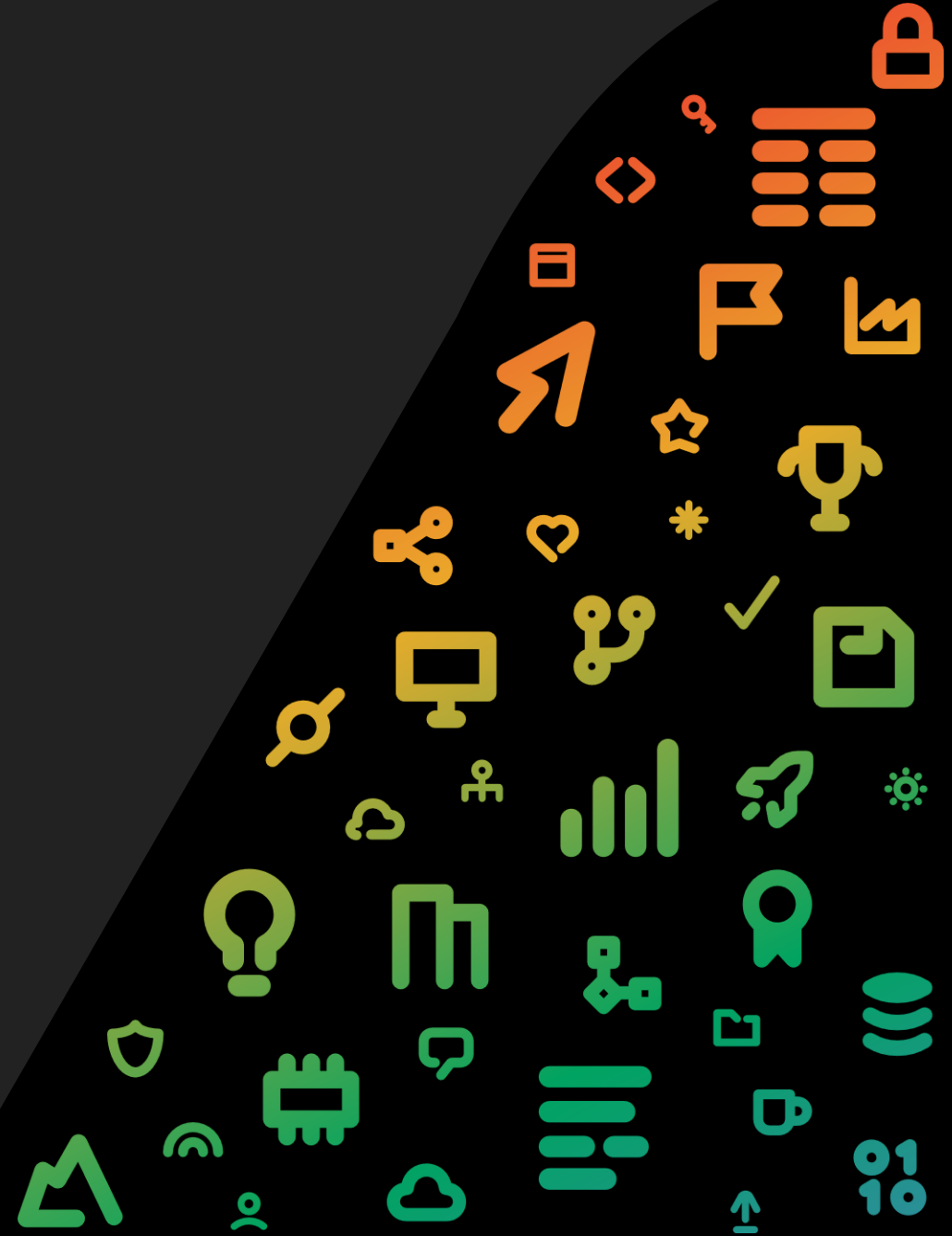


Using Power BI with a Lot of Data

Paul Turley

Director, Competency Lead


3Cloud Solutions



Paul Turley



Director, Solution Architect
3Cloud Solutions

 @paul_turley

 [SqlServerBI.blog](https://sqlserverbi.blog)

 [Linkedin.com/pturley](https://linkedin.com/pturley)

Love learning, volunteering,
teaching, mentoring, speaking
& leading within the Microsoft
data platform community

What's this about?

Can we use Power to deal with real-time data sources, very large tables, and transactional details? Yes, but it requires some planning and proper design. Out of the box, Power BI is super-fast with moderate data volumes and data schemas optimized for analytic reporting. How big is your data? Millions or billions of rows? Gigabytes or terabytes? - we can handle that.

Moving beyond the basics, learn to use DirectQuery alongside in-memory Import mode in composite models to access very large tables with real-time results; models and reports that combine interactive "dashboard-style" reports and drill-through to transactional details. Take a journey and learn to use the Power BI platform to achieve the best of both worlds: performance and scale.

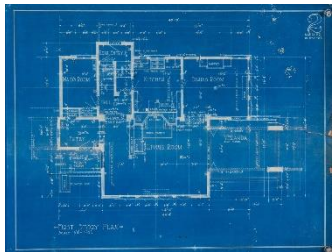
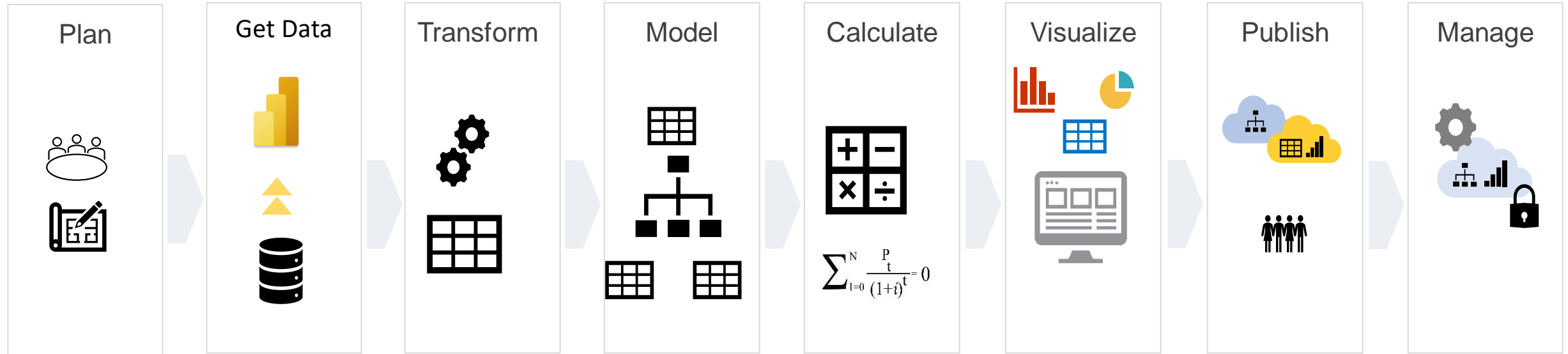
Sample files and examples

Due to the large volume of data used in demos, some sample files for this presentation are too large to be posted for download.

