Arquitectura de los procesos ETL para carga de avances

Proyecto: Contrato de prestación de servicios n.° 202500136 de 2025 celebrado entre la corporación Ruta N - Medellín y Guane Enterprises S.A.S

Fecha de elaboración: 2026-08-12

Índice

Resumen Ejecutivo 3
1. Objeto del Documento 3
2. Principio rector del planteamiento 3
3. Visión General de la Arquitectura 4
4. Actores y Responsabilidades 5
4.1 Actores de gobernanza y decisión. 5
4.2 Actor humano operativo - Funcionario DAP. 6
4.3 Componentes técnicos e identidades de servicio. 7
5. Descripción por Nodos 8
5.1 Nodo de Usuario Institucional 8
5.2 Nodo de Fuentes en Intranet 9
5.3 Nodo de Procesamiento en Azure 10
6. Flujo Operativo 12
7. Control de Acceso 14
7.1 Identidades de servicio - autenticación sin contraseñas entre componentes Azure 14
7.2 Roles operativos para Funcionarios DAP 14
7.3 Cuentas de servicio para acceso a sistemas on-premise 16
7.4 Principios de seguridad de red 16
8. Política de retención de datos 17
8.1 Horizonte y principio general 17
8.2 Retención por zona 17
8.3 Controles de reproceso e idempotencia 17
8.4 Almacenamiento de bajo acceso 17
9. Políticas de triggers y ejecuciones automáticas 17
9.1 ETL de archivos / variables (ETL-CARGA-VARIABLES) 18
9.2 ETL de capas (ETL-CARGA-CAPAS) 18
9.3 Reglas comunes 18
10. Volúmenes y costos 18
10.1 Síntesis para decisión 19
10.2 Alcance y premisas 20
10.3 El sistema y los volúmenes 21
10.3.1 Variables: 61 archivos 22

10.3.2 Capas: lo medido y lo proyectado 22
10.3.3 Acumulado por zona y volumen del modelo 23
10.3.4 El modelo de escenarios en detalle 24
10.4 Estructura de costos 24
10.5 Tarifas validadas 26
10.6 Escenarios de dimensionamiento de la base de datos 27
10.6.1 E1 - Mínimo (B1ms) 27
10.6.2 E2 - Tier actual (B2s) 27
10.6.3 E3 / E3b - Intermedio (B2ms o D2ds_v5) 27
10.6.4 E4 - Producción con alta disponibilidad (D4ds_v5 + HA) 28
10.7 Evidencia de medición real (Fase 1) 29
10.7.1 Cuello de botella #1 - vistas analíticas (CPU) 29
10.7.2 Cuello de botella #2 - render de teselas MVT “on-the-fly” 29
10.8 Proyección anual a 12 años 30
10.9 Frente 2 - Almacenamiento del modelo de escenarios 30
10.10 Palancas de ahorro y simulación 31
10.11 Decisiones del Distrito 32
10.12 Sensibilidad y riesgos 33
10.13 Confianza y trazabilidad 34
10.14 Recomendación 35
10.15 Glosario de costos y catálogo de SKU 35
10.15.1 Glosario de costos 35
10.15.2 Catálogo de SKU y metodología 36
11. Conclusiones 37
Anexos 38
Anexo A - Diagrama de despliegue implementado 38
Anexo B - Operación manual y monitoreo institucional 39
Anexo C - Ingesta ETL desde fuentes on-premise 40
Anexo D - Flujo extremo a extremo de ingesta ETL 41
Anexo E - Glosario de arquitectura 42

Resumen Ejecutivo

Este documento describe la arquitectura de los procesos ETL que cargan los avances del POT Medellín (variables tabulares y capas geoespaciales), desde las fuentes institucionales on-premise hasta la

publicación en la Zona Oro. Azure Synapse Analytics orquesta el flujo; el Self-Hosted Integration Runtime integra los nodos en red privada con el lago de datos y la base de publicación en Azure.

El documento está orientado a la operación y al gobierno del ETL: nodos de despliegue, actores y roles, flujo operativo, control de acceso, retención de datos y políticas de disparo (manual y programado). Además de la estimación de volúmenes y costos (horizonte a doce años) para sustentar el dimensionamiento de la plataforma, de modo que la arquitectura operativa y el criterio económico queden en un solo entregable.

El detalle de los tres nodos y del flujo Bronce → Plata → Oro se desarrolla en las secciones 3 a 6.

1. Objeto del Documento

El presente documento tiene como objeto describir la arquitectura de los procesos ETL de carga de avances del sistema del POT (variables tabulares y capas geoespaciales), con énfasis en la organización de nodos, componentes, flujo operativo, estructura de pipelines de Azure Synapse Analytics, modelo de seguridad y control de acceso, política de retención, políticas de ejecución automática y estimación de volúmenes y costos. La descripción incorpora los nombres reales de los recursos y artefactos. Establece la referencia técnica para la operación y el seguimiento del proceso de ingesta.

El alcance de esta arquitectura se limita al despliegue y operación del proceso ETL. No incluye la capa de aplicación de visualización de datos, cuya definición corresponde al Distrito en una fase posterior.

2. Principio rector del planteamiento

El principio rector de esta arquitectura es separar tres responsabilidades: orquestar el flujo, transformar los datos y almacenarlos por zonas. Azure Synapse Analytics coordina los pipelines, la validación (notebooks) y las transformaciones tabulares (data flows). El Data Lake guarda los insumos en zonas progresivas (Bronce → Plata). La base Azure PostgreSQL guarda el resultado (Zona Oro). El nodo de usuario institucional supervisa y controla la operación desde la interfaz web de Synapse.

En la práctica, los pipelines ETL coordinan el orden de las actividades, y las reglas de negocio viven en los notebooks y data flows de Synapse. El pipeline calcula el periodo a procesar, toma los insumos del servidor institucional mediante el agente de integración (Self-Hosted Integration Runtime), los copia al Data Lake en la Zona Bronce (copia cruda), ejecuta la validación hacia la Zona Plata y publica variables y capas en la Zona Oro, dejando trazabilidad de cada corrida. Las credenciales de la base de datos se obtienen en el momento de la ejecución desde Key Vault, con Identidad Administrada, sin guardar secretos en el código.

3. Visión General de la Arquitectura

Arquitectura de despliegue, procesos ETL para carga de información de avances de POT

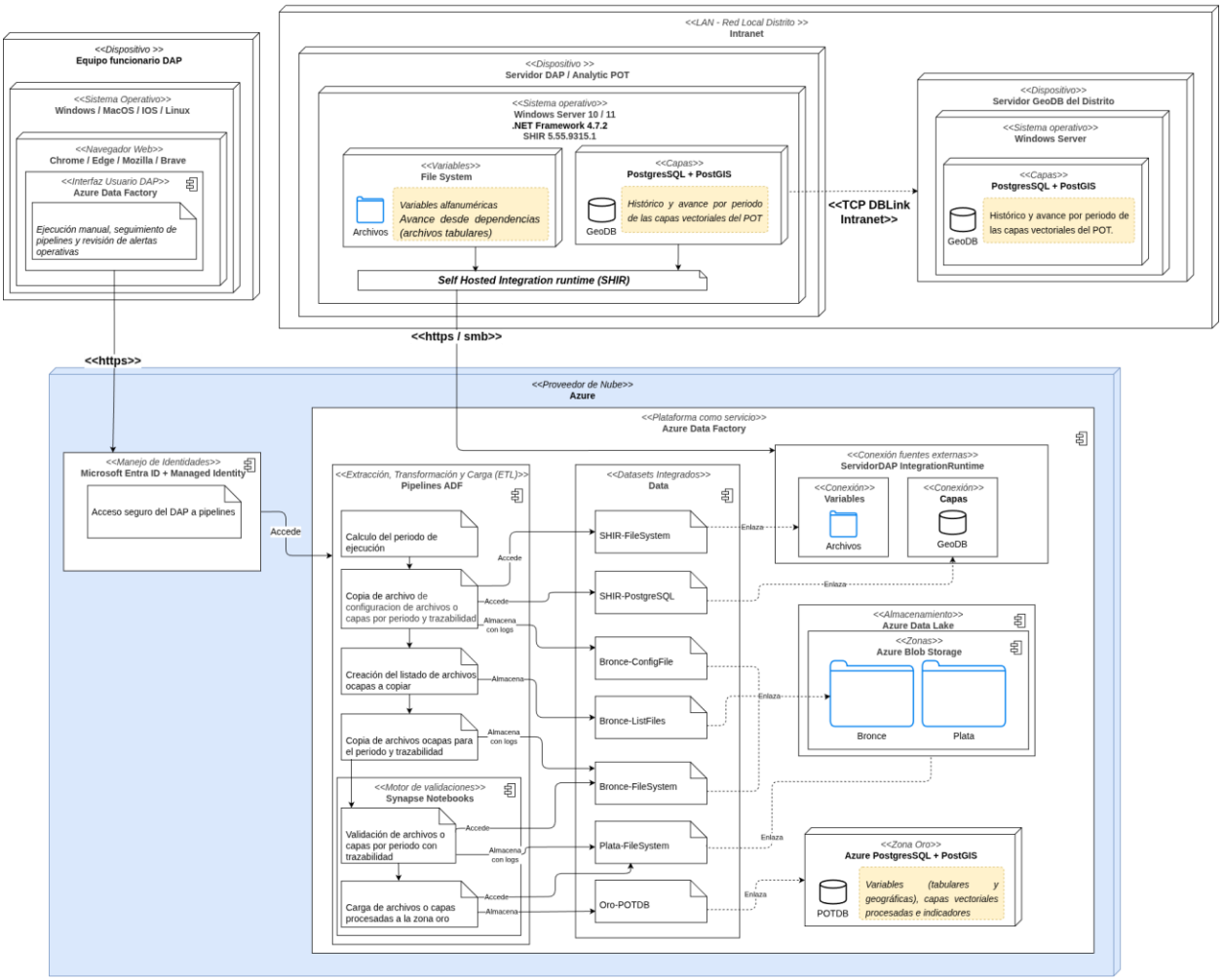


Figura 1. Diagrama de despliegue de la solución ETL.

