



**Optimize Model Performance**

# Learning Objectives

By the end of this lesson, you will be able to:

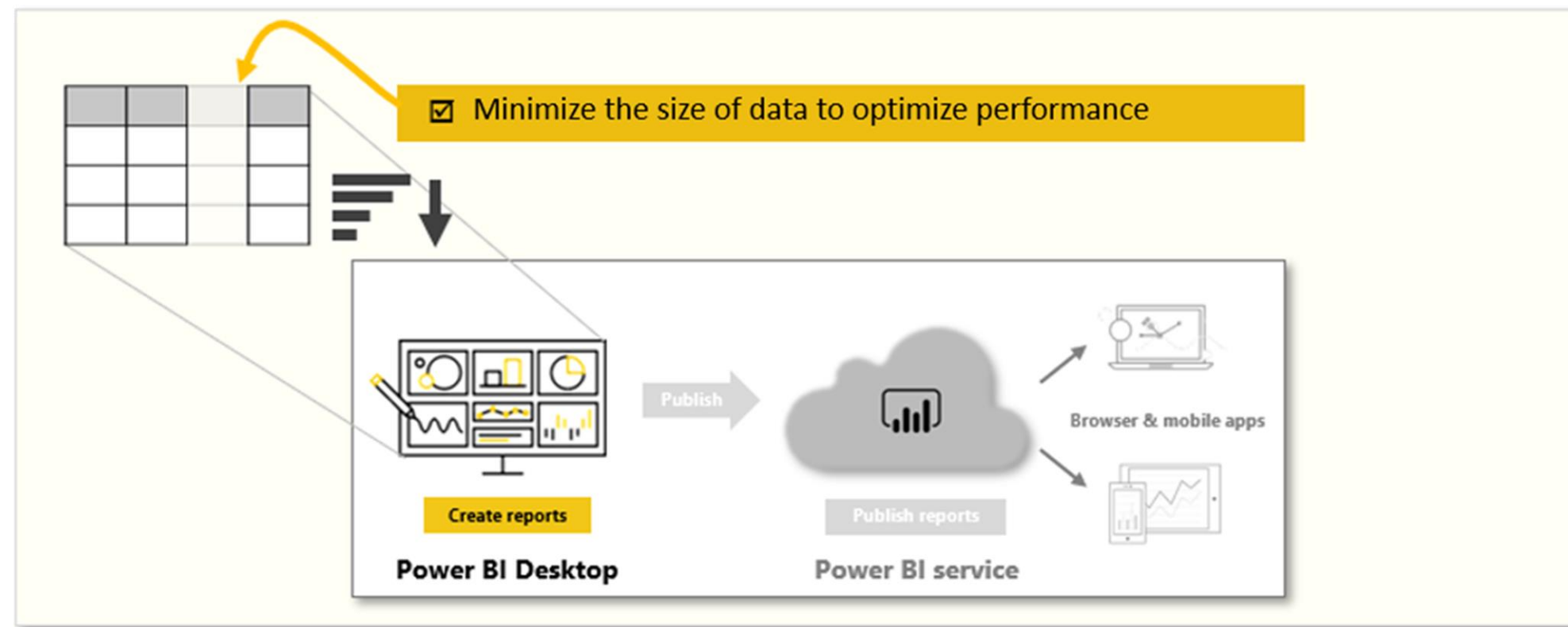
- 🕒 List the tasks involved in performance optimization
- 🕒 Review performance results
- 🕒 Identify and reduce cardinality levels
- 🕒 Create and manage aggregations



## Data Model Performance Optimization

# Introduction to Performance Optimization

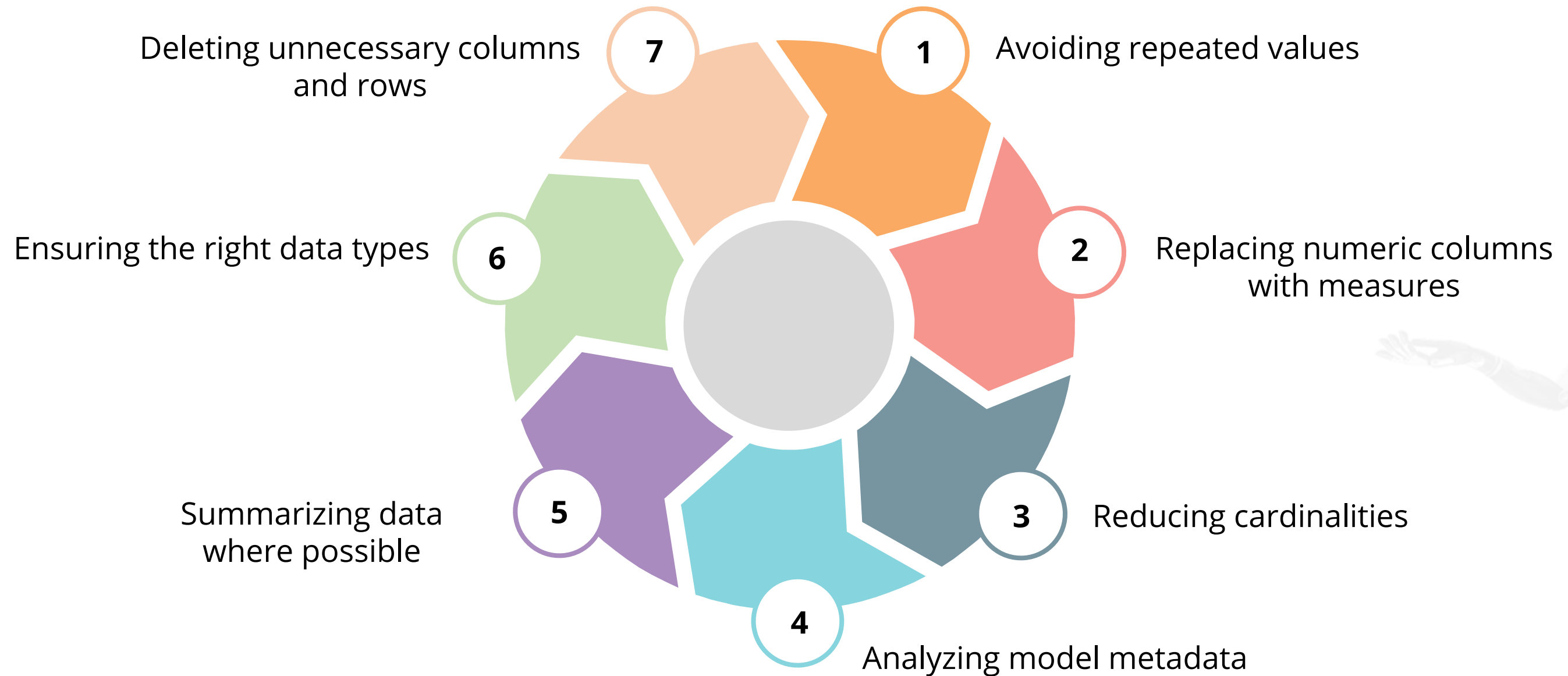
Performance optimization involves making changes to the current state of the data model to run more efficiently.