Some examples may be found on my blog, at:

<https://sqlserverbi.blog/presentations>

The Business Intelligence Process



Reality Check...

EMBRACE THE TECHNOLOGY

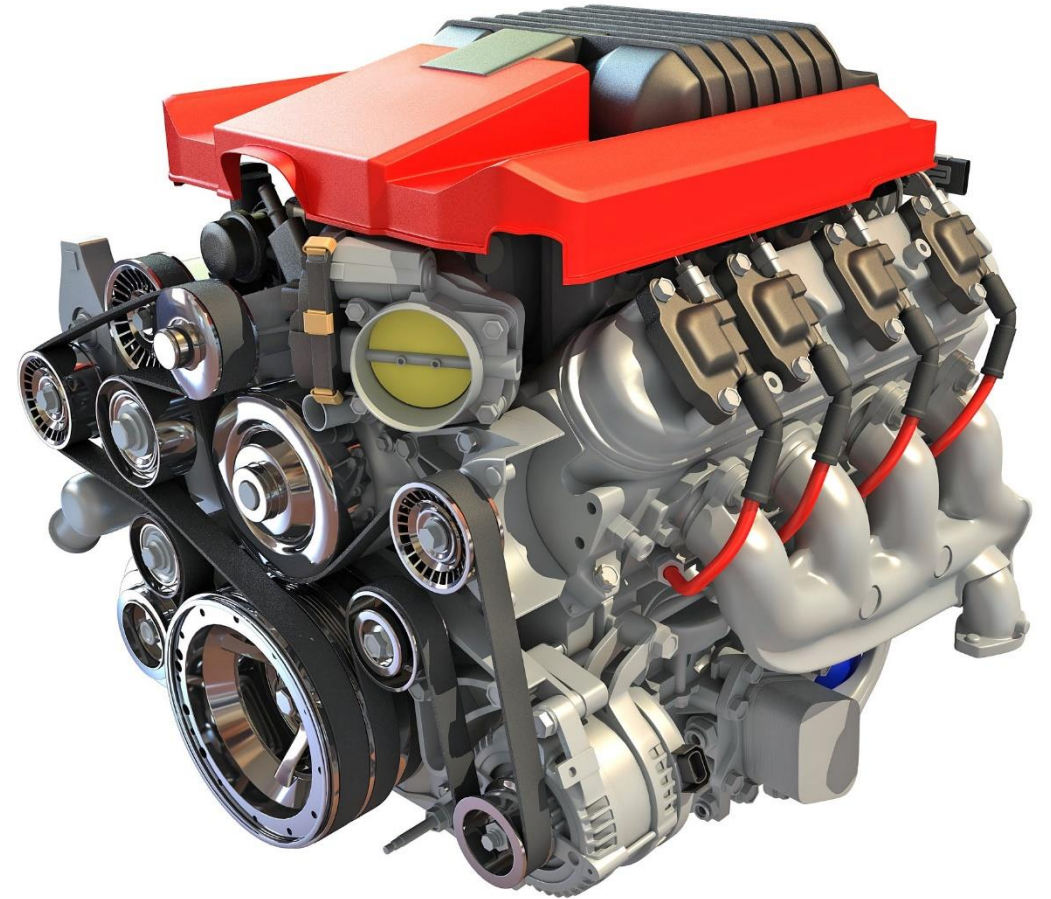
USE THE RIGHT TOOL FOR THE JOB



- Power BI natively is optimized for analytical reporting, not for transactional reporting
- ...BUT we CAN manage very large datasets and transactional details using the right features
- Enterprise Power BI = Premium
- Star schema design is nearly always the right solution
- Import mode is nearly always fastest

DATA MODEL ENGINE Vertipaq

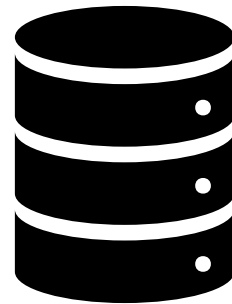
- In-memory cache
- Can be refreshed frequently
- Columnar compression
- Internal storage optimization:
 - Value encoding
 - Hash encoding
 - Run Length encoding
- Storage engine, Formula engine



How Much Data Do You Have?