La arquitectura se organiza en tres nodos principales, acorde a la Figura 1, el del usuario institucional, el de la Intranet con dos servidores on-premise y el de Azure con la infraestructura de orquestación, transformación y almacenamiento.

El nodo de Intranet comprende el Servidor DAP / Analytic POT como origen de los archivos tabulares de variables y del archivo de configuración (config.json), y el Servidor GeoDB del Distrito como origen activo de las capas vectoriales oficiales del territorio. Ambos servidores son fuentes de solo lectura; ninguno recibe datos ni es modificado por el proceso ETL.

Dentro del dominio Azure se distinguen cuatro capacidades: Azure Synapse Analytics como plataforma central de orquestación y transformación ETL; Azure Data Lake Storage Gen2 como soporte de

almacenamiento estructurado en las zonas Bronce y Plata; Azure Database for PostgreSQL Flexible Server como repositorio de datos de la Zona Oro; y Azure Key Vault como bóveda centralizada de credenciales.

El flujo inicia con el disparo del proceso por parte del Funcionario DAP desde la interfaz web de Synapse (web.azuresynapse.net), manual o mediante trigger programado según la política de la Sección 9, continúa con la extracción de insumos desde el recurso compartido del Servidor DAP a través del Self-Hosted Integration Runtime, el aterrizaje en la Zona Bronce, la validación y transformación hacia la Zona Plata mediante los notebooks de Synapse, y la publicación de datos en la Zona Oro. Los artefactos de Synapse (pipelines, datasets, linked services, integration runtimes y notebooks) se almacenan y versionan en el repositorio de GitHub, conectado al workspace mediante la configuración Git nativa del servicio.

4. Actores y Responsabilidades

La arquitectura involucra actores en tres planos complementarios: el de gobernanza y decisión (quién define el presupuesto, la modalidad y el dimensionamiento de la solución), el de operación humana (quién deposita insumos, ejecuta el proceso, monitorea y consume resultados) y el de componentes técnicos e identidades de servicio (qué elementos de la plataforma orquestan, transforman, almacenan y se autentican entre sí). Distinguir estos planos clarifica quién decide, quién opera y qué autentica, y delimita responsabilidades sin solapamientos. Las decisiones estratégicas de dimensionamiento y costeo se sustentan en la Sección 10 (Volúmenes y costos), de la cual se toma el desglose de responsabilidades de decisión de la siguiente tabla.

4.1 Actores de gobernanza y decisión.

Definen el marco presupuestal, operativo y de infraestructura dentro del cual opera la plataforma. Sus decisiones no se ejecutan en el flujo ETL, pero lo condicionan.

Actor	Responsabilidad principal	Decisiones a su cargo
Distrito - Secretaría de Innovación	Aprobación presupuestal y de modelo de servicio	Presupuesto plurianual (horizonte 12 años); modalidad de la Zona Oro (nube Azure o en tierra / DAP); régimen de disponibilidad (24/7 u horario 8×5); ventana de riesgo (holguras 30 %–100 %)

Distrito - DAP (área funcional)	Definición del uso y validación de resultados	Perfil de uso y concurrencia esperada; SKU de la Zona Oro (E1–E4, decidido con carga medida); frecuencia de cargas; validación funcional de los datos publicados; depósito de insumos por las dependencias
Guane Enterprises (equipo técnico)	Implementación, operación y soporte de la plataforma ETL	Aprovisionamiento de la infraestructura; mantenimiento del código en el repositorio dap-pipelines; medición de carga real; recomendación de SKU y palancas de ahorro; operación y monitoreo del flujo

Tabla 1. Actores de gobernanza y decisión.

4.2 Actor humano operativo - Funcionario DAP.

Es el único actor que interactúa directamente con el flujo ETL en producción. Desde la interfaz de Azure Synapse Analytics, en su equipo institucional, consulta estados, inicia ejecuciones (principalmente en el rol Ejecutor) y verifica el avance de la ingesta. Su actuación se organiza en tres roles diferenciados, cuyos permisos se detallan en la Sección 7.2.

Rol operativo	Responsabilidad en el flujo	Alcance
Editor	Depósito de archivos .xlsx de variables en \\10.0.0.4\ADF y actualización de config.json (archivos_variables y matrícula capas_estandarizadas); puede iniciar los pipelines maestros cuando se requiera, sin que ello sea su tarea principal	Interviene al comienzo del proceso (insumos y matrícula); la ejecución de ETL es secundaria; no modifica la plataforma
Ejecutor	Inicio de las ejecuciones de los pipelines maestros (ETL-CARGA-VARIABLES y ETL-CARGA-CAPAS), monitoreo de cada actividad, gestión de errores, reintento de sub-pipelines fallidos y consulta de logs de trazabilidad	Es quien inicia y opera las corridas ETL; no modifica la plataforma

Visualizador	Consulta de los datos publicados en la Zona Oro	No interviene en el ETL; consume el resultado

Tabla 2. Roles operativos del Funcionario DAP.

Como actores de soporte, las dependencias del Distrito (61 dependencias responsables) depositan mensualmente sus reportes de variables en las carpetas correspondientes del recurso compartido del servidor, previo a cada ejecución del proceso. No acceden a la plataforma Azure.

4.3 Componentes técnicos e identidades de servicio.

En síntesis, la plataforma se compone de orquestación (Synapse), puente al servidor institucional (SHIR), almacenamiento por zonas (Data Lake), publicación (PostgreSQL) y custodia de secretos (Key Vault).

Los componentes esenciales son: el workspace synapsedap (centro de orquestación); el agente DAP-IntegrationRuntime1 (puente seguro hacia el servidor institucional); las conexiones AzureDataLakeStorage, DAPKeyVault, SHIRFileSystem y SHIRCapasDB; los pipelines maestros ETL-CARGA-VARIABLES y ETL-CARGA-CAPAS (cada uno con etapas 01 Bronce, 02 Plata y 03 Oro); los notebooks de validación y carga; los data flows VariablesDataFlow y CapasDataflow; el pool SparkPoolDAPP1; el Data Lake (zonas Bronce y Plata); Key Vault; la base postgresql-dap (Zona Oro); Microsoft Entra ID; y el repositorio GitHub donde se versiona el código.

La gestión de identidades mediante Microsoft Entra ID y la Identidad Administrada del workspace synapsedap garantiza el acceso seguro del Funcionario DAP al workspace de Synapse y la autenticación sin contraseñas entre los servicios Azure. Las credenciales de los sistemas on-premise del Distrito se custodian exclusivamente en Azure Key Vault o como credenciales cifradas en los linked services y se inyectan en tiempo de ejecución; nunca residen en el código ni en archivos de configuración versionados. El modelo completo de identidades de servicio, roles operativos y cuentas de acceso a sistemas on-premise se desarrolla en la Sección 7.

5. Descripción por Nodos

A continuación se describe qué hay en cada nodo, equipo del funcionario, servidores institucionales y nube Azure, y cuál es su papel en el proceso de carga.

5.1 Nodo de Usuario Institucional

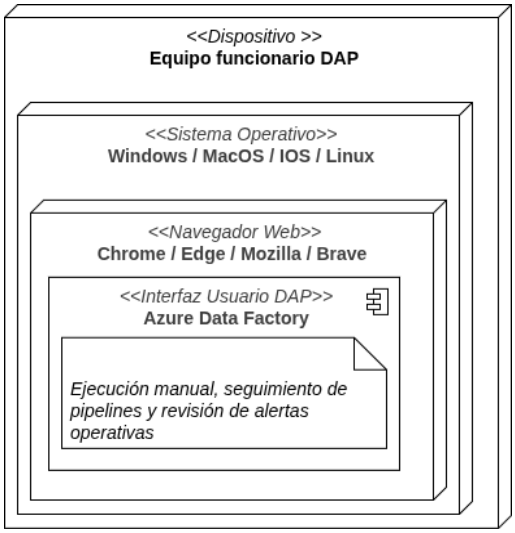


Figura 2. Nodo de usuario institucional (recorte del diagrama de despliegue).

En la Figura 2, el nodo Equipo Funcionario DAP representa el entorno desde el cual se realiza la supervisión operativa del proceso ETL. El dispositivo puede ejecutar Windows, MacOS o Linux, con acceso vía navegador web (Chrome, Edge, Mozilla Firefox o Brave). Desde este punto se accede a la interfaz de Azure Synapse Analytics (web.azuresynapse.net, workspace synapsedap) para consultar corridas, revisar alertas operativas, ejecutar los pipelines maestros ETL-CARGA-VARIABLES o ETL-CARGA-CAPAS con el parámetro de periodo p_periodo en formato YYYY_MM, y verificar el avance del flujo de ingesta.

El acceso a la interfaz de Synapse está habilitado por Microsoft Entra ID (el directorio de usuarios institucionales). En la operación cotidiana, el rol Ejecutor inicia las corridas de los pipelines maestros y hace seguimiento desde Synapse Studio; el rol Editor puede iniciarlas si se requiere, pero su tarea principal es preparar los insumos en el recurso compartido. Ningún funcionario manipula los servidores del Distrito desde Azure. Si el parámetro p_periodo se deja vacío, el pipeline calcula automáticamente el periodo del mes anterior respecto a la fecha de ejecución.

