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

1. Maintain a data analytics solution

25–30%

Implement security and governance

- Los roles de área de trabajo (Admin, Member, Contributor y Viewer) se asignan a usuarios o grupos y se aplican a todos los elementos. Si alguien está en varios grupos, obtiene el rol más alto. Viewer lee contenido y datos con T-SQL, pero no usa Spark ni escribe; Contributor crea y ejecuta; Member además añade personas; solo Admin elimina el área y conecta Git.
- Compartir un elemento da acceso a ese elemento sin rol en el área de trabajo y siempre concede Read. Se pueden añadir ReadData (T-SQL), ReadAll (Spark y OneLake), Build o Write en modelos semánticos, y Share, que solo permite conceder los permisos propios. Los permisos heredados de un rol no se quitan desde el elemento; para afinar en un almacén de datos se usa GRANT.
- En el almacén de datos, la RLS usa CREATE SECURITY POLICY con funciones insertadas con valores de tabla. La CLS se aplica con GRANT sobre columnas y la OLS con GRANT, REVOKE y DENY. En modelos semánticos, RLS y OLS se definen en roles y solo afectan a Viewers. La seguridad de OneLake concede Read por carpeta, tabla, vista o columna, con el rol DefaultReader para usuarios con ReadAll.
- Las etiquetas de confidencialidad de Microsoft Purview se aplican desde la cabecera o la configuración del elemento y requieren permisos de edición. Pueden heredarse al crear de forma descendente (sin sobrescribir etiquetas manuales ni rebajar la restricción) y desde orígenes en modelos de importación. En el servicio no limitan el acceso salvo con directivas de protección; el cifrado viaja en las exportaciones compatibles.
- La aprobación ayuda a encontrar contenido fiable mediante tres insignias. Promoted la aplica cualquiera con permisos de escritura; Certified y Master data solo los usuarios autorizados por el administrador de Fabric, que además necesitan escritura. Se puede promocionar o certificar todo salvo los paneles de Power BI, y Master data solo en elementos con datos.

Maintain the analytics development lifecycle

- La integración con Git de Fabric 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 desde la rama completa. Solo se versionan definiciones (no datos) y, para aislar el desarrollo, cada desarrollador puede bifurcar a su propia área de trabajo.
- Un proyecto de Power BI (PBIP) guarda el informe y el modelo semántico como carpetas de archivos de texto listos para Git: definition.pbir referencia al modelo y la carpeta definition contiene el modelo en TMDL. El .gitignore predeterminado excluya cache.abf y localSettings.json. Publicar desde Power BI Desktop envía también la caché de datos; Git y las API solo implementan metadatos.
- Una canalización de implementación asigna un área de trabajo a cada fase (por defecto Development, Test y Production) y copia metadatos, no datos, entre elementos emparejados. Las reglas de origen de datos, de parámetros o de almacén de lago 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.
- El análisis de impacto muestra los elementos y áreas de trabajo que pueden verse afectados al cambiar un elemento, con hijos directos o todos los elementos aguas abajo, y permite avisar a sus contactos con Notify contacts. Requiere permiso de escritura sobre el elemento, y en los orígenes de datos solo muestra hijos directos. La vista de linaje no muestra elementos aguas abajo de otras áreas de trabajo.
- La conexión XMLA expone cada área de trabajo como un servidor de Analysis Services: es de solo lectura por defecto y debe habilitarse para lectura y escritura en la capacidad para implementar modelos, ejecutar TMSL o hacer actualizaciones detalladas por partición. Los colaboradores equivalen a administradores de base de datos; las credenciales se configuran en el servicio, no mediante XMLA.