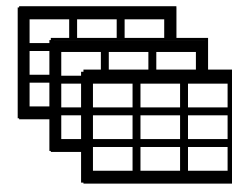


Data sources
Data warehouse
Data lake
Files



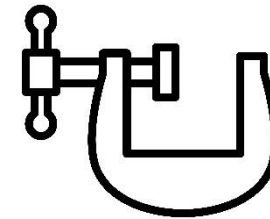
1 TB

Table & Column
selection



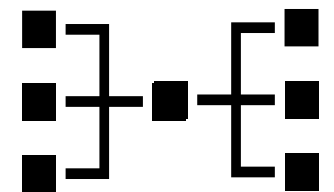
30 GB

Compression



2 GB

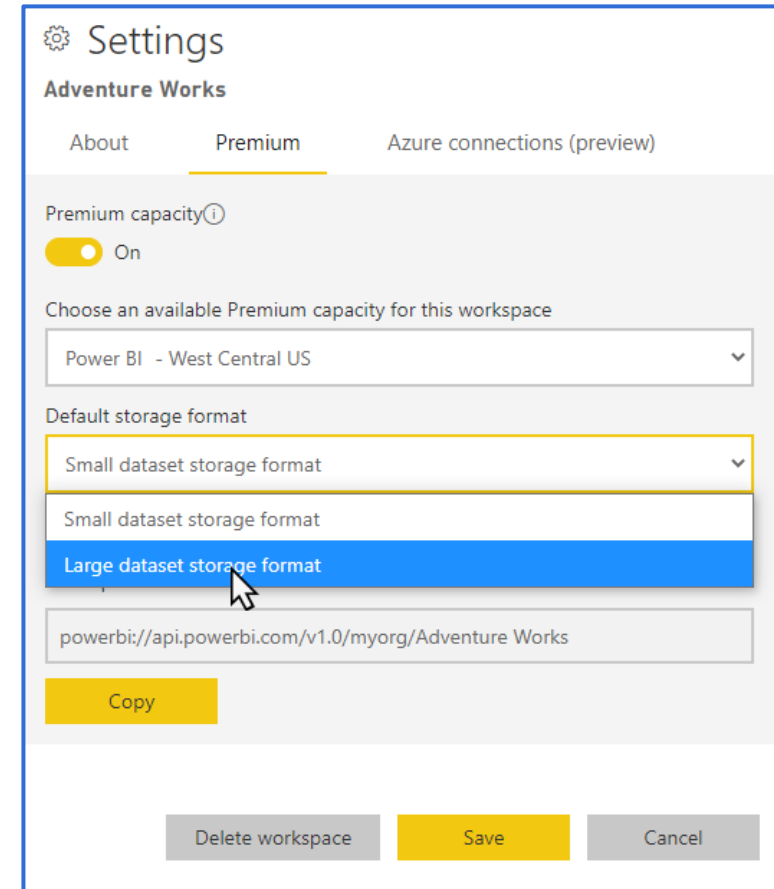
Semantic
Model



**100 MB (repo)
2 GB (service)**

Use Large dataset storage format in Premium workspaces

- Introduced in Premium Gen2
- Optimized for 1GB+ datasets
- No/little negative impact for smaller datasets
- On-demand load
- Dataset eviction
- Cannot be downloaded from the service



The screenshot shows the 'Settings' window for an 'Adventure Works' workspace. The 'Premium' tab is selected, showing 'Premium capacity' as 'On'. Below this, a dropdown menu shows 'Power BI - West Central US'. The 'Default storage format' dropdown is open, showing 'Small dataset storage format' and 'Large dataset storage format', with the latter highlighted in blue. A 'Copy' button is visible below the dropdown. At the bottom, there are 'Delete workspace', 'Save', and 'Cancel' buttons.

Settings
Adventure Works

About Premium Azure connections (preview)

Premium capacity ⓘ
On

Choose an available Premium capacity for this workspace
Power BI - West Central US

Default storage format
Small dataset storage format
Large dataset storage format

powerbi://api.powerbi.com/v1.0/myorg/Adventure Works

Copy

Delete workspace Save Cancel

Dedicated Capacity Resource Limits

Capacity SKUs	Total v-cores	Backend v-cores	Frontend v-cores	Max memory per dataset (GB)	DirectQuery/ Live connection (per second)	Max memory per query (GB)	Model refresh parallelism
EM1/A1	1	0.5	0.5	3	3.75	1	5
EM2/A2	2	1	1	5	7.5	2	10
EM3/A3	4	2	2	10	15	2	20
P1/A4	8	4	4	25	30	6	40
P2/A5	16	8	8	50	60	6	80
P3/A6	32	16	16	100	120	10	160
P4/A7	64	32	32	200	240	10	320
P5/A8	128	64	64	400	480	10	640

<https://learn.microsoft.com/en-us/power-bi/enterprise/service-premium-gen2-what-is#memory-allocation>

Large Dataset Options

- Date range parameters
- Import mode with partitioning
- DirectQuery
- Dual mode
- Aggregations
- Composite & hybrid data models

Data Source Scale

Does the data source support query-folding in Power Query and DirectQuery?

- File (CSV, JSON, CSV...)
- Folder
- Data Lake (files)
- REST APIs
- SharePoint
- OneDrive
- SQL Server / Azure SQL
- Relational & SPARK sources
- Cloud-based data warehouses:
 - Synapse Analytics
 - Snowflake
 - Redshift
 - Delta Lake, SQL Connector

Optimize for Compression

- Eliminate wide text columns
- Store distinct values with low cardinality
- Store numbers as numbers
- Use the most conservative data type

Decimal number	Fixed decimal	Fixed decimal
Precision: 15	Decimal(19,4)	Cleansed: 2 decimals
123.456700000012	123.4567	123.46
123.456700000013	123.4567	123.46
123.456700000015	123.4567	123.46
2345.12000000023	2345.12	2345.12
2345.12000000022	2345.12	2345.12
2345.12000000024	2345.12	2345.12
2345.12000000023	2345.12	2345.12
45.0020525001121	45.002	45
45.0020525001105	45.002	45
45.0020525000026	45.002	45

Developer:

I need a manageable dataset size

- Keep .PBIX files small
- Fast refresh
- Fast deployment
- Manageable
version control & CI/CD



Demo 1:

Parametrized date range filters

Partitioning Large Tables

```
{
  "create": {
    "parentObject": {
      "database": "Contoso Sales",
      "table": "Fact Online Sales"
    },
    "partition": {
      "name": "2022Q1",
      "mode": "import",
      "source": {
        "type": "policyRange",
        "start": "2022-01-01T00:00:00",
        "end": "2022-04-01T00:00:00",
        "granularity": "quarter",
        "refreshBookmark": "1/1/1901"
      }
    }
  }
}
```

The screenshot displays the TOM Explorer application interface. The main pane shows a tree view of the database structure. Under 'Fact Online Sales', the 'Partitions' folder is expanded, showing a list of partitions: 2019, 2020, 2021, 2022Q1 (selected), 2022Q2, 2022Q307, 2022Q308, and 2022Q309. Below the tree, the 'Properties' pane is visible, showing details for the selected '2022Q1' partition.

TOM Explorer

Enter text to search...

- > Fact Exchange Rate
- > Fact Online Sales
 - > Partitions
 - 2019
 - 2020
 - 2021
 - 2022Q1**
 - 2022Q2
 - 2022Q307
 - 2022Q308
 - 2022Q309
 - CustomerKey
 - DateKey

TOM Explorer | Data Refresh | Macros | Best Practice A...