5.2 Nodo de Fuentes en Intranet

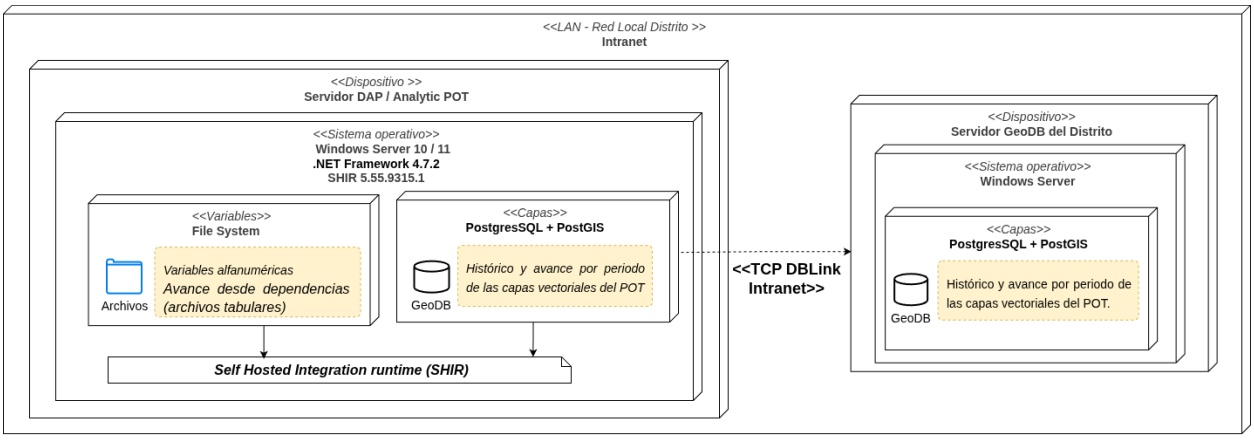


Figura 3. Nodo de fuentes en intranet (recorte del diagrama de despliegue).

El Servidor DAP / Analytic POT es el origen principal de los insumos institucionales on-premise. Su sistema operativo es Windows Server con .NET Framework 4.7.2. En este nodo residen la fuente de archivos y el componente de conectividad.

Fuentes de datos activas en el flujo ETL:

- File System (recurso compartido \\10.0.0.4\ADF): repositorio institucional organizado por dependencia. Las dependencias depositan los reportes de variables en formato .xlsx con cadencia mensual, en subcarpetas por periodo YYYY_MM. En la raíz del share reside el archivo config.json, que define el patrón de periodo, la lista de archivos de variables esperados (archivos_variables) y la matrícula de capas a extraer (capas_estandarizadas: dependencia, capa y query SQL). El share no almacena las geometrías de capas: sólo la configuración que orienta la extracción dinámica desde GeoDB.
- PostgreSQL + PostGIS (Servidor GeoDB del Distrito, 10.0.0.6): base de datos local con las capas vectoriales del territorio. Es la fuente activa del ETL de capas, referenciada mediante el linked service SHIRCapasDB y el dataset SHIRPostgreSqlTable. A partir de la matrícula en config.json (arreglo capas_estandarizadas), el sub-pipeline 01-CARGA-CAPAS-BRONCE ejecuta CapasDataflow y, por cada capa registrada, una copia dinámica (CopyCapaPostgreSQL) que corre la query configurada contra GeoDB y aterriza el resultado en la Zona Bronce como Parquet. No se ingieren capas como archivos planos desde el File System.

El Self-Hosted Integration Runtime DAP-IntegrationRuntime1 está instalado en el Servidor DAP / Analytic POT y registrado en el workspace synapsedap. Habilita, en red privada, la lectura del File System (variables y config.json) y la consulta al Servidor GeoDB (capas) sin exponer esos orígenes a Internet.

Ambos servidores on-premise forman parte de la Intranet de la organización. La comunicación entre la Intranet y Azure se realiza a través del Self-Hosted Integration Runtime, sin exponer los motores de base de datos ni el file system directamente a Internet.

5.3 Nodo de Procesamiento en Azure

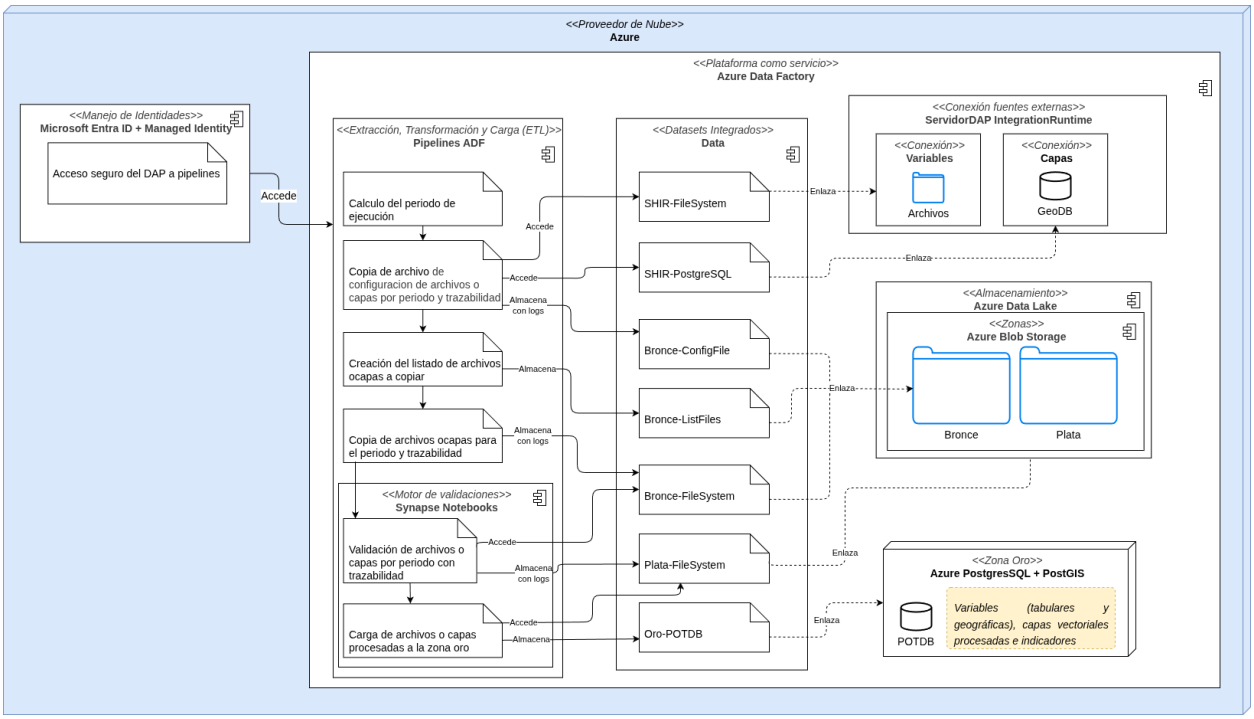


Figura 4. Nodo de procesamiento en Azure (recorte del diagrama de despliegue).

El dominio Azure del desarrollo, grupo de recursos QTE-DMDE-DAPCMP02, región West US 2, concentra cuatro funciones: orquestación y transformación (Azure Synapse Analytics), almacenamiento por zonas (Azure Data Lake Storage Gen2), custodia de credenciales (Azure Key Vault) y publicación de resultados (Azure Database for PostgreSQL).

Azure Synapse Analytics: plataforma central. Centraliza pipelines, notebooks, data flows, linked services e integration runtimes. El Funcionario DAP opera el workspace desde Synapse Studio (web.azuresynapse.net); el endpoint de desarrollo es https://synapsedap.dev.azuresynapse.net. El workspace opera con Identidad Administrada asignada por el sistema.

Repositorio dap-pipelines (GitHub, Git Configuration de Synapse Studio): control de versiones de los artefactos del workspace en la rama de colaboración (main). La acción Publish en Synapse Studio despliega las definiciones al entorno operativo. Los artefactos de conectividad y orquestación del workspace se resumen en la Tabla 3. El detalle de parámetros y rutas de lago permanece en el repositorio dap-pipelines.

Artefacto	Tipo	Función en el ETL
AutoResolveIntegrationRuntime	Integration Runtime (Azure)	Actividades en la nube, llamadas Web a Key Vault y data flows

DAP-IntegrationRuntime1	Integration Runtime (Self-Hosted)	Conectividad segura hacia el Servidor DAP (share) y el Servidor GeoDB (capas)
AzureDataLakeStorage	Linked Service	Lectura y escritura sobre las zonas Bronce y Plata del lago
DAPKeyVault	Linked Service	Obtención de secretos en tiempo de ejecución
SHIRFileSystem	Linked Service	Lectura del share institucional (config.json y archivos de variables) vía SHIR
SHIRCapasDB (+ dataset SHIRPostgreSqlTable)	Linked Service / Dataset	Fuente activa de capas: extracción dinámica desde GeoDB (queries de capas_estandarizadas)
ETL-CARGA-VARIABLES	Pipeline maestro	Orquesta la carga mensual de variables desde el share (01 → 02 → 03)
ETL-CARGA-CAPAS	Pipeline maestro	Orquesta la carga semestral de capas desde GeoDB (01 → 02 → 03)
01-*-BRONCE / 02-*-PLATA / 03-*-ORO	Sub-pipelines	Bronce: archivos (variables) o extracción GeoDB→Parquet (capas); luego validación y publicación
VariablesDataFlow / CapasDataflow	Data flow	Variables: manifiesto de archivos; Capas: matrícula capas_estandarizadas para la copia dinámica desde GeoDB
SHIRConfigFile, BlobConfig*, BlobList*, BlobFileSystem / BlobParquetCapaDinamica	Datasets	Configuración y listados; Parquet dinámico de capas (resultado de la extracción desde GeoDB)
ETLValidar* / ETLCarga*Oro / ETLCargarGeodatosOro	Notebooks	Reglas de negocio de validación y publicación en Zona Oro
VISUALIZAR-LOGS, VISUALIZAR-DATOS, VALIDAR-CONEXION	Pipelines de apoyo	Consulta de logs, verificación de datos y prueba de conectividad

Tabla 3. Catálogo de artefactos del nodo Azure (Synapse).

Cada maestro encadena tres sub-pipelines en secuencia Bronce → Plata → Oro.

Variables: 01-CARGA-ARCHIVOS-BRONCE →
02-VALIDAR-ARCHIVOS-PLATA →
03-CARGA-VARIABLES-ORO.

Capas: 01-CARGA-CAPAS-BRONCE →
02-VALIDAR-CAPAS-PLATA →
03-CARGA-CAPAS-ORO.

Los notebooks aplican las reglas de negocio; los pipelines de apoyo facilitan la supervisión operativa.

Azure Data Lake Storage Gen2: la cuenta del workspace provee el almacenamiento intermedio en dos zonas. Los datos fluyen siempre en una dirección: Bronce → Plata → Oro (PostgreSQL).