An optimized data model performs better during execution.

# Introduction to Performance Optimization

It involves minimizing the size of the data model that includes:





# Performance Analyzer

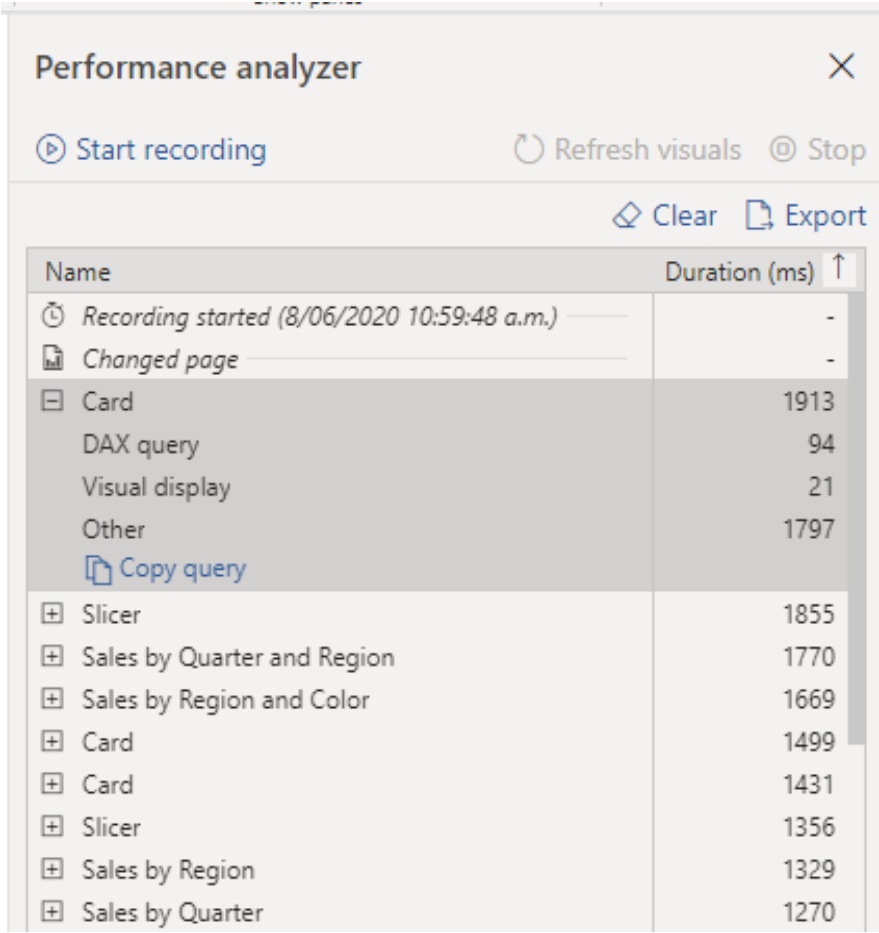
Performance analyzer identifies measures, relationships, and visuals whose performance is poor. These are later optimized for improving the performance.



- Identify the performance of each report element
- Measure report elements during user interaction
- Detect which aspects are least or most resource intensive

# Review Performance Results

The performance test results are reviewed in the performance analyzer pane.



The Performance analyzer pane displays a table of tasks and their durations. The table has two columns: 'Name' and 'Duration (ms)'. The tasks are listed in descending order of duration. The first task is 'Recording started (8/06/2020 10:59:48 a.m.)' with a duration of '-'. The second task is 'Changed page' with a duration of '-'. The third task is 'Card' with a duration of 1913 ms. The fourth task is 'DAX query' with a duration of 94 ms. The fifth task is 'Visual display' with a duration of 21 ms. The sixth task is 'Other' with a duration of 1797 ms. The seventh task is 'Slicer' with a duration of 1855 ms. The eighth task is 'Sales by Quarter and Region' with a duration of 1770 ms. The ninth task is 'Sales by Region and Color' with a duration of 1669 ms. The tenth task is 'Card' with a duration of 1499 ms. The eleventh task is 'Card' with a duration of 1431 ms. The twelfth task is 'Slicer' with a duration of 1356 ms. The thirteenth task is 'Sales by Region' with a duration of 1329 ms. The fourteenth task is 'Sales by Quarter' with a duration of 1270 ms.

Name	Duration (ms)
Recording started (8/06/2020 10:59:48 a.m.)	-
Changed page	-
Card	1913
DAX query	94
Visual display	21
Other	1797
Slicer	1855
Sales by Quarter and Region	1770
Sales by Region and Color	1669
Card	1499
Card	1431
Slicer	1356
Sales by Region	1329
Sales by Quarter	1270

