

Power BI Performance tips



Ing. Alejandro Cabanchik

Consultor BI

<https://www.smartbi.com.ar/>

alejandro.cabanchik@smartbi.com.ar

+54 11 5113 8649

@ale_cabanchik

 smart BI

Microsoft
Partner

Silver Data Analytics
Silver Data Platform



Argentina Edition
Junio 16 2020

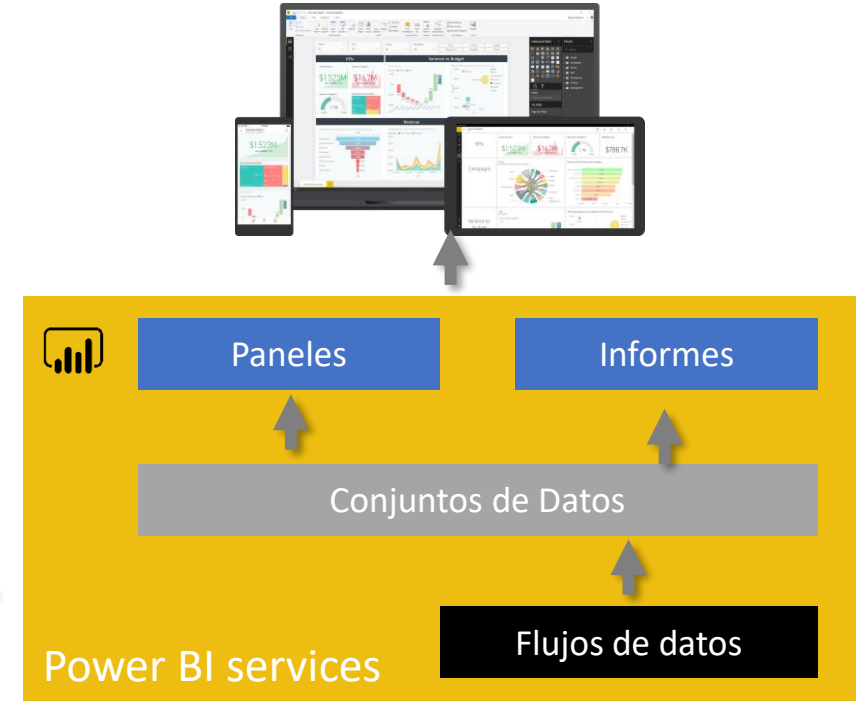


Arquitectura



Cloud data sources

Scheduled refresh/ Live connection/ DirectQuery



On-prem data sources



Power BI gateway

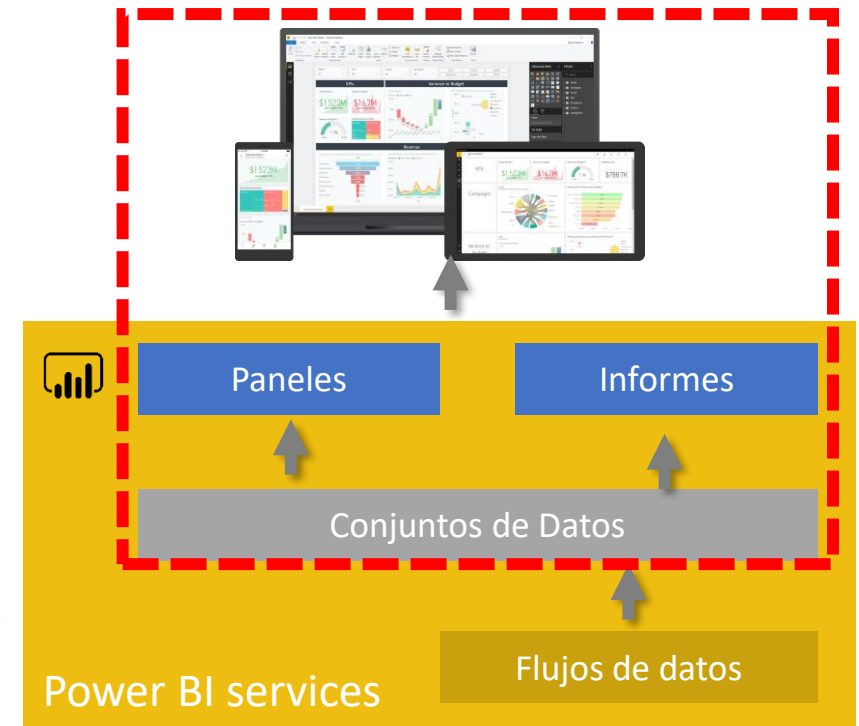
Scheduled refresh/ Live connection/ DirectQuery