- Zona Bronce (bronce): aterrizaje sin transformación de negocio. Conserva los archivos .xlsx de variables tal como fueron entregados, la copia de config.json, los manifiestos generados por los data flows y, para capas, los Parquet obtenidos por extracción dinámica desde GeoDB (una copia cruda por capa/query). Es el registro de auditoría del sistema.
- Zona Plata (plata): datos validados y estandarizados producidos por los notebooks de validación (variables y capas), organizados por periodo. Los rechazos quedan registrados en el log de ejecución con la causa.

Azure Key Vault: custodia el secreto de conexión a la base de indicadores (linked service DAPKeyVault). Synapse lo lee en tiempo de ejecución mediante Managed Identity (rol Key Vault Secrets User); ninguna contraseña reside en el código ni en archivos de configuración.

Azure Database for PostgreSQL Flexible Server: destino final de variables, indicadores y capas publicadas (Zona Oro), organizados por periodo. Éste opera en West US 2 (PostgreSQL 18, SKU Standard_B1ms, 32 GB); el FQDN se consulta en el portal del recurso. Es la fuente de verdad publicada desde la cual el Distrito consulta resultados.

6. Flujo Operativo

El flujo operativo de cada proceso ETL (variables o capas) se desarrolla en siete etapas:

- Primera etapa - Preparación de insumos. Para variables, las dependencias depositan en \\10.0.0.4\ADF los archivos .xlsx del periodo (cadencia mensual). Para capas, las geometrías permanecen en el Servidor GeoDB; el rol Editor actualiza en config.json la matrícula del periodo (capas_estandarizadas con dependencia, capa y query) de modo que el ETL sepa qué capas extraer. El Editor verifica patrón de periodo, archivos de variables esperados y capas activas antes del disparo.
- Segunda etapa - Disparo del proceso. El Funcionario DAP accede a Azure Synapse Analytics y ejecuta el pipeline maestro correspondiente (ETL-CARGA-VARIABLES o ETL-CARGA-CAPAS), con p_periodo en formato YYYY_MM (o vacío para el mes anterior). La ejecución puede

ser manual (Trigger now) o automática según la Sección 9. A partir de ese momento, la plataforma encadena los tres sub-pipelines en secuencia.

- Tercera etapa - Ingesta y aterrizaje en Bronce. En variables, 01-CARGA-ARCHIVOS-BRONCE lee la configuración mediante DAP-IntegrationRuntime1, ejecuta VariablesDataFlow, enumera orígenes y copia los .xlsx al lago. En capas, 01-CARGA-CAPAS-BRONCE copia config.json al lago, ejecuta CapasDataflow para materializar la matrícula (capas.csv) y, en un ForEach, ejecuta CopyCapaPostgreSQL: cada query de capas_estandarizadas se corre contra GeoDB vía SHIRCapasDB y el resultado se escribe en Bronce como Parquet. En ambos casos se registra trazabilidad en los logs del periodo.
- Cuarta etapa - Validación y transformación a Plata. El sub-pipeline 02-VALIDAR-ARCHIVOS-PLATA o 02-VALIDAR-CAPAS-PLATA recupera la cadena de conexión desde Key Vault y ejecuta ETLValidarArchivosVariables o ETLValidarCapasGeograficas sobre SparkPoolDAPP1; escribe los datos aprobados en la Zona Plata.
- Quinta etapa - Publicación en Zona Oro. El sub-pipeline 03-CARGA-VARIABLES-ORO o 03-CARGA-CAPAS-ORO ejecuta ETLCargaVariablesOro o ETLCargarGeodatosOro, que lee Plata y publica en postgresql-dap, con control de idempotencia por lock.json y registro de trazabilidad por periodo.
- Sexta etapa - Supervisión operativa. El Funcionario DAP consulta el estado de la corrida y las alertas desde Synapse Studio. Puede reintentar sub-pipelines fallidos individualmente. Los pipelines VISUALIZAR-LOGS y VISUALIZAR-DATOS facilitan la consulta de logs y la verificación de datos cargados; VALIDAR-CONEXION verifica conectividad de fuentes y destinos.
- Séptima etapa - Cargas complementarias. Fuera del ciclo ordinario, los notebooks ETLCargaInstrumentoPOT y ETLCargarHistoricoIndicadores atienden la carga del instrumento de indicadores y la importación de series históricas desde Bronce hacia PostgreSQL.

La operación desde el nodo de usuario institucional (disparo, monitoreo y reintentos) se ilustra en el Anexo B. El desglose de la ingesta desde fuentes on-premise y de los sub-pipelines se muestra en el Anexo C. El flujo extremo a extremo, con la participación de todos los actores, se representa en el Anexo D.

7. Control de Acceso

7.1 Identidades de servicio - autenticación sin contraseñas entre componentes Azure

Los servicios de Azure de la plataforma se autentican entre sí con Identidades Administradas (Managed Identities), de modo que no se intercambian contraseñas entre componentes internos. Las credenciales de los sistemas on-premise del Distrito se guardan en Key Vault o cifradas en las conexiones (linked services)

y se usan solo en tiempo de ejecución; nunca van en el código. La Identidad Administrada del workspace synapsedap tiene los accesos resumidos en la siguiente tabla.

Identidad	Recurso al que accede	Nivel de acceso (rol RBAC)	Alcance del permiso
Workspace Synapse (synapsedap)	Lago de datos ADLS (synapsedap)	Storage Blob Data Contributor	Solo la cuenta de almacenamiento del proyecto
Workspace Synapse (synapsedap)	Bóveda de credenciales (synapsedap)	Key Vault Secrets User + Key Vault Reader	Solo el Key Vault del proyecto
Workspace Synapse (synapsedap)	Workspace (synapsedap)	Contributor	Solo el workspace del proyecto
Agente de despliegue (Guane DevOps)	Bóveda de credenciales	Key Vault Secrets Officer	Solo durante el despliegue de infraestructura
Agente de despliegue (Guane DevOps)	Grupo de recursos (QTE-DMDE-DAPCMP02)	Contributor	Solo el grupo de recursos del proyecto

Tabla 4. Identidades de servicio y roles RBAC.

7.2 Roles operativos para Funcionarios DAP

La operación distingue tres roles. Editor y Ejecutor intervienen directamente en el proceso ETL: el Ejecutor inicia las ejecuciones de los pipelines maestros; el Editor también puede iniciarlas, pero su tarea principal es el depósito y preparación de insumos. Visualizador no interviene: consume exclusivamente los indicadores publicados en la Zona Oro.

Rol	Acciones permitidas	Acciones no permitidas
Editor	Depositar .xlsx de variables en \\10.0.0.4\ADF; actualizar config.json (variables y matrícula de capas_estandarizadas); iniciar ETL-CARGA-VARIABLES o ETL-CARGA-CAPAS desde	Modificar pipelines, notebooks, data flows o configuraciones de la plataforma

	Synapse Studio cuando se requiera (no es su tarea principal)	
Ejecutor	Iniciar las ejecuciones de los pipelines maestros; monitorear el estado de cada actividad; gestionar errores; reintentar sub-pipelines fallidos; consultar logs de trazabilidad	Modificar la plataforma (pipelines, notebooks, data flows o configuraciones)
Visualizador	Consultar indicadores publicados en la Zona Oro (base indicadores)	Acceder a las zonas Bronce o Plata; iniciar procesos; modificar configuraciones

Tabla 5. Acciones permitidas y no permitidas por rol.

La tabla siguiente resume, en lenguaje de permisos de la plataforma, qué puede hacer cada rol en Synapse, en el recurso compartido, en el Data Lake y en la base de datos. La asignación concreta de usuarios a estos roles se gestiona con el procedimiento institucional de altas (fuera del alcance de este documento).

Rol	Synapse Studio	Share \\10.0.0.4\ADF	ADLS synapsedap	PostgreSQL indicadores
Editor	Synapse Artifact User + Synapse Compute Operator	Escritura (NTFS)	Sin acceso directo	Sin acceso
Ejecutor	Synapse Artifact User + Synapse Compute Operator	Sin escritura	Storage Blob Data Reader	SELECT (verificación)
Visualizador	Sin acceso	Sin acceso	Sin acceso	SELECT (consumo analítico)

Tabla 6. Permisos por herramienta según rol operativo.

7.3 Cuentas de servicio para acceso a sistemas on-premise

Cuenta	Sistema al que accede	Permiso otorgado	Restricción explícita
Cuenta de dominio DAP (en SHIRFileSystem)	File System \\10.0.0.4\ADF - Servidor DAP	Solo lectura sobre carpetas de dependencias y config.json	Sin escritura; sin acceso a otros directorios

Usuario PostgreSQL on-premise (en SHIRPostgreSqlTable)	Servidor GeoDB del Distrito (10.0.0.6)	Solo lectura sobre las capas matriculadas (ETL-CARGA-CAPAS activo)	Sin acceso a tablas administrativas; sin escritura
Secreto postgres-connection-string-indicadores (Key Vault)	Azure PostgreSQL postgresql-dap - base indicadores (Zona Oro)	Inserción y actualización desde los notebooks	Sin exposición en código; sin operaciones de eliminación

Tabla 7. Cuentas de servicio hacia sistemas on-premise.

7.4 Principios de seguridad de red

Principio	Implementación
Ninguna contraseña en el código	Las credenciales de PostgreSQL residen en Key Vault synapsedap; los pipelines las consultan en runtime mediante actividad Web con Managed Identity. Las credenciales del share están cifradas en el linked service SHIRFileSystem.
Ningún puerto de entrada abierto en el DAP	El DAP-IntegrationRuntime1 establece únicamente conexiones salientes hacia Azure por HTTPS (443); el DAP no recibe tráfico desde internet.
Mínimo privilegio en todos los niveles	Cada identidad y cuenta de servicio tiene acceso solo al recurso específico que su función requiere, según las matrices de las secciones 7.1 a 7.3.

Tabla 8. Principios de seguridad de red.

En el entorno actual de desarrollo, Synapse, PostgreSQL y la cuenta de almacenamiento operan con acceso de red pública controlado por autenticación (Entra ID), permisos (RBAC) y reglas de firewall. En síntesis: hoy la seguridad se basa en identidad y permisos; para producción se endurece la red (redes privadas/Private Endpoints, restringir IPs institucionales, alta disponibilidad en PostgreSQL y protección contra borrado en Key Vault). Estas medidas no cambian el flujo ETL ni el código de los pipelines.