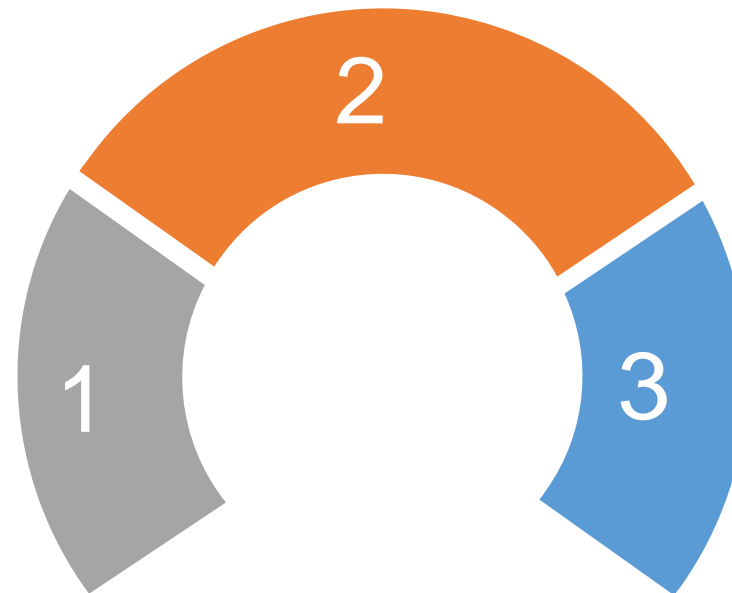
To review the tasks in the order of duration, right-click the **Sort** icon next to the **Duration (ms)** column header, and then select **Total time** in descending order.

# Review Performance Results

The log information shows the different durations to complete the following tasks:

## DAX Query

The total time taken by the visual to send a query and the analysis services to return the results



## Other

The time taken for the visual to prepare queries, wait for other visuals to complete, or perform other background processing tasks

## Visual Display

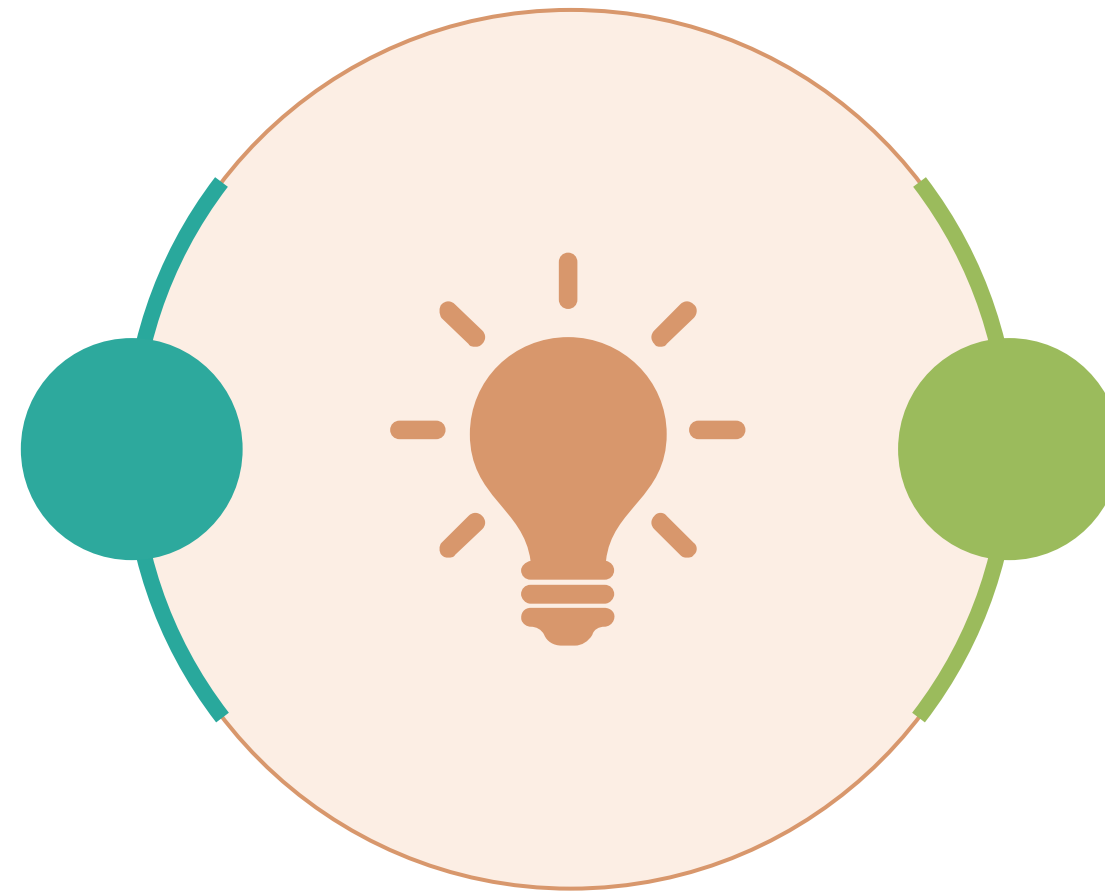
The time taken for the visual to render on screen



# Removal of Unnecessary Rows and Columns

Power Query includes the below options:

**Unnecessary Columns**  
Evaluates the need for each column



**Unnecessary Rows**  
Checks the first few rows for null values or any unnecessary data

## Cardinality Reduction

# Cardinality

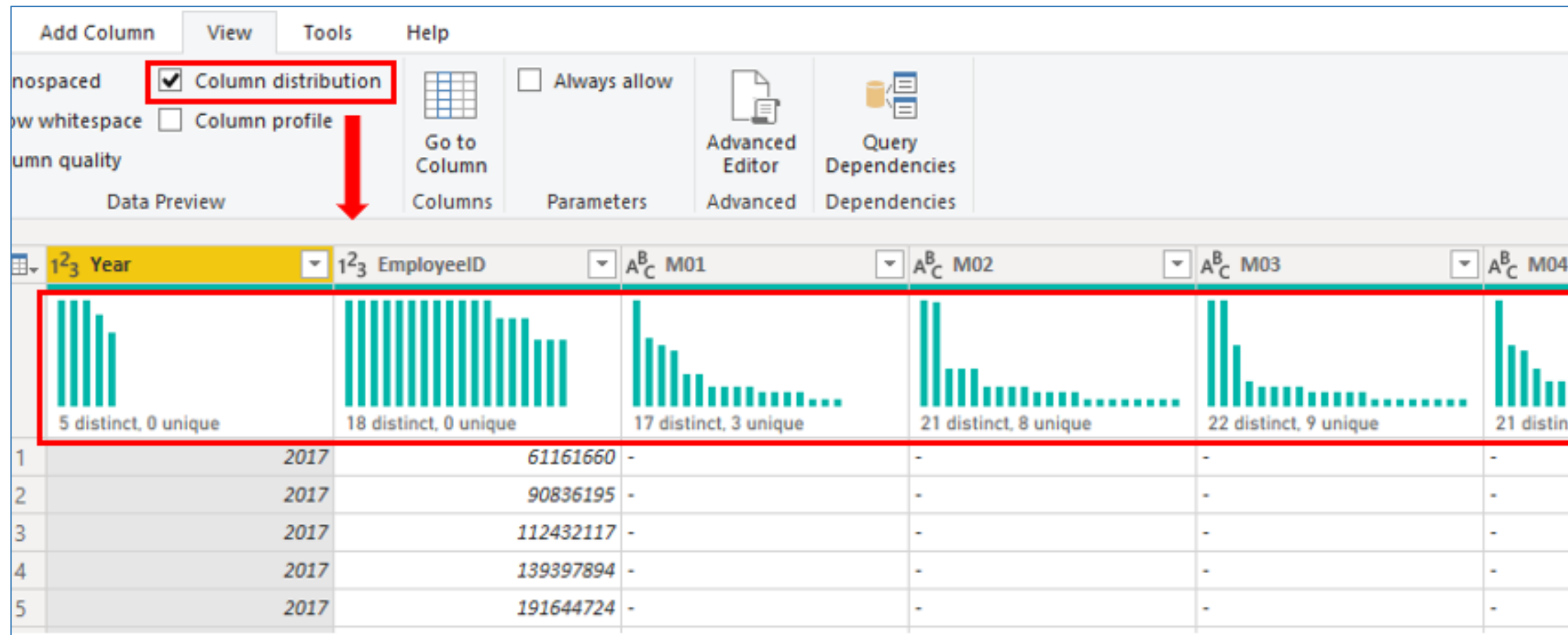
Cardinality describes the uniqueness of the values in a column.



It is also used in the context of the relationships between two tables to describe its direction.

# Cardinality Levels

The column distribution on the **View** tab displays statistics on unique and distinct items present in each column of the data.



# Cardinality Levels

Values distributed in a column are of two types:

## Distinct Values Count

- Number of different values in a column

## Unique Values Count

- Number of values that only appear once in a column

A column that has a lot of repeated values in this range has low level of cardinality.

A column that has a lot of unique values in this range has high level of cardinality.

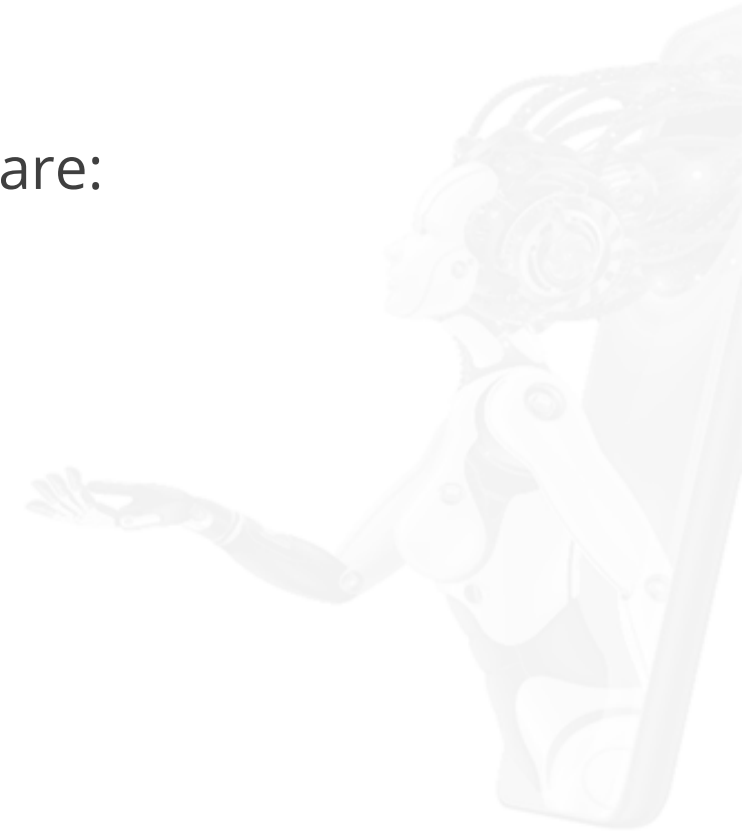
# Cardinality Relationship

Relationships between the imported tables are necessary to calculate accurate results and depict the correct information.



The different cardinality options are:

- Many-to-one (\*:1)
- One-to-one (1:1)
- One-to-many (1:\*)
- Many-to-many (\*:\*)

