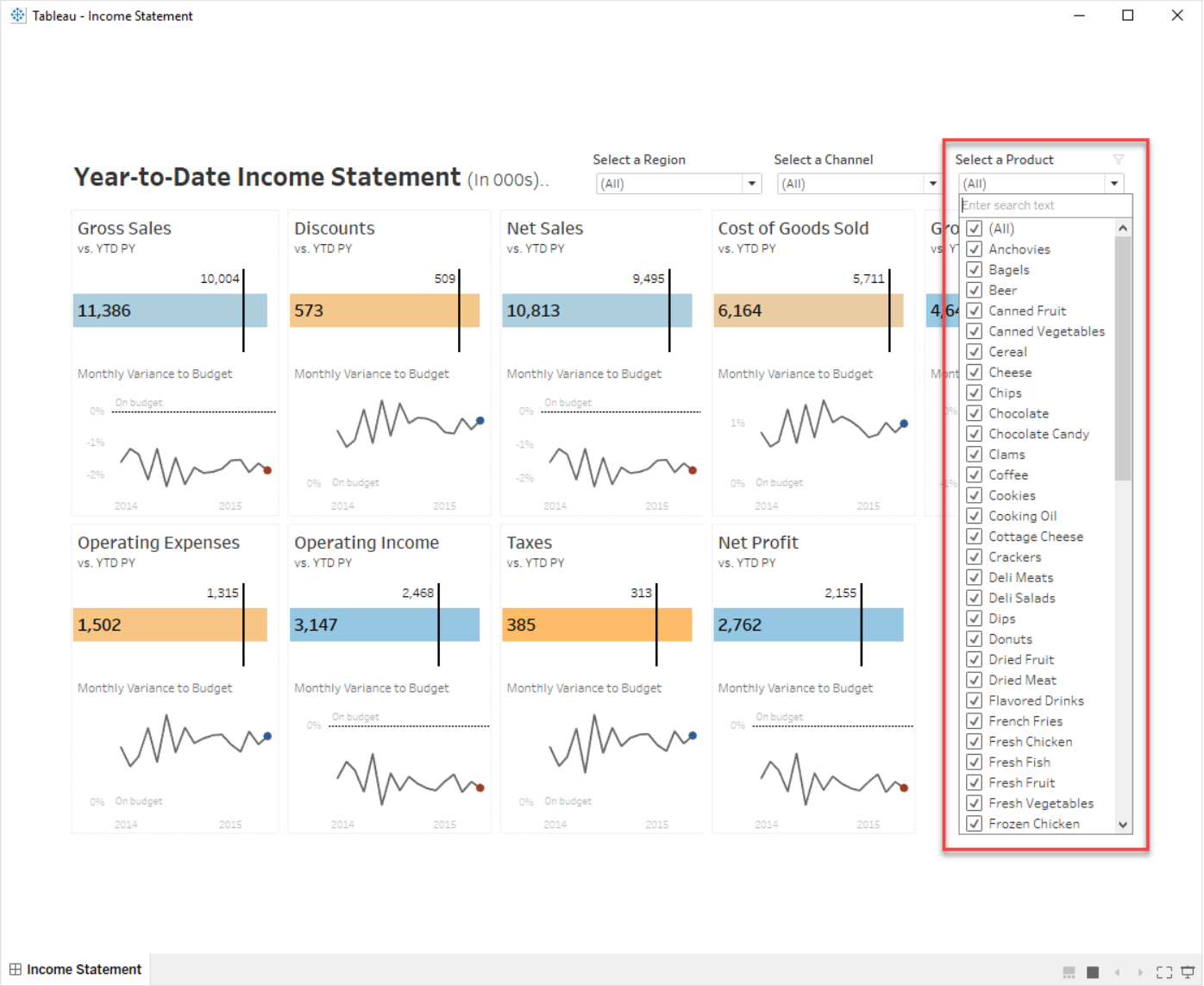
# Dashboards—Best Practices

**Minimise the complexity of each zone**

## Keep worksheets as simple as possible

## Avoid:

- Large, enumerated filters, parameter lists, legends
- Large, high resolution images
- Complex embedded web parts

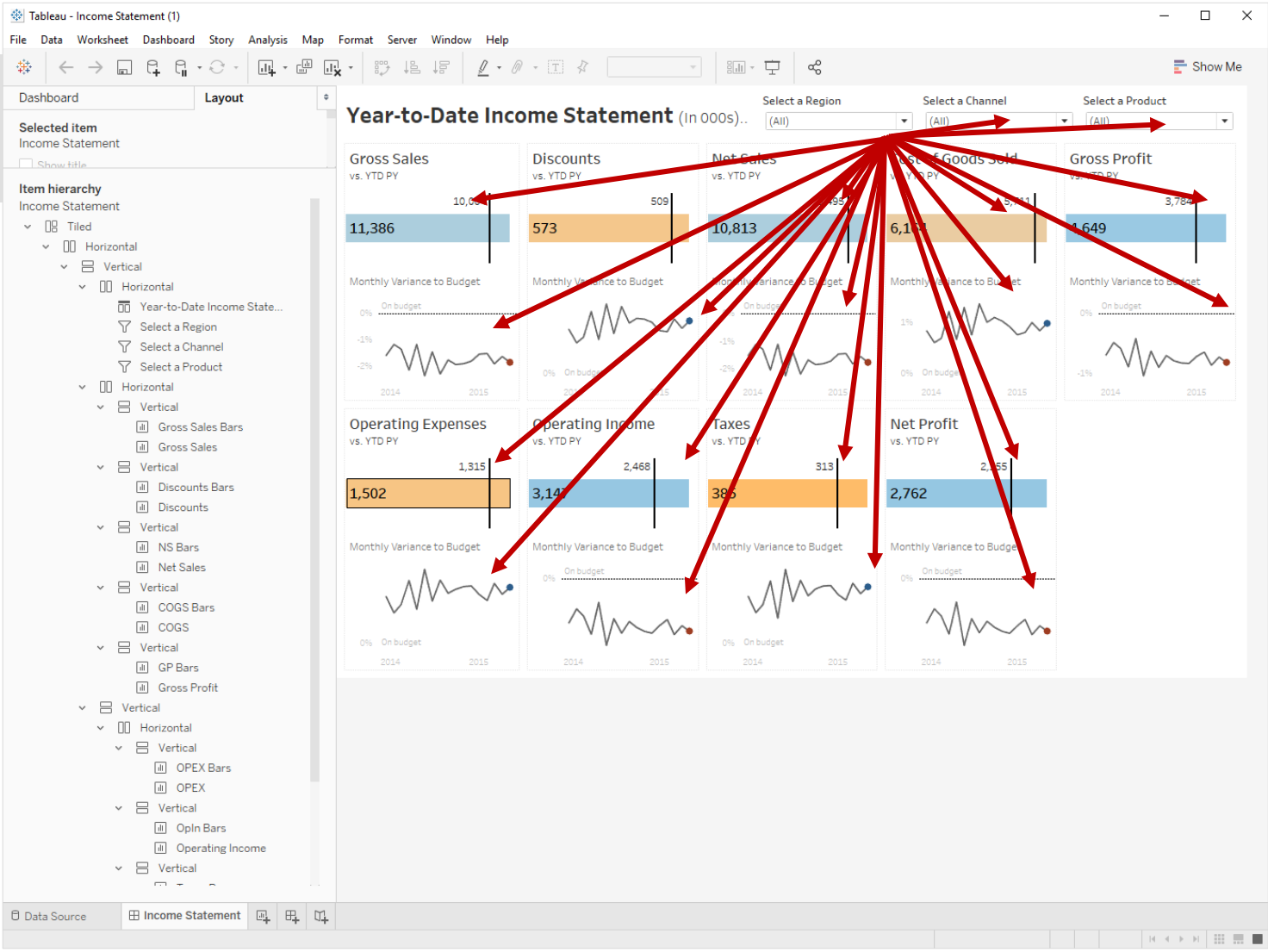


# Dashboards—Best Practices

**Minimise the scope and impact of updates**

## Be conscious of the impact of:

- Global filters
- Filters that apply to all sheets on a dashboard
- Filters with “show relevant values”



# Dashboards—Best Practices

**Maximise the effectiveness of caches**

## Try to:

Set dashboards and stories to fixed size

Use client-side rendering where possible

SQL Editor | Graphical Query Builder

Previous queries

```
select http_referer from http_requests where http_referer like '%size%'
```

Output pane

Data Output | Explain | Messages | History

| | http_referer character varying(255) |
|---|---|
| 1 | http://localhost/vizql/showadminview/tabbed admin views/BackgroundTasksforNonExtracts?:size=1272,1055&:embe |
| 2 | http://localhost/vizql/showadminview/tabbed admin views/BackgroundTasksforNonExtracts?:size=1272,1055&:embe |
| 3 | http://localhost/vizql/showadminview/tabbed admin views/BackgroundTasksforNonExtracts?:size=1272,1055&:embe |
| 4 | http://localhost/vizql/showadminview/tabbed admin views/BackgroundTasksforNonExtracts?:size=1272,1055&:embe |
| 5 | http://localhost/vizql/showadminview/tabbed admin views/BackgroundTasksforNonExtracts?:size=1272,1055&:embe |
| 6 | http://localhost/vizql/showadminview/tabbed admin views/BackgroundTasksforNonExtracts?:size=1272,1055&:embe |
| 7 | http://localhost/vizql/showadminview/tabbed admin views/BackgroundTasksforNonExtracts?:size=1272,1055&:embe |
| 8 | http://localhost/vizql/showadminview/tabbed admin views/BackgroundTasksforNonExtracts?:size=1272,1055&:embe |
| 9 | http://localhost/views/Superstore/Overview?:size=1263,531&:embed=y&:showVizHome=n&:bootstrapWhenNotified=y& |
| 10 | http://localhost/views/Superstore/Overview?:size=1263,531&:embed=y&:showVizHome=n&:bootstrapWhenNotified=y& |
| 11 | http://localhost/views/Superstore/Overview?:size=1263,531&:embed=y&:showVizHome=n&:bootstrapWhenNotified=y& |
| 12 | http://jswickard4/views/Attribute-Customer-Name/Sheet1?:size=1247,521&:embed=y&:showVizHome=n&:bootstrapWhe |
| 13 | http://jswickard4/views/Attribute-Customer-Name/Sheet1?:size=1247,521&:embed=y&:showVizHome=n&:bootstrapWhe |
| 14 | http://jswickard4/views/Attribute-Customer-Name/Sheet1?:size=1247,521&:embed=y&:showVizHome=n&:bootstrapWhe |
| 15 | http://jswickard4/views/Attribute-Customer-Name/Sheet1?:size=1247,521&:embed=y&:showVizHome=n&:bootstrapWhe |
| 16 | http://jswickard4/views/Attribute-Customer-Name/Sheet1?:size=1247,521&:embed=y&:showVizHome=n&:bootstrapWhe |
| 17 | http://jswickard4/views/Attribute-Customer-Name/Sheet1?:size=1903,931&:embed=y&:showVizHome=n&:bootstrapWhe |
| 18 | http://jswickard4/views/Attribute-Customer-Name/Sheet1?:size=1903,931&:embed=y&:showVizHome=n&:bootstrapWhe |
| 19 | http://jswickard4/views/Attribute-Customer-Name/Sheet1?:size=1903,931&:embed=y&:showVizHome=n&:bootstrapWhe |
| 20 | http://jswickard4/views/Attribute-Customer-Name/Sheet1?:size=1903,931&:embed=y&:showVizHome=n&:bootstrapWhe |

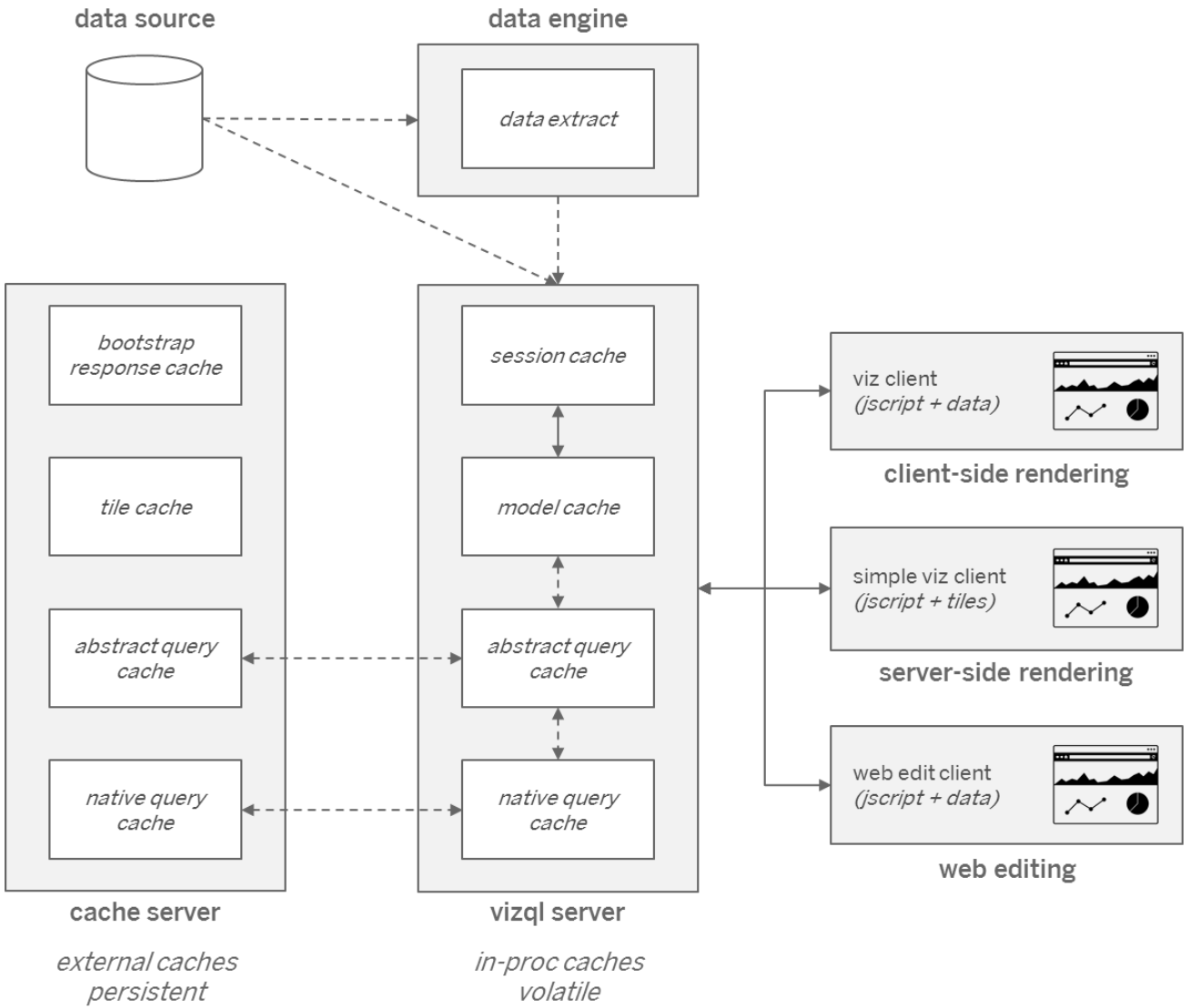
OK.