8. Política de retención de datos

Esta sección define por cuánto tiempo se conservan los datos en cada zona del lago y en la base de publicación (Zona Oro). La premisa se alinea con el horizonte de doce años de la Sección 10 y con los controles del ETL que evitan recargar un periodo ya publicado.

8.1 Horizonte y principio general

Premisa operativa: el histórico ingerido se conserva y acumula durante el horizonte de planeación de doce (12) años. No se contempla depuración automática del histórico publicado en la Zona Oro ni de los insumos de auditoría en Bronce dentro de ese horizonte, salvo decisión explícita del Distrito.

8.2 Retención por zona

Zona Bronce: conserva la copia cruda de archivos de variables, los Parquet de capas extraídos desde GeoDB, configuración, manifiestos y logs de copia por periodo. Es la evidencia de auditoría de lo recibido.

Zona Plata: conserva los productos validados (p. ej. Parquet/GeoJSON procesados) por periodo. Los rechazos quedan en logs de ejecución con causa.

Zona Oro (PostgreSQL): conserva variables e indicadores publicados y capas cargadas, con crecimiento acumulado por periodo. El volumen residente proyectado cabe en el dimensionamiento de disco de los escenarios de la Sección 10; la restricción dominante es de cómputo, no de retención.

8.3 Controles de reproceso e idempotencia

Para no duplicar cargas, los procesos de publicación en la Zona Oro dejan un archivo de bloqueo por periodo (lock.json). Si ese bloqueo existe, una nueva carga del mismo periodo se detiene, salvo que la operación autorice explícitamente omitir. Así se protege la consistencia de los datos ya publicados.

8.4 Almacenamiento de bajo acceso

Para históricos de poco acceso, en particular el frente del modelo de escenarios, la Sección 10 contempla el uso de nivel Cool LRS como palanca de costo sin alterar la política de conservación del horizonte de 12 años.

9. Políticas de triggers y ejecuciones automáticas

Esta sección define cuándo se ejecutan los procesos ETL de archivos (variables) y de capas: de forma manual o con programación automática. A la fecha, la operación se dispara desde Synapse Studio; la política indica la periodicidad y las reglas para cuando se activen disparadores programados.

9.1 ETL de archivos / variables (ETL-CARGA-VARIABLES)

Periodicidad: mensual, tras el depósito de los archivos de variables por las dependencias en \\10.0.0.4\ADF y la verificación de la configuración del periodo.

Parámetros: p_periodo (YYYY_MM; vacío = mes anterior), p_env, p_correo_usuario. Encadena 01-CARGA-ARCHIVOS-BRONCE → 02-VALIDAR-ARCHIVOS-PLATA → 03-CARGA-VARIABLES-ORO.

Disparo: (a) manual por el Funcionario DAP (Trigger now) durante la ejecución y validaciones; (b) trigger programado mensual en ventana operativa acordada, una vez estabilizado el depósito de insumos.

9.2 ETL de capas (ETL-CARGA-CAPAS)

Periodicidad: semestral (dos cargas al año), alineada a la disponibilidad de capas actualizadas en GeoDB y a la matrícula vigente en config.json.

Parámetros: p_periodo, p_env, p_correo_usuario, p_approval_status. Encadena 01-CARGA-CAPAS-BRONCE → 02-VALIDAR-CAPAS-PLATA → 03-CARGA-CAPAS-ORO.

Disparo: (a) manual, previa verificación de la matrícula de capas en config.json y de la conectividad de lectura a GeoDB vía SHIRCapasDB; (b) trigger programado semestral en las ventanas de actualización institucional. La publicación a Oro respeta el estado de aprobación cuando el parámetro p_approval_status lo exige.

9.3 Reglas comunes

No deben solaparse dos corridas del mismo proceso sobre el mismo periodo mientras el bloqueo del periodo esté activo, salvo un reproceso autorizado. El seguimiento se hace en el monitor de ejecuciones de Synapse Studio. Si una etapa falla y las anteriores quedaron correctas, se puede reintentar solo la etapa fallida.

10. Volúmenes y costos

Esta sección presenta la estimación de volúmenes y costos del ETL de variables y capas del POT, integrada a la arquitectura para sustentar las decisiones de dimensionamiento. Explica qué se costea, cuánto pesan los datos, cómo se estructura el gasto, qué escenarios de base de datos existen y qué debe decidir el Distrito. Nota de alcance: la capa de visualización queda fuera del ETL; la evidencia de carga del visor (subsección 10.7) se usa solo para dimensionar el SKU de la Zona Oro, porque esa base también servirá al consumo institucional futuro.

Cómo leer esta sección.

Tesis central: el costo de la solución lo determina el tamaño (SKU) de la base de datos siempre encendida -la Zona Oro-, no el volumen de datos; y ese SKU debe elegirse a partir de la carga medida. Se presentan escenarios como opciones, no se prescribe uno. Para quién es: funcionarios de Innovación y DAP que deben decidir presupuesto, perfil de uso y SKU de la Zona Oro, frecuencia de cargas, régimen de disponibilidad (24/7 u horario) y modalidad (nube o en tierra); y el equipo técnico de Guane. Moneda: USD, con equivalente en COP a una TRM de referencia de $4.000 COP/USD. Horizonte y retención: 12 años sin depuración; todo dato ingerido se conserva y acumula. Validación de tarifas: Azure Retail Prices API, región West US 2, 1 de junio de 2026 (Anexo A).

Tabla 9. Guía de lectura de la sección de volúmenes y costos.

10.1 Síntesis para decisión

El costo de operar la solución en Azure lo domina la base de datos PostgreSQL siempre encendida (la Zona Oro): cerca del 74 % del gasto del ETL. El almacenamiento es prácticamente despreciable (~1 %). En consecuencia, el costo total depende casi por completo del SKU de esa base de datos. Conservar doce años de histórico añade apenas decenas de dólares; el sistema maneja gigabytes, no terabytes.

Por eso esta sección presenta cinco escenarios de dimensionamiento (E1–E4, con una variante E3b). Entre el más pequeño y el de producción con alta disponibilidad, el costo a 12 años varía en un factor cercano a ×27. La medición real de carga (subsección 10.7) muestra que los tiers menores no sostienen la concurrencia de un visor geoespacial, por lo que cada escenario se acompaña del perfil de uso que soporta.

Costo total a 12 años por escenario (base, sin margen)

Esc.	SKU	vCore / RAM / Disco (GB)	BD/año (USD)	12 años (USD)	+30 % (USD)	COP +30 %
E1	B1ms	1 / 2 / 32	193	3.156	4.103	16,4 M
E2	B2s	2 / 4 / 64	684	9.047	11.761	47,0 M
E3	B2ms	2 / 8 / 128	1.368	17.255	22.432	89,7 M
E3b	D2ds_v5	2 / 8 / 128	1.736	21.670	28.171	112,7 M
E4	D4ds_v5 +HA	4 / 16 / 256	6.944	84.163	109.412	437,6 M

Tabla 10. Costo total a 12 años por escenario (base, sin margen).

Cada total incluye los componentes del ETL que no dependen del SKU (Synapse Spark, Data flow, Data Lake y Key Vault ~839 USD a 12 años), invariantes entre escenarios. La única palanca que se mueve en esta tabla es el SKU de la base de datos. El frente del modelo de escenarios (solo almacenamiento) se costea aparte (subsección 10.9).

Total combinado de los dos frentes (POT por escenario + modelo Hot LRS ~623 USD), USD a 12 años, base

Total combinado (12 años)	POT	Modelo	Combinado	Combinado/año	Combinado COP
E1 - B1ms	3.156	623	3.779	315	15,1 M
E2 - B2s	9.047	623	9.670	806	38,7 M
E3 - B2ms	17.255	623	17.878	1.490	71,5 M
E3b - D2ds_v5	21.670	623	22.293	1.858	89,2 M
E4 - D4ds_v5 +HA	84.163	623	84.786	7.066	339,1 M

Tabla 11. Total combinado POT + modelo a 12 años.

Es decir, el frente del modelo añade ~623 USD a 12 años (~52 USD/año) a cualquier escenario; en promedio el combinado va de ~315 USD/año (E1) a ~7.066 USD/año (E4). El almacenamiento del modelo es marginal frente al cómputo de la base de datos en todos los escenarios.

Las tres conclusiones que importan
• Es un sistema pequeño: ~21–38 GB acumulados al año 12 en el frente del POT y ~419 GB en el del modelo: gigabytes, no terabytes. • El costo es disponibilidad, no volumen: la base de datos 24/7 concentra ~74 % del gasto del ETL; el almacenamiento, ~1 %. • La decisión de mayor impacto es el SKU de la Zona Oro, y debe tomarse con carga medida: entre E1 y E4 el costo a 12 años va de ~3.156 a ~84.163 USD (×27).

Tabla 12. Las tres conclusiones que importan.

10.2 Alcance y premisas

La estimación cubre los costos en Microsoft Azure de dos líneas independientes: (1) el ETL de seguimiento y evaluación del POT -variables mensuales y capas semestrales, con la Zona Oro sobre PostgreSQL en Azure- y (2) el modelo de escenarios físico-espaciales -~35 GB de entrada más ~16 GB por ejecución semestral, acumulados-. Se delimita así para dimensionar el presupuesto (anual y plurianual), sustentar la respuesta a Innovación sobre volúmenes y recursos, y habilitar la decisión sobre el dimensionamiento de la Zona Oro.

Premisas de cálculo (validadas el 1 de junio de 2026):

- Horizonte: 12 años. Sin rotación: todo dato ingerido se conserva y acumula.
- Moneda: USD; equivalente en COP con TRM 4.000 COP/USD.
- Tarifas: Azure Retail Prices API, región West US 2, 1 de junio de 2026 (tipo Consumption).
- Margen de contingencia: se reportan holguras del 30 %, 50 %, 70 % y 100 % sobre el costo base.

