

Power BI Incremental Refresh

14 de abril de 2020

Ing. Alejandro Cabanchik
@ale_cabanchik
alejandro.cabanchik@smartbi.com.ar



Microsoft
Partner

Silver Data Analytics
Silver Data Platform



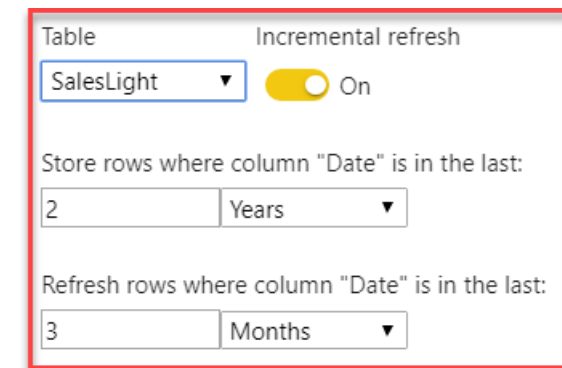
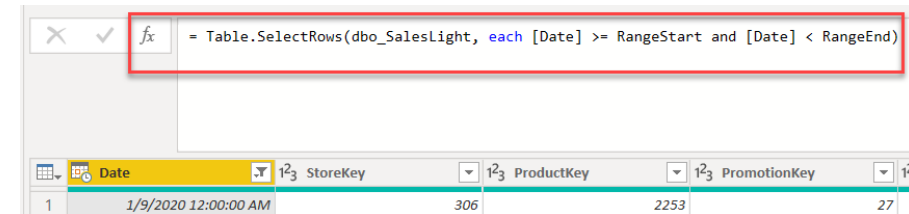
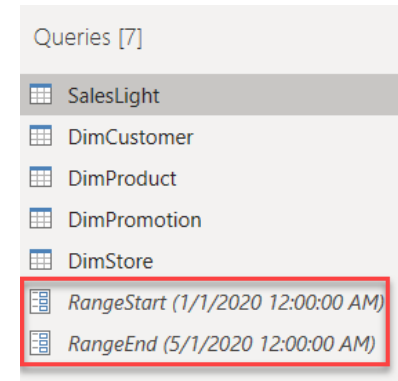
Moderador: Mariano Kovo

Incremental Refresh

- *“Permite actualizar tablas “importadas” por rango de fechas”*
 - *Particiona la tabla por rangos*
- ¡¡¡Ahora en Pro!!!
 - Inicialmente solo Power BI Premium
 - Desde Febrero 2020 en Power BI Pro
 - **Votado en Power BI ideas.**
- Ayuda a refrescar “datasets” grandes con los siguientes beneficios
 - Menor tiempo procesamiento
 - Mayor confiabilidad
 - Menor consumo de recursos

Configuración

- Power BI Desktop
 - Definición de Filtros
 - Definición de Políticas
 - Siempre pisa definiciones (Excepto Premium – ALM Toolkit)
- Power BI Services
 - Creación de particiones
 - En la primera actualización
 - Ejecución de políticas de actualización
- Parámetros RangeStart – RangeEnd
 - Tipo Datetime
 - Filtro en Power Query
 - \geq RangeStart
 - $<$ RangeEnd
 - Power BI Services modifica valores durante la actualización



- ☐ Detect data changes [Learn more](#)
- ☐ Only refresh complete months [Learn more](#)

Datos duplicados

- El campo Datetime debe ser una fecha estable (¿Fecha de transacción?)
 - Si utilizamos fecha de modificación podemos duplicar registros
- RangeEnd es un limite externo
 - Valor del inicio del siguiente limite
 - Fin + 1
 - WHERE ColumnaFecha \geq RangeStart AND ColumnaFecha $<$ RangeEnd