Modo Importado - Consultas



Cloud data sources

Scheduled refresh/ Live connection/ DirectQuery



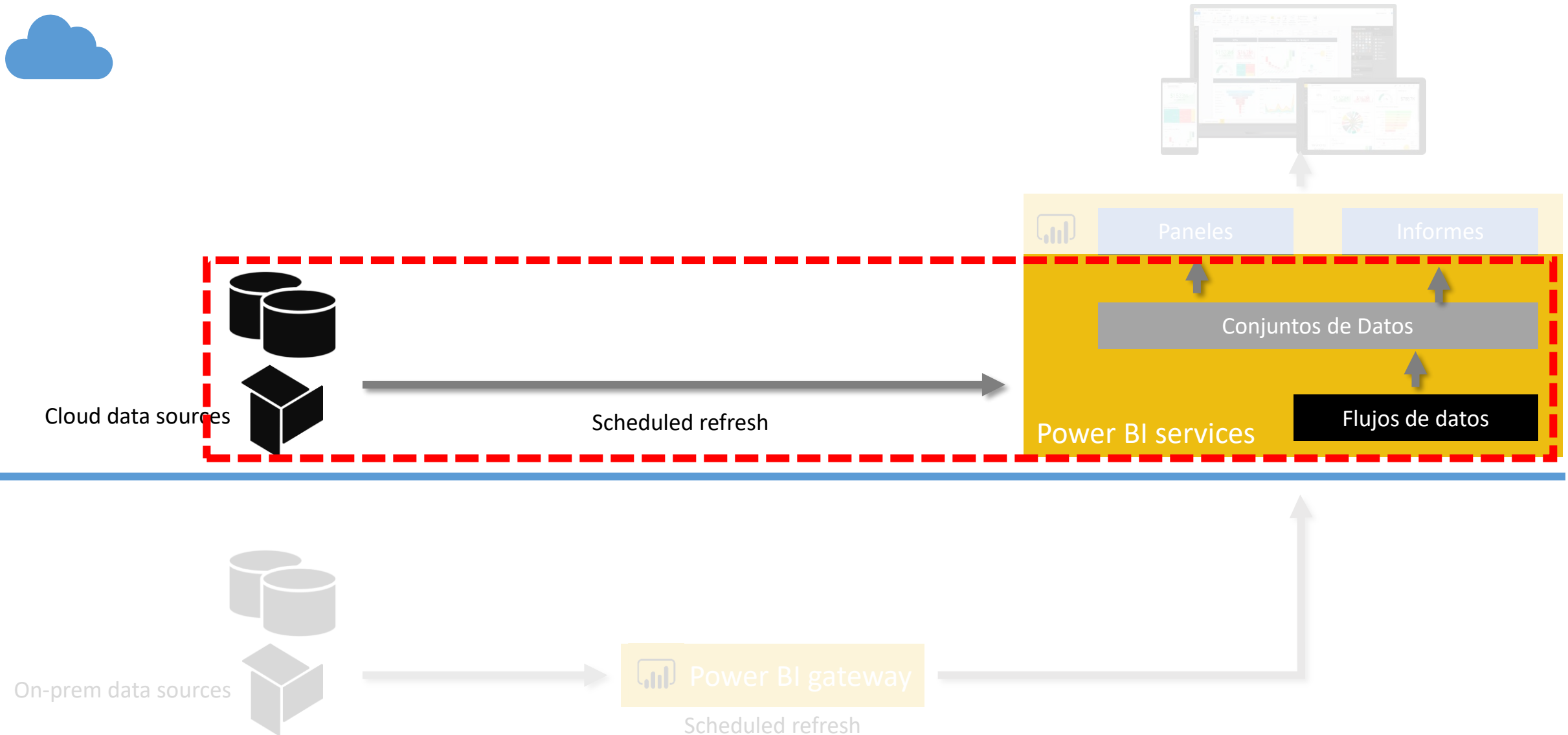
On-prem data sources



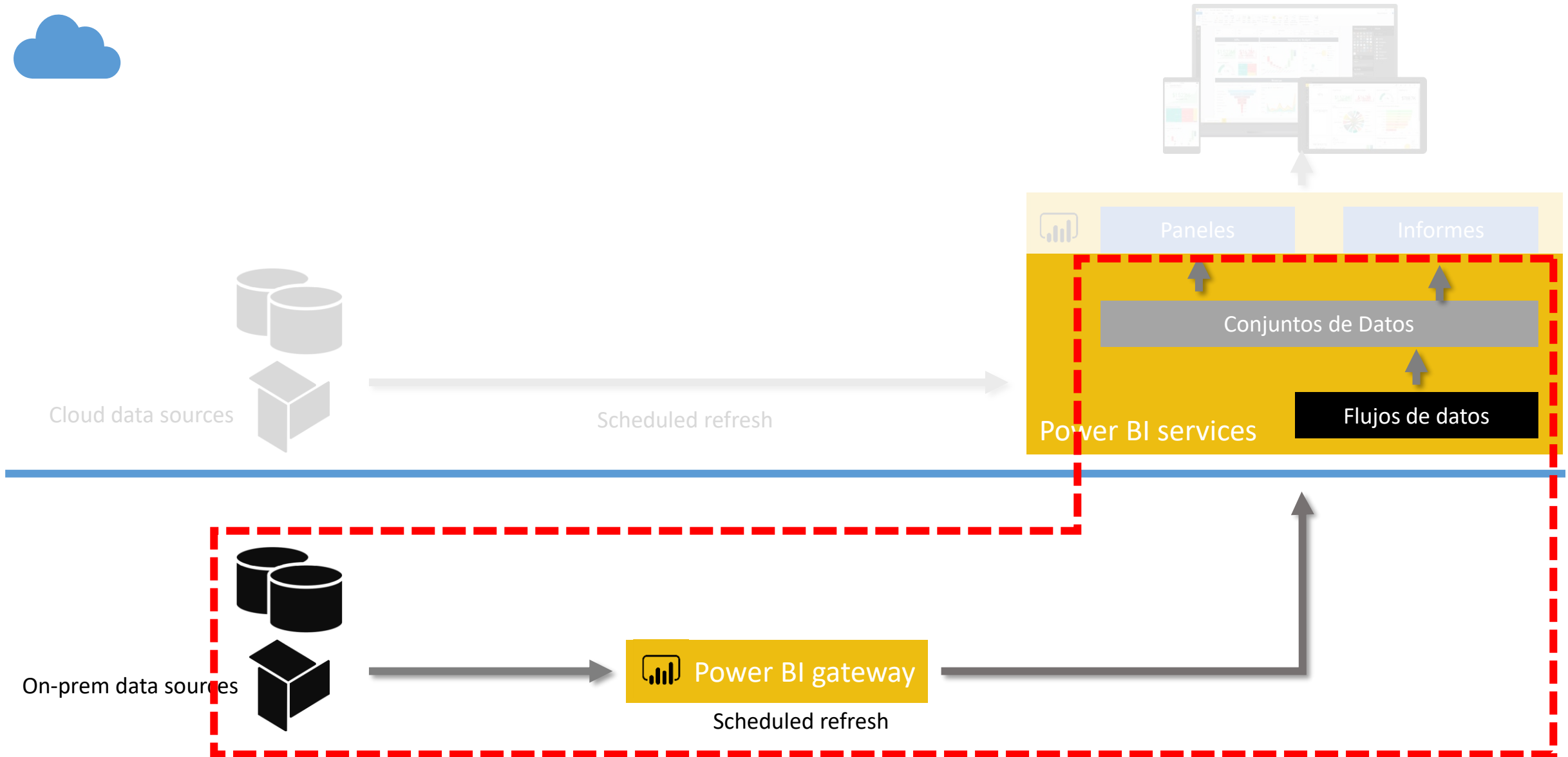
Power BI gateway

Scheduled refresh/ Live connection/ DirectQuery

Modo Importado – Actualización



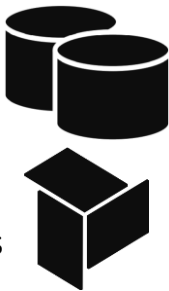
Modo Importado – Actualización



Modo DirectQuery/Live



Cloud data sources



DirectQuery/Live



Paneles