Qué entra y qué no entra en el costeo:

Sí está incluido	No está incluido
Servicios Azure del ETL: PostgreSQL, Synapse Spark, Data flow, Data Lake y Key Vault.	Infraestructura propia del DAP (escenario "en tierra"): no se costea aquí.
Los dos frentes: variables + capas (con su proyección) y el almacenamiento del modelo.	Aplicación web o tableros de consulta para la ciudadanía.
El histórico acumulado de 12 años y holguras de 30 %, 50 %, 70 % y 100 %.	Salarios del personal del Distrito y licencias de software de escritorio.
Cinco escenarios de SKU de base de datos (E1–E4, E3b) y palancas de ahorro.	Egress y backups de PostgreSQL: se enuncian con su tarifa, pero no se suman a la base.
El cómputo del modelo como job efímero (~1 USD/año).	Descuentos Enterprise (EA/CSP) y reservas, salvo en la simulación de palancas (sección 10).

Tabla 13. Alcance del costeo: qué se incluye y qué no.

10.3 El sistema y los volúmenes

El seguimiento del POT recibe información de dos clases y dos ritmos. Cada mes se procesan archivos estructurados de variables (principalmente Excel) desde el File System institucional; dos veces al año se extraen capas geográficas desde el Servidor GeoDB del Distrito, según la matrícula definida en config.json. Esa información recorre tres zonas del lago: Bronce conserva la copia en bruto (archivos de variables y Parquet de capas extraídos), Plata la versión validada y estandarizada, y Oro la publica para consumo institucional sobre una base de datos PostgreSQL en Azure.

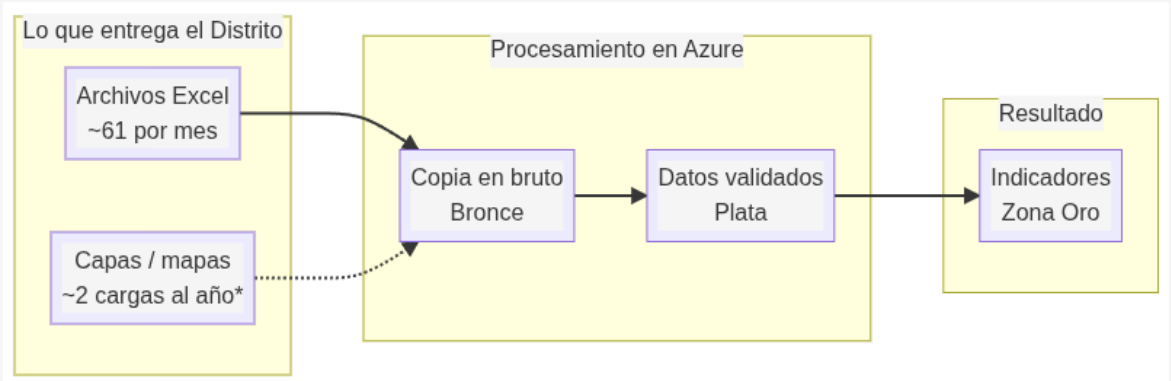


Figura 5. Flujo y frecuencia de carga (Bronce → Plata → Oro) del seguimiento y evaluación del POT.

10.3.1 Variables: 61 archivos

La batería oficial tiene 984 variables (783 del Programa de Ejecución y 201 del modelo de ocupación), pero no viajan en 984 archivos: llegan agrupadas en los reportes de las 61 dependencias responsables. El volumen lo determina el número de archivos (61), no el de variables. El peso por archivo -~1 MB- es un supuesto operativo acordado (la muestra real ronda 10 KB, elevada a 1 MB para cubrir hojas, tablas dinámicas y formatos).

Concepto	Valor	De dónde sale / confianza
Fuentes que entregan archivo	61	Una por dependencia - acuerdo de reunión.
Variables en la batería oficial	984	783 + 201; hoja Variables-DATOS - confirmado.
MB por archivo (Bronce)	~1 MB	Supuesto operativo; muestra real ~10 KB.
Entra por corrida mensual	~61 MB	61 archivos × 1 MB - calculado.
Acumulado a 12 años (Bronce)	~8,6 GB	732 MB/año × 12, sin borrar - calculado.

Tabla 14. Premisas de volumen de variables.

10.3.2 Capas: lo medido y lo proyectado

Se distingue siempre lo medido de lo proyectado. Lo medido es el Acuerdo 048: 54 capas que pesan 80 MB (~1,48 MB/capa, verificado el 27-05-2026). Lo proyectado es un supuesto: cerca del 30 % de las 984 variables podría geolocalizarse (~295 capas), llevando cada carga a ~437 MB. Ese 30 % es la mayor fuente de incertidumbre del volumen y debe sustentarse con el inventario capa a capa.

Escenario	Por carga	Al año (2 cargas)	Acumulado 12 años (Bronce)
A - Medido (Acuerdo 048)	80 MB	~160 MB	~1,9 GB
B - Proyección (30 % espacializable)	~437 MB	~874 MB	~10,2 GB

Tabla 15. Volúmenes de capas: medido vs proyección.

Para no sobredimensionar, una referencia tangible: una capa liviana (aseo) pesa pocos KB; una media (clasificación de suelo) ~5 MB; una densa (nomenclatura, ~500 mil filas) ~15 MB; y solo un caso extremo (catastro completo) llegaría a ~600 MB, un techo puntual y no la carga típica. Una capa pesa por su número

de filas, no porque la cartografía sea inmanejable; incluso en la proyección, el conjunto se mantiene en gigabytes.

10.3.3 Acumulado por zona y volumen del modelo

Bronce guarda el bruto; Plata comprime en Parquet (factor ~÷3 en variables); Oro publica indicadores calculados y capas de consumo. El frente del POT queda entre ~21,5 GB (medido) y ~38,2 GB (proyección) al año 12, con crecimiento lineal. El modelo de escenarios sigue otra lógica: ~35 GB de entrada (carga única) más ~16 GB por ejecución semestral, de modo que crece ~32GB/año hasta ~419 GB (~0,4 TB) al año 12.

Zona / frente	Variables	Capas (medido → proy.)	Total año 12
Bronce	~8,6 GB	~1,9 → ~10,2 GB	~10,5 → ~18,8 GB
Plata (Parquet ~÷3)	~2,8 GB	~1,9 → ~10,2 GB	~4,7 → ~13,1 GB
Oro (publicación)	~5,6 GB	~0,7 GB	~6,3 GB
Frente POT (suma)	-	-	~21,5 → ~38,2 GB
Frente modelo (entradas+salidas)	-	-	~419 GB

Tabla 16. Volumen acumulado por zona y frente a 12 años.

Nota de coherencia con el dimensionamiento: el volumen residente en la Zona Oro (~6,3 GB proyectado) cabe holgadamente incluso en el disco del escenario mínimo (E1, 32 GB). Es decir, el disco no es la restricción; lo es la capacidad de cómputo de la base de datos, como demuestra la medición de la subsección 10.7.

10.3.4 El modelo de escenarios en detalle

El modelo requiere ~35 GB de información de entrada -variables, capas e indicadores base de cuatro módulos- que se cargan una sola vez. Cada ejecución genera ~16 GB de salida (resultados y escenarios derivados). Con frecuencia semestral son dos ejecuciones al año, de modo que el volumen crece ~32GB anuales. Tanto las frecuencias como los volúmenes son estimaciones del nivel actual de definición del modelo y se refinaron al precisar sus conjuntos de datos.

Módulo	Entrada (carga única)	Salida (por ejecución)
Modelo Multipropósito	~20 GB	~5 GB
Zonas Homogéneas	~10 GB	~5 GB

Estudios Socioeconómicos	~5 GB	~5 GB
Estudio de Riesgo	-	~1 GB
Total	~35 GB	~16 GB

Tabla 17. Volúmenes del modelo de escenarios por módulo.

Volumen acumulado por año (entradas + salidas): ~67 GB (año 1), ~131 GB (año 3), ~227 GB (año 6), ~323 GB (año 9) y ~419 GB (año 12). El crecimiento, igual que en el POT, es lineal.

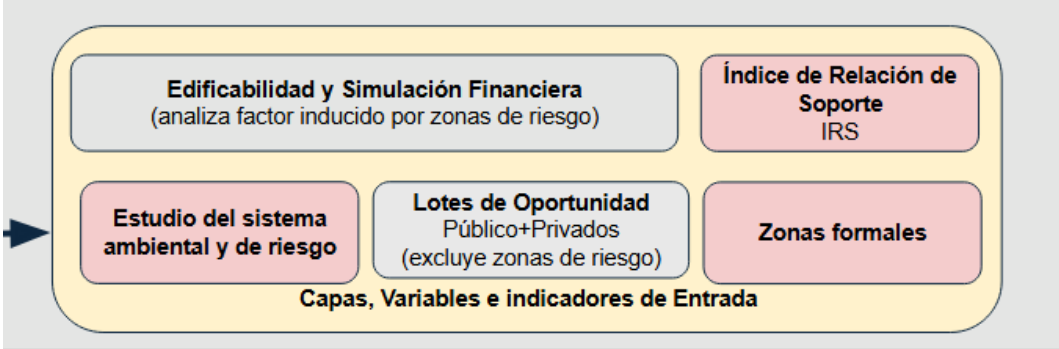


Figura 6. Capas, variables e indicadores de entrada del modelo de escenarios físico-espaciales.

10.4 Estructura de costos

El ETL se apoya en cinco servicios de Azure con dos lógicas de cobro: lo que se paga de forma continua (la base de datos, encendida todo el mes) y lo que se paga solo durante la ejecución (el procesamiento). Esa distinción explica por qué el gasto se concentra donde se concentra.

Servicio	Para qué sirve	Cómo se paga	% del costo del ETL
Azure PostgreSQL	Base de datos de la Zona Oro (variables y capas publicadas)	Continuo (24/7)	~74 %
Synapse Spark	Valida y carga variables y capas	Solo al correr el ETL	~22 %
Synapse Data flow	Arma la lista de archivos a copiar por corrida	Por corrida	~3 %

Data Lake (ADLS)	Carpetas Bronce y Plata (histórico)	Por GB guardado	~1 %
Key Vault	Credenciales	Uso mínimo	< 1 %

Tabla 18. Estructura de costos por servicio Azure.