Properties

Enter text to search...

Error Message	
> Extended Properties	0 extended properties
Last Processed	9/12/2022 8:53 PM
Object Type	Partition (Policy Range)
State	Ready

Options

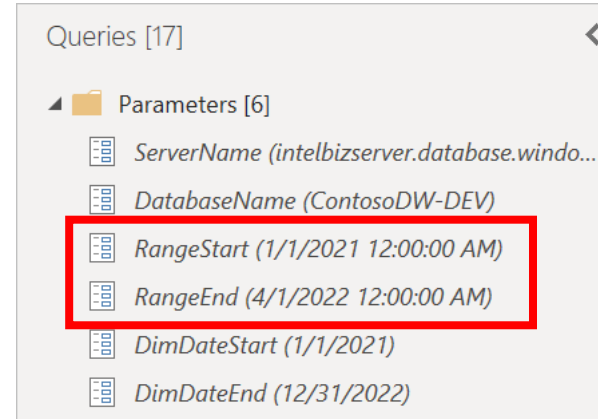
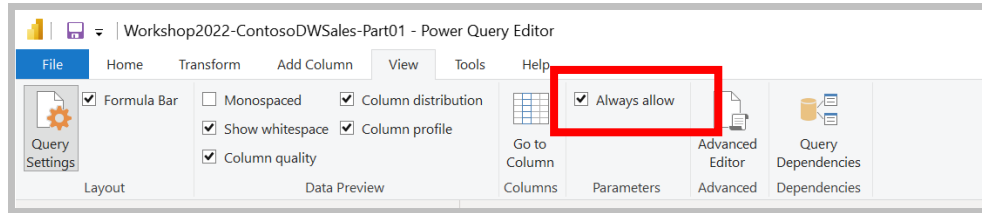
Data View	Default
Mode	Import
Query Group	
Source Type	PolicyRange

Refresh Policy

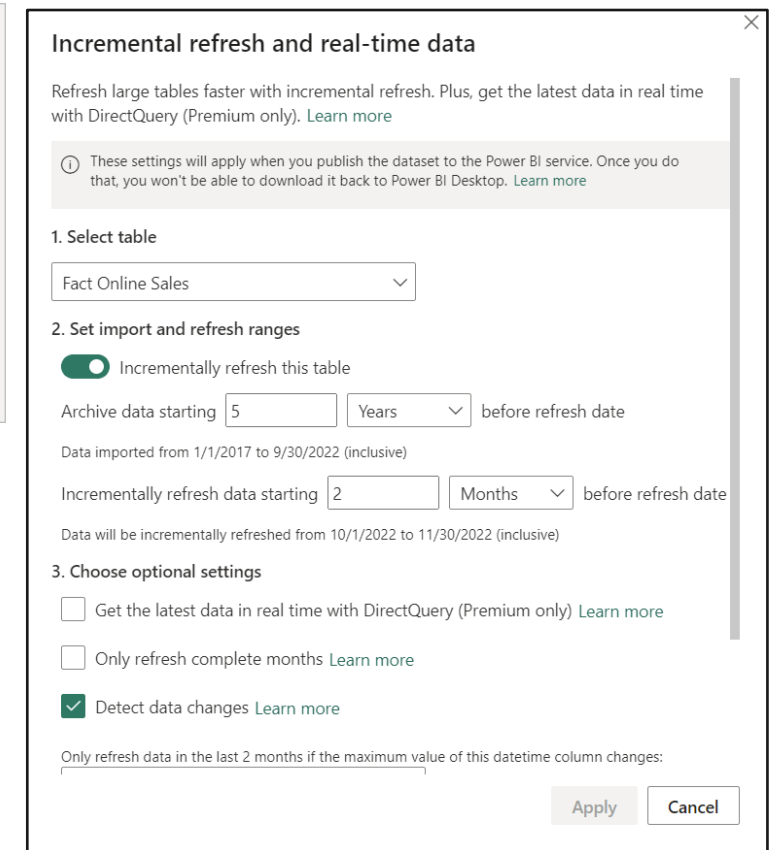
End	4/1/2022
Granularity	Quarter
RefreshBookmark	1/1/1901
Start	1/1/2022

Partition "2022Q1" | Enterprise Workshop DEV

Implementing Incremental Refresh



- Allow parameterization
- Create RangeStart & RangeEnd parameters
- Add date range filter to fact table
- Create Incremental Refresh policy



Data Administrator:

Optimize & partition large tables

- Improve refresh speed
- Prevent reloading history
- Capture updated history
- Reduce database resource load