# A Sidebar on Caching…

## Tableau tries to reuse as much of its work as it can

## There are many caches:

- Browser data cache
- Model cache
- Tile cache
- Query cache
- Bootstrap response cache



# A Sidebar on Query Optimisation…

## Tableau is very clever and will try to reduce the number and complexity of queries it has to run

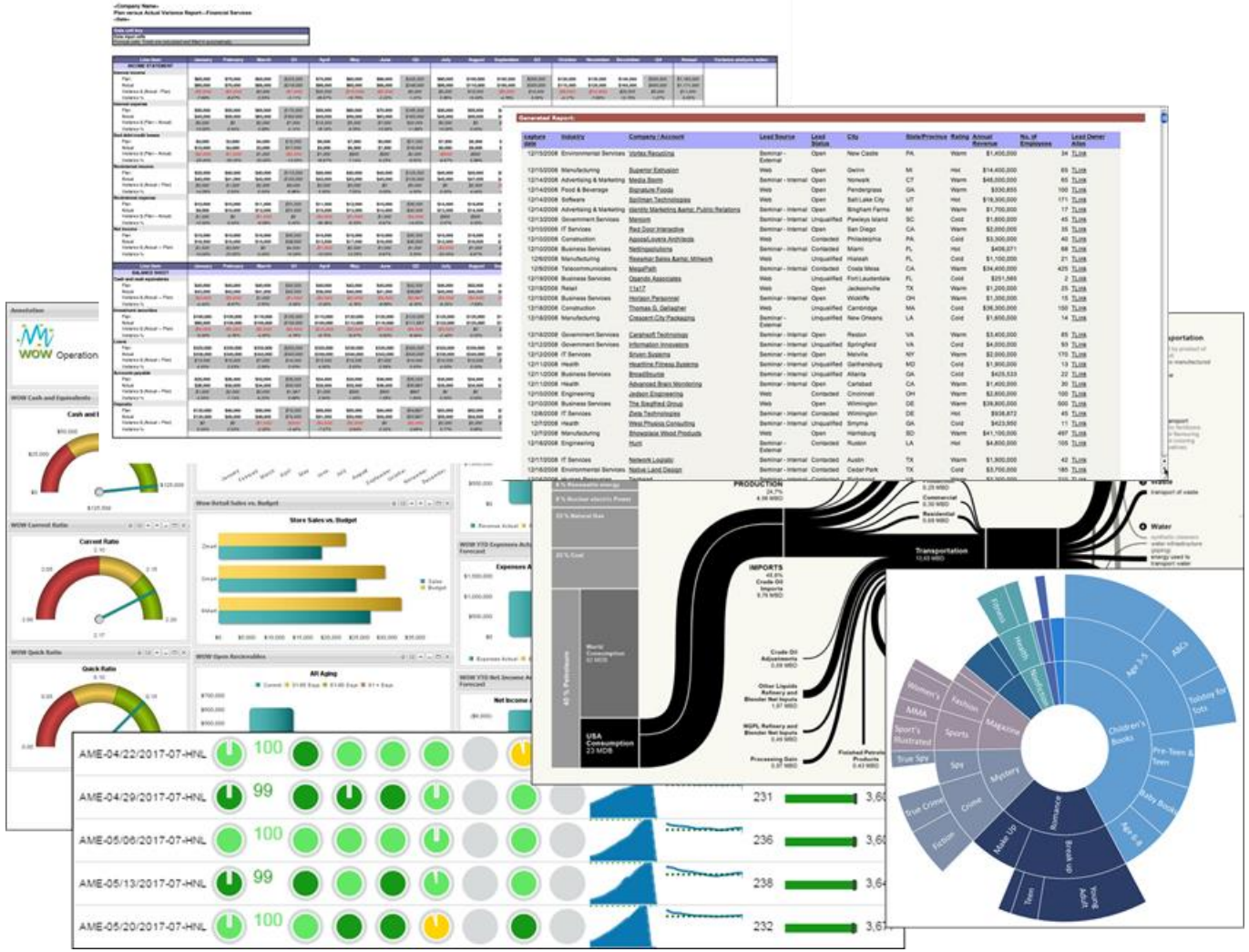
## It uses:

- Abstract queries
- Parallel queries
- Query fusion
- Constant folding
- …

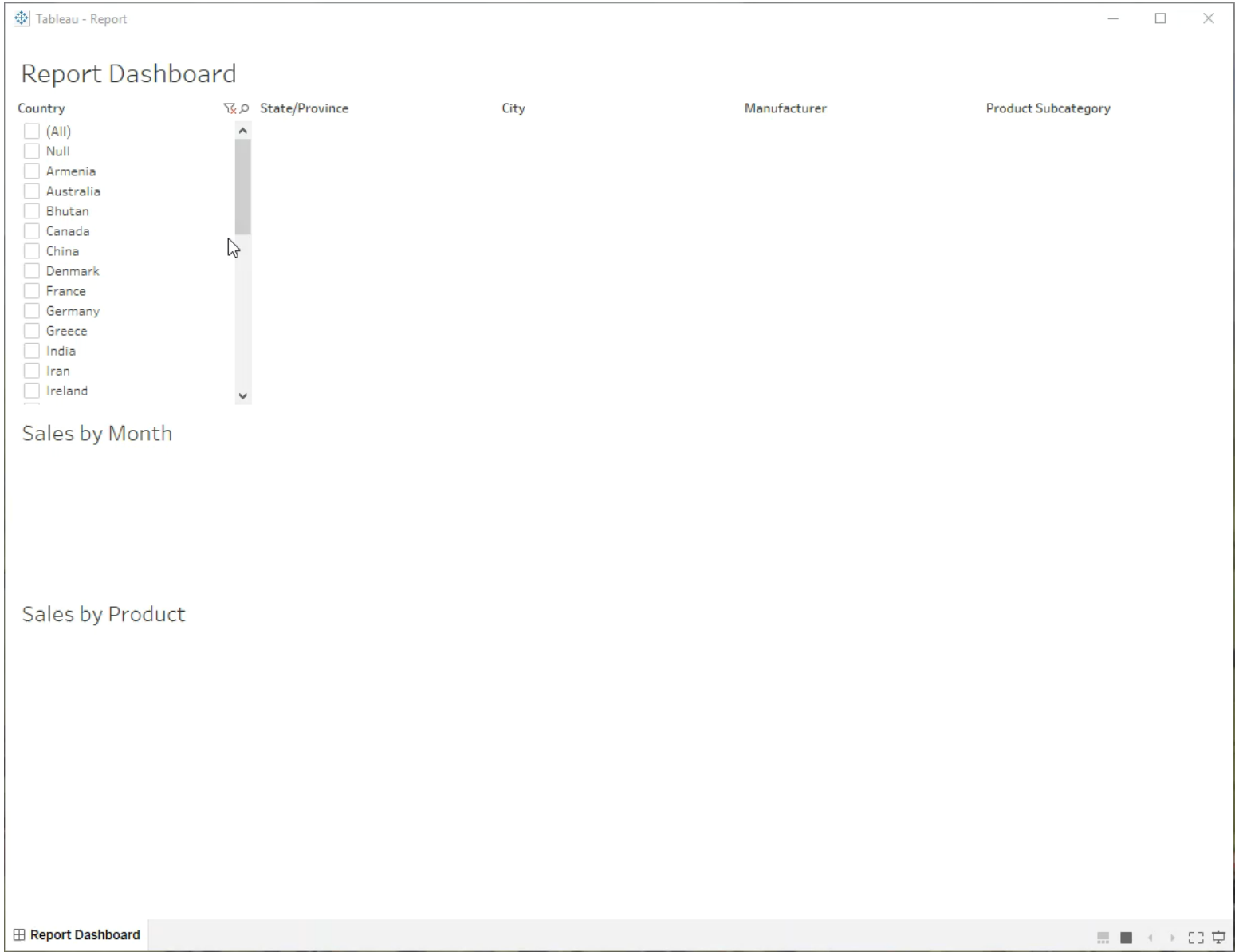
# Dashboard Design: by the example

Poor choice in dashboard design is the most common cause of poor dashboard performance

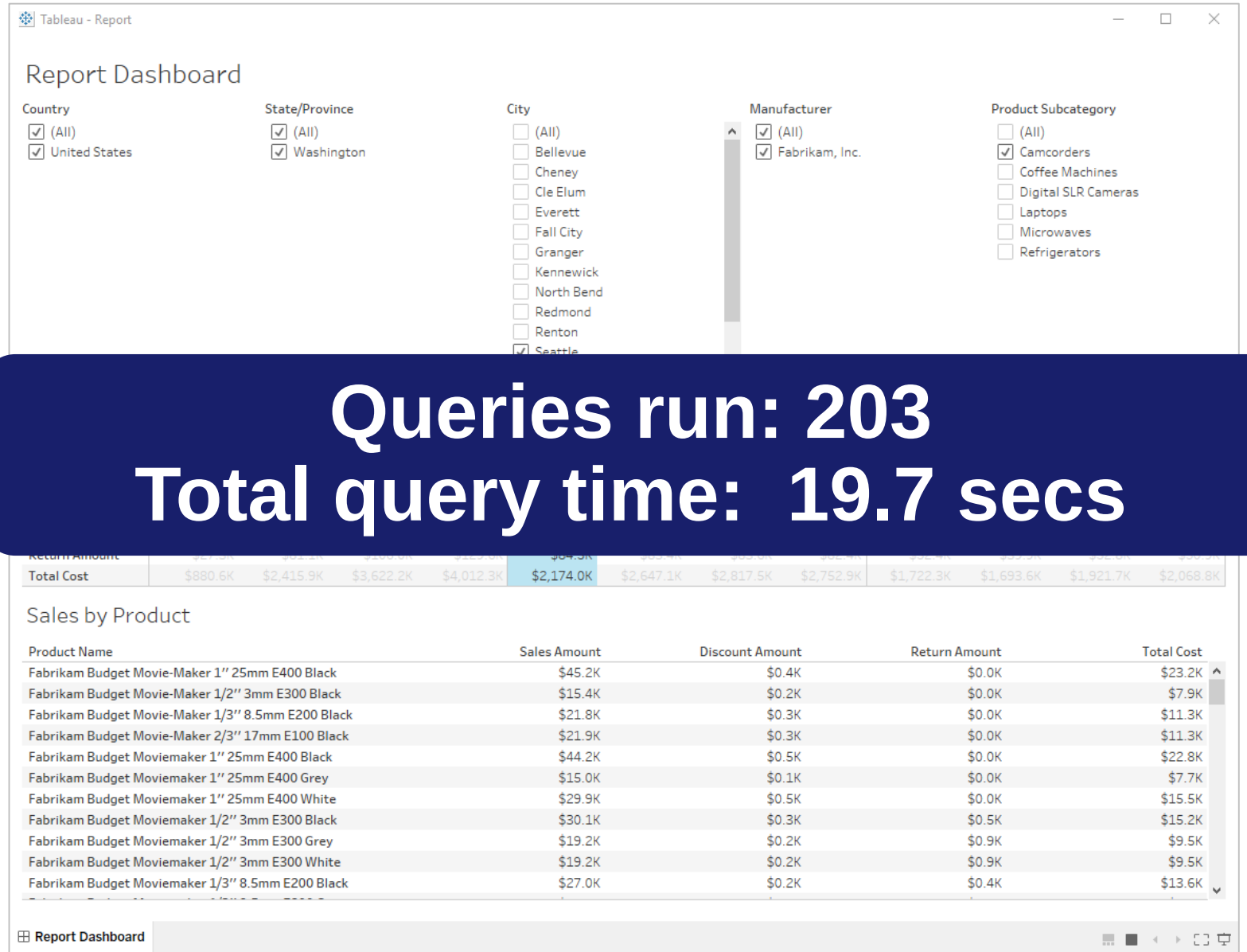
Fortunately, it’s the easiest thing to fix (and avoid in future)



# Dashboard Design: by the example

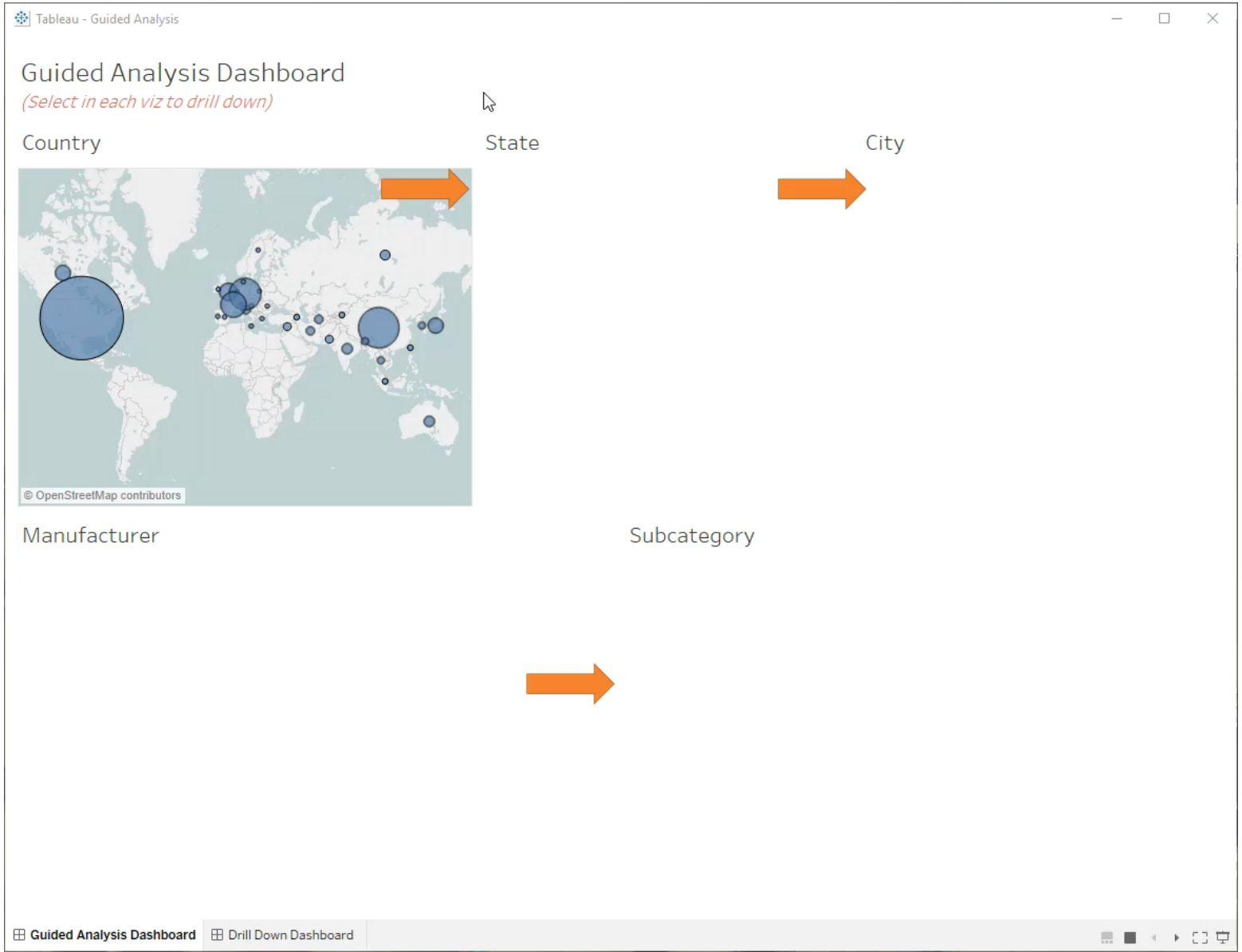


# Dashboard Design: by the example

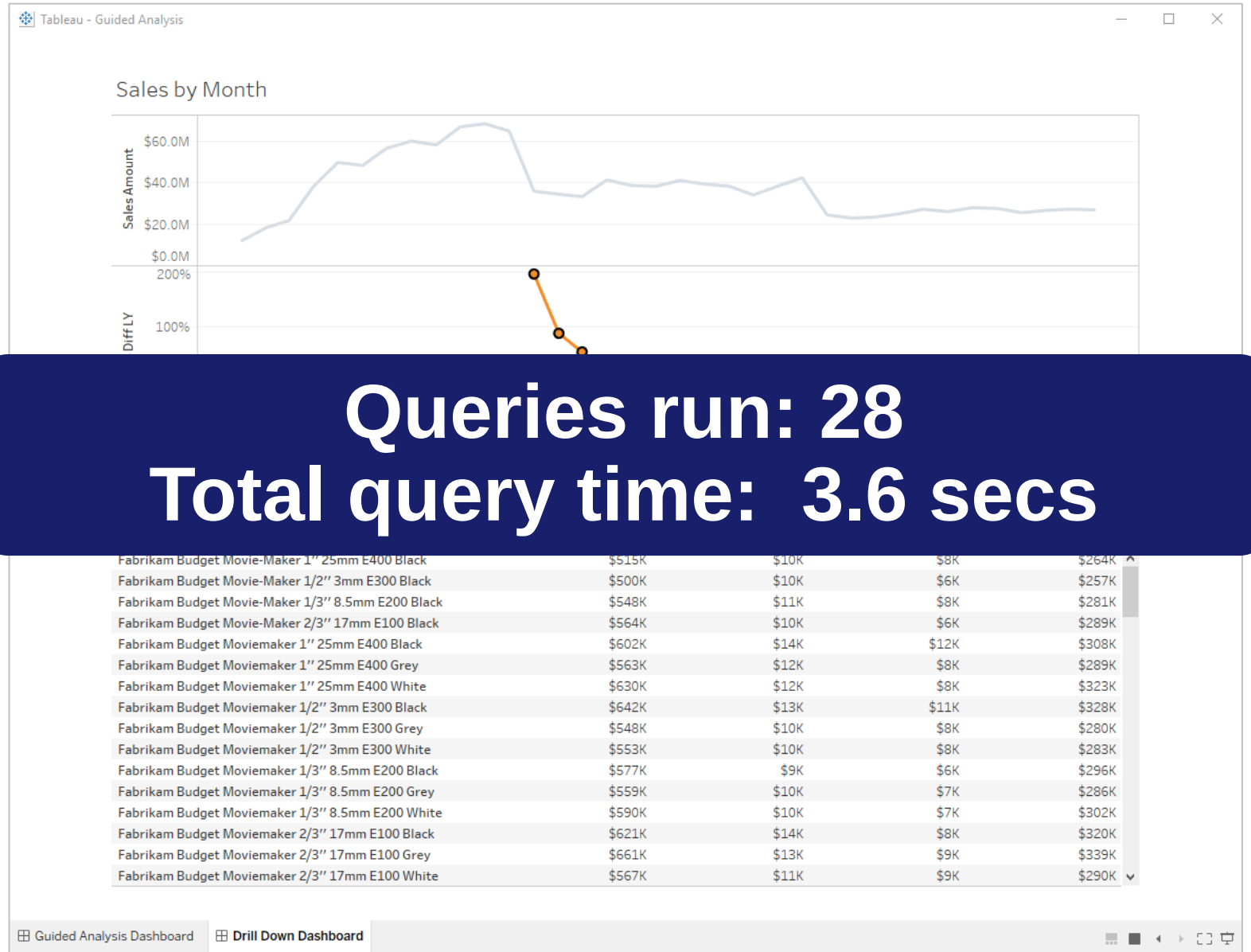



**Queries run: 203**
**Total query time: 19.7 secs**

# Dashboard Design: by the example



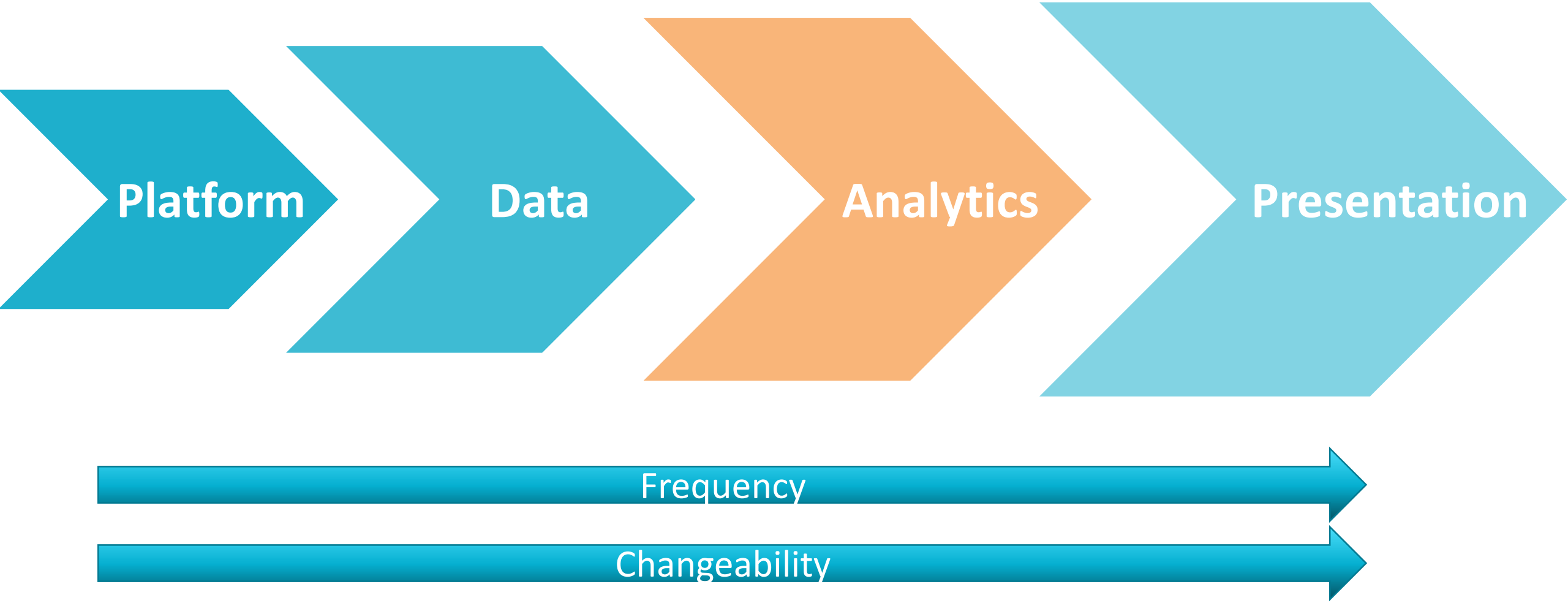
# Dashboard Design: by the example




| | | | | |
|---|---|---|---|---|
| Fabrikam Budget Movie-Maker 1'' 25mm E400 Black | $515K | $10K | $8K | $264K |
| Fabrikam Budget Movie-Maker 1/2'' 3mm E300 Black | $500K | $10K | $6K | $257K |
| Fabrikam Budget Movie-Maker 1/3'' 8.5mm E200 Black | $548K | $11K | $8K | $281K |
| Fabrikam Budget Movie-Maker 2/3'' 17mm E100 Black | $564K | $10K | $6K | $289K |
| Fabrikam Budget Moviemaker 1'' 25mm E400 Black | $602K | $14K | $12K | $308K |
| Fabrikam Budget Moviemaker 1'' 25mm E400 Grey | $563K | $12K | $8K | $289K |
| Fabrikam Budget Moviemaker 1'' 25mm E400 White | $630K | $12K | $8K | $323K |
| Fabrikam Budget Moviemaker 1/2'' 3mm E300 Black | $642K | $13K | $11K | $328K |
| Fabrikam Budget Moviemaker 1/2'' 3mm E300 Grey | $548K | $10K | $8K | $280K |
| Fabrikam Budget Moviemaker 1/2'' 3mm E300 White | $553K | $10K | $8K | $283K |
| Fabrikam Budget Moviemaker 1/3'' 8.5mm E200 Black | $577K | $9K | $6K | $296K |
| Fabrikam Budget Moviemaker 1/3'' 8.5mm E200 Grey | $559K | $10K | $7K | $286K |
| Fabrikam Budget Moviemaker 1/3'' 8.5mm E200 White | $590K | $10K | $7K | $302K |
| Fabrikam Budget Moviemaker 2/3'' 17mm E100 Black | $621K | $14K | $8K | $320K |
| Fabrikam Budget Moviemaker 2/3'' 17mm E100 Grey | $661K | $13K | $9K | $339K |
| Fabrikam Budget Moviemaker 2/3'' 17mm E100 White | $567K | $11K | $9K | $290K |

**Queries run: 28**
**Total query time: 3.6 secs**

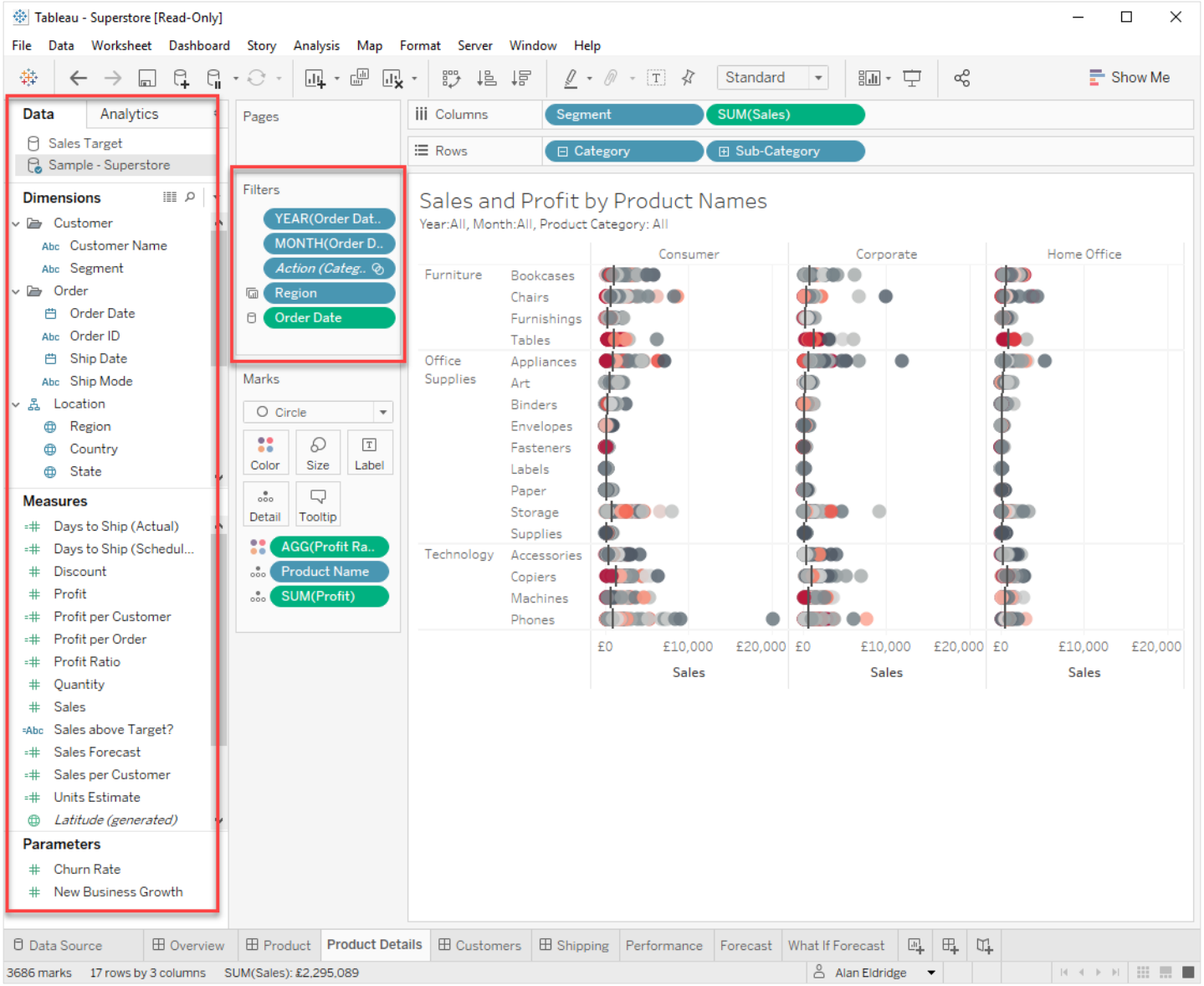
# Analytics


Platform
Data
Analytics
Presentation
Frequency
Changeability

# Analytics Layer

## Concerned with:

- Data elements
- Filters
- Calculations
- Working across data sources

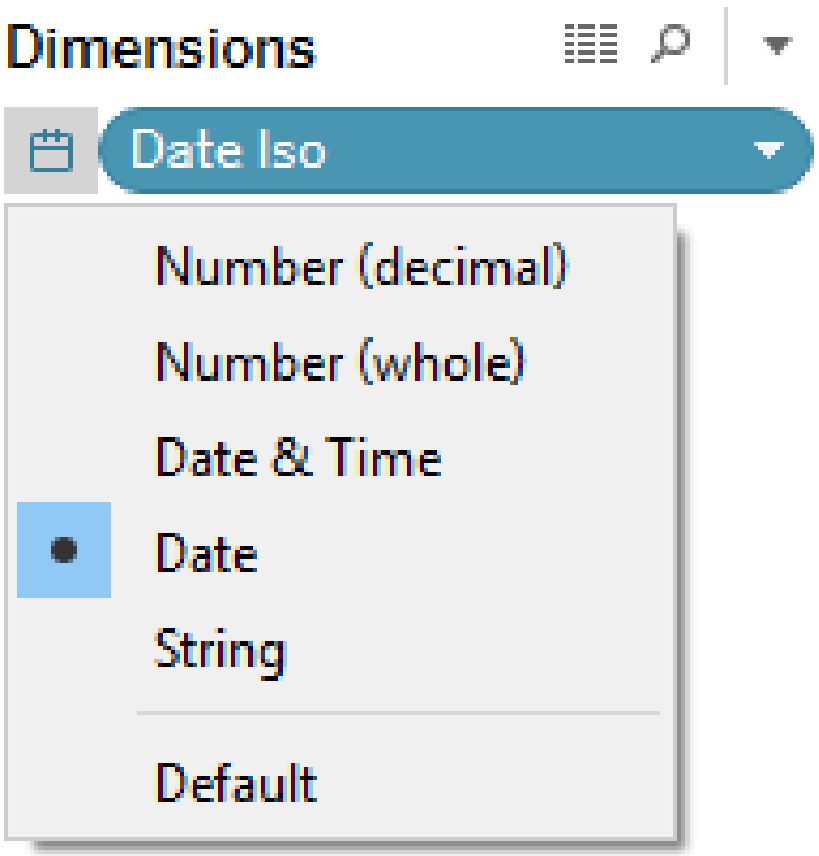


# Data Elements

## Native vs. Converted Type

**Minimise data type casting**

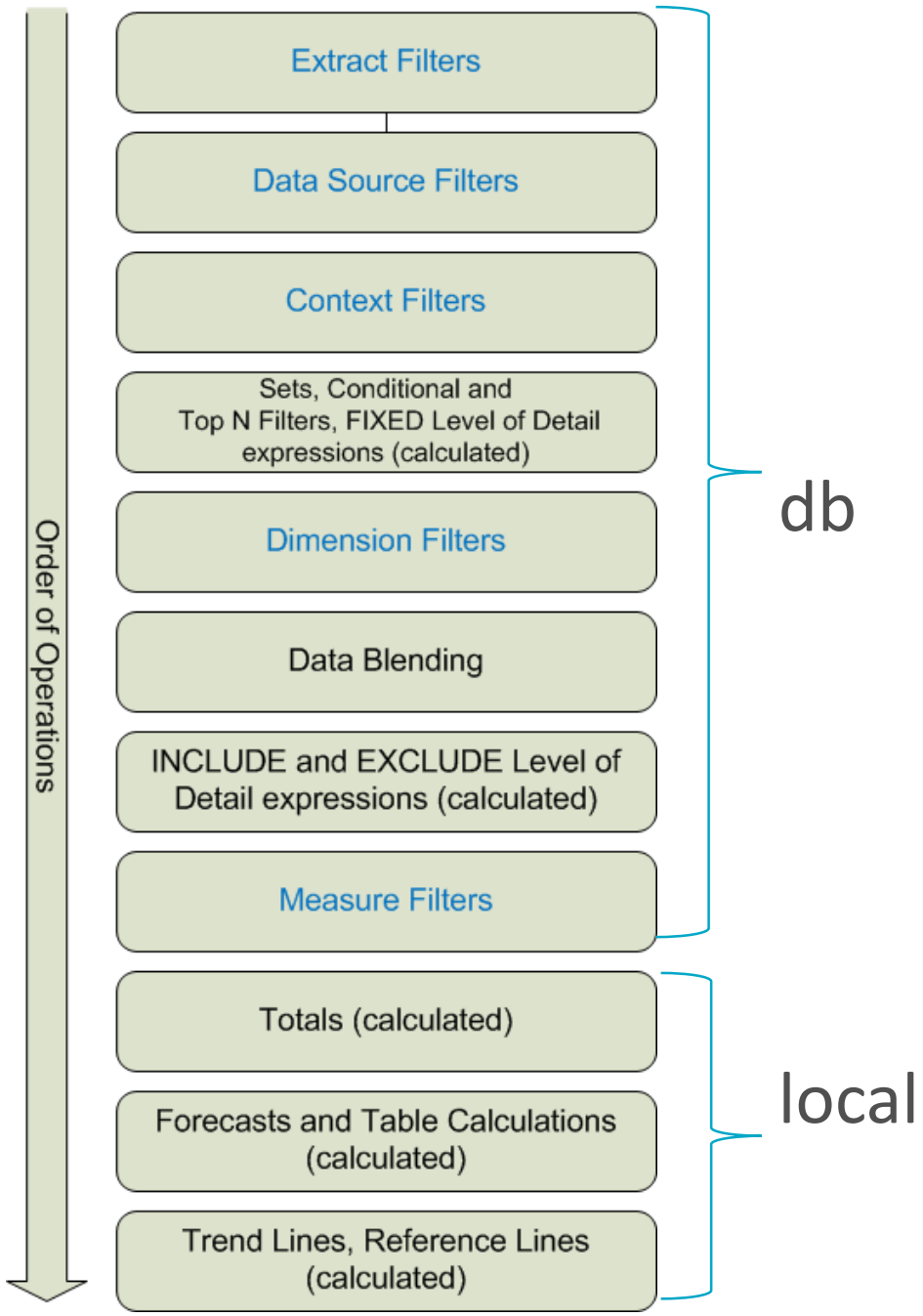
**Convert at the data source if possible**



# Filters

Are applied at different stages within the computation pipeline

Some pass through to the data source, others are applied after



# Filters

**Use the most appropriate type of filters**

**Discrete vs. ranged**

**Enumerated vs. non-enumerated**

**Aggregate vs. row level**

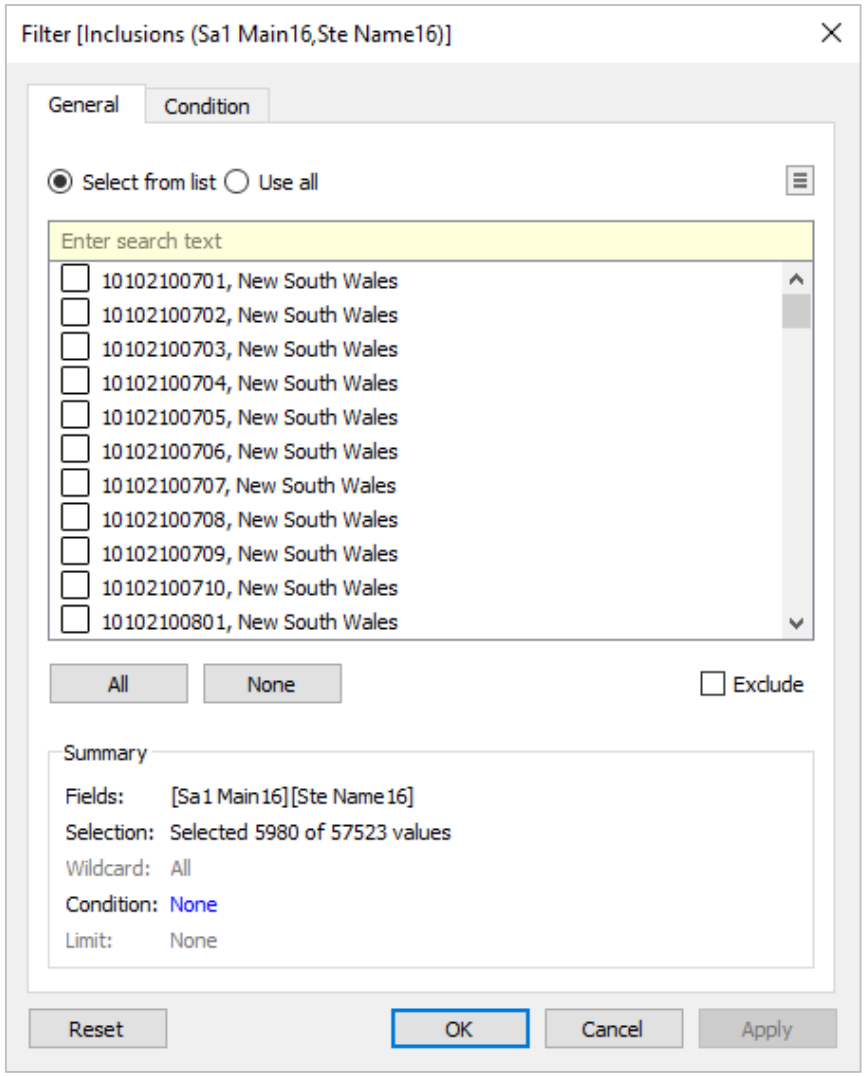
**Dates!**

**Slicing filters**

**Sets**

**Cross data source**

**Customize > show “Apply” button**




***“Begin with the End in Mind”***

# Filters

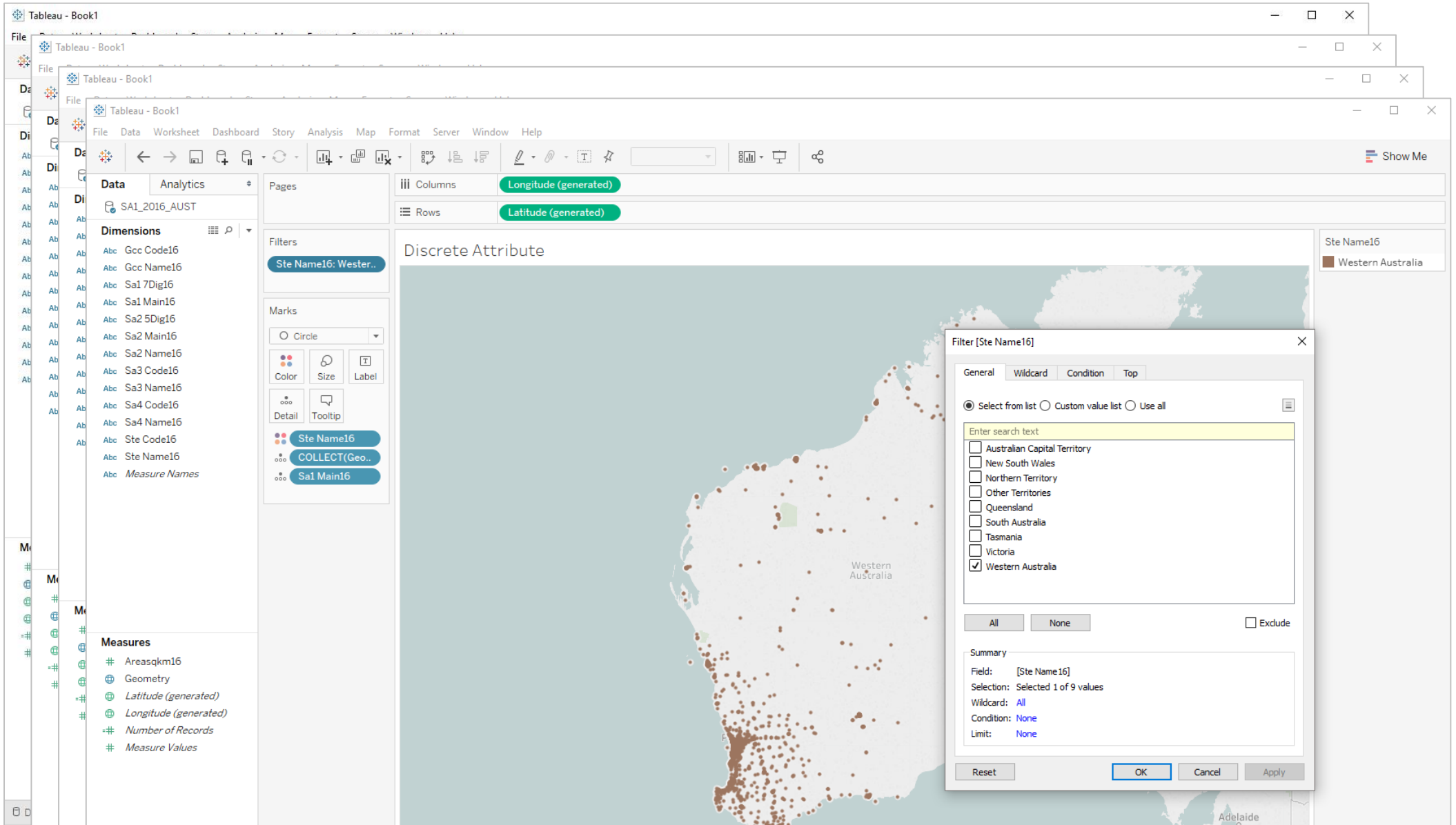


Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data
Analytics
SA1_2016_AUST
Dimensions
Gcc Code16
Gcc Name16
Sa1 7Dig16
Sa1 Main16
Sa2 5Dig16
Sa2 Main16
Sa2 Name16
Sa3 Code16
Sa3 Name16
Sa4 Code16
Sa4 Name16
Ste Code16
Ste Name16
Measure Names
Measures
Areasqkm16
Geometry
Latitude (generated)
Longitude (generated)
Number of Records
Measure Values
Pages
Filters
Ste Name16: Wester..
Marks
Circle
Color
Size
Label
Detail
Tooltip
Ste Name16
COLLECT(Geo..
Sa1 Main16
Columns
Longitude (generated)
Rows
Latitude (generated)
Discrete Attribute
Ste Name16
Western Australia
Western Australia
Adelaide
Filter [Ste Name16]
General
Wildcard
Condition
Top
Select from list
Custom value list
Use all
Enter search text
Australian Capital Territory
New South Wales
Northern Territory
Other Territories
Queensland
South Australia
Tasmania
Victoria
Western Australia
All
None
Exclude
Summary
Field: [Ste Name16]
Selection: Selected 1 of 9 values
Wildcard: All
Condition: None
Limit: None
Reset
OK
Cancel
Apply

# Filters

**Use *"Only Relevant Values"* Sparingly**

**Filter list is refreshed when any other filter is updated**

**Queries can span all tables in the model**

Reduce effectiveness of Join Culling

**Consider filtering from smaller data source**

Apply to >> All Using Related Data Source

- All Using Related Data Sources
- All Using This Data Source
- Selected Worksheets...
- Only This Worksheet

**January 2018 feature: All Values in Hierarchy (Cascading Filters)**

- Only Relevant Values
- All Values in Hierarchy
- All Values in Database

- Edit Filter...
- Apply to Worksheets
- Format Filters...
- Customize
- Show Title
- Edit Title...
- Single Value (list)
- Single Value (dropdown)
- Single Value (slider)
- Multiple Values (list)
- Multiple Values (dropdown)
- Multiple Values (custom list)
- Wildcard Match
- Only Relevant Values
- All Values in Database
- Include Values
- Exclude Values
- Floating
- Fixed Height
- Edit Height...
- Select Layout Container
- Deselect
- Remove from Dashboard

# Filters

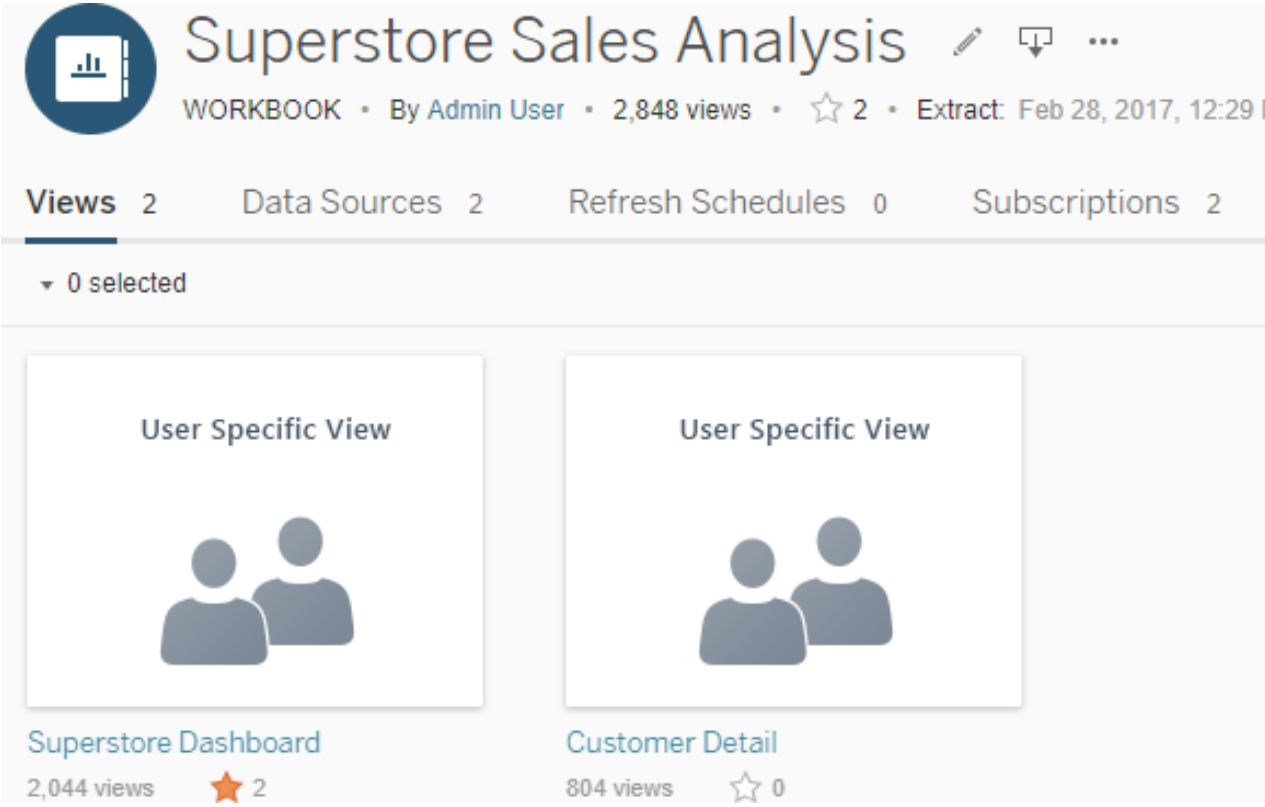
**Understand impact of User Filters**

## Apply filters based on the user's identity

"Create user filter…"
or ISMEMBEROF()

Can have a dramatic effect on scalability as caching cannot be shared

Use prudently!



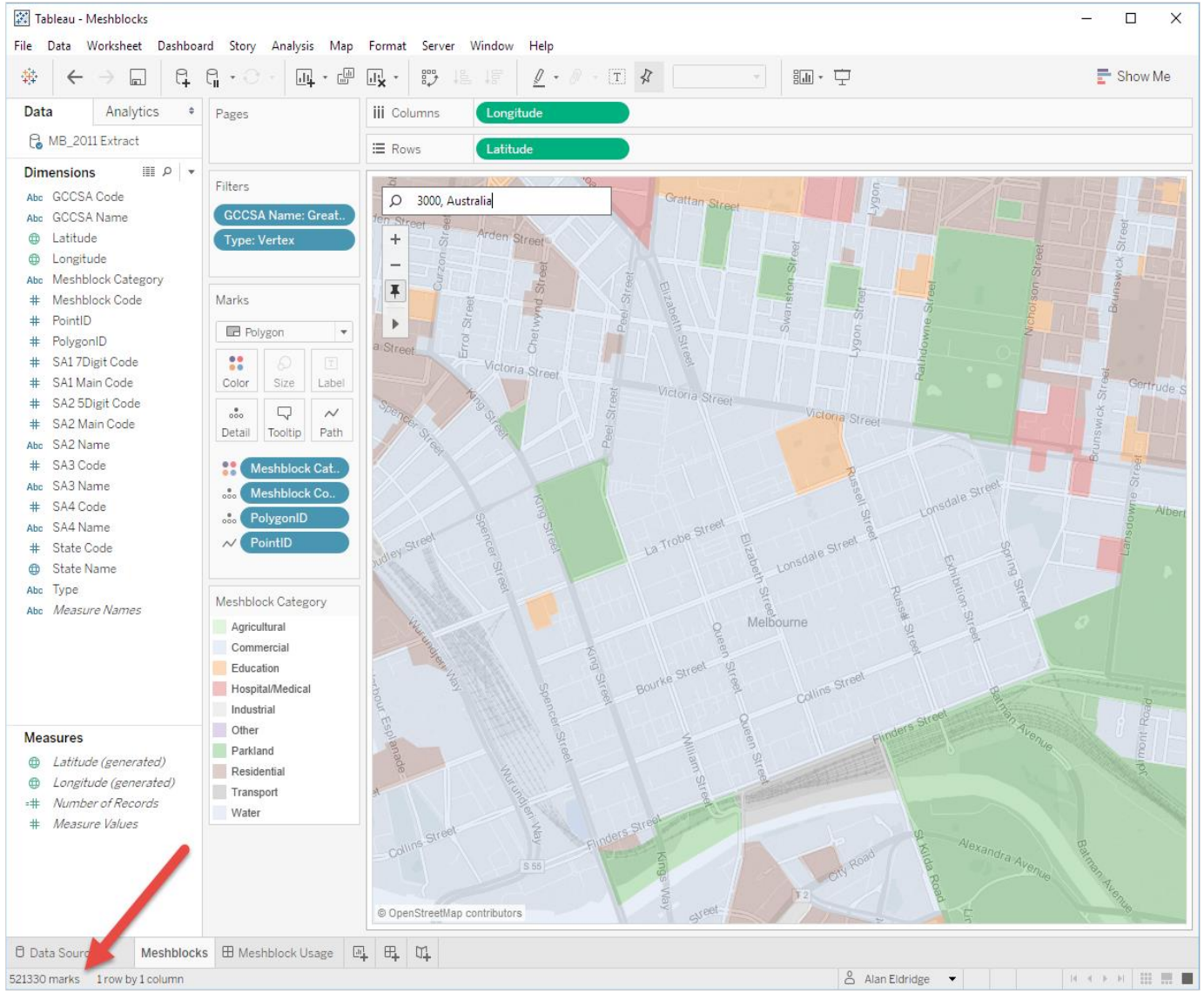
# Filters

## Filtering

Reduces the result set

## Zooming

Doesn’t change the result set, just the marks being rendered



# Calculations

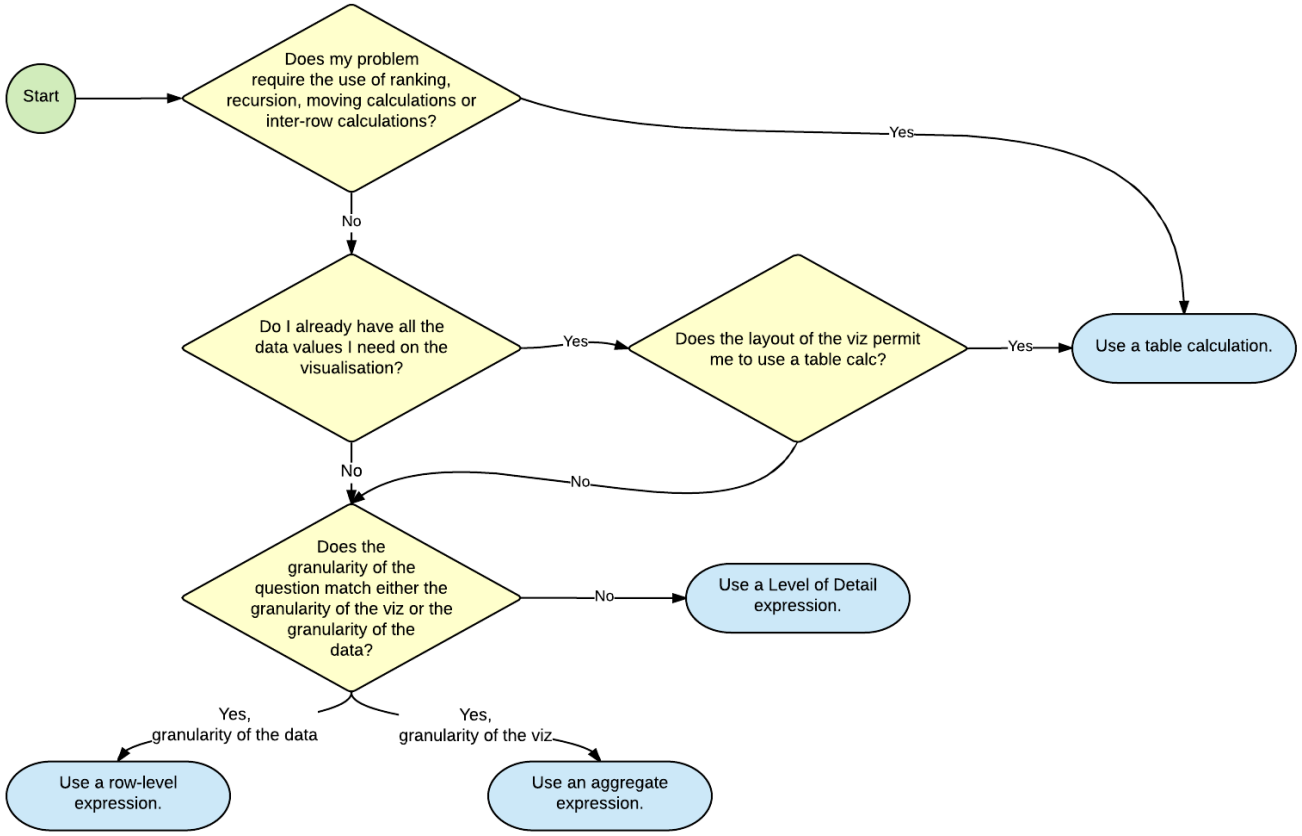
**Understand where calculations are computed**

## Many types of calculations

- Row-level, aggregate, level-of-detail, table calcs
- All apart from table calcs are passed to the underlying DB

## External function calls (R/Python/MatLab) can be slow

- Data is serialised to/from
- Correct “compute using” can have dramatic impact



# Calculations

**Understand Impact of Data Types and Functions**

## Numbers > dates, Booleans > strings

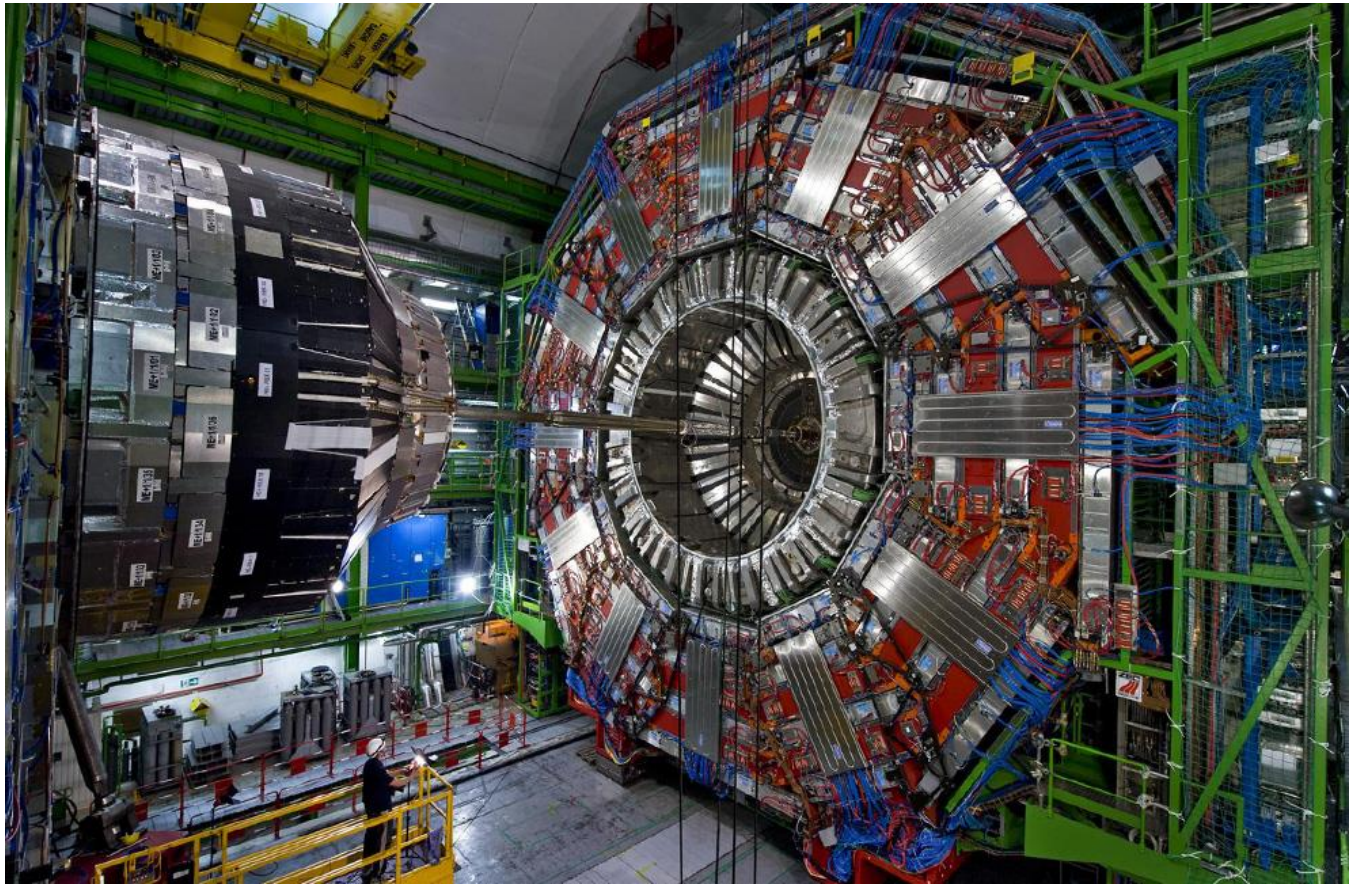
Expensive aggregation types: `COUNTD()`, `Percentile`, etc.

`MIN(), MAX() > ATTR()`

## LODs execute in the database

Create subqueries

Can be expensive over large data



# Calculations

Choice

```
case [Choose]
  when "A" then SUM([Sales Amount])
  when "B" then AVG([Discount Qty])
  when "C" then MIN([Sales Qty])
end
```

The calculation is valid.

Sheets Affected Apply OK

**We always evaluate all of them!**

```
SELECT SUM([FactSales].[SalesAmount]) AS [TEMP(Calculation_2537215499680038912)],
  (MIN([FactSales].[SalesQuantity])) AS [TEMP(Calculation_2537215499680038912)],
  COUNT_BIG([FactSales].[DiscountQuantity]) AS [TEMP(Calculation_2537215499680038912)],
  SUM(CAST([FactSales].[DiscountQuantity]) as BIGINT)) AS
[TEMP(Calculation_2537215499680038912)]
FROM [dbo].[FactSales] [FactSales]
GROUP BY ()
```

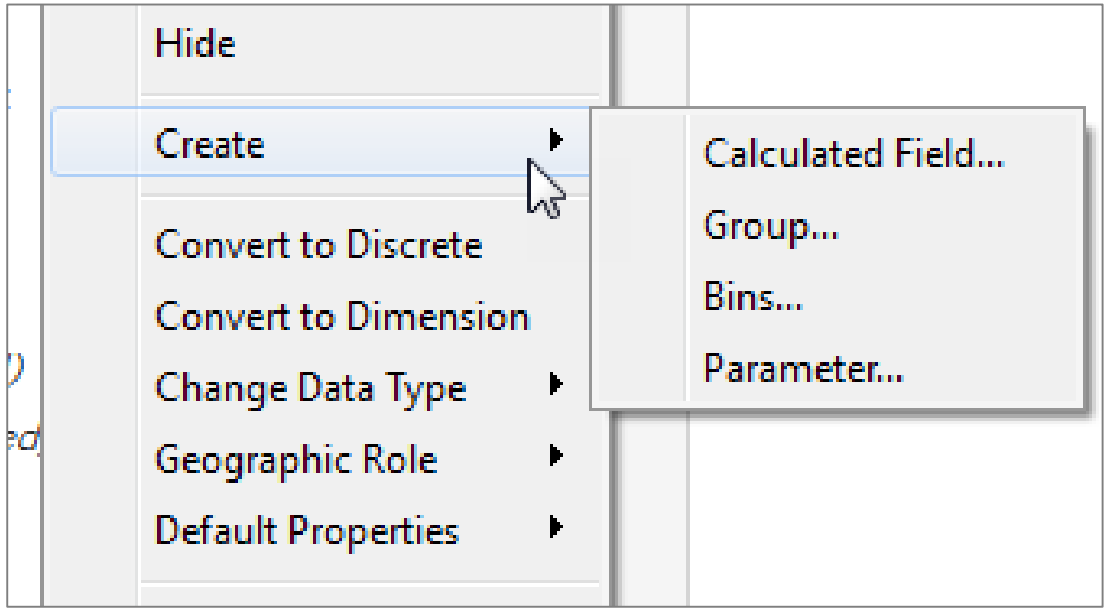
# Calculations

```
case [Parameters].[Data]
  when 'Enrollment (Headcount)' then [Count]
  when 'Enrollment (Seats)' then if count({Fixed [SurCl], [Sur P] :
max([Sur P])})> 10 then count({Include [SurCl], [Sur P]: max([Sur P])})
END
  when 'Headcount (Running Sum)' then RUNNING_SUM(Sum(IF {Fixed [SurP (usv
FACT STDNT ENRL)], [Filter 3]: MIN([SurT (usv FACT STDNT ENRL)])} =
              {Fixed [SurT (usv FACT STDNT ENRL)], [Filter 3]: MIN([SurT
(usv FACT STDNT ENRL)])} THEN 1 ELSE 0 END))
  when 'Class Sections' then countd([SurCl])
  when 'Instructors' then countd([Instructor ID])
  when 'FTE' then round(sum({Fixed [Sur PIYCT], [SurCl]: min([Units
Taken])/15}),1)
  when 'Weighted SCH' then round(sum({Fixed [Sur PIYCT], [SurCl]:
min([Weighted SCH])}),0)
  when 'Tuition and Fees' then round(sum({Fixed [Sur PIYCT]: min([Tuition
and Fees])}),0)
  when 'Tuition and Fees (Average)' then round(avg({Fixed [Sur PIYCT] :
min([Tuition and Fees])}),0)
  when 'FTE/Instructor' then round(sum({Fixed [Sur PIYCT], [SurCl]:
max([Units Taken])/15})/countd([Instructor ID]),1)
  when 'Age Value' then round(avg({Fixed [Sur PIYCT]: min([Age
Value])}),1)
  when 'FTE/Enrollment' then round(sum({Fixed [Sur PIYCT], [SurCl]:
min([Units Taken])/15})/[Count],3)
  when 'HS GPA' then round(avg({Fixed [Sur PIYCT] : min(case
[HS_GPA_Exists]
  when 'Y' then [HS GPA]end )}),2)
  when 'ACT_COMP' then round(avg({Fixed [Sur PIYCT] : min(if [ACT_COMP] >
0 then [ACT_COMP] end )}),2)
  when 'Full-time' then round(countd(case [Academic Load]
  when 'Full-time' then [Sur P] end)/[Count],3)   when 'Current Score'
then round(AVG([Cv Current Score]),2)
  when 'DFWI' then  round(countd(case [Dfwi]
  when 1 then [Sur P] end)/[Count],3)
  when 'Math completion this year' then round (COUNTD(case [College Math
Completion]
  when 'Completed College Math this year' then [surp] end)/[Count],3)
  when 'CUR_GPA' then round(avg({Fixed [Sur PIYCT] : min([CUR_GPA])}),2)
  when 'Semester GPA Variance' then round(VARP({Fixed [Sur PIYCT]:
min([CUR_GPA])}),2)
  when 'Cum_GPA' then round(avg({Fixed [Sur PIYCT] : min([Cum_GPA])}),2)
  when 'Good Academic Standing' then round(countd(case [Academic Standing]
  when 'Good' then [Sur P] end)/[Count],3)
  when 'Financial Aid' then round(sum({Fixed [Sur PIYCT]:
min([Amount])}),0)
  when '1-term Retention' then round([Count 1-term ret]/[Count],3)
  when '1-term Retention or Graduation' then round([Count 1-term ret  or
grad]/[Count],3)
  when '1-yr Retention' then round([Count 1-yr ret]/[Count],3)
  when '2-yr Retention' then round([Count 2-yr ret]/[Count],3)
  when '3-yr Retention' then round([Count 3-yr ret]/[Count],3)
  when '4-yr Retention' then round([Count 4-yr ret]/[Count],3)
  when '4-yr Graduation' then round([Count 4-yr grad]/[Count],3)
  when '5-yr Graduation' then round([Count 5-yr grad]/[Count],3)
  when '6-yr Graduation' then round([Count 6-yr grad]/[Count],3)
END
```

# Calculations

Use native features over calculations

Groups
Sets
Bins
Custom date fields
Combined fields
Aliases
Reference lines
Analytics (clustering, trend lines, forecasts)



# Calculations

Consider efficiency of your calculations

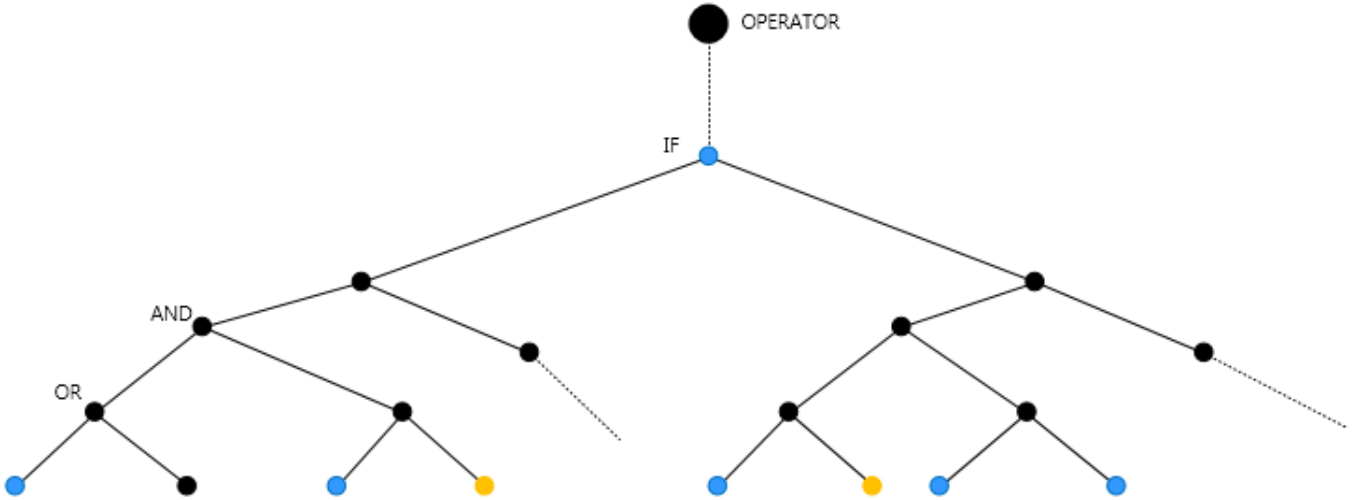
CASE or Groups > IF .. IF … IF…

ELSEIF > ELSE IF

REGEX for strings!

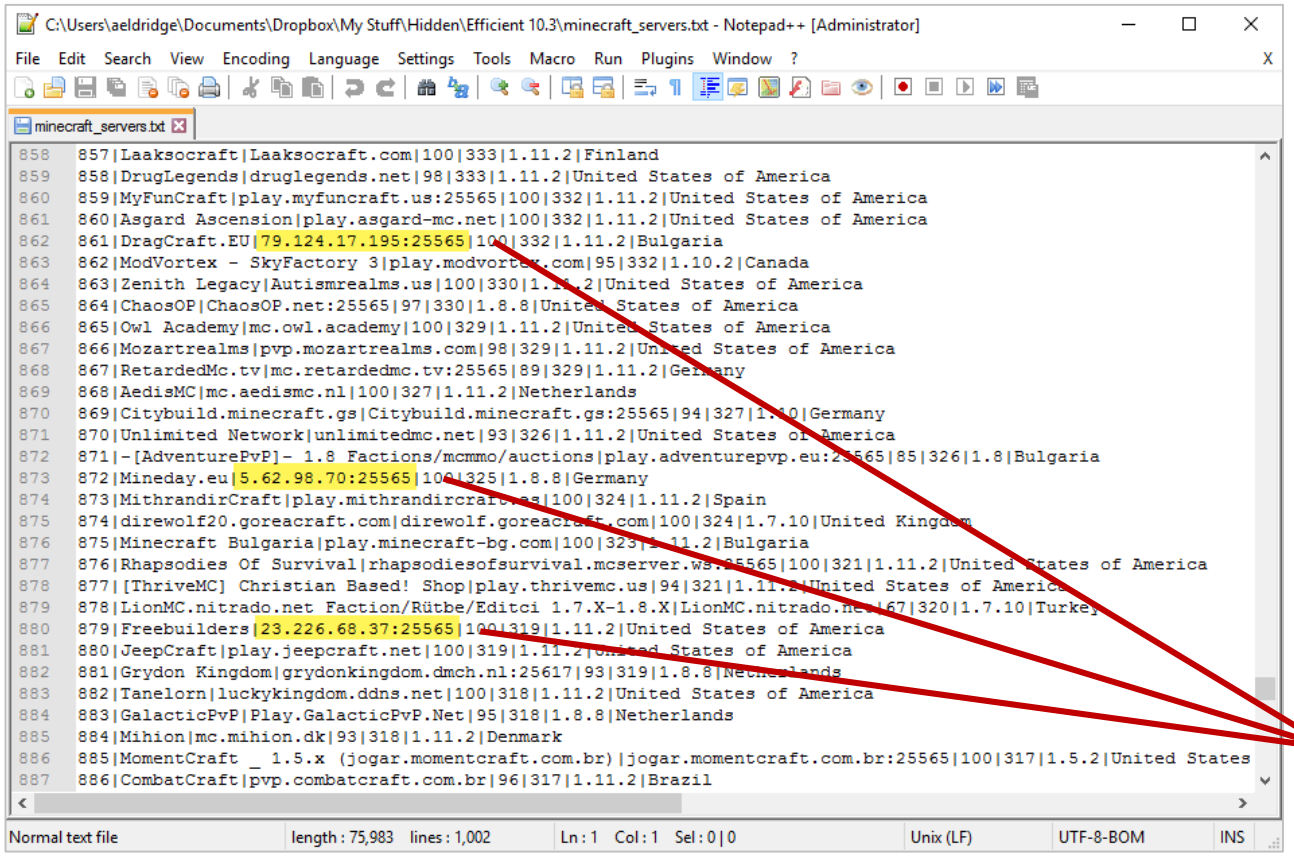
Sets

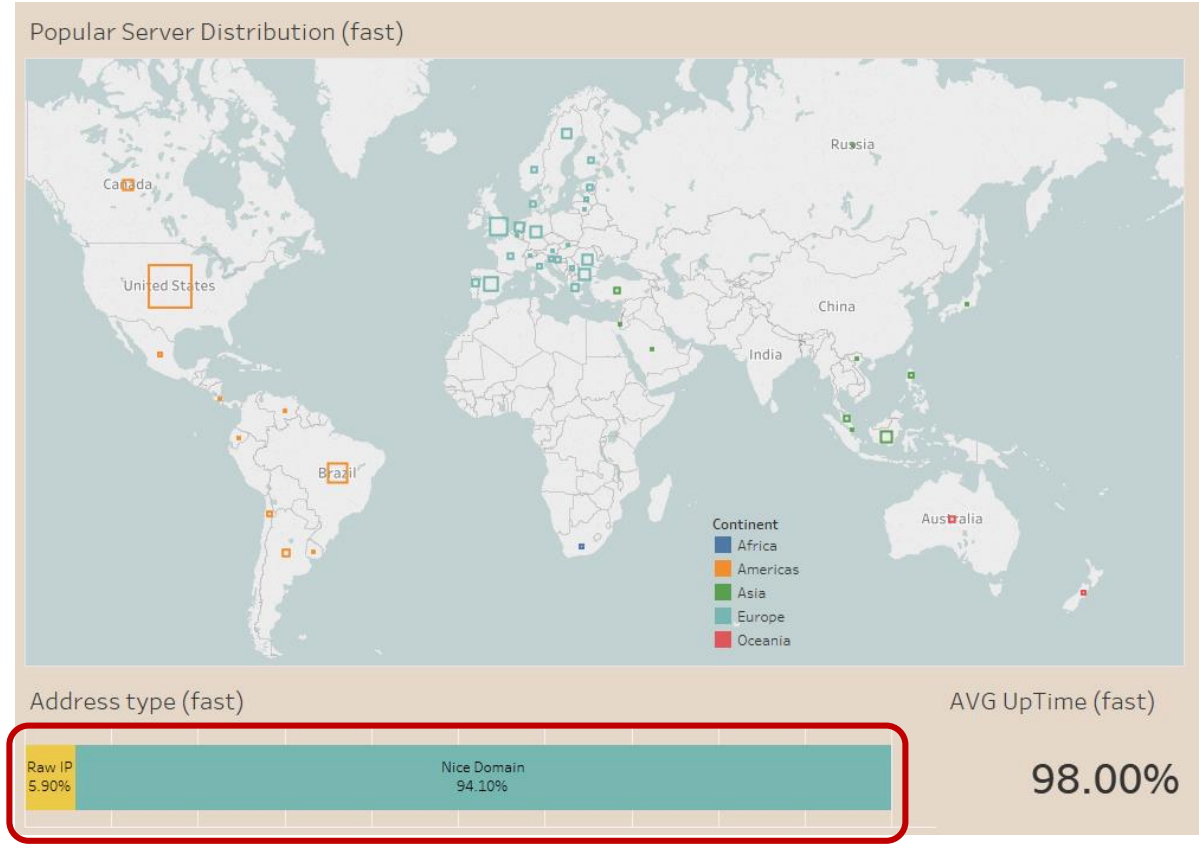
Calculation “Explosion”