Desglose en USD para el escenario de referencia E1 (B1ms), a 12 años

Qué es (en simple)	Servicio	USD a 12 años	% del total	Por qué cuesta eso
Base de datos de variables y capas	PostgreSQL B1ms	~2.317	~74 %	Encendida 24 h + 32 GB aprovisionados.
Motor de procesamiento	Synapse Spark	~703	~22 %	Se paga solo cuando corre el ETL.
Transformación de configuración	Data flow	~99	~3 %	Paso automático de cada corrida.
Almacenamiento del lago	Data Lake (Bronce/Plata)	~20 a 42	~1 %	Crece con la acumulación; mínimo.
Bóveda de contraseñas	Key Vault	~6	< 1 %	Uso mínimo.

Tabla 19. Desglose de costos del escenario E1 (12 años).

Los montos en dólares por servicio dependen del SKU; arriba se muestran para E1 (total ~3.156 USD). Al subir de escenario, solo cambia la fila de PostgreSQL: el resto de servicios (~839 USD a 12 años) es invariante.

El hallazgo estructural
No es un gasto de “big data en la nube”: es, sobre todo, una base de datos pequeña encendida 24/7. Acumular 12 años de histórico añade apenas decenas de dólares. Por eso la decisión sobre el SKU de la Zona Oro pesa tanto, y todo lo demás del ETL (~839 USD a 12 años) es invariante frente a esa decisión.

Tabla 20. Hallazgo estructural del costeo.

10.5 Tarifas validadas

Cada tarifa se consultó el 1 de junio de 2026 contra la API oficial de precios de Azure (West US 2, tipo Consumption). Las seis opciones de SKU de PostgreSQL Flexible Server se resumen abajo; el catálogo completo y la metodología están en la subsección 10.15.2. El almacenamiento de la base de datos cuesta 0,115 USD/GB-mes y el del lago (ADLS Hot LRS) 0,0184 USD/GB-mes.

SKU	Serie	vCore	RAM	USD/h	USD/mes	Uso (escenario)
B1ms	Burstable	1	2 GB	0,017	12,4	Mínimo; dev baja concurrencia (E1)
B2s	Burstable	2	4 GB	0,068	49,6	Equivalente al tier actual; dev/demo (E2)
B2ms	Burstable	2	8 GB	0,136	99,3	Producción concurrencia moderada (E3)
D2ds_v5	General Purpose	2	8 GB	0,178	129,9	E3 sostenido, cómputo dedicado (E3b)
D4ds_v5	General Purpose	4	16 GB	0,356	259,9	Producción / Gemelo Digital (E4)
E2ds_v5	Memory Optimized	2	16 GB	0,250	182,5	Alternativa si la RAM es el límite

Tabla 21. Tarifas de cómputo PostgreSQL por SKU.

Las tarifas de cómputo B-series son por instancia; las D y E se cobran por vCore (Ddsv5 0,089 USD/vCore-h; Edsv5 0,125 USD/vCore-h).

Dimensionamiento de la ejecución del ETL (no depende del SKU de la base de datos): el procesamiento con Synapse Spark (Memory Optimized) usa ~12 vCores durante ~1,1 h por corrida, y el Data flow (Basic) ~8 vCores durante ~20 min por corrida. Sobre esa base se obtienen los ~703 USD (Spark) y ~99 USD (Data flow) a 12 años del desglose de la subsección 10.4. El backup de PostgreSQL (LRS, 0,095 USD/GB-mes) se enuncia pero no se suma al costo base.

10.6 Escenarios de dimensionamiento de la base de datos

Cada escenario se describe por el perfil de carga que soporta, su uso y su limitación principal. El costo a 12 años (subsección 10.1) es la consecuencia directa del SKU; aquí se justifica el porqué de cada uno.

10.6.1 E1 - Mínimo (B1ms)

- SKU: B1ms - 1 vCore burstable, 2 GB RAM, 32 GB de disco.
- Soporta: uso interno de muy baja concurrencia (1–2 usuarios), y solo si las vistas pesadas están materializadas y las teselas cacheadas.
- Limitación: 1 vCore serializa cualquier consulta analítica; sin esas correcciones el visor no es interactivo (subsección 10.7).

10.6.2 E2 - Tier actual (B2s)

- SKU: B2s - 2 vCore burstable, 4 GB RAM. Equivalente en Azure del tier vivo medido (1 vCPU dedicado / 3,75 GB).
- Soporta: desarrollo y demostración; consultas analíticas puntuales sin concurrencia.
- Limitación: el burst no sostiene carga continua.

10.6.3 E3 / E3b - Intermedio (B2ms o D2ds_v5)

- SKU: B2ms (2 vCore burstable, 8 GB) o D2ds_v5 (2 vCore dedicado, sostenido).
- Soporta: concurrencia moderada (< 20 usuarios) tras materializar vistas y cachear teselas. Primer escalón viable para un visor real.
- Limitación: no dimensionado para servir al Gemelo Digital ni picos altos; sin alta disponibilidad.

10.6.4 E4 - Producción con alta disponibilidad (D4ds_v5 + HA)

- SKU: D4ds_v5 - 4 vCore dedicado, 16 GB, 256 GB SSD, con HA zona-redundante (duplica cómputo y almacenamiento).
- Soporta: concurrencia alta, servir teselas al Gemelo Digital de Medellín y reejecutar el modelo en paralelo al servicio.
- Limitación: es el escenario de mayor costo.

Presupuesto por escenario y holgura (USD a 12 años)

Escenario (12 años)	Base	+30 %	+50 %	+70 %	+100 %
E1 - B1ms	3.156	4.103	4.734	5.365	6.312
E2 - B2s	9.047	11.761	13.571	15.380	18.094
E3 - B2ms	17.255	22.432	25.883	29.334	34.510
E3b - D2ds_v5	21.670	28.171	32.505	36.839	43.340
E4 - D4ds_v5 +HA	84.163	109.412	126.245	143.077	168.326

Tabla 22. Presupuesto por escenario y holgura (USD a 12 años).

En pesos (TRM 4.000), el escenario E1 va de ~12,62M (base) a ~25,25 M (+100 %); el E3, de ~69,02 M a ~138,04 M; el E4, de ~336,65 M a ~673,30 M. Las holguras cubren incertidumbre tarifaria y de alcance, no un cambio de SKU, que se modela como escenario propio.

Cómo elegir entre: volumen vs. rendimiento
El salto de SKU es una decisión de rendimiento y concurrencia, no de volumen. El tier mínimo (E1, B1ms) ofrece el mejor costo (~3.156 USD/12a) y su disco de 32 GB cubre con holgura la Zona Oro proyectada (~6,3 GB). Conviene mantenerse en un tier burstable pequeño (E1/E2) mientras la Zona Oro siga por debajo de ~25–28 GB y la concurrencia de consumidores sea baja. Subir a E3 o superior se justifica cuando una medición real evidencie presión de cómputo, memoria o concurrencia -como la documentada en la sección 7-, ya que multiplica el costo del frente sin cambiar el volumen que cabe.

Tabla 23. Criterio de elección: volumen vs rendimiento.

10.7 Evidencia de medición real (Fase 1)

Las cifras de carga se midieron el 1 de junio de 2026 sobre el despliegue vivo del visor y explican por qué los escenarios difieren en el perfil que soportan. El tier medido equivale en Azure a un B2s (E2): 1 vCPU dedicado / 3,75 GB, disco de 10 GB. La base pesa 1,84 GB (88 % catastro estático): el almacenamiento es y seguirá siendo despreciable. El cuello de botella es de cómputo (CPU), no de almacenamiento, y son dos problemas distintos.

10.7.1 Cuello de botella #1 - vistas analíticas (CPU)

La vista model.v_irs_tema_manzana cuesta 4,06 s por un count(*) con todo en caché, lo que implica 100 % de CPU. Sobre 1 vCPU, el tiempo de respuesta crece linealmente con la concurrencia: lo que tarda un usuario se multiplica al sumar consultas simultáneas, porque un único núcleo serializa el trabajo analítico.

10.7.2 Cuello de botella #2 - render de teselas MVT "on-the-fly"

Las teselas vectoriales se renderizan por petición, sin caché. En caliente, un único usuario espera entre 8 y 41 s por tesela según la capa y el nivel de zoom. Una prueba de carga con k6 (6 usuarios concurrentes, 3 minutos) evidenció la saturación:

Recurso medido (k6, 6 usuarios, 3 min)	Resultado
Teselas de lotes	p95 = 60 s (0/7 respondieron: timeout)
Teselas de manzanas	p95 = 59,95 s
Features IRS	p95 = 44,11 s
Peticiones fallidas (global)	39,13 %
Render por tesela (1 usuario, en caliente)	8–41 s (umbral interactivo ~1,5 s)

Tabla 24. Resultados de la prueba de carga k6.

Lectura de la medición
Implicación para el dimensionamiento: como el cuello es de CPU, los tiers menores (E1/E2) solo aplican a uso de baja concurrencia. Un visor con concurrencia real requiere E3 o superior, y además exige dos correcciones de ingeniería -materializar las vistas analíticas y cachear las teselas-, sin las cuales ningún SKU entrega una experiencia interactiva.

Tabla 25. Lectura de la medición de carga.

10.8 Proyección anual a 12 años

Conservando todo el histórico, el costo acumulado crece de forma aproximadamente lineal: el cómputo always-on domina y el almacenamiento aporta poco. La tabla muestra el acumulado año por año (USD, base sin margen).

Año	E1	E2	E3	E3b	E4
1	263	754	1.438	1.806	7.014
2	526	1.508	2.876	3.612	14.027
3	789	2.262	4.314	5.418	21.041
4	1.052	3.016	5.752	7.223	28.054
5	1.315	3.770	7.190	9.029	35.068
6	1.578	4.524	8.628	10.835	42.082

7	1.841	5.277	10.065	12.641	49.095
8	2.104	6.031	11.503	14.447	56.109
9	2.367	6.785	12.941	16.253	63.122
10	2.630	7.539	14.379	18.058	70.136
11	2.893	8.293	15.817	19.864	77.150
12	3.156	9.047	17.255	21.670	84.163

Tabla 26. Proyección acumulada de costo por año y escenario.