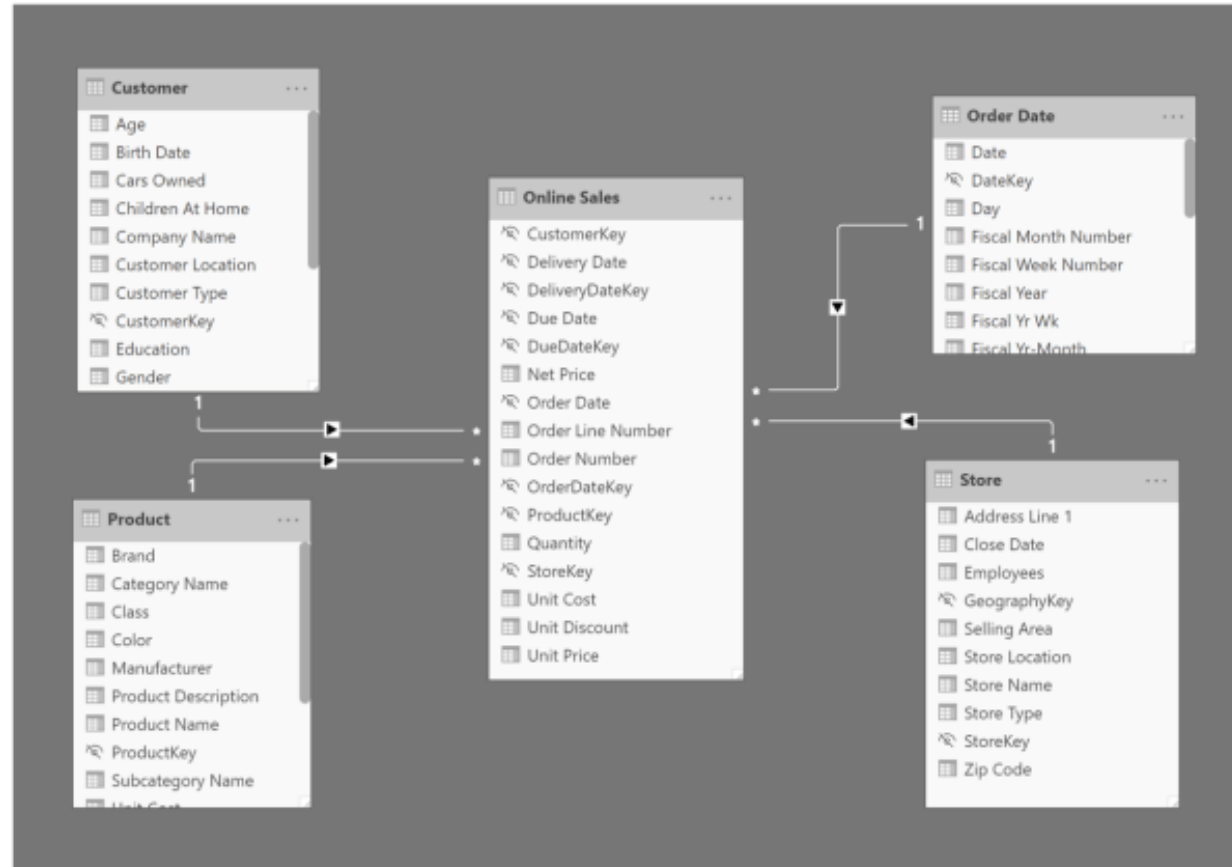


Demo 2:

Incremental Refresh

Table partitions

Hybrid tables

[illegible][illegible]

Business Users:

We need to see the details

- Drill-down in context
- Transactions
- Detail records
- See recent data changes

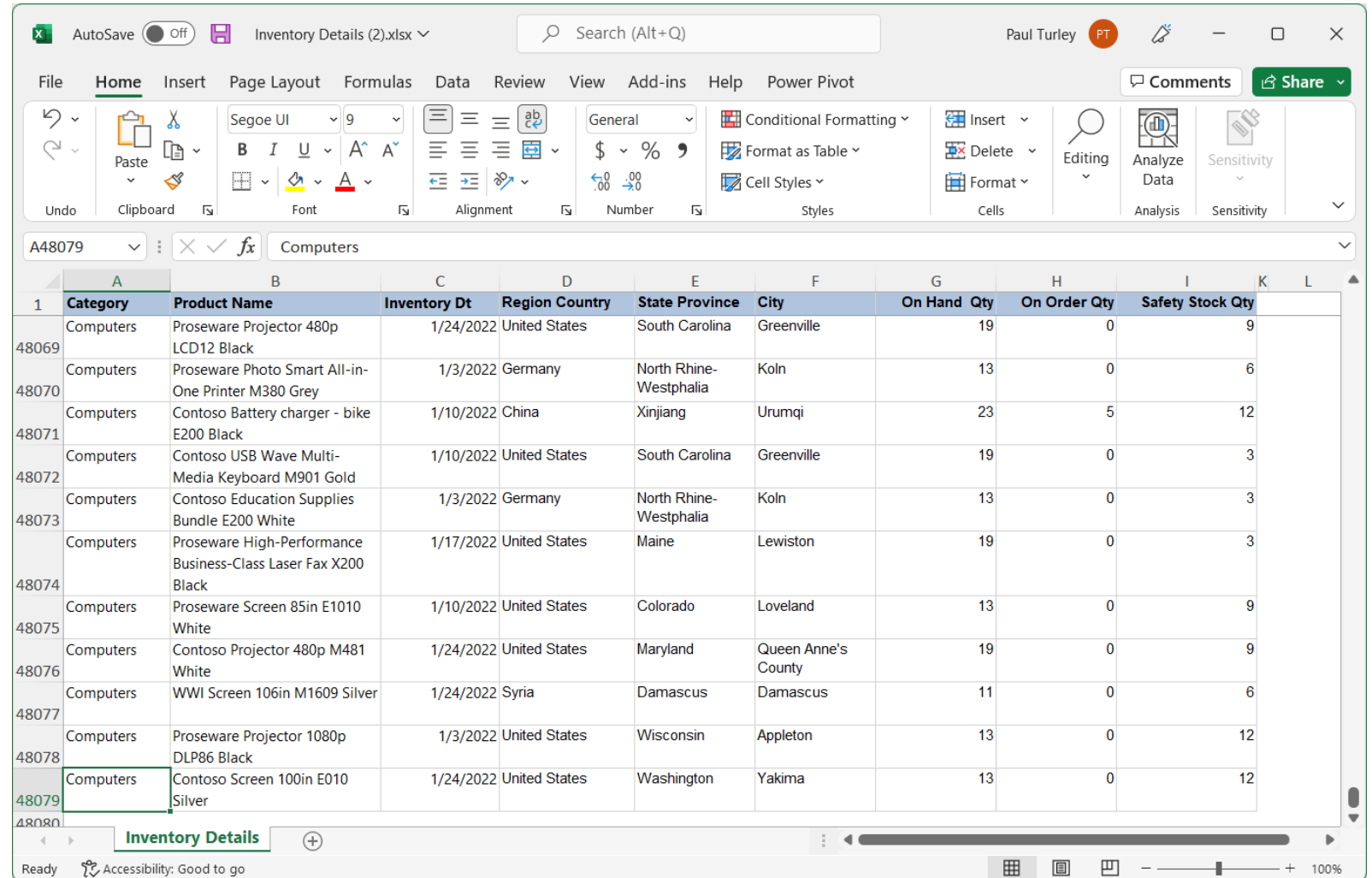


Demo 3:

Impact of importing a big, wide table

DirectQuery Details

- Wide table
- Complex source query
- Frequently changing data



The screenshot shows the Microsoft Excel interface with a DirectQuery table named 'Computers'. The table is displayed in a wide format, spanning multiple columns. The interface includes the Excel ribbon, a search bar, and a status bar at the bottom.

	A	B	C	D	E	F	G	H	I	K	L
1	Category	Product Name	Inventory Dt	Region Country	State Province	City	On Hand Qty	On Order Qty	Safety Stock Qty		
48069	Computers	Proseware Projector 480p LCD12 Black	1/24/2022	United States	South Carolina	Greenville	19	0	9		
48070	Computers	Proseware Photo Smart All-in-One Printer M380 Grey	1/3/2022	Germany	North Rhine-Westphalia	Koln	13	0	6		
48071	Computers	Contoso Battery charger - bike E200 Black	1/10/2022	China	Xinjiang	Urumqi	23	5	12		
48072	Computers	Contoso USB Wave Multi-Media Keyboard M901 Gold	1/10/2022	United States	South Carolina	Greenville	19	0	3		
48073	Computers	Contoso Education Supplies Bundle E200 White	1/3/2022	Germany	North Rhine-Westphalia	Koln	13	0	3		
48074	Computers	Proseware High-Performance Business-Class Laser Fax X200 Black	1/17/2022	United States	Maine	Lewiston	19	0	3		
48075	Computers	Proseware Screen 85in E1010 White	1/10/2022	United States	Colorado	Loveland	13	0	9		
48076	Computers	Contoso Projector 480p M481 White	1/24/2022	United States	Maryland	Queen Anne's County	19	0	9		
48077	Computers	WWI Screen 106in M1609 Silver	1/24/2022	Syria	Damascus	Damascus	11	0	6		
48078	Computers	Proseware Projector 1080p DLP86 Black	1/3/2022	United States	Wisconsin	Appleton	13	0	12		
48079	Computers	Contoso Screen 100in E010 Silver	1/24/2022	United States	Washington	Yakima	13	0	12		

Drill-through to Details with DirectQuery

Computers	\$ 264,760,665.48	\$ 24,148,655.66
2021	\$ 215,329,826.66	\$ 19,541,225.62
2022	\$ 49,430,838.83	\$ 4,607,430.04
January	\$ 15,617,765.48	\$ 1,457,813.13
February	\$ 16,678,813.13	\$ 1,541,225.62
March	\$ 17,134,260.04	\$ 1,541,225.62
Games and Toys	\$ 13,276,900.00	\$ 1,276,900.00
Home Appliances	\$ 408,336,435.00	\$ 39,833,643.50
Music, Movies and Audio Books	\$ 13,797,889.00	\$ 1,379,788.90
TV and Video	\$ 126,109,709.00	\$ 12,610,970.90
Total	\$ 1,124,153,171.00	\$ 112,415,317.10

- Show as a table
- Include
- Exclude
- Drill through
- Group
- Summarize
- Copy

Inventory Details

Inventory Details for Computers

Total Inventory On Hand Qty by Store Type

Store Type	Total Inventory On Hand Qty
Online	0.8M
Reseller	0.2M
Catalog	0.1M

Total Inventory On Hand Qty by Region Country Name

Region Country Name	Total Inventory On Hand Qty
United States	0.5M
Germany	0.1M
China	0.05M
United Kingdom	0.05M
Canada	0.05M

Total Inventory On Hand Qty by Product Subcategory Name

Product Subcategory Name	Total Inventory On Hand Qty
Computers Accessories	0.2M
Laptops	0.15M
Projectors & Scanners	0.1M
Printers, Scanners	0.05M

Category	Product Name	Inventory Date	Region Country Name	State Province Name	City Name	On Hand Quantity	On Order Quantity	Max Day In Stock	Min Day In Stock
Computers	Adventure Works CRT15 E101 Black	1/3/2022	Japan	Hokkaido	Sapporo	10	0	81	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	Syria	Damascus	Damascus	10	0	92	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	India	National Capital Territory of Delhi	New Delhi	10	0	115	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	United States	Washington	Seattle	12	0	92	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	Russia	Moskovskaya oblast	Moscow	13	0	72	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	Canada	British Columbia	Vancouver	15	0	116	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	United States	Virginia	Martinsville	19	0	77	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	United States	Virginia	Norfolk	19	0	90	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	United States	New Jersey	Cape May	19	0	108	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	United States	New Jersey	East Orange	19	0	118	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	China	Beijing	Beijing	20	0	65	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	United States	Maryland	North Harford	20	1	83	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	United States	Wisconsin	Oregon	20	1	112	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	Canada	Alberta	Calgary	21	0	84	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	United States	Wisconsin	Milwaukee	21	0	102	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	United States	Colorado	Greeley	22	3	112	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	United States	Texas	Piano	22	17	64	
Computers	Adventure Works CRT15 E101 Black	1/3/2022	United States	Maryland	Back River	24	5	78	

Demo 4:

**Composite model with DirectQuery
Drillthrough**

Everyone:
How do we speed this up?

We want the best of
both worlds:

- DirectQuery details
- Import aggregates



Aggregations

- Two tables in the model:
- Details: DirectQuery
- Summary: Import
- Aggregations map fields between the tables

Manage aggregations

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

Aggregation table: Sales Agg Precedence: 0

AGGREGATION COLUMN	SUMMARIZATION	DETAIL TABLE	DETAIL COLUMN	
OrderDateKey	GroupBy	Sales	OrderDateKey	🗑️
CustomerKey	GroupBy	Sales	CustomerKey	🗑️
ProductSubcategoryKey	GroupBy	Product	ProductSubcategory...	🗑️
SalesAmount_Sum	Sum	Sales	SalesAmount	🗑️
UnitPrice_Sum	Sum	Sales	UnitPrice	🗑️

Apply all Cancel

Demo 5:

Aggregations and summary table

The art and science of aggregations

- Capture DAX query on Performance Analyzer
- Execute in DAX Studio with Server Timings
- Analyze trace for aggregation hits

The screenshot displays the DAX Studio interface (version 3.0.5) with a DAX query in the main editor and its execution plan and server timings in the bottom pane.

DAX Query:

```
TREATAS({ "January" }, 'Dim Date'[Month Name])
VAR __DS0Core =
SUMMARIZECOLUMNS(
    ROLLUPADISSUBTOTAL('Inventory Details'[Region Country Name], "IsGrandTotalRowTotal"),
    __DS0FilterTable,
    __DS0FilterTable2,
    __DS0FilterTable3,
    "Total_Inventory_On_Hand_Qty", '_Measures'[Total Inventory On Hand Qty],
    "v_Total_Inventory_On_Hand_Qty_FormatString", IGNORE('__Measures'[_Total Inventory On Hand Qty FormatString])
)
VAR __DS0Primarywindowed =
TOPN(
    502,
    __DS0Core,
    [IsGrandTotalRowTotal],
    0,
    [Total_Inventory_On_Hand_Qty],
    0,
    'Inventory Details'[Region Country Name],
    1
)
VAR __DS0CoreNoInstanceFiltersNoTotals =
CALCULATE(
    __DS0Core,
    [IsGrandTotalRowTotal] = FALSE
)
```

Server Timings Summary:

Category	Item	Value
Total		23,914 ms
FE		18 ms (0.1%)
SE CPU		23,896 ms (99.9%)
SE Queries		3
SE Cache		0 (0.0%)

Execution Plan Table:

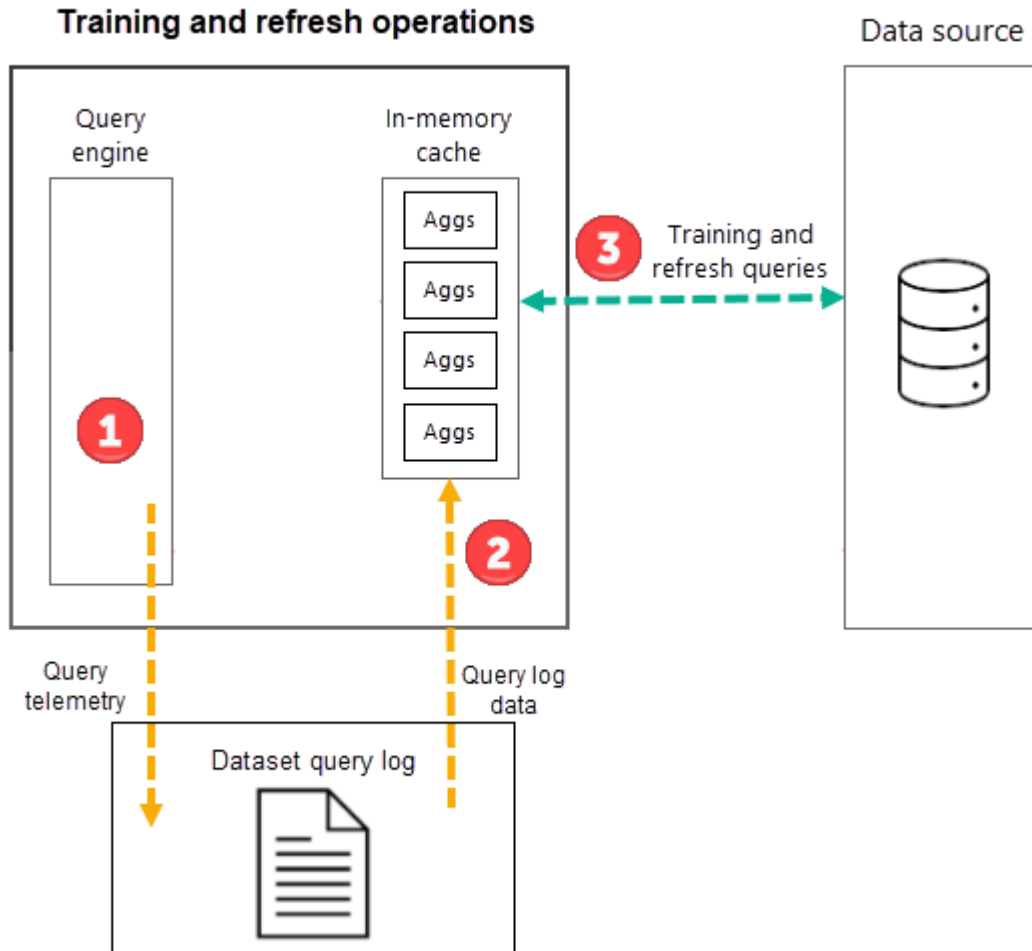
Line	Subclass	Duration	CPU	Par.	Rows	KB	Waterfall	Query
2	Scan	0	0		2,520	20		SELECT 'Dim Product'[Pro...
4	Scan	0	0		733	6		SELECT 'Dim Date'[Date] F
5	RewriteAtter	0						<attemptFailed>
6	SQL	23,896	23,896	x1.0				SELECT TOP (1000001) * F

Query Text (Right Pane):

```
{
  "table": "Inventory Details",
  "matchingResult": "attemptFailed",
  "failureReasons": [
    {
      "alternateSource": "Inventory Summary Aggs",
      "reason": "no column mapping",
      "column": "Inventory Details[Region Country Name]"
    }
  ],
  "dataRequest": [
    {
      "aggregation": "Sum",
      "table": "Inventory Details",
      "column": "On Hand Quantity"
    },
    {
      "table": "Inventory Details",
      "column": "Inventory Date"
    }
  ]
}
```

Status Bar: I Ln 52, Col 1 | localhost:55528 | 16.0.53.24 | 722 | 1 row | 00:23.9

Auto Aggregations



Aggregation creation can be automated through usage metrics based on the dataset query log

<https://learn.microsoft.com/en-us/power-bi/enterprise/aggregations-auto>

Drill-through to Paginated Report

PASSDC-LotsOfData-Contoso Sales Drillthrough			
File Export Share Chat in Teams Get insights Edit			
Category Name	Combined Sales Amt	Store Sales Amt	Inventory Report Link
Audio	\$ 14,980,466.13	\$ 1,416,215.21	
Cameras and camcorders	\$ 213,059,663.74	\$ 18,204,103.00	
Cell phones	\$ 69,831,441.34	\$ 5,836,959.79	
Computers	\$ 264,760,665.48	\$ 24,148,655.66	
2021	\$ 215,329,826.66	\$ 19,541,225.62	
2022	\$ 49,430,838.83	\$ 4,607,430.04	
January	\$ 15,617,765.43	\$ 1,537,843.48	
February	\$ 16,678,813.21	\$ 1,367,816.25	
March	\$ 17,134,260.18	\$ 1,701,770.31	
Games and Toys	\$ 13,276,900.61	\$ 1,127,459.23	
Home Appliances	\$ 408,336,435.12	\$ 32,975,363.06	
Music, Movies and Audio Books	\$ 13,797,889.47	\$ 1,204,159.75	
TV and Video	\$ 126,109,709.19	\$ 11,416,870.37	
Total	\$ 1,124,153,171.08	\$ 96,329,786.07	



Inventory Details - Power BI

https://app.powerbi.com/groups/6658929d-9b68-487b-b370-c7cc6e39ad40...