# Calculations

## Calculation “Explosion”



# Calculations

## Calculation “explosion”

**Address_IsValidIP** — minecraft_servers

```
NOT(ISNULL([Address_Segment1_AsInt])) AND
NOT(ISNULL([Address_Segment2_AsInt])) AND
NOT(ISNULL([Address_Segment3_AsInt])) AND
NOT(ISNULL([Address_Segment4_AsInt])) AND
[Address_Segment1_AsInt] >= 0 AND
[Address_Segment1_AsInt] <= 255 AND
[Address_Segment2_AsInt] >= 0 AND
[Address_Segment2_AsInt] <= 255 AND
[Address_Segment3_AsInt] >= 0 AND
[Address_Segment3_AsInt] <= 255 AND
[Address_Segment4_AsInt] >= 0 AND
[Address_Segment4_AsInt] <= 255 AND
[Address_Segment5_IsMissing]
```

The calculation is valid. Sheets Affected

**Address_Segment5** — minecraft_servers

**Address_Segment4** — minecraft_server

**Address_Segment3** — minecraft_

**Address_Segment2** — mine

**Address_Segment1**

```
IF FIND([Address_RemoveProtoc
  [Address_RemoveProtocolAndP
ELSE
  MID([Address_RemoveProtocol
END
```

The calculation is valid. Sheets Affected Apply OK

**lress_RemoveProtocolAndPort** — minecraft_servers

**Address_RemoveProtocol** — minecraft_servers

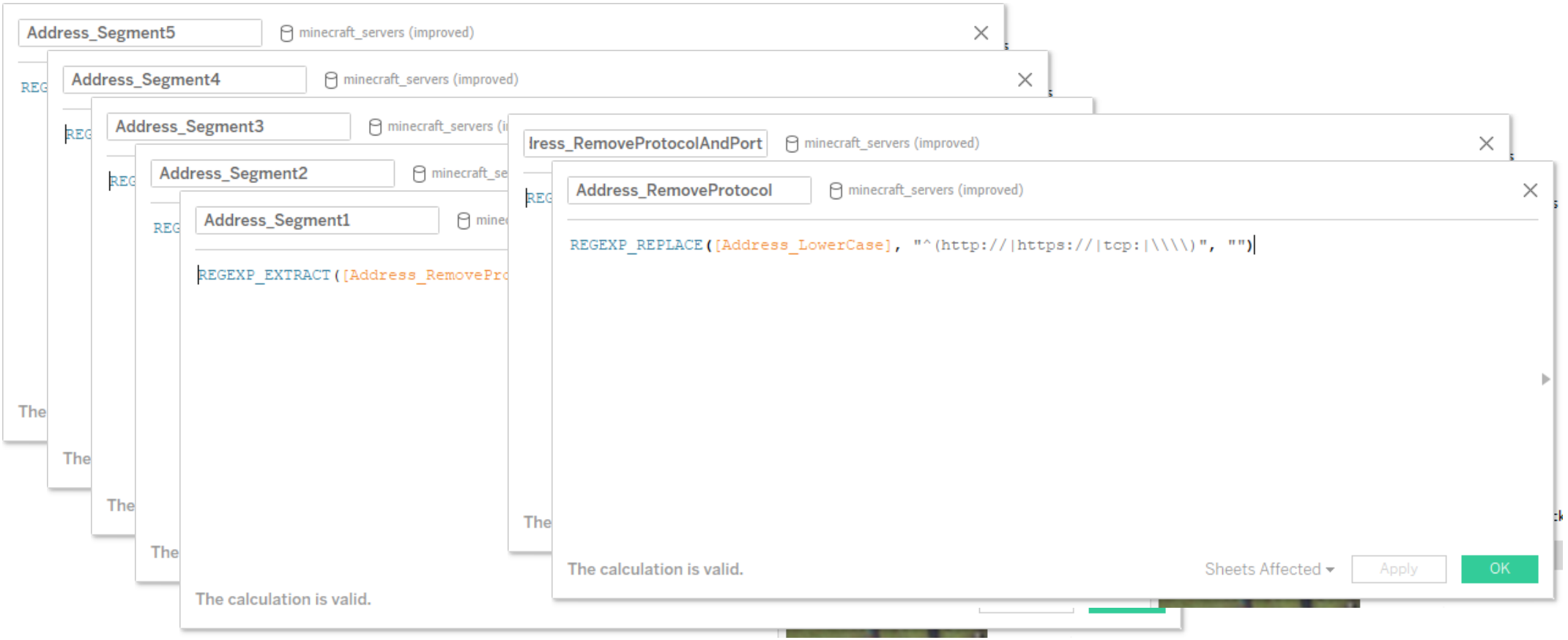
**Address_LowerCase** — minecraft_servers

```
IF STARTSWITH([Address], "{") AND ENDSWITH([Address],"}") THEN
  LOWER(MID([Address], 2, LEN([Address]) - 2))
ELSEIF STARTSWITH([Address], "ip=") THEN
  LOWER(MID([Address], 4, LEN([Address]) - 4))
ELSE
  LOWER([Address])
END
```

The calculation is valid. Sheets Affected Apply OK

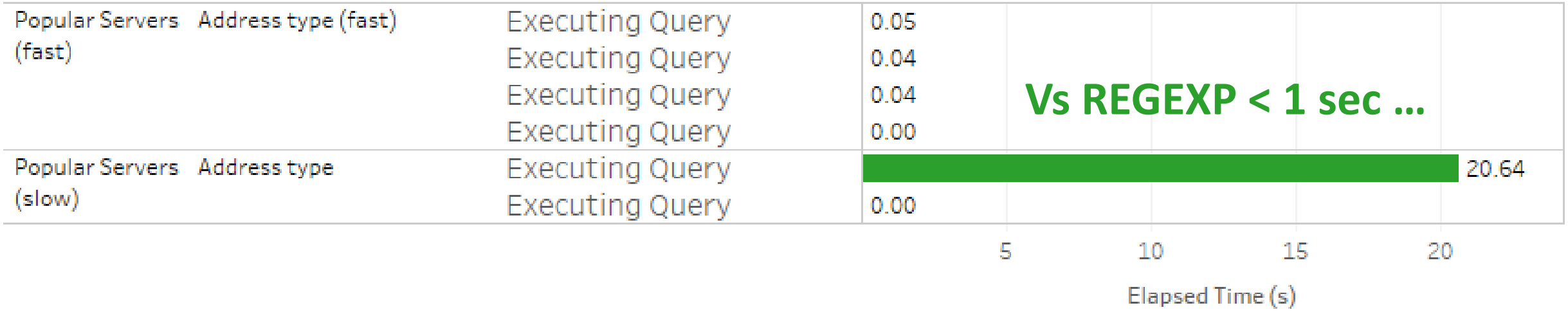
# Calculations

## Calculation efficiency



# Calculations

## Calculation efficiency



# Working Across Data Sources

## Use data blending wisely

## Blending

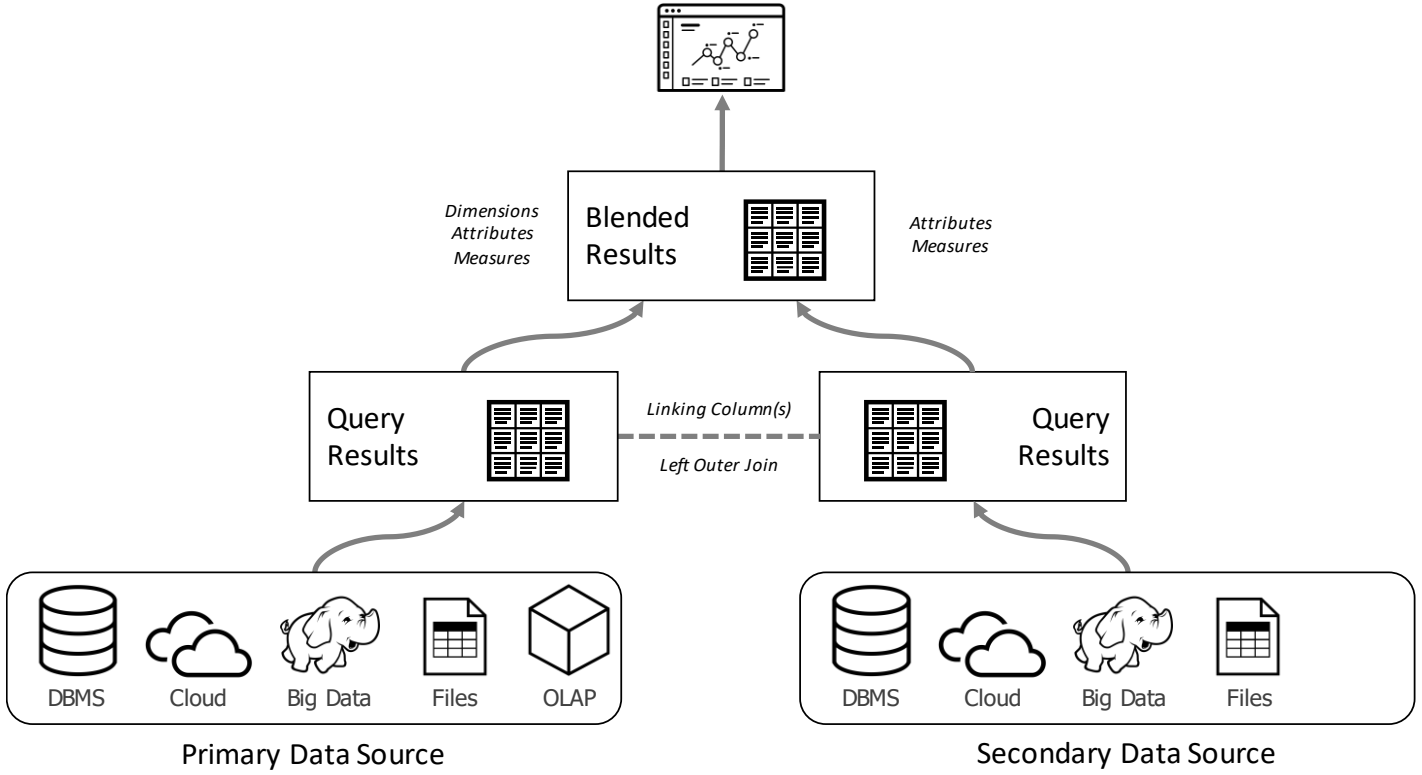
Aggregate, then join

Different to (cross-database) joins which are at the record level

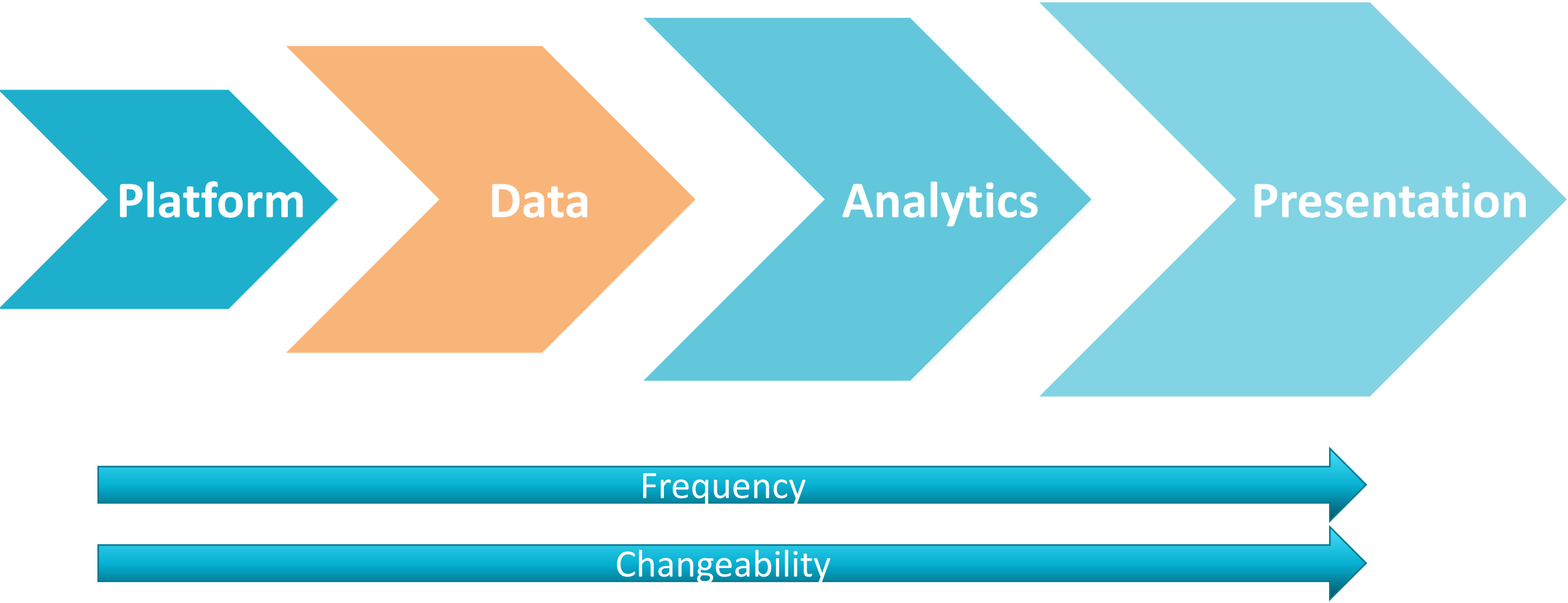
## Don't blend on high-cardinality dimensions

Consider a cross-database join instead

## Use primary groups/aliases to eliminate blending for label lookups



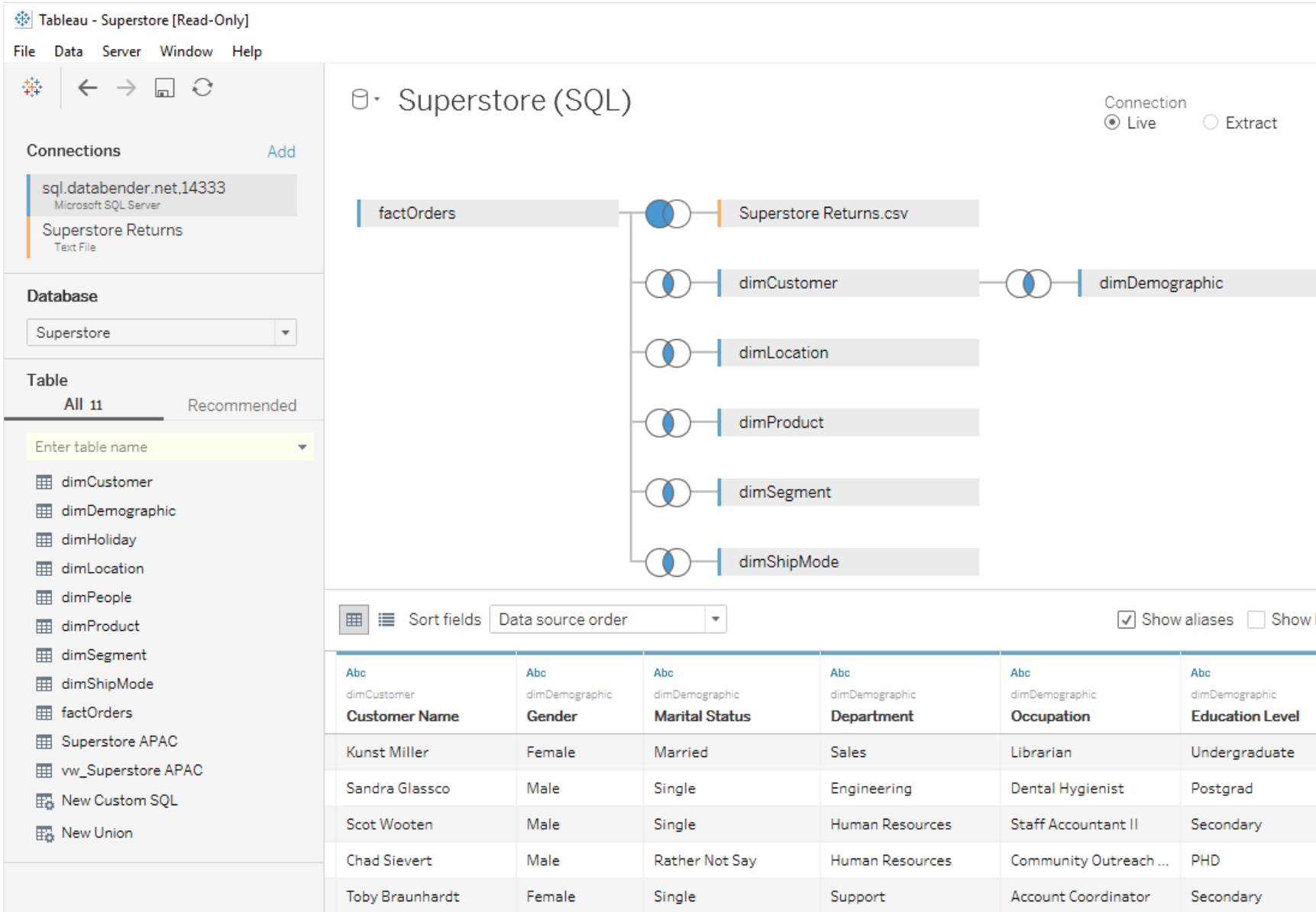
# Data


Platform
Data
Analytics
Presentation
Frequency
Changeability

# Data Layer

## Concerned with:

- Data connections
- Connection models
- Extracts



# Connection Types

**Connection type performance varies**

## Many data source types

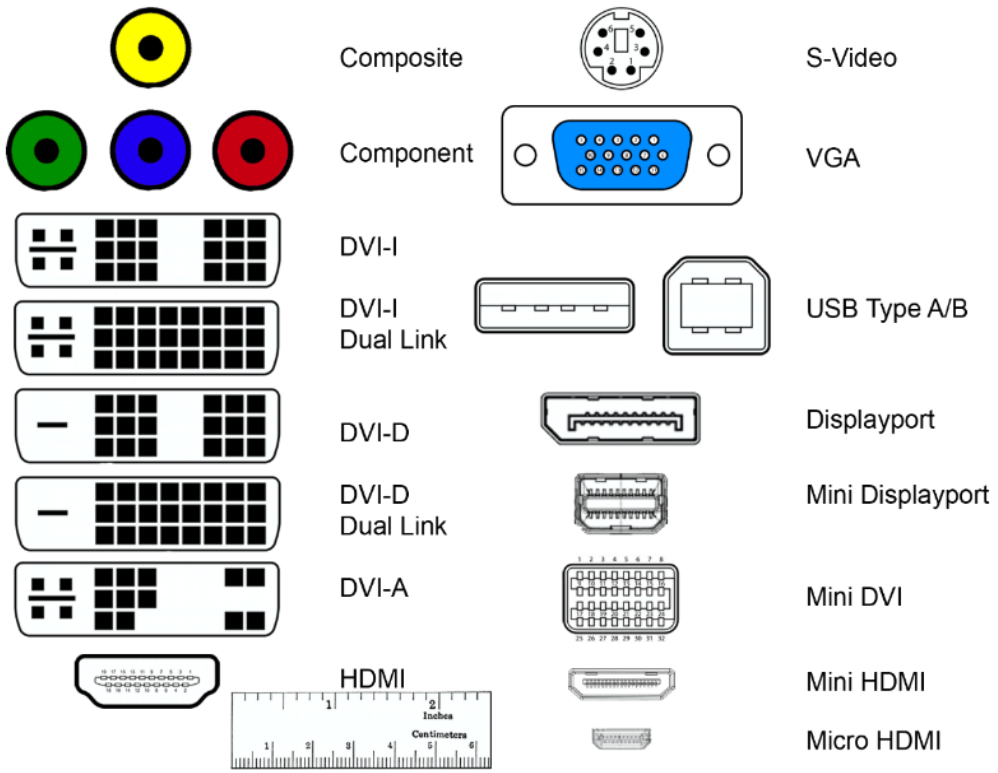
- Not all created equal
- E.g., Native > ODBC

## Embedded > published?

- Embedded can be faster
- Acceleration views

## Direct vs. bridged (Online)

- Query bridge can be extract or live
- Data engine

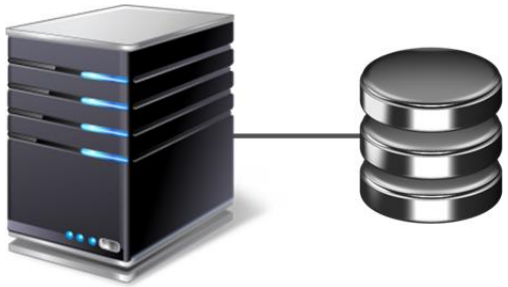


# Data Connections

**Try not to move too much data**

## Having your data in one place is faster than pulling it from multiple sources

- Joining
- Blending
- Data integration (cross-database joins)



# Connection Models

**Trust Tableau to do its job**

## Avoid custom SQL and stored procedures where possible

If you can, deconstruct monolithic custom SQL to be discrete, table-focused custom SQL statements

Use an extract to materialise them so the custom SQL is run only once

# Custom SQL

Quarterly

Custom SQL Query

```
/* TYPE ABCD */
SELECT 'ABCD' AS CATEGORY, REGION, FLAG, YEAR, QUARTER, MEASURE,
MEASURE_ID, AMOUNT
FROM   DB.FACT_TABLE
WHERE  CATEGORY NOT IN ('Not available','Unknown','Not applicable')
  AND (MEASURE_ID IN (301,604))
  AND  AS_AT_DATE = <Parameters.As At Date>
GROUP BY CATEGORY, REGION, FLAG, YEAR, QUARTER, MEASURE, MEASURE_ID

UNION ALL

SELECT 'ABCD' AS CATEGORY, REGION, FLAG, YEAR, QUARTER, 'AA ' ||
MEASURE, MEASURE_ID, AMOUNT
FROM   DB.FACT_TABLE
WHERE  CATEGORY IN (SELECT CATEGORY FROM DB.DIM_TABLE_A)
```

```
/* TYPE WXYZ */
SELECT 'WXYZ' AS CATEGORY, REGION, FLAG, YEAR, QUARTER, MEASURE,
MEASURE_ID, AMOUNT
FROM   DB.FACT_TABLE
WHERE  CATEGORY NOT IN ('Not available','Unknown','Not applicable')
  AND (MEASURE_ID = 302
   OR  MEASURE_ID = 605)
  AND  AS_AT_DATE = <Parameters.As At Date>
GROUP BY CATEGORY, REGION, FLAG, YEAR, QUARTER, MEASURE, MEASURE_ID

UNION ALL

SELECT 'NWXYZ' AS CATEGORY, REGION, FLAG, YEAR, QUARTER, 'CC ' ||
MEASURE, MEASURE_ID, AMOUNT
```

```
FROM DB.DIM_TABLE_A)

t Date>
R, QUARTER, 'CC ' || MEASURE,
```

**Time to open TWB: ~5 mins**

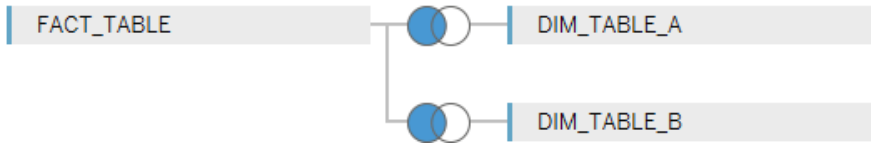
```
FROM   DB.FACT_TABLE
WHERE  CATEGORY IN (SELECT CATEGORY FROM DB.DIM_TABLE_B)
  AND  MEASURE_ID = 504
  AND  AS_AT_DATE = <Parameters.As At Date>
GROUP BY CATEGORY, REGION, FLAG, YEAR, QUARTER, 'BB ' || MEASURE,
MEASURE_ID

UNION ALL
```

```
SELECT 'WXYZ' AS CATEGORY, REGION, FLAG, YEAR, QUARTER, 'DD ' ||
MEASURE, MEASURE_ID, AMOUNT
FROM   DB.FACT_TABLE
WHERE  CATEGORY IN (SELECT CATEGORY FROM DB.DIM_TABLE_B)
  AND (MEASURE_ID = 505)
  AND  AS_AT_DATE = <Parameters.As At Date>
GROUP BY CATEGORY, REGION, FLAG, YEAR, QUARTER, 'DD ' || MEASURE,
MEASURE_ID
```

# Custom SQL

Quarterly




Category 2

Measure 2

```
//* Type ABCD */
if ([category] <> 'Not available' and [category] <> 'Unknown' and [category] <> 'Not applicable')
    and ([Measure Id] = 301 or [Measure Id] = 604)
    then [Measure]
elseif not ISNULL([category (DIM TABLE A)])
    and [Measure Id] = 504
    then 'AA ' + [Measure]
elseif not ISNULL([category (DIM TABLE B)])
    and [Measure Id] = 504
```

```
    then 'DD ' + [Measure]
end
```

The calculation is valid.

Apply OK

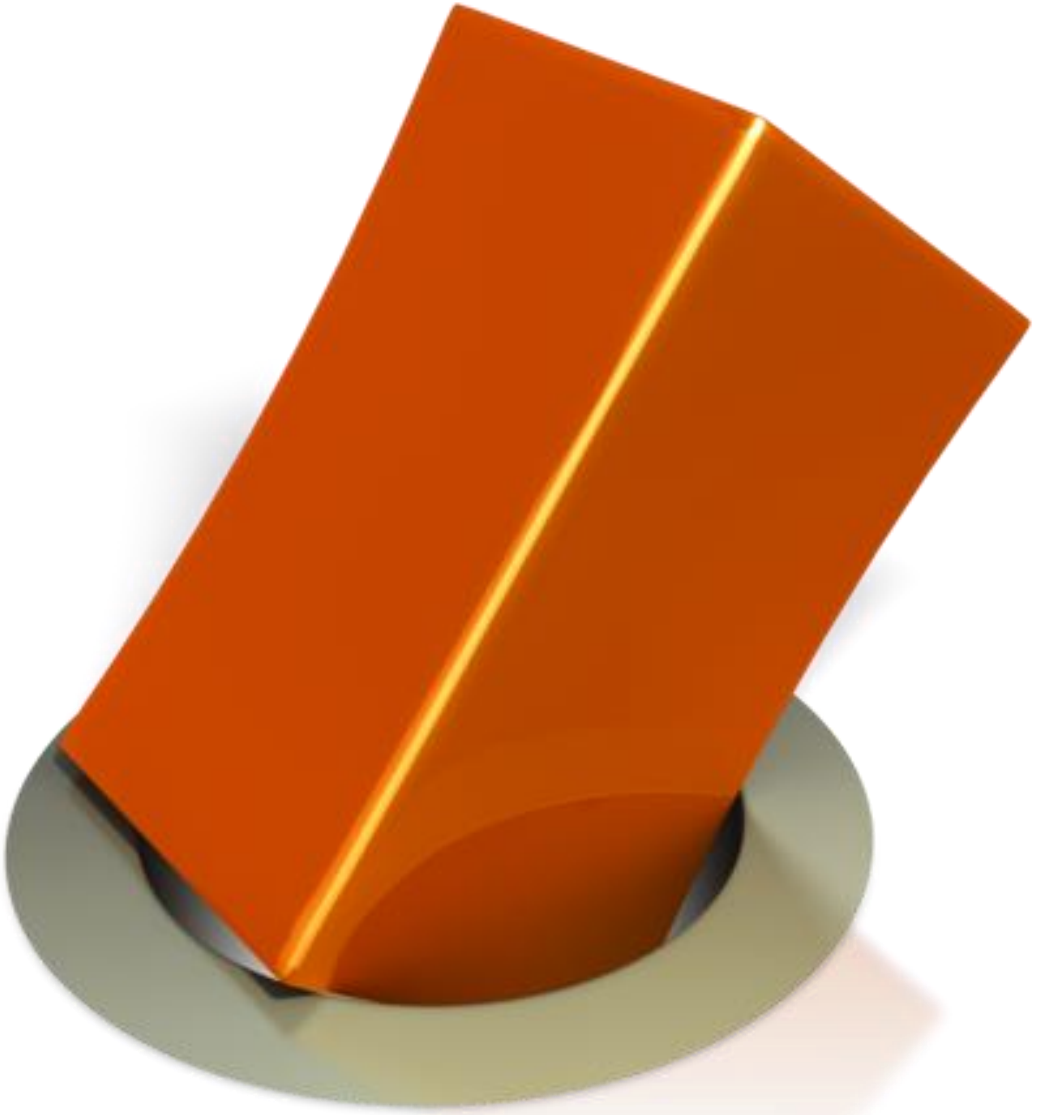
**Time to open TWB: < 10 secs (30x!)**

# Connection Models

**Try to have your data in the best "shape" for your analysis**

## The less you have to manipulate your data, the faster it will be

- Union
- Pivot
- Calculations



# Data Connections

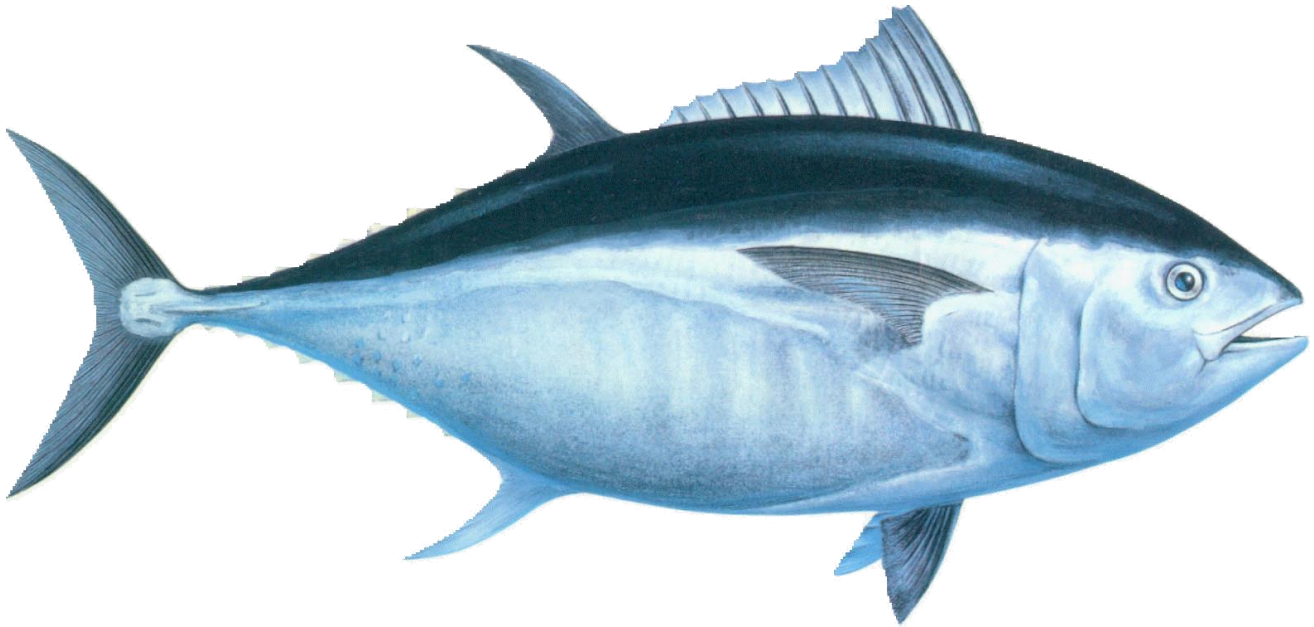
**Tune your database for optimal performance**

**Indexes on joining/filtering dimensions**

**Define columns as NOT NULL when possible**

**Referential integrity = join culling**

**If no “hard” RI, then “Assume Referential Integrity” on Data menu**



# Connection Models

**Use Data Server for governance**

**Have data experts optimise the connections; share with business users as published data connections**

**Share extracts across multiple workbooks**

# Connection Models

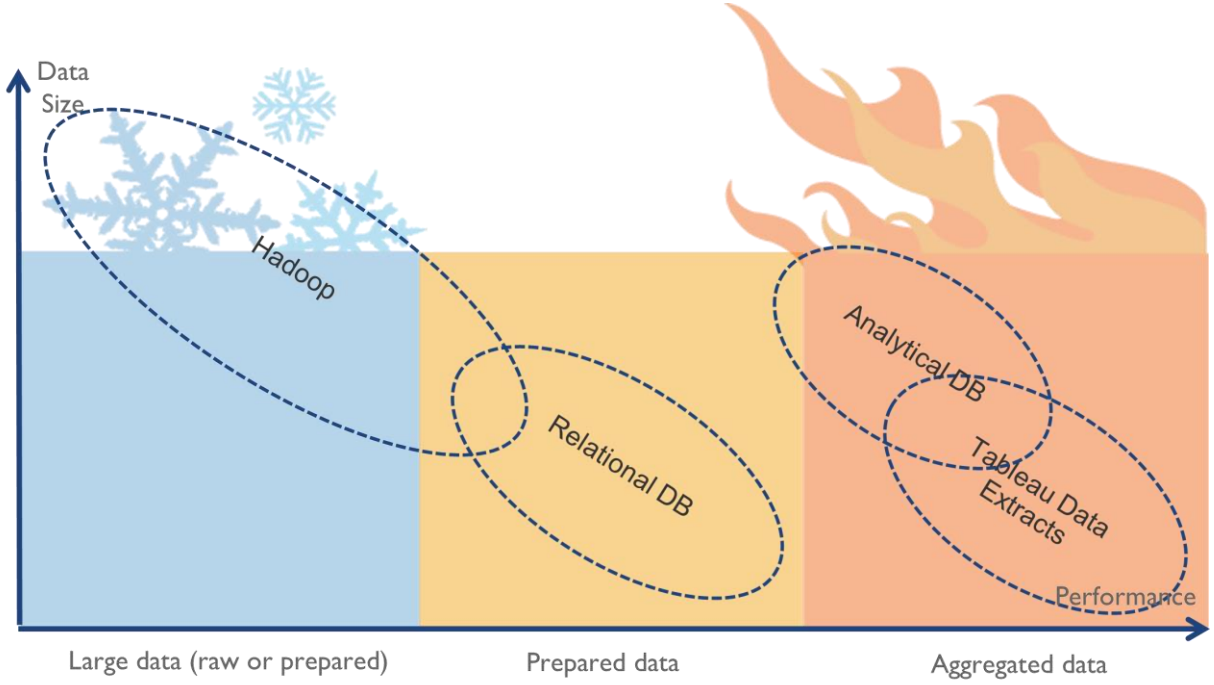
**Use cold/warm/hot strategies for big data**