Informes

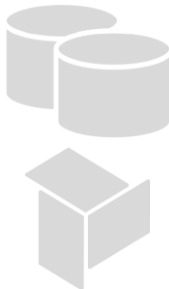
Conjuntos de Datos

Flujos de datos

Power BI services



On-prem data sources



Power BI gateway

DirectQuery/Live

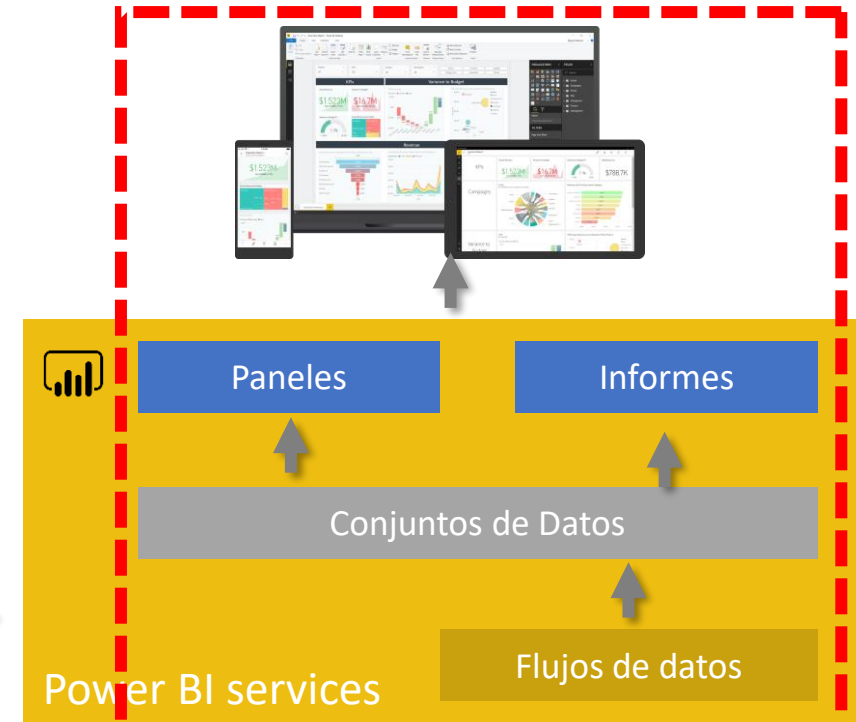
Modo DirectQuery/Live



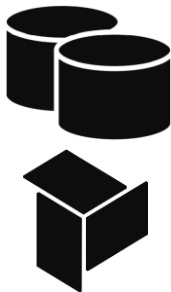
Cloud data sources



DirectQuery/Live



On-prem data sources



Power BI gateway

DirectQuery/Live



Importar, DirectQuery o Live

1. Importación

- Siempre Primera opción
- En memoria
- Máxima velocidad
- DAX sin límites

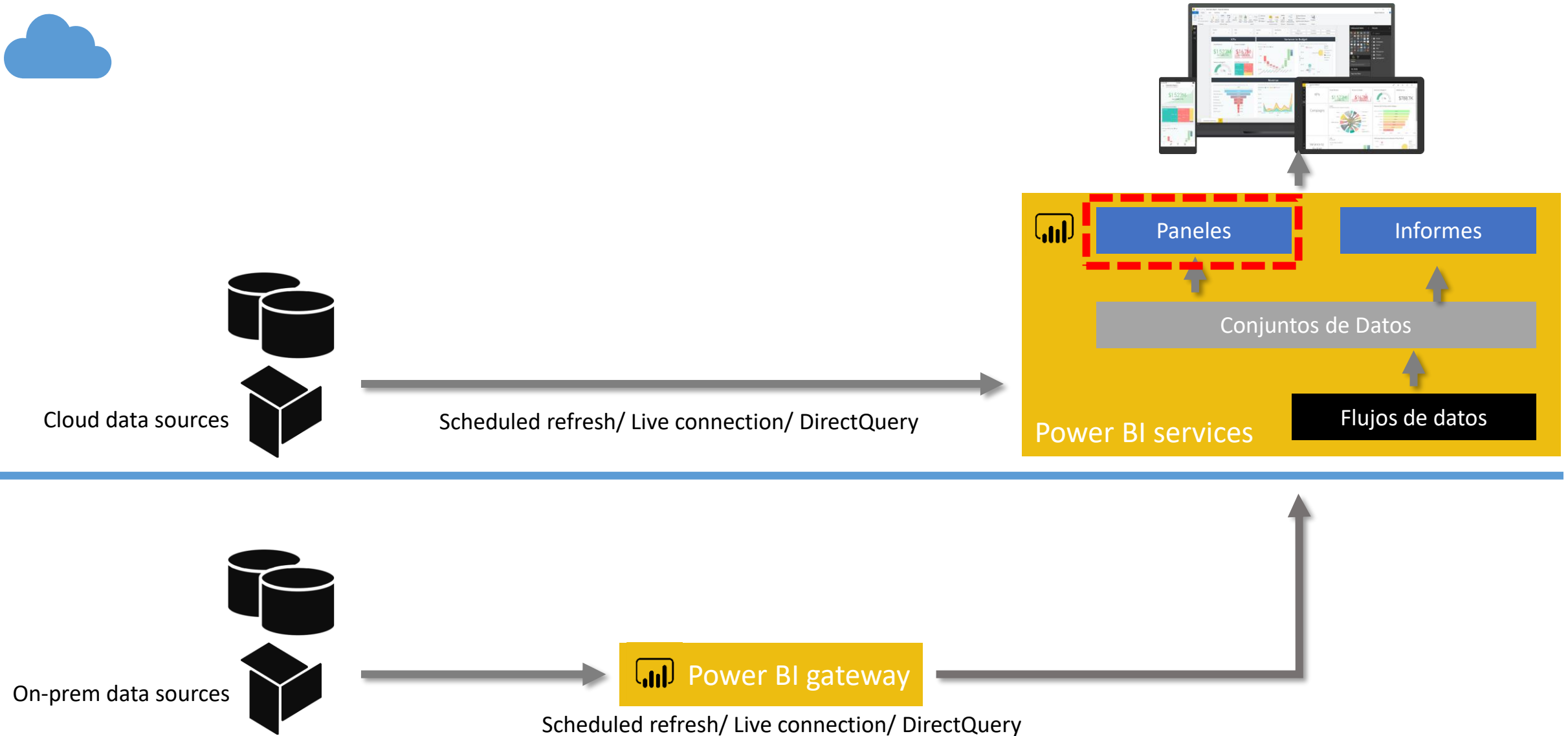
2. Live (Análisis Services)

- Inversión existente, modelos maduros
- Control en particiones y actualización de datos
- Requisitos de soberanía regulatoria y de datos

3. DirectQuery

- Volúmenes de datos extremadamente grandes
- Necesidad de acceso casi en tiempo real a los datos de la fuente
- Inversión existente en DW o OLAP externo
- Requisitos de soberanía regulatoria y de datos