Por construcción, el acumulado del año 12 coincide con el costo total a 12 años de la subsección 10.1. El crecimiento lineal confirma que la retención íntegra del histórico no genera saltos de costo.

10.9 Frente 2 - Almacenamiento del modelo de escenarios

El frente del modelo se costea solo por almacenamiento, dominado por el histórico acumulado (~419 GB al año 12). Con la tarifa validada de Data Lake (Hot LRS), el almacenamiento a 12 años es del orden de 623 USD (~52 USD/año). Para datos históricos de poco acceso, el nivel Cool LRS reduce el costo a casi la mitad (~338 USD). El cómputo del modelo, como job efímero, es ~1 USD/año.

Concepto (acum. 12 años)	Base	+30 %	+50 %	+70 %	+100 %
Almacenamiento Hot LRS (USD)	623	810	935	1.059	1.246
Almacenamiento Hot LRS (COP, TRM 4.000)	2,49 M	3,24 M	3,74 M	4,24 M	4,98 M
Optimizado Cool LRS (USD)	338	439	507	575	676

Tabla 27. Costo de almacenamiento del modelo a 12 años.

Método de devengo: el almacenamiento se devenga mes a mes sobre el volumen residente; para el reporte anual se valora con el volumen de cierre de cada año (USD/año = volumen × 0,0184 × 12). Se declara este único método para evitar ambigüedad: según el mes exacto de carga, el acumulado del modelo a 12 años oscila entre ~580 y ~644 USD; se reporta ~623 USD.

Por año (Hot LRS, base): ~15 USD (año 1), ~29 (año 3), ~50 (año 6), ~71 (año 9) y ~93 USD (año 12). El promedio anual es ~52 USD.

Si el modelo llegara a requerir base de datos o cómputo propios siempre encendidos, su costo sería adicional -una línea equivalente a E2/E3- y se incorporaría en una versión posterior.

10.10 Palancas de ahorro y simulación

Además del SKU, existen palancas que reducen el costo sin cambiar el dimensionamiento. Las dos principales son mutuamente excluyentes (no se combinan):

- Reserva de capacidad a 3 años: compromiso 24/7 con descuento, −51 % del cómputo (−38 % a 1 año).
- Apagado fuera de horario (8×5): la base corre 173 h/mes en vez de 730, reduce el cómputo ~76 %. Aplica solo a uso interno sin SLA 24/7.
- Cool LRS para el histórico del lago: tier frío para datos no consultados (modelo: 623 → 338 USD/12a).
- Descuento Enterprise (EA/CSP): variable, sujeto al acuerdo Microsoft del Distrito.

Simulación del TCO a 12 años (USD, base) por estrategia de operación de la base de datos

Esc.	Base (24/7)	Reserva 3 años	Apagado 8×5	Ahorro reserva	Ahorro 8×5
E1	3.156	2.245	1.792	-29 %	-43 %
E2	9.047	5.401	3.593	-40 %	-60 %
E3	17.255	9.964	6.347	-42 %	-63 %
E3b	21.670	12.127	7.393	-44 %	-66 %
E4	84.163	45.992	27.055	-45 %	-68 %

Tabla 28. Simulación de TCO con palancas de ahorro.

El ahorro porcentual es sobre el total (incluye los componentes ETL fijos, que no se reducen). Las cifras son base; para presupuesto, aplicar el margen elegido (30 %–100 %) sobre la estrategia seleccionada. La reserva y el apagado son alternativas, no acumulables.

10.11 Decisiones del Distrito

La elección entre escenarios depende del perfil de uso que defina el Distrito. La siguiente tabla ofrece la correspondencia entre perfil y escenario que lo soporta, sin prescribir una opción.

Perfil de uso	Escenario	Estrategia de ahorro aplicable
Uso interno, 1–2 usuarios, sin SLA	E1 / E2	Apagado 8×5
Visor con concurrencia moderada (< 20)	E3 / E3b	Reserva 3 años
Concurrencia alta o servir al Gemelo Digital	E4 (con HA)	Reserva 3 años

Tabla 29. Perfil de uso vs escenario recomendable.

Junto a la elección de SKU, hay decisiones transversales que también mueven el costo:

Decisión	Opciones	Impacto cuantificado
¿La Zona Oro se implementa en la nube (Azure) o en tierra (DAP)?	Nube / En tierra	En nube, la base de datos cuesta ~193 USD/año en E1 (hasta ~6.944 en E4). En tierra, esa parte sale de Azure pero se suma infraestructura propia (servidor 24/7, licencias, operación y respaldo), no estimada aquí y habitualmente mayor. Fuera del alcance Azure, pero es una decisión real.
¿La base de datos puede apagarse fuera de horario?	24/7 / 8 h × 5 d	Reduce el cómputo ~76 %: p. ej. E3 de ~17.255 a ~6.347 USD/12a. Solo uso interno sin SLA.
¿Se contrata reserva de capacidad a 3 años?	Sí / No	−51 % del cómputo: p. ej. E3 de ~17.255 a ~9.964 USD/12a. Para operación 24/7 estable.

¿Qué nivel de almacenamiento para el histórico?	Hot / Cool LRS	En el modelo, Cool reduce de ~623 a ~338 USD/12a. En el POT el efecto es marginal (~1 %).

Tabla 30. Decisiones transversales del Distrito e impacto.

La decisión requiere fijar la concurrencia objetivo y, en paralelo, ejecutar las correcciones de ingeniería identificadas en la medición (subsección 10.7), que permiten sostener un escenario menor.

10.12 Sensibilidad y riesgos

El resultado depende sobre todo del SKU y de cómo se opere la base de datos. Las palancas de cómputo mueven el costo de verdad; las de volumen, no.

Variable	Cambio	Efecto estimado
SKU de la base de datos (E1 → E4)	B1ms → D4ds_v5 +HA	TCO 12 años ~3.156 → ~84.163 USD (×27).
Apagado fuera de horario (8×5)	173 h/mes vs 730	−~76 % del cómputo (p. ej. E3 −63 % del total).
Reserva de capacidad a 3 años	Compromiso 24/7	−51 % del cómputo (p. ej. E3 −42 % del total).
TRM 4.000 → 4.500 COP/USD	+12,5 %	Sube el equivalente en pesos; USD sin cambio.
Volumen del modelo ×2	Duplica entradas/salidas	Almacenamiento 623 → ~1.246 USD/12a (sigue marginal).
Capas espacializables 30 % → 50 %	+20 puntos	Sube el volumen; el costo del ETL apenas varía (~1 %).
1 MB → 2 MB por archivo	×2 variables	Variables 8,6 → 17 GB/12a; efecto en costo despreciable.

Tabla 31. Sensibilidad del costo ante cambios de variables.

10.13 Confianza y trazabilidad

Dato	Valor	Estado

Variables batería oficial	984 (783 + 201)	Confirmado
Corridas variables / año	12 (mensual)	Confirmado
Fuentes / archivos por corrida	61	Acuerdo de reunión
MB por archivo (Bronce)	~1 MB (muestra real ~10 KB)	Acuerdo de reunión
Acuerdo 048: capas / peso	54 / 80 MB	Confirmado
Capas proyectadas (30 %)	~295	Calculado (supuesto)
Cargas de capas / año	2 (semestral)	Acuerdo de reunión
Modelo (entrada / salida)	35 / 16 GB	Acuerdo de reunión, a refinar
Componentes ETL no-BD (12 años)	~839 USD	Calculado (invariante por SKU)
Volumen año 12 (POT / modelo)	~21,5–38,2 GB / ~419 GB	Calculado
Tier vivo medido (equiv. Azure)	B2s (1 vCPU / 3,75 GB)	Confirmado (medición Fase 1)
Carga: vista IRS / teselas	4,06 s CPU / 8–41 s, p95 60 s	Confirmado (medición k6)
Tarifas de los 6 SKU	Según Anexo A	Confirmado (API, West US 2, 1-06-2026)
TCO 12 años E1 / E3 / E4 (base)	~3.156 / ~17.255 / ~84.163 USD	Calculado
Almacenamiento modelo 12 años	~623 USD (Hot) / ~338 (Cool)	Calculado

Tabla 32. Trazabilidad y nivel de confianza de los datos.

10.14 Recomendación

Las opciones quedan disponibles con su costo a 12 años (subsecciones 10.1, 10.6 y 10.8) y su variante con palancas (subsección 10.10). Para decidir, se recomienda:

- Fijar primero la concurrencia objetivo del visor: es la variable que determina el escenario mínimo viable (E1/E2 para uso interno; E3/E3b para un visor real; E4 para el Gemelo Digital).
- Ejecutar en paralelo las correcciones de ingeniería de la subsección 10.7 (materializar vistas y cachear teselas): permiten sostener un escenario menor y, por tanto, un costo menor.
- Operar en la nube (Azure) y aplicar la palanca acorde al uso: apagado 8×5 para uso interno sin SLA, o reserva a 3 años para operación 24/7 estable.

- Conservar todo el histórico (el almacenamiento es ~1 %); evaluar Cool LRS para el histórico antiguo del modelo.
- Adoptar una holgura explícita entre 30 % y 100 % según la ventana de riesgo, y re-verificar las tarifas contra la API antes de cerrar el presupuesto.

10.15 Glosario de costos y catálogo de SKU

10.15.1 Glosario de costos

Término	Significado
Bronce / Plata / Oro	Zonas del lago: bruto, validado/estandarizado y publicado para consumo.
ETL	Extract, Transform, Load: extracción, transformación y carga de datos.
SKU	Referencia comercial del tamaño de la base de datos (vCore, RAM, serie).
Burstable (B) / General Purpose (D) / Memory Optimized (E)	Series de cómputo de Azure: económica con créditos de CPU; dedicada sostenida; optimizada en RAM.
vCore-hora	Unidad de cobro de cómputo: un núcleo virtual durante una hora.
HA (alta disponibilidad)	Réplica zona-redundante; duplica cómputo y almacenamiento.
MVT / tesela	Mapbox Vector Tile: fragmento vectorial de un mapa que se renderiza por petición.
Vista materializada	Resultado de una consulta pre calculado y almacenado para responder rