

DP-600 · CHEAT SHEET · CERTIFYDECK

Microsoft Certified: Fabric Analytics Engineer Associate

Dominios, pesos oficiales y habilidades clave del temario de DP-600, para repasar antes del examen.

Nivel: Intermediate (Associate) Duración: 100 min Nota de corte: 700/1000 Fuente: guía de estudio oficial de Microsoft Learn

25–30%

- **Saint-Emilion** nos brinda la funcionalidad de **funciona a nivel de área de trabajo**: solo un administrador la conecta o desconecta de una rama de Azure DevOps o GitHub, y después se confirman los cambios del área de trabajo o se actualiza.
- **Desarrollo de Datos**: El **SPDR** guarda el informe y el modelo en una carpeta de archivos de texto para cada definición de referencia al modelo y la carpeta **definition** contiene el modelo en TMDL. El **.gitignore** excluye el modelo.
- **Procesamiento de datos**: El **SPDR** usa un **archivo de trabajo** para cada fase de desarrollo. El **cache** de **datos** y las **definiciones de datos** se emparejan. Las reglas de origen de datos, de **parámetros** o de **almacenamiento** predeterminado se crean en la fase de destino y se aplican en la siguiente implementación. Para implementar hay que ser administrador de la canalización y colaborador en ambas áreas de trabajo.
- **Formato de escritura**: El **SPDR** usa un **archivo de trabajo** para cada fase de desarrollo. El **cache** de **datos** y las **definiciones de datos** se emparejan. Las reglas de origen de datos, de **parámetros** o de **almacenamiento** predeterminado se crean en la fase de destino y se aplican en la siguiente implementación. Para implementar hay que ser administrador de la canalización y colaborador en ambas áreas de trabajo.

45–50%

- El comando **GROUP BY** agrupa los grupos de Fabric por nivel en la tabla compuesta de puntos de conexión de análisis SQL y bases de datos reflejadas sin código, usando la vista de diagrama de Power Query: arrastrar tablas, combinarlas, filtrar y agrupar.
- En un primer paso de datos SELECT punto de conexión de análisis SQL de Fabric se filtra con WHERE antes de guardar el resultado como una tabla GROUPED TABLE para luego poder trabajar sobre las columnas no agregadas, y ROLLUP o CUBE añaden más niveles de agregación.
- En los pasos de filtro con WHERE se filtran las columnas combinando la suma agregada con summarize del paso de binning para agregar fechas al análisis SQL y al mismo tiempo se puede crear la función Dcount() (estimación) o arg_max() para obtener la máxima fecha de cada grupo.
- La consulta RAW muestra los datos en su estado original y es el resultado de una expresión de cálculo sin que se aplique ningún tipo de modificación ni de formato. Se define por medio de la cláusula de consulta SUMMARIZECOLUMNS agrupa y agrega omitiendo las filas en blanco, TOPN puede devolver más filas si hay empates y CALCULATE sustituye los filtros de la misma columna salvo que se use KEEPFILTERS.

3. Implement and manage semantic models

25–30%

Design and build semantic models

- Importación copia los datos en memoria y necesita actualizaciones; DirectQuery consulta el origen y sirve para volúmenes enormes o datos casi en tiempo real, con límites en M y DAX; Dual combina ambos para dimensiones. Direct Lake carga tablas Delta de OneLake con el motor VertiPaq y su actualización solo copia metadatos: on OneLake admite varios orígenes sin recurrir a DirectQuery, y on SQL usa un único origen y puede recurrir a él.
- En un modelo semántico, el tipo de tabla lo decide la cardinalidad de la relación: el lado uno es dimensión y el lado varios, hecho. Conviene desnormalizar las dimensiones en copo de nieve en una sola tabla, crear una tabla por rol para las dimensiones de rol múltiple y cargar las dimensiones de tipo 2 ya versionadas desde un almacén de datos, porque Power Query no puede generarlas.
- Para relacionar dos dimensiones de varios a varios se usa una tabla puente con relaciones de uno a varios y una relación bidireccional, ocultando la tabla puente. Los hechos de mayor granularidad se relacionan con una relación de varios a varios en una sola dirección. Las relaciones de varios a varios y entre grupos de origen son limitadas: usan combinación interna y no admiten RELATED.
- Las variables (VAR) se evalúan una sola vez y fuera de los filtros que aplica RETURN, lo que mejora rendimiento y legibilidad. En CALCULATE conviene usar filtros booleanos con KEEPFILTERS y reservar FILTER para condiciones con medidas. WINDOW, OFFSET e INDEX trabajan con posiciones dentro de particiones, y funciones de información como ISINSCOPE o HASONEVALUE detectan el contexto.
- Los grupos de cálculo aplican expresiones reutilizables con SELECTEDMEASURE solo a medidas explícitas; la prioridad decide cómo se anidan varios grupos. Las cadenas de formato dinámicas cambian el formato según el contexto sin convertir la medida en texto, a diferencia de FORMAT. Los parámetros de campo permiten al lector elegir en una segmentación qué medidas o columnas usa un objeto visual.

Optimize enterprise-scale semantic models

- Para acelerar informes, limita los objetos visuales por página y usa obtención de detalles e información sobre herramientas, aplica filtros N superior en tablas grandes y desactiva totales innecesarios en DirectQuery. La reducción de consultas desactiva el filtrado cruzado y añade botones Aplicar. El Analizador de rendimiento desglosa cada objeto visual en consulta DAX, Direct query, visualización y otros.
- Para mejorar el rendimiento de DAX, usa variables para no evaluar dos veces la misma expresión, DIVIDE cuando el denominador puede ser cero o BLANK y COUNTROWS para contar filas. Pasa filtros booleanos a CALCULATE (con KEEPFILTERS si hay que conservar filtros) y reserva FILTER para condiciones con medidas. Evita IFERROR y no conviertas BLANK en cero.
- Un modelo Direct Lake se actualiza mediante enmarcado, que solo copia metadatos y fija la versión Delta que ven las consultas; las columnas se cargan en memoria a petición. Las actualizaciones automáticas enmarcan al detectar cambios y pueden desactivarse para esperar al ETL. En Direct Lake on SQL, DirectLakeBehavior decide si la reserva a DirectQuery es automática, prohibida o forzada.
- Direct Lake on OneLake lee tablas Delta de uno o varios elementos de Fabric mediante las API de OneLake, aplica OneLake security y nunca recurre a DirectQuery; es la opción recomendada para modelos nuevos. Direct Lake on SQL usa un único origen, comprueba permisos en el punto de conexión de análisis SQL y recurre a DirectQuery con vistas, RLS de SQL o barreras superadas.
- La actualización incremental filtra la tabla con los parámetros Fecha/Hora RangeStart y RangeEnd, que el servicio reemplaza para crear particiones de actualización e históricas en una ventana móvil; conviene que el filtro se pliegue en el origen. Opciones: detectar cambios con una columna de auditoría, solo días completos y una partición DirectQuery en tiempo real. Volver a publicar borra las particiones: usa XMLA para cambios solo de metadatos.