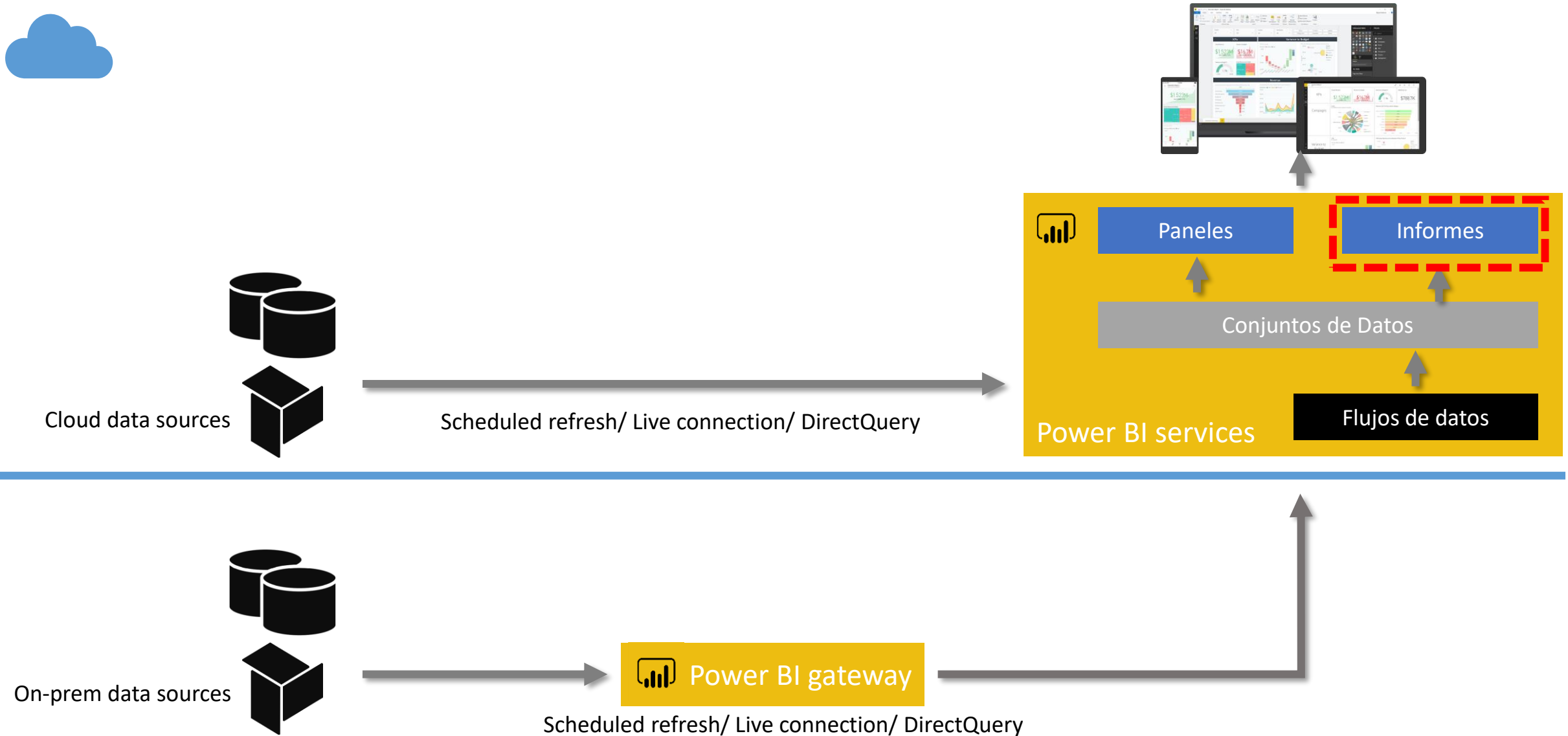
Paneles



Paneles

- Pensar Paneles como un punto de partida
 - Vista resumen ágil
 - Detalle en reportes
- Evitar anclar Reportes
 - Se ejecutan a demanda
 - No manejan cache
 - Anclar Visualizaciones
- Ajustar actualización de la cache
 - En modo importado, automático al actualizar datos
 - En DirectQuery y Live predeterminado cada una hora

Informes



Informes (Visualizaciones)

- Limitar la cantidad visualizaciones por hoja (20)
 - Cada elemento visual desencadena al menos una consulta de datos
 - Usar "Marcadores" para ocultar elementos visuales
- Evitar tablas/matrices con complejas (muchas columnas y medidas)
 - Rediseñar a múltiples hojas (resumen + "Obtener detalle")
 - Mover medidas menos críticas a "Tooltips".
- Remover Iteraciones visuales innecesarias
 - Deshabilitar "filtros cruzados".
 - Ajustar iteraciones.

Informes (Segmentaciones)

- Evitar “Segmentaciones” con muchos elementos (miles de filas)
 - Reducir cantidad de valores (Filtros de Visualización)
 - Mover a filtros
- Asignar valor predeterminado en “Segmentaciones”
 - Seleccionar valor por defecto
 - Ajustar tipo de selección (simple/múltiple)
- Usar “Segmentaciones” sincronizados con prudencia
 - Usar filtros en lugar de “Segmentaciones”
 - Limitar cantidad de valores
 - Cambiar a “Obtener detalle”

Informes (Varios)

- Utilizar características de “Reducción de consultas”
 - Deshabilitar “resaltado y filtrado cruzado” predeterminado
 - Agregar botones de “Aplicar”
- Evitar la exportación intensiva de datos
 - Limitar la exportación en tiempo de diseño
 - Usar reportes paginados
- Medir prestaciones de las visualizaciones “no nativas”
 - Probar de forma aislada
 - Usar visualizaciones certificadas
 - Reemplazar con visualizaciones nativas

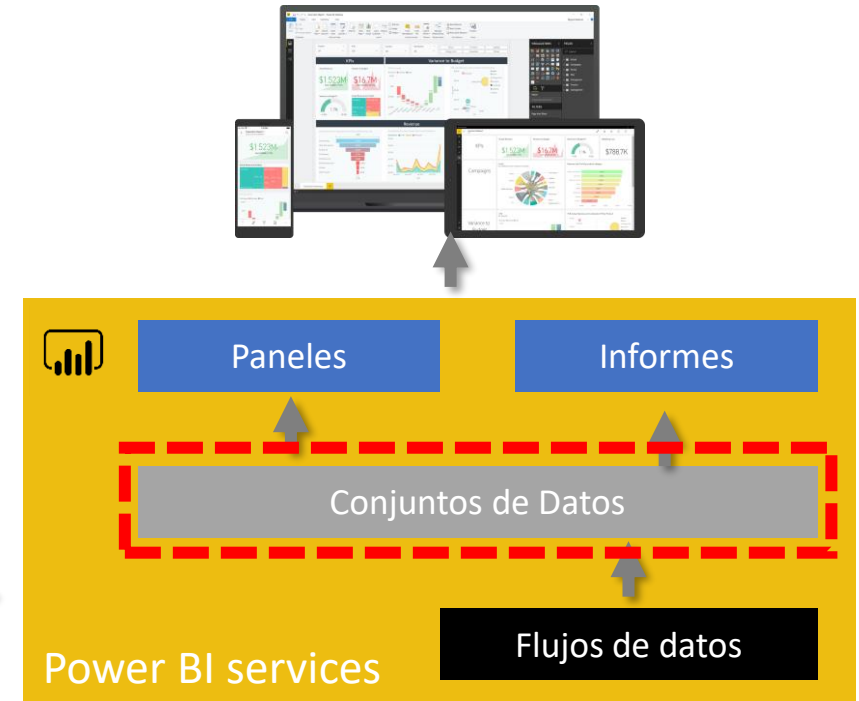
Demo

Conjunto de datos



Cloud data sources

Scheduled refresh/ Live connection/ DirectQuery



On-prem data sources



Power BI gateway

Scheduled refresh/ Live connection/ DirectQuery

Modelo de Datos (1)

- Modelado
 - “Esquema en Estrella” (Modelado Dimensional)
 - Estructura los datos específicamente para el análisis
 - Reduce “Joins” e incrementa “Usabilidad”
 - <https://docs.microsoft.com/en-us/power-bi/guidance/star-schema>
 - “Bus matrix”
 - Ayuda a priorizar y comunicar alcance/progreso
 - Ayuda a identificar dimensiones conformadas/compartidas
 - Candidatas a dataflows?
- <https://www.kimballgroup.com/>



Modelo de Datos (2)

- Eliminar columnas y tablas innecesarias
 - Incrementa tamaño del modelo
 - Incrementa tiempos de carga en memoria
 - Incrementa tiempos de actualización
- Remover o reducir columnas de alta precisión/cardinalidad
 - Tienen poca compresión
 - Incrementan tiempos de carga en memoria
 - Incrementan tiempos de actualización
- Utilizar números enteros en lugar de cadenas de texto
 - Es conveniente revisar tipos de datos y ajustar

Modelo de Datos (3)

- Utilizar claves substitutas tipo Entero, ordenadas
 - El ordenamiento maximiza compresión
- Usar relaciones bidireccionales solo si son necesarias
 - Con muchas relaciones podrían afectar performance
 - A veces pueden generar confusión al usuario
- Evitar las relaciones M2M
 - Son relaciones débiles, no performance como las 1-M
 - Utilizar una "bridge table"
- Definir tipo de agregación predeterminada
 - Establecer agregación en "Ninguna" cuando corresponda.