Transaction...	TransactionDatetime	Transaction...	Transaccio
1	2019-01-01 00:00:00.000	1.0000	2019-01-0
2	2019-02-01 00:00:00.000	1.0000	2019-02-0
3	2019-03-01 00:00:00.000	1.0000	2019-03-0
4	2019-04-01 00:00:00.000	1.0000	2019-04-0
5	2019-05-01 00:00:00.000	1.0000	2019-05-0
6	2019-06-01 00:00:00.000	1.0000	2019-06-0
7	2019-07-01 00:00:00.000	1.0000	2019-07-0
8	2019-08-01 00:00:00.000	1.0000	2019-08-0
9	2019-09-01 00:00:00.000	1.0000	2019-09-0
10	2019-10-01 00:00:00.000	1.0000	2019-10-0
11	2019-11-01 00:00:00.000	1.0000	2019-11-0
12	2019-12-01 00:00:00.000	1.0000	2019-12-0
13	2020-01-01 00:00:00.000	1.0000	2020-01-0
14	2020-02-01 00:00:00.000	1.0000	2020-02-0
15	2020-03-01 00:00:00.000	1.0000	2020-03-0
16	2020-04-01 00:00:00.000	1.0000	2020-04-0

10	2019-10-01 00:00:00.000	1.0000	2019-10-01 00:00:00.000
11	2019-11-01 00:00:00.000	1.0000	2019-11-01 00:00:00.000
12	2020-01-01 00:00:00.000	1.0000	2019-12-01 00:00:00.000
13	2020-01-01 00:00:00.000	1.0000	2020-01-01 00:00:00.000

TransactionID	TransactionDatetime	TransactionQuantity
1	1/1/2019 12:00:00 AM	1.
2	2/1/2019 12:00:00 AM	1.
3	3/1/2019 12:00:00 AM	1.
4	4/1/2019 12:00:00 AM	1.
5	5/1/2019 12:00:00 AM	1.
6	6/1/2019 12:00:00 AM	1.
7	7/1/2019 12:00:00 AM	1.
8	8/1/2019 12:00:00 AM	1.
9	9/1/2019 12:00:00 AM	1.
10	10/1/2019 12:00:00 AM	1.
11	11/1/2019 12:00:00 AM	1.
12	12/1/2019 12:00:00 AM	1.
12	4/1/2020 12:00:00 AM	1.
13	1/1/2020 12:00:00 AM	1.
14	2/1/2020 12:00:00 AM	1.

Campo No Datetime

- Castear en Power Query

- Date

- CampoFecha >= Datetime.Date(RangeStart) And
CampoFecha < Datetime.Date(RangeEnd)

```
fx = Table.SelectRows(dbo_PBI_Incr_Refresh, each [TransactionDate] >= DateTime.Date(RangeStart) and [TransactionDate] < DateTime.Date(RangeEnd))
```

- Numérico

- Date.Year(RangeStart) * 10000 + Date.Month(RangeStart)
* 100 + Date.Day(RangeStart)

Queries [4]

- PBI_Incr_Refresh
- RangeStart (1/1/2019 12:00:00 AM)
- RangeEnd
- fx fnDateTok

```
fx = (x as datetime) => Date.Year(x)*10000 + Date.Month(x)*100 + Date.Day(x)
```

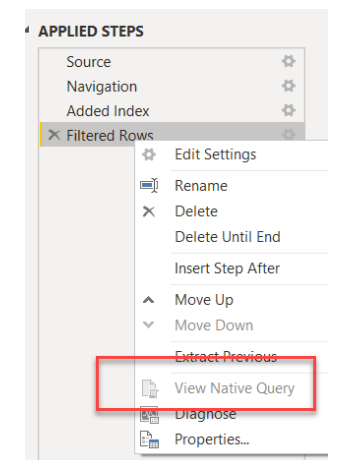
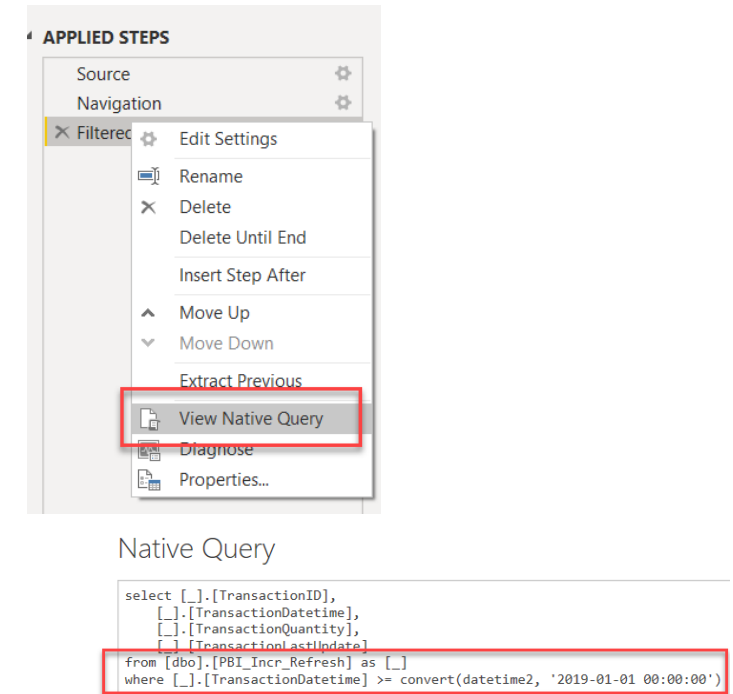
Enter Parameter

x

```
= Table.SelectRows(dbo_PBI_Incr_Refresh, each [TransactionDateKey] >= fnDateToKey RangeStart) and [TransactionDateKey] < fnDateToKey(RangeEnd))
```

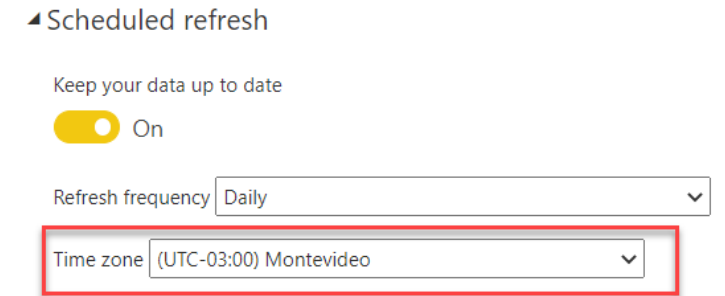
Query folding

- Es importante empujar el filtro a la fuente de datos
- Si la fuente no soporta Query folding
 - Flat files, Blob, Web, oData
 - El filtro se produce localmente
 - Viajan todos los datos y se filtran a posterior
 - No mejorarían tiempos
 - Puede agotar recursos de procesamiento
- Power BI Desktop verifica existencia de Query folding
 - En algunos tipos de orígenes debemos verificar con un Trace



Current Date

- La fecha actual es determinada en Power BI Service
- Utiliza el Time Zone definido en la planificación del dataset.
 - En actualizaciones manuales y automáticas



Detect data changes

- Permite reducir las particiones a procesar
 - Genera consultas al origen para determinar el procesamiento
- Debe ser un campo fecha diferente
 - Fecha de Modificación
- Requiere que la columna este en el modelo
- Podemos reducir tamaño por
 - Guardar el máximo valor (Power Query)
 - Modificar precisión del campo
 - Definir un custom query para detectar cambios (Premium)

Table: PBI_Incr_Refresh

Incremental refresh: On

Store rows where column "TransactionDatetime" is in the last: 1 Years

Refresh rows where column "TransactionDatetime" is in the last: 2 Months

☒ Detect data changes [Learn more](#)

Only refresh data in the last 2 months if the maximum value of this column is less than or equal to the value of this column: TransactionLastUpdate

☐ Only refresh complete months [Learn more](#)

Only refresh complete periods

- Permite evitar incorporar periodos en curso
- Tiene un alcance limitado a la frecuencia de la actualización
 - Por ejemplo, cierre de mes el día 10.
 - No debemos actualizar días previos.