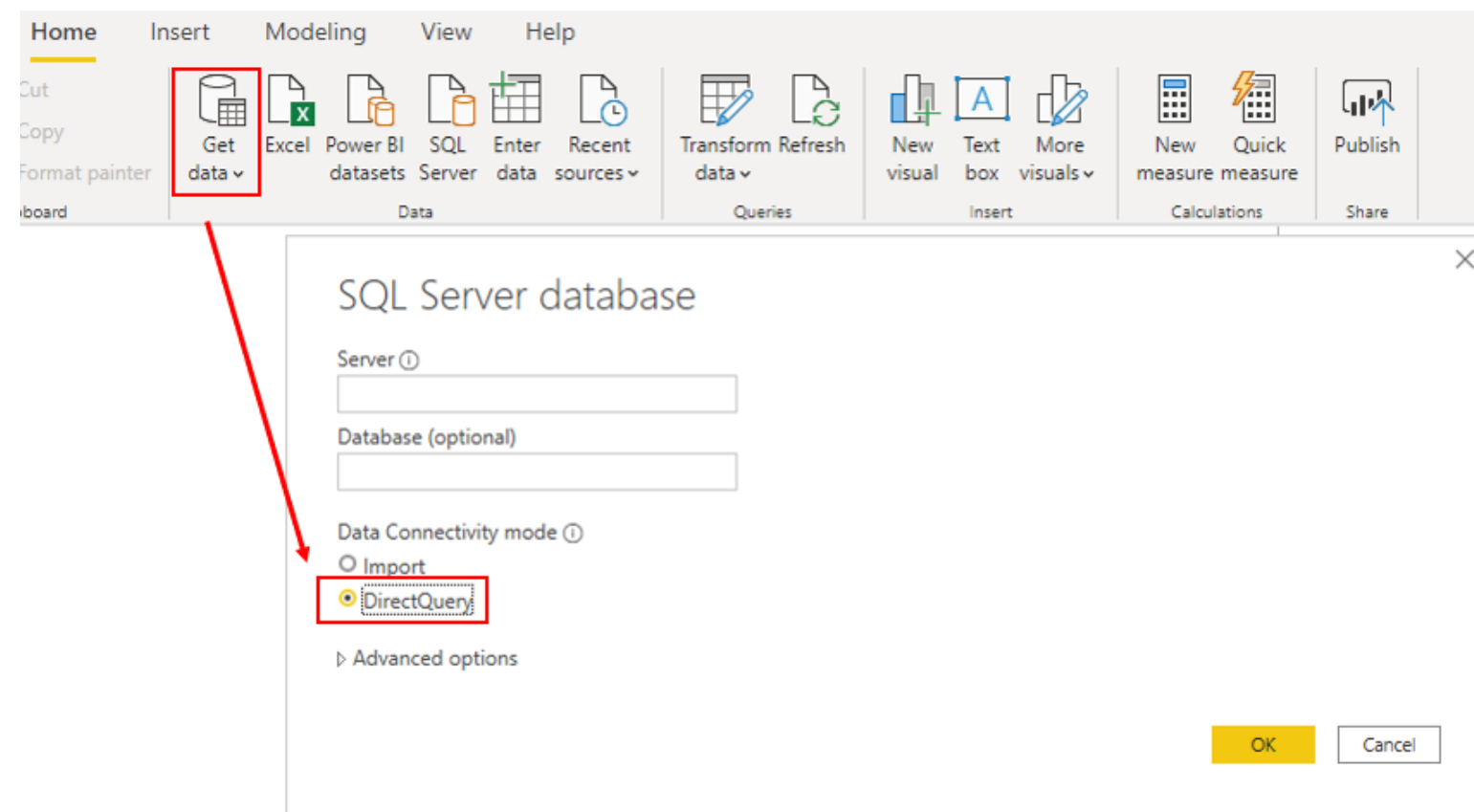




## Optimize DirectQuery Model

# Introduction to DirectQuery

DirectQuery is a method of connecting data directly to its source repository from the Power BI Desktop.



It is an alternative to importing data into Power BI Desktop.

# Benefits of DirectQuery

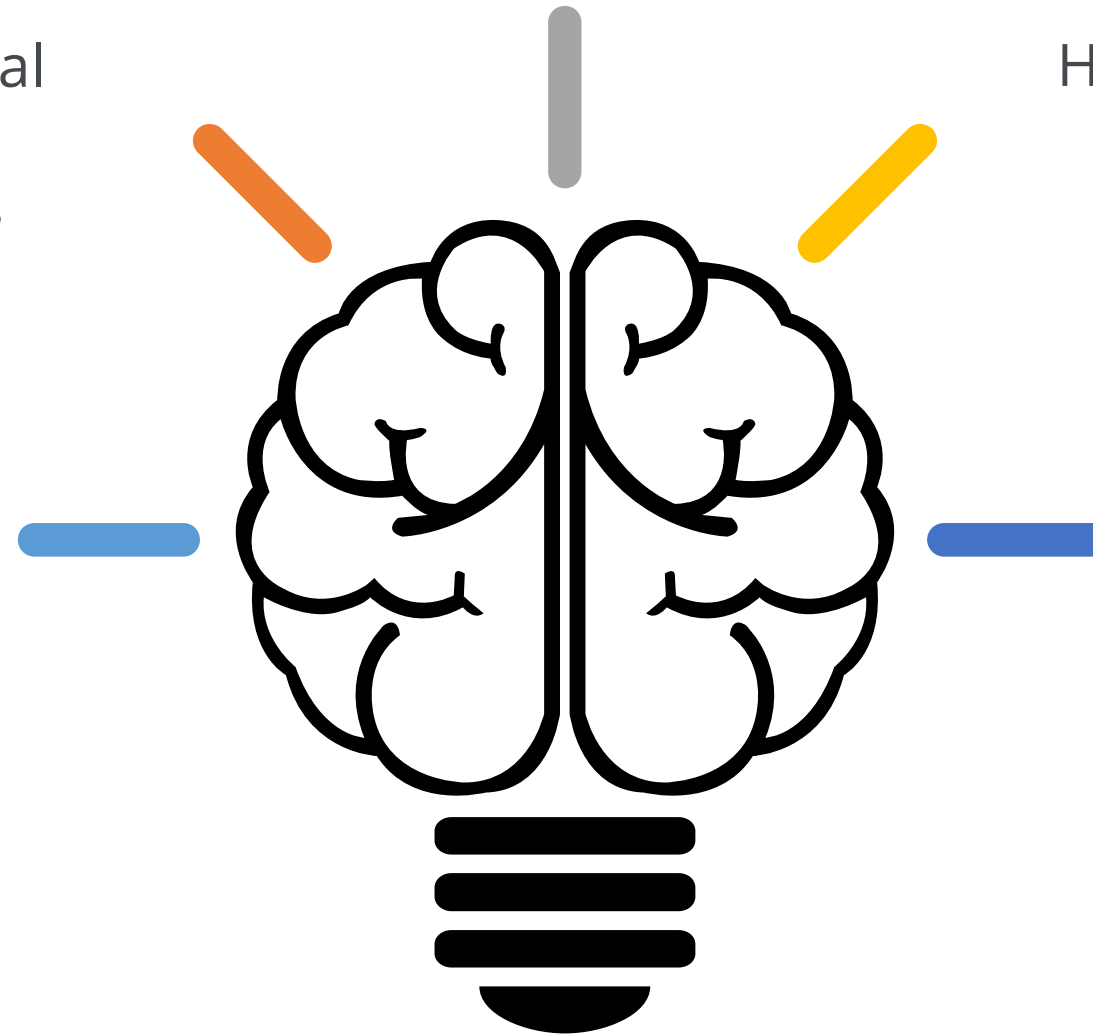
Applies data sovereignty restrictions to comply with legal requirements

Is used with multi-dimensional data sources that contain measures like SAP business

Handles a large amount of data without the need of pre-aggregate

Allows to create visualizations over large datasets

Is suitable in situations where data changes frequently and real-time reporting is required



# Limitations of DirectQuery

## Transformation

There are limitations to data transformation techniques.