Modelo de Datos (4)

- Mover columnas calculadas al origen
 - Con alta cardinalidad No comprimen como columnas físicas
- Utilizar subconjuntos para modelos muy grandes
 - Evaluar uso de modelos compuestos y agregaciones
 - Construir tablas resumen con medidas inteligentes
- En DirectQuery aplicar Modelos compuestos y agregaciones
 - Dimensiones en modo Importado, hechos en DirectQuery
 - Agregaciones para consultas mas frecuentes

DAX (1)

- Utilizar variables en lugar de repetir medidas o expresiones
 - Reducen la cantidad de cálculos
 - Código mas legible y mejora tiempos
- Utilizar la función DIVIDE en lugar de /
 - DIVIDE verifica que numerador no sea 0 ni BLANK
 - Si el denominador no puede dar 0 utilizar /
- No reemplazar valores BLANK con 0 u otro valor
 - Power BI utiliza los valores BLANK para descartar filas
 - Si se reemplaza incrementa tamaño de respuesta y tiempos

Demo

DAX (2)

- Calcular "ratios" eficientemente
 - $(a-b)/b$ en lugar de $a/b-1$
 - El segundo caso generar resultados cuando no hay información
 - Utilizar variables
- Utilizar `SELECTEDVALUE()` en lugar de `HASONVALUE() + VALUES() + IF()`
- No utilizar `IFERROR()` e `ISERROR()`
 - Gran impacto en las prestaciones
 - Anticipar el error
 - `DIVIDE`, `FIND`, `SEARCH` permiten parámetros error

DAX (3)

- Evitar expresiones en SUMMARIZE
 - Solo Tabla y columnas de agrupación
 - No utilizar parámetros de expresiones
 - Utilizar SUMMARIZECOLUMNS
- Utilizar ISBLANK() o == BLANK() en lugar de = BLANK()
- Utilizar las funciones DISTINCT y VALUES de forma consistente
 - DISTINCT solo valores existentes
 - VALUES agrega valor BLANK por problemas de integridad
- Evitar la manipulación de cadenas de texto y FORMAT en medidas
 - Hace lentas las medidas

Demo

DirectQuery (1)

- Empujar las transformaciones la fuente de datos
 - Evitar transformaciones en M
- Activar Integridad referencial de la relación
 - Evitar el OUTER JOIN
- Evitar relaciones por columnas calculadas
 - Utilizar COMBINEVALUES(), empuja al origen
 - Crear columnas en origen
 - Usar vistas materializadas
- El uso de RLS podría ayudar

DirectQuery (2)

- Optimizar la fuente de datos
 - No siempre es suficiente optimizar en Power BI
 - Analizar consultas y planes
 - Comprender capacidades del origen
 - Indices, In-Memory, Columnar, Vistas materializadas, replicas de solo lectura
- Ajustar cantidad de conexiones al origen
- Considerar modelos compuestos y agregaciones
 - Dims en modo importado
 - Facts en DirectQuery y agregaciones

Demo

Seguridad de filas

- Evitar filtros de seguridad sobre tabla de hechos
 - Definir reglas de filtro en las dimensiones
- Evitar condicionales y manipulación de textos
 - Las expresiones deben ser lo mas performance posible
- Crear tablas de seguridad compactas
- Tabla de hechos con múltiples criterios de seguridad en diferentes dimensiones
 - Crear Junk dimension

Power Query

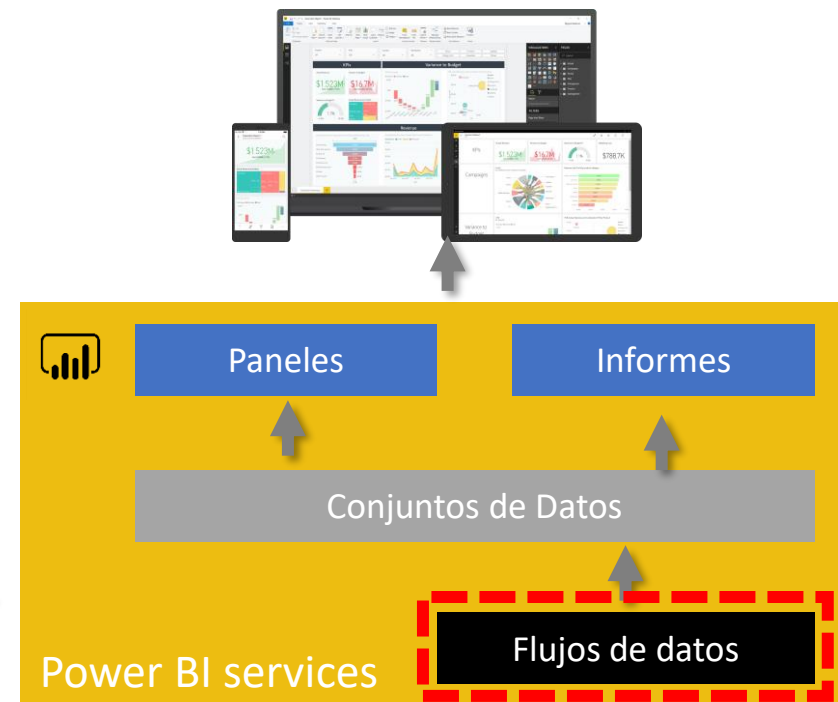
- Maximizar el “Query folding” (consultas nativas o en origen)
 - Mover al inicio todos los pasos que permiten “folding”
- Deshabilitar la configuración de privacidad
 - Un uso incorrecto puede forzar a evitar “folding”
- Evitar transformaciones que escanean tablas completas
 - Toda la tabla cargarse en memoria antes de pasar al siguiente paso
- No cargar en memoria consultas intermedias
 - Ocupa memoria, ocultarla no resuelve el problema
- Las referencias de Querys no son “tablas temporales” necesariamente
 - Table.Buffer
- Deshabilitar la fecha/hora automática
 - Crea tantas tablas de fechas como columnas fechas tengamos