Table: PBI_Incr_Refresh

Incremental refresh: ☒ On

Store rows where column "TransactionDatetime" is in the last: 1 Years

Refresh rows where column "TransactionDatetime" is in the last: 2 Months

☐ Detect data changes [Learn more](#)

☒ Only refresh complete months [Learn more](#)

Power BI Premium

- XMLA endpoint
 - Actualizar desde Management Studio (SSMS)
 - Actualizar particiones históricas
 - Modificar el comportamiento de la actualización
 - TMSL + TOM
 - applyRefreshPolicy → False → Procesamiento full
 - effectiveDate → modificar fecha de hoy
- Query custom para detector cambios
 - TMSL + TOM
 - Permitiría evitar tener la columna "Fecha modificación" en el modelo
 - pollingExpression → expresión M que se aplica a cada partición
- Desplegar solo Metadata
 - Power BI Desktop siempre actualiza modelo completo, forzando procesamiento full

Use partitions to divide a table into logical parts that can be processed independently.

Table: SalesLight

Partitions

Partition Name	# Rows	Last Processed
2018	3615870	4/14/2020 10:06:43 AM
2019	4579586	4/14/2020 10:06:48 AM
2020Q101	408394	4/14/2020 10:06:08 AM
2020Q102	407737	4/14/2020 10:05:42 AM
2020Q103	469698	4/14/2020 10:06:08 AM
2020Q204	213740	4/14/2020 10:06:07 AM

XMLAQuery1.xml...ctor@pepsico.com)

```
{
  "refresh": {
    "type": "full",
    "objects": [
      {
        "database": "PBI Inc - 2 - 12MM - Incr",
        "table": "SalesLight",
        "partition": "2018"
      }
    ]
  },
  "createOrReplace": {
    "object": {
      "database": "PBI Inc - 2 - 12MM - Incr",
      "table": "SalesLight",
      "partition": "2020Q102"
    },
    "partition": {
      "name": "2020Q102",
      "mode": "import",
      "source": {
        "type": "policyRange",
        "start": "2020-02-01T00:00:00",
        "end": "2020-03-01T00:00:00",
        "granularity": "month"
      }
    }
  }
}
```

Power BI dataflow

- Incremental en Premium
- Es un poco diferente
 - No permite definir criterio de filtro
- Dataflow no es una fuente que soporte Query folding

Query settings >

Name
SalesLight

Entity type ⓘ
Custom

Applied steps
Source
X Navigation

☒ On

Choose a DateTime field to filter by
Date

Store rows from the past
2 Years

Refresh rows from the past
2 Months

☐ Detect data changes [Learn more](#)

Only refresh data if the maximum value in this field changes
Choose a field

☐ Only refresh complete months [Learn more](#)

When you save these settings, data from the past 2 years will be loaded to your dataflow storage the next time this dataflow is refreshed. Subsequent refreshes will update only data that's changed in the past 2 months.

Query settings >

Name
SalesLight

Entity type ⓘ
Custom

Applied steps
Source
Navigation
X SalesLight-446174...

`Table.SelectRows(Navigation, each DateTime.From([Date]) >= RangeStart and DateTime.From([Date]) < RangeEnd)`

	StoreKey	ProductKey	PromotionKey	CustomerKey	SalesQuantity	SalesAmount	UpdateDate
2:00:00 ...	306	782	10	333	1	10.36	
2:00:00 ...	306	782	10	334	1	10.36	
2:00:00 ...	306	782	10	335	1	10.36	
2:00:00 ...	306	782	10	336	1	10.36	
2:00:00 ...	306	782	10	337	1	10.36	
2:00:00 ...	306	1701	10	333	1	3.98	