Power BI Workshop Inventory Details

File Export View Page 1 of 1415 Parameters Subscribe Share

Product Category: Computers Year: 2022 Month: January

11/1/2022 1:21:08 AM Page 1 of 1,415 pages

Inventory Details

Category: **Computers** For: **January 2022**

Includes: **48,078** inventory records

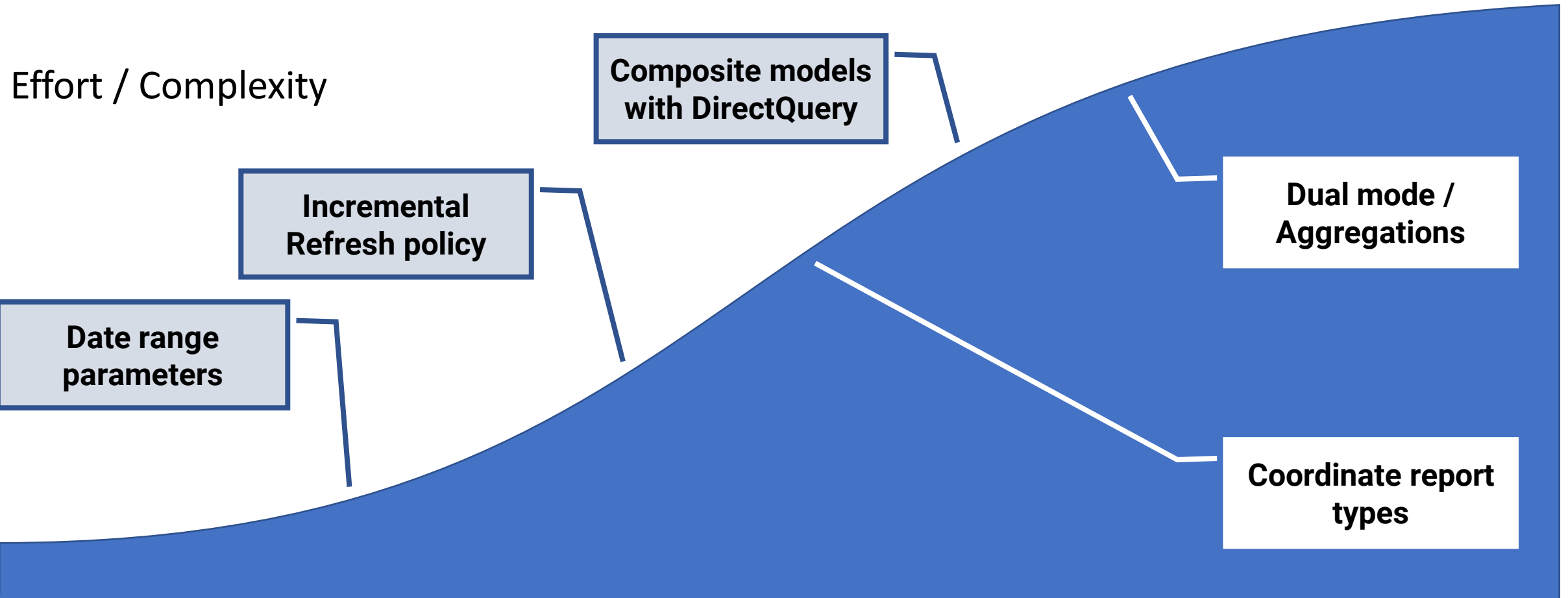
Category	Product Name	Inventory Dt	Region Country	State Province	City	On Hand Qty	On Order Qty	Safety Stock Qty
Computers	Contoso Battery charger - bike E200 Black	1/17/2022	United States	Washington	Seattle	19	0	9
Computers	WWI LCD20 M220 Black	1/3/2022	United States	Washington	Seattle	17	0	6
Computers	Contoso Smart Battery M901 Blue	1/24/2022	United States	Washington	Seattle	20	1	3
Computers	Proseware LCD17 E200 White	1/3/2022	United States	Washington	Seattle	19	0	3
Computers	Fabrikam Laptop13.3W M3080 Red	1/3/2022	United States	Washington	Seattle	20	1	6
Computers	Contoso Car power adapter M90 Black	1/3/2022	United States	Washington	Seattle	24	5	9
Computers	WWI Projector 1080p DLP86 Silver	1/31/2022	United States	Washington	Seattle	19	0	3
Computers	Proseware LCD15 E103 White	1/3/2022	United States	Washington	Seattle	19	0	9
Computers	Contoso Screen 100in E010 Silver	1/31/2022	United States	Washington	Seattle	15	0	9
Computers	Proseware Laser Jet All in one X300 Black	1/3/2022	United States	Washington	Seattle	21	0	9
Computers	Proseware Photo smart All-in-One Printer M380 Black	1/31/2022	United States	Washington	Seattle	12	0	3
Computers	Adventure Works Laptop8.9 E0890 Silver	1/31/2022	United States	Washington	Seattle	51	0	9
Computers	Contoso Screen 113in M251 Black	1/3/2022	United States	Washington	Seattle	14	0	3
Computers	Contoso Battery charger -	1/10/2022	United States	Washington	Seattle	19	0	9

Page 1 of 1415

Demo 6:

Drill-through to Paginated Report

Summary: Data Scaling Options



Summary

- Optimize models with dimensional schemas
- Use date range parameters to manage table volume
- Partition & use incremental refresh when warranted
- Keep transactional tables & flat tables out of the model
- Use DirectQuery for drillthrough
- Drillthrough using Paginated Reports

Thank you

Please support your local data community

Paul Turley

@paul_turley

[Linkedin.com/pturley](https://www.linkedin.com/in/pturley)

[SqlServerBi.blog](https://sqlserverbi.blogspot.com/)



Session evaluation

Your feedback is important to us

Evaluate this session at:

www.PASSDataCommunitySummit.com/evaluation