## Modeling

There are limitations to modeling capabilities or are not supported.

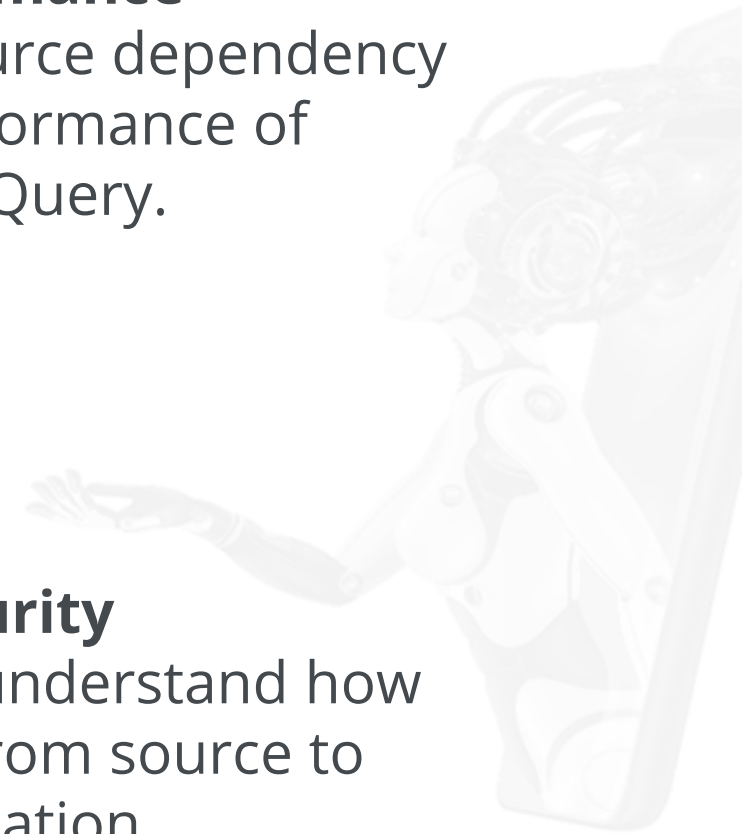


## Performance

There is data source dependency for the performance of DirectQuery.

## Security

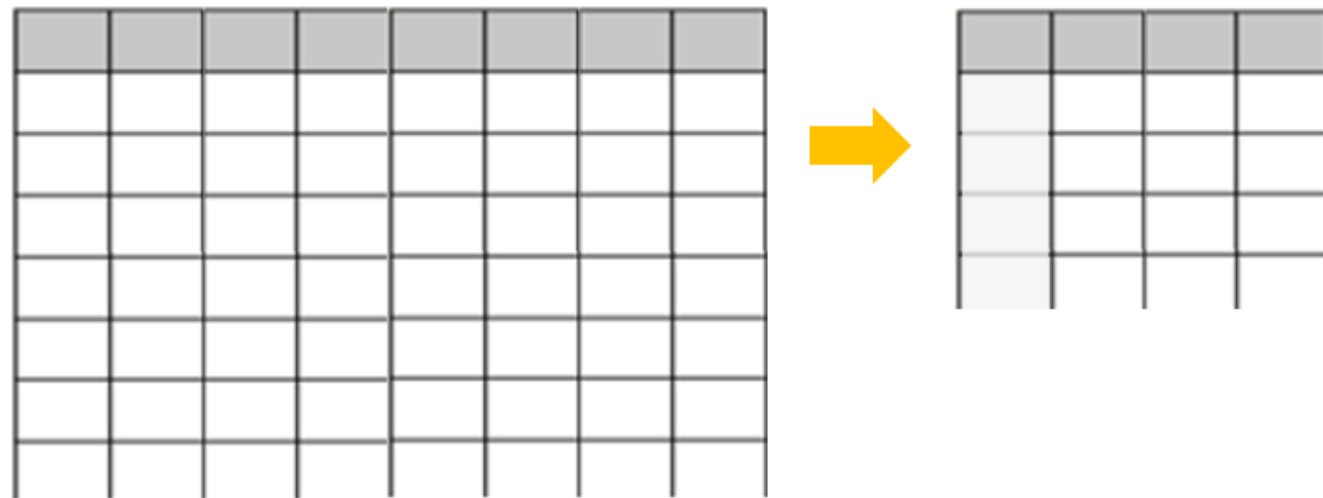
It is difficult to understand how data travels from source to destination.



## Create and Manage Aggregations

# Introduction to Aggregations

Aggregation refers to summarizing data and presenting it at a higher grain or level.