**Aggregated extracts for high-level analysis**

**Live/extract connection to DBMS for mid-level analysis**

**Live connection to Big Data for detail reports**

**Start at high level and drill down to detail**



# Extracts

Extracts are an easy way to make things go faster

**Hide unused fields!**

**Aggregated**

**Filtered/sampled**

**Materialise row-level calculations**

# Extracts

**Hyper data engine keeps getting better!**

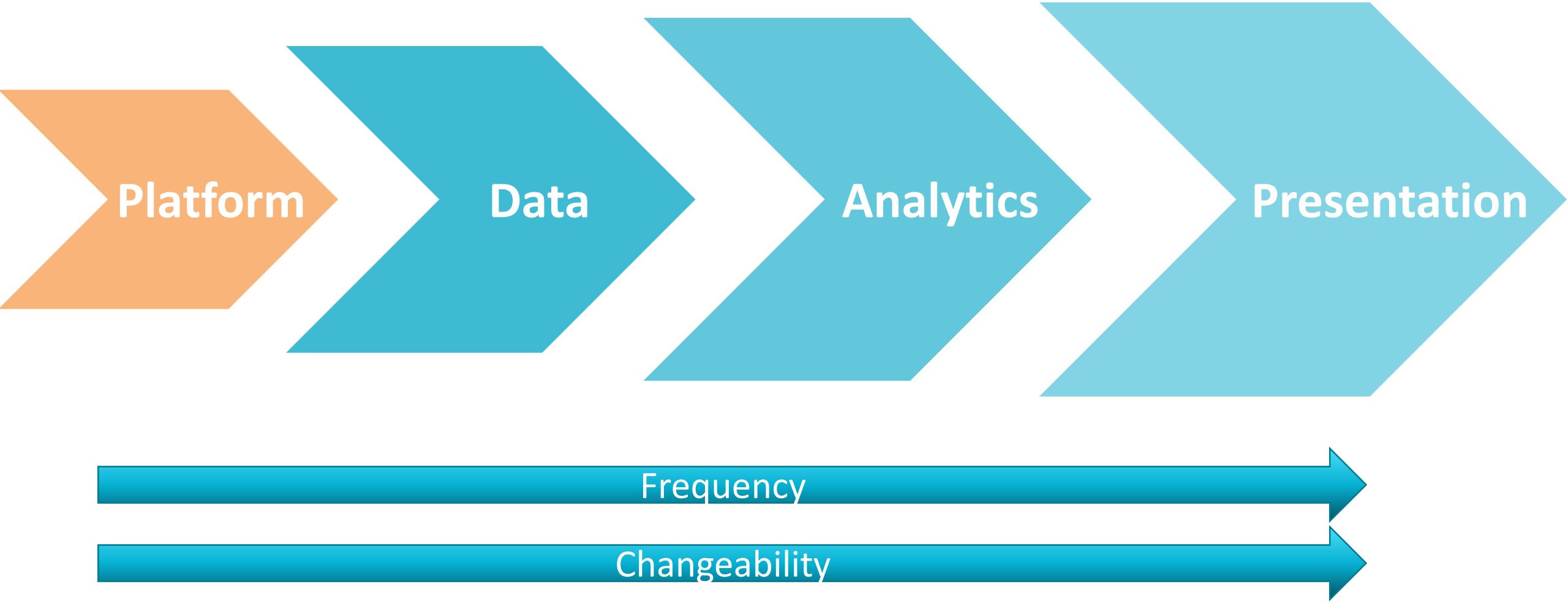
**Faster extract creation and refresh**

**Larger extracts**

**Improved query performance**

**Normalised extracts!**

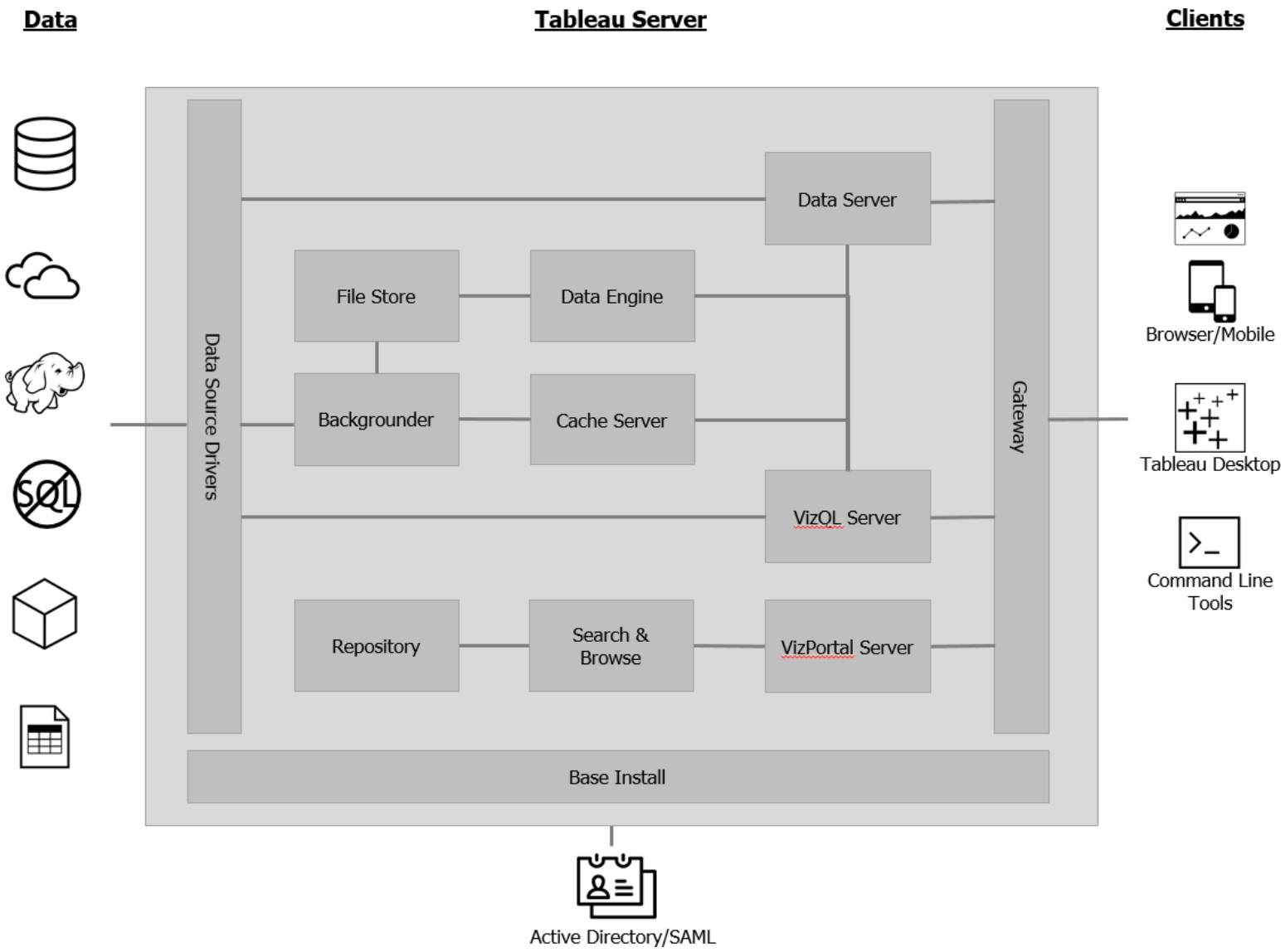
# Concept: Visual Pipeline



# Platform Layer

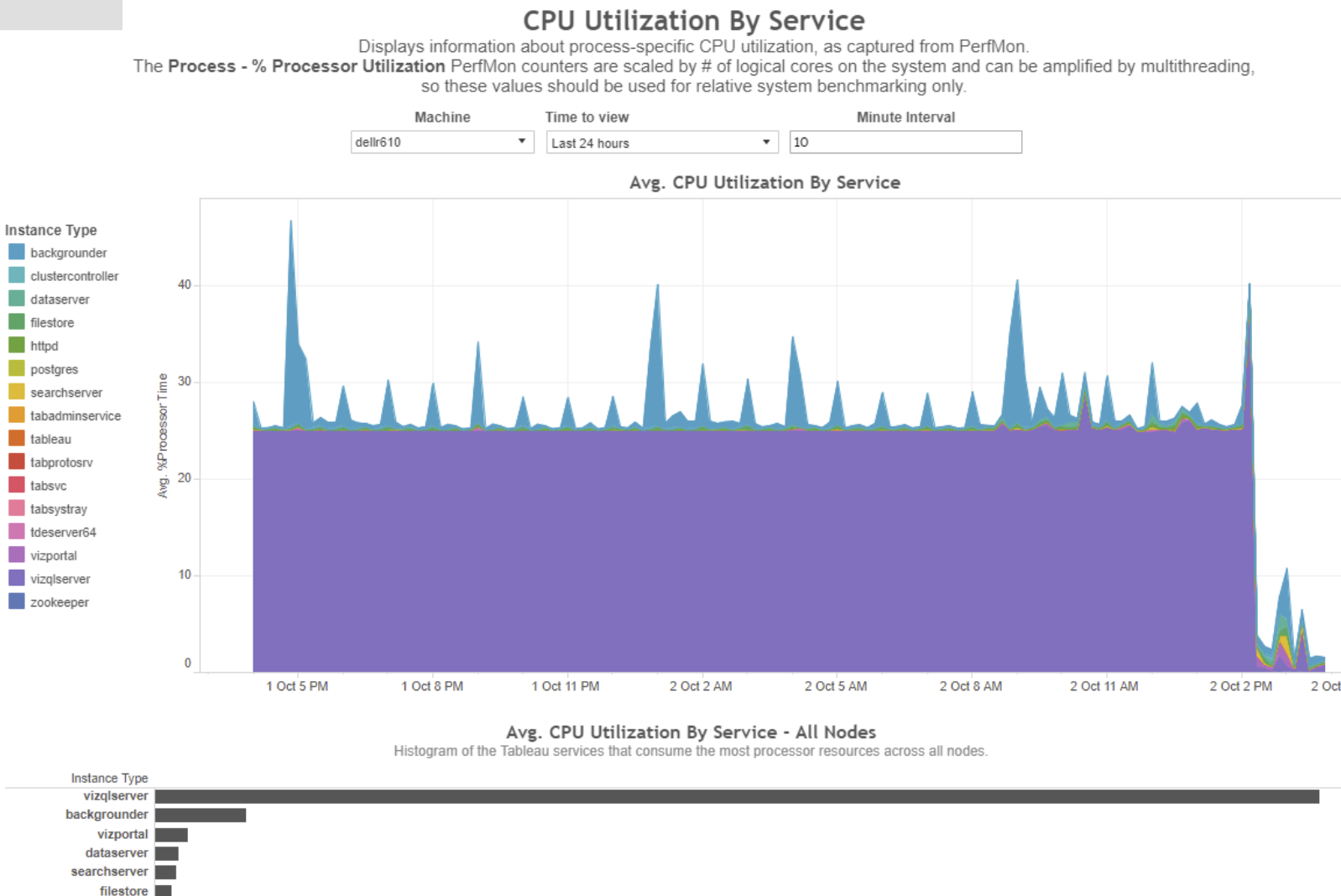
## Concerned with:

- Environment configuration
- OS
- Hardware



# Environment

## Monitor!



# Environment

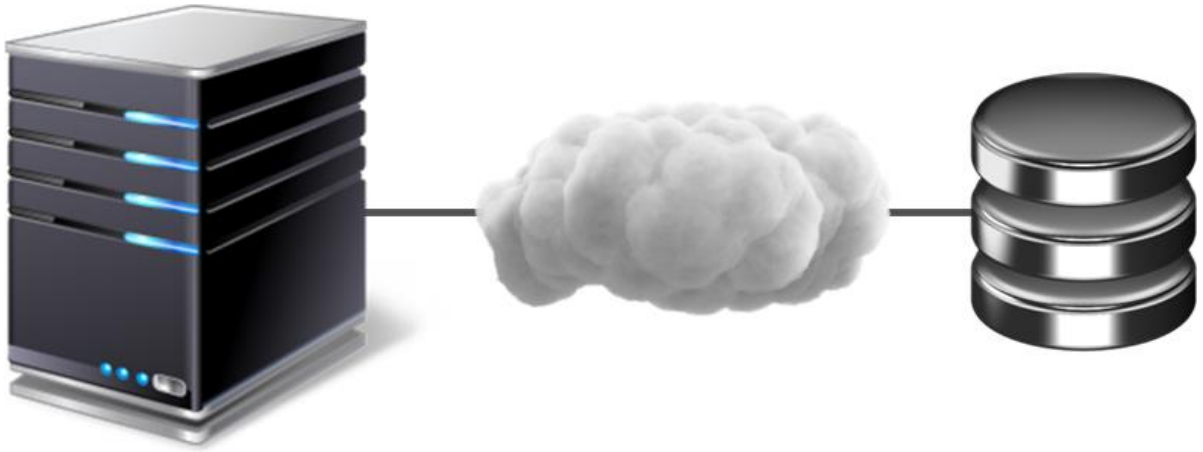
**Be prepared to test both on the server and remotely**

## On the server:

- Closer to the data
- Find configuration, data and workbook design issues

## Remote to the server:

- True end-user experience
- Find network issues



**Slow on Desktop ➔ Slow on Server**

# Environment

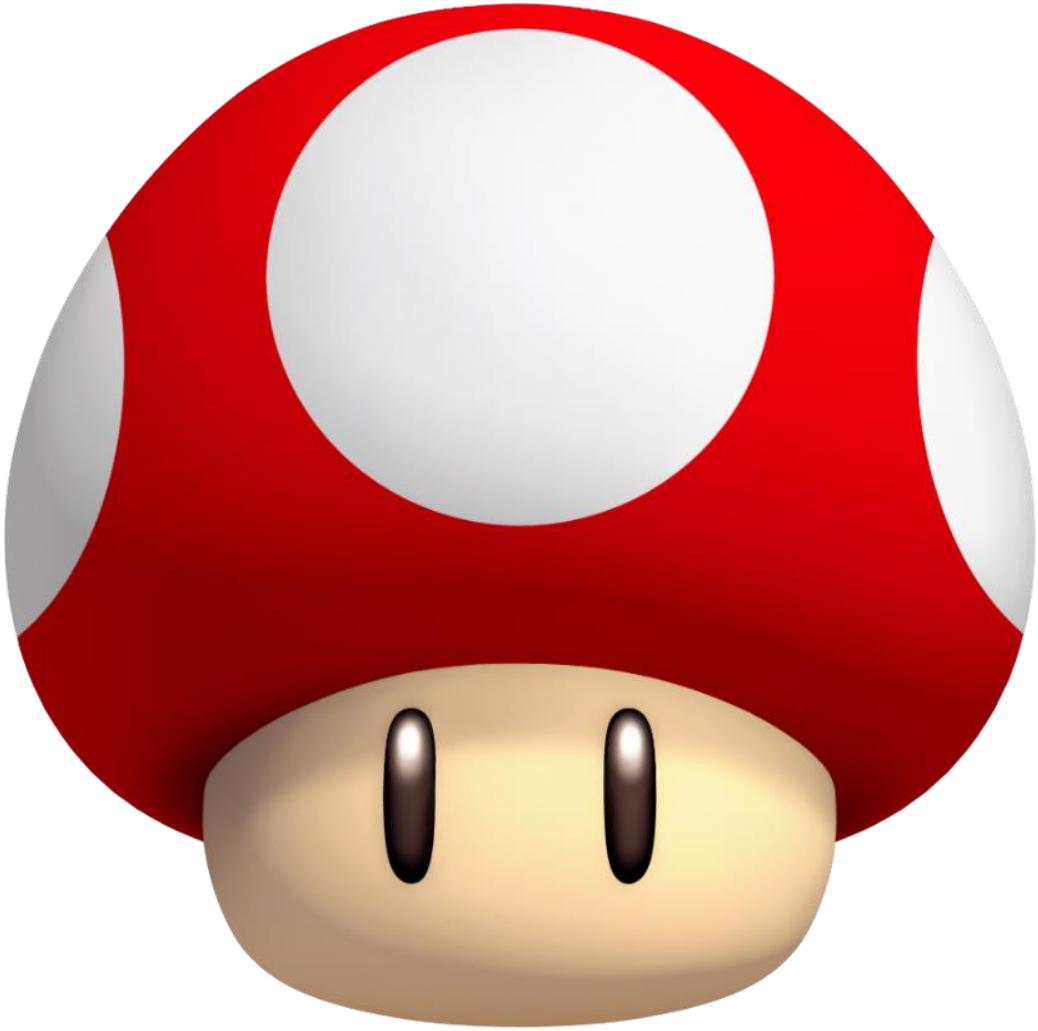
**Upgrade!**

## Software:

- Desktop and Server
- Performance improvements
- Bug fixes

## Hardware

- RAM, CPU
- Fast disk, esp. with extracts
- Fast network



# Environment

**You gotta keep ‘em separated**

## Keep interactive users and extract refreshes separated

- VizQL vs. backgrounder
- Consider a dedicated extract processing node

# Environment

## Know virtualisation

Don’t run on oversubscribed virtual machines (CPU or RAM)

Virtual has a performance overhead (10-20%) vs physical infrastructure

If using cloud VMs (AWS, Azure, GCP) use the right instance and disk types

# QUESTIONS?

Thank you !

# Agenda

**09:30 – 10:00 -** Updates on the Tableau Community at BNP Paribas

**10:00 – 10:30 -** Newest and upcoming features of Tableau

**10:30 – 10:45 -** Break

**10:45 – 12:15 -** BNPP Testimonials: BNP Paribas Fortis, BDDF, Real Estate

**12:15 – 13:30 -** Lunch

**13:30 – 14:45 -** Workshop of your choice
Tableau Desktop Hands-on (beginner) **OR** Designing efficient workbooks (advanced)

**15:00 – 16:15 -** Workshop of your choice
Tableau Prep Hands-on (beginner) **OR** Data Modelling (advanced)