Flujo de datos



Cloud data sources

Scheduled refresh/ Live connection/ DirectQuery



On-prem data sources



Power BI gateway

Scheduled refresh/ Live connection/ DirectQuery

Flujo de datos

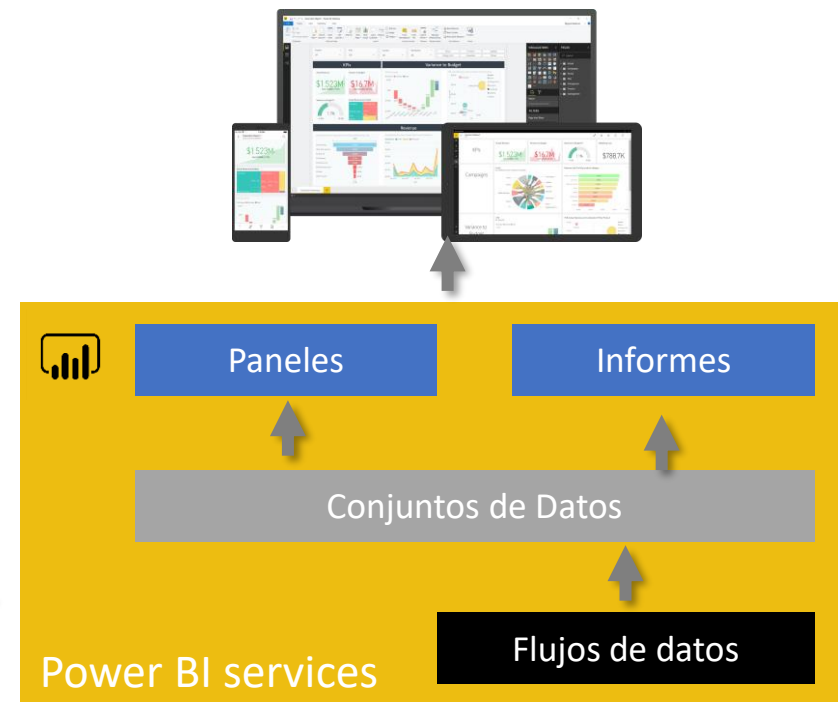
- Crear flujos separados para fuentes locales y en la nube
- Separar la Ingestión de la Transformación (P)
 - Crear flujos de datos separados
- Ajustar tamaño y memoria máxima del contenedor (P)
 - Mejora el paralelismo
 - Puede afectar otras cargas de trabajo

Puerta de enlace



Cloud data sources

Scheduled refresh/ Live connection/ DirectQuery



On-prem data sources

Scheduled refresh/ Live connection/ DirectQuery



Power BI gateway

Puerta de enlace (1)

- Planificar la actualización de datos y cache
 - Procesar cuando es realmente necesario
- Instalar lo mas cerca de la fuente de datos
- Puertas de enlace dedicadas según uso
 - Puerta de enlace para Actualización programada
 - Puerta de enlace para Consultas DirectQuery/Live
- Utilizar almacenamiento rápido/SSD
 - Alojamiento local de la puerta de enlace en almacenamiento rápida

Puerta de enlace (2)

- Reducir la latencia de red
 - Co-localizar servicios
 - Replica en la nube
 - Azure ExpressRoute
- Supervisar y escalar si es necesario
 - 8 cores y 16 GB RAM
 - Cluster
- Prestar atención a límites de paralelismo y ajustar
 - Predeterminados
 - DirectQuery: 20 (MashupDQPoolContainerMaxCount)
 - Actualizaciones simultaneas: 6 (MashupDefaultPoolContainerMaxCount)
 - 2 x core
- Habilitar Auto-scaling
 - Permite escalar en forma automática al incorporar CPUs.
 - MashupDisableContainerAutoConfig

Herramientas

- Informes
 - Ver → Analizador de Rendimiento
- Modelo de Datos/DAX
 - Ver → Analizador de Rendimiento
 - [DAX Studio](#)
- PowerQuery
 - PowerQuery → Diagnostico de Consultas
- Gateway
 - Gateway Performance Monitor
 - QueryExecutionReportOn
 - SystemCounterReportOn
 - ReportFilePath
 - [Power BI Template](#)

Muchas gracias