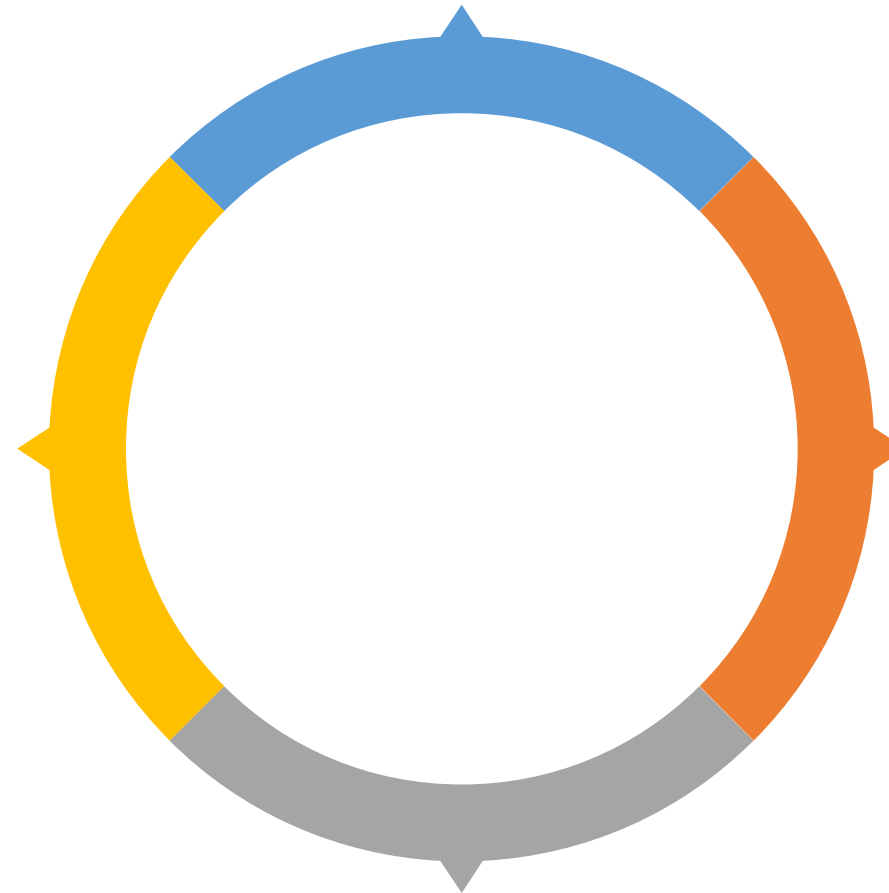
- Aggregation reduces the table sizes in the data model and helps improve query performance.



# Aggregation Use Cases

While dealing with large amounts of data, aggregation provides better query performance and helps derive insights.

When the size of the data model increases, aggregation helps to reduce the issues of performance, refresh issues, and overall query problems.



During slow refreshes, aggregation helps speed up the refresh process.

While using large data models, aggregation helps reduce and maintain the size of the model.

# Creating Aggregations

Power Query Editor creates aggregations in Power BI Desktop.

Queries [1]  
ResellerSales\_202006

	1 SalesOrderNumber	2 SalesOrderLineNumber	3 OrderDate	4 DueDate	5 ShipDate
1	SO71691	2	1/06/2020	11/06/2020	
2	SO71691	4	1/06/2020	11/06/2020	
3	SO71774	1	1/06/2020	11/06/2020	
4	SO71774	2	1/06/2020	11/06/2020	
5	SO71775	1	1/06/2020	11/06/2020	
6	SO71775	2	1/06/2020	11/06/2020	
7	SO71775	3	1/06/2020	11/06/2020	
8	SO71776	1	2/06/2020	12/06/2020	
9	SO71777	1	2/06/2020	12/06/2020	
10	SO71777	2	2/06/2020	12/06/2020	
11	SO71778	1	2/06/2020	12/06/2020	
12	SO71778	2	2/06/2020	12/06/2020	
13	SO71778	3	2/06/2020	12/06/2020	
14	SO71778	4	2/06/2020	12/06/2020	
15	SO71779	1	2/06/2020	12/06/2020	
16	SO71779	2	2/06/2020	12/06/2020	
17	SO71779	3	2/06/2020	12/06/2020	
18	SO71779	4	2/06/2020	12/06/2020	
19	SO71779	5	2/06/2020	12/06/2020	
20	SO71779	6	2/06/2020	12/06/2020	
21	SO71779	7	2/06/2020	12/06/2020	
22	SO71779	8	2/06/2020	12/06/2020	
23	SO71779	9	2/06/2020	12/06/2020	
24	SO71779	10	2/06/2020	12/06/2020	
25					

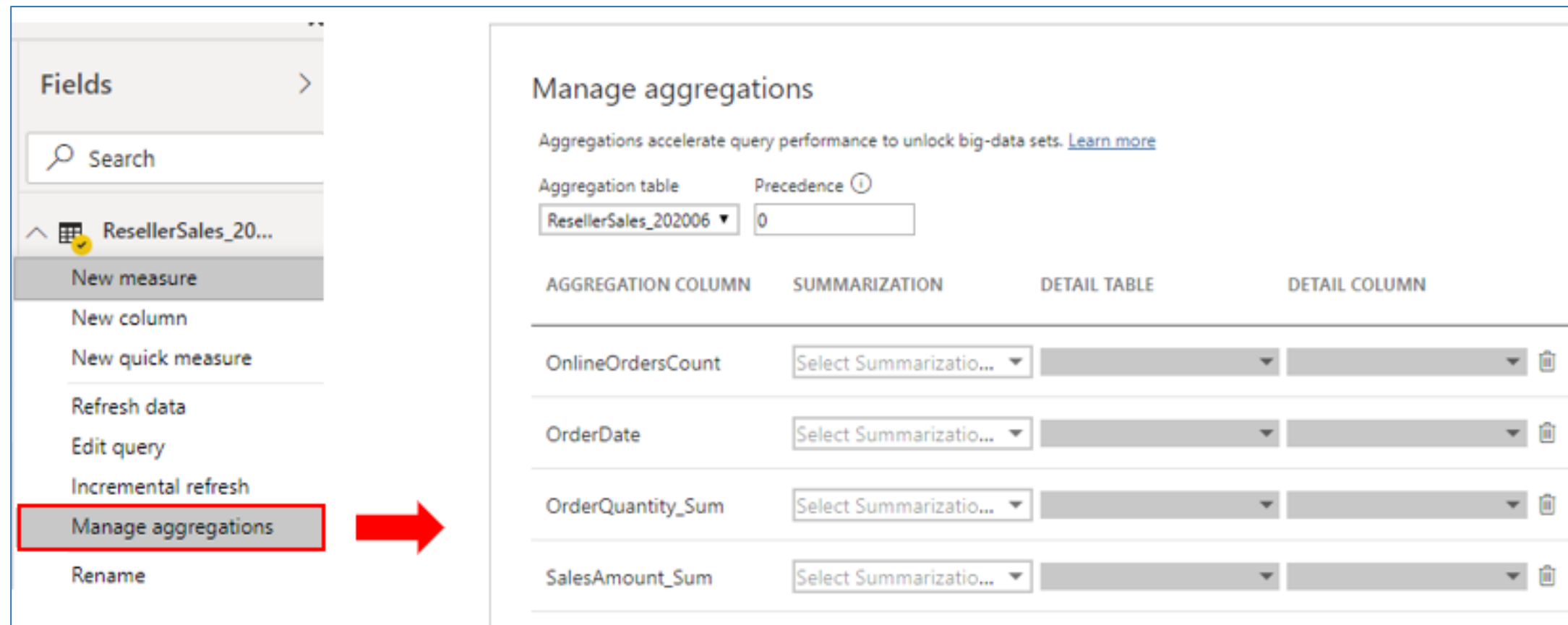
14 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

	22	22/06/2020	55	73935.41	129
	23	23/06/2020	116	191212.91	371
	24	24/06/2020	20	11193.33	26
	25	25/06/2020	62	65857.75	183

4 COLUMNS, 30 ROWS Column profiling based on top 1000 rows

# Managing Aggregations

In the **Fields** pane, right-click the table and then select **Manage aggregations**.