2. Prepare data

45–50%

Get data

- Las conexiones se crean en Administrar conexiones y puertas de enlace y tienen su propia lista de acceso (Usuario, Usuario con posibilidad de volver a compartir, Propietario). Los datos locales necesitan la puerta de enlace estándar; el modo personal solo sirve para import en Power BI. La identidad del área de trabajo evita gestionar secretos y permite el acceso de confianza a almacenamiento con firewall.
- El catálogo de OneLake centraliza la detección, Explorar enumera los elementos accesibles con filtros, dominios y detalles (esquema, linaje); Gobernar muestra el estado de gobernanza y Proteger los roles de seguridad. Real-Time hub, disponible en todo Inquilino, recibe flujos, tablas KQL, conectores de streaming y eventos de Fabric, de Azure y empresariales.
- Para acceder sin copiar se usan accesos directos de OneLake (internos o a ADLS, S3, GCS…), con caché opcional para reducir la salida. Para ingerir: creación de réplicas reflejadas para replicación continua de solo lectura, Copy job para copias masivas e incrementales sin canalizaciones, canalizaciones para orquestación, Dataflow Gen2 para Power Query, flujos de eventos para streaming y Spark para código.
- Eventhouse sirve para eventos de streaming y análisis detallado con KQL; Cosmos DB en Fabric para NoSQL; la base de datos SQL en Fabric para OLTP; el almacén de datos para almacenamiento empresarial en T-SQL con transacciones de varias tablas, y el almacén de lago para Spark y datos no estructurados, con un punto de conexión SQL de solo lectura. Todos están en OneLake, y el KQL es opcional.
- La disponibilidad en OneLake de un Eventhouse crea una copia lógica en formato Delta de una base de datos o tabla KQL, legible por Spark, almacenes de lago y Direct Lake; mientras está activa no se pueden renombrar tablas, cambiar tipos de columna ni purgar datos. La integración con OneLake de un modelo semántico escribe sus tablas de importación como tablas Delta tras cada actualización.

Transform data

- En Fabric Data Warehouse, con el punto de conexión de análisis SQL puedes crear vistas, funciones con valores de tabla insertadas y procedimientos almacenados; el punto de conexión es de solo lectura, así que la lógica que modifica tablas va en un almacén de datos. No hay vistas materializadas: T-SQL para resultados persistidos se usan vistas materializadas de lago, y en KQL, funciones almacenadas.
- Para enriquecer datos en Power Query se usan columnas personalizadas (fórmula M), condicionales, a partir de ejemplos o de índice, y combinaciones de consultas, con coincidencia aproximada en columnas de texto. En un almacén de datos, ALTER TABLE solo agrega columnas que aceptan NULL y las tablas nuevas se crean con CTAS, SELECT INTO o INSERT…SELECT, en KQL, extend agrega columnas calculadas al resultado.
- Un esquema de estrella separa tablas de hechos, con claves de dimensión y medidas, de tablas de dimensión desnormalizadas con clave suplente. En Fabric Warehouse la clave suplente suele ser BIGINT IDENTITY, la de fecha un INT YYYYMMDD, y las claves principales y externas se declaran NOT ENFORCED, por lo que el ETL carga primero las dimensiones y comprueba la integridad.
- Las tablas de dimensión casi siempre se desnormalizan aplanando jerarquías como categoría y subcategoría porque su redundancia compensa por rendimiento y facilidad de uso; el copo de nieve solo conviene para relacionar hechos de mayor nivel o seguir historial en niveles superiores. Se desnormaliza con vistas que combinan tablas, combinaciones de Power Query con expansión de columnas o lookup en KQL.
- En Dataflow Gen2, Agrupar por resume filas con operaciones como Suma, Contar filas, Contar valores distintos o Todas las filas, y admite agrupación aproximada para textos con variantes. En T-SQL de Fabric, ROLLUP, CUBE y GROUPING SETS añaden subtotales y total general, y GROUP BY ALL agrupa por todo lo no agregado. En KQL, summarize agrega por grupos y arg_max devuelve la fila completa del valor máximo.

Query and analyze data

- El editor de consultas visual de Fabric permite consultar almacenes, 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. Solo ejecuta SELECT de solo lectura con pasos que se pliegan; muestra el T-SQL generado y guarda el resultado como vista o como tabla (CREATE TABLE AS SELECT).
- En un almacén de datos o un punto de conexión de análisis SQL de Fabric se filtra con WHERE antes de agrupar y con HAVING después; GROUP BY ALL agrupa por todas las columnas no agregadas, y ROLLUP o CUBE añaden subtotales. Los nombres de tres partes permiten combinar elementos de la misma área de trabajo, y el punto de conexión de análisis SQL admite vistas pero no crea tablas ni DML.
- En KQL se filtra con where, se eligen columnas con project y se agrega con summarize … by, usando bin() para agrupar fechas en intervalos y funciones como count(), countif() o dcount() (estimación) o arg_max() para obtener la fila completa del máximo. Más sencillos y más rápidos que contains, que busca subcadenas), top ordena de forma descendente por defecto y take no garantiza qué filas devuelve.
- Una consulta DAX necesita al menos un EVALUATE seguido de una expresión de tabla (un escalar va entre llaves) y admite un único bloque DEFINE con medidas de ámbito 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.