**Fields**

Search

ResellerSales\_20...

- New measure
- New column
- New quick measure
- Refresh data
- Edit query
- Incremental refresh
- Manage aggregations**
- Rename

**Manage aggregations**

Aggregations accelerate query performance to unlock big-data sets. [Learn more](#)

Aggregation table: ResellerSales\_202006 Precedence: 0

AGGREGATION COLUMN	SUMMARIZATION	DETAIL TABLE	DETAIL COLUMN
OnlineOrdersCount	Select Summarization...		
OrderDate	Select Summarization...		
OrderQuantity_Sum	Select Summarization...		
SalesAmount_Sum	Select Summarization...		

## Query Diagnostics

# Analyze Query Plans

In the **Fields** pane, right-click the table and then select **Manage aggregations**.

[-] Sales by Year	270
DAX query	2754
Visual display	57
Other	160
Copy query	

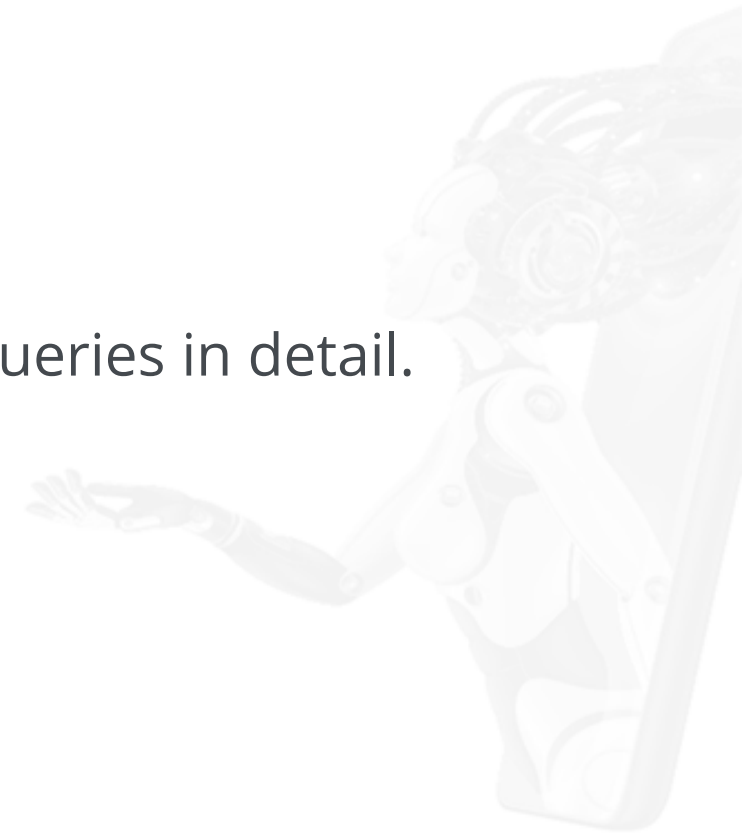
- When you examine the results in the performance analyzer pane, it shows the duration taken by the Power BI Desktop engine to evaluate each query.

# DAX Query

Performance analyzer highlights potential issues, but it does not provide a solution to improve.



- DAX Studio is used to investigate queries in detail.





# DAX Query: Example

[-] Sales by Year	270
DAX query	2754
Visual display	57
Other	160
Copy query	

```
Count Customers =  
CALCULATE ( DISTINCTCOUNT ( Order[ProductID] ), FILTER ( Order, Order[OrderQty] >= 5 ) )
```

[-] Sales by Year	270
DAX query	54
Visual display	57
Other	160
Copy query	

```
Count Customers =  
CALCULATE ( DISTINCTCOUNT ( Order[ProductID] ), KEEPFILTERS (Order[OrderQty] >= 5 ) )
```



## Knowledge Check

## Knowledge Check

1

**When does an optimized model perform better?**

- A. During execution
- B. After execution
- C. Before execution
- D. During initiation



## Knowledge Check

1

When does an optimized model perform better?

- A. During execution
- B. After execution
- C. Before execution
- D. During initiation



The correct answer is **A**

**An optimized model performs better during execution.**

## Knowledge Check

2

What is the term used to refer to the uniqueness of the values in a column?

- A. DX Query
- B. Cardinality
- C. Aggregation
- D. DirectQuery



Knowledge  
Check

2

What is the term used to refer to the uniqueness of the values in a column?

- A. DX Query
- B. Cardinality
- C. Aggregation
- D. DirectQuery



The correct answer is **B**

**Cardinality** is the term used to refer to the uniqueness of the values in a column.



## Knowledge Check

3

Which of the following refers to summarizing data and presenting it at a higher level?

- A. DX Query
- B. Cardinality
- C. Aggregation
- D. DirectQuery



Knowledge  
Check

3

Which of the following refers to summarizing data and presenting it at a higher level?

- A. DX Query
- B. Cardinality
- C. Aggregation
- D. DirectQuery



The correct answer is **C**

**Aggregation refers to summarizing data and presenting it at a higher level.**

## Key Takeaways

- Performance optimization involves making changes to the current state of the data model to run more efficiently.
- A column that has a lot of repeated values in this range has a low level of cardinality.
- The different cardinality relationships are many-to-one, one-to-one, one-to-many, and many-to-many.
- DirectQuery is a method of connecting data directly to its source repository from the Power BI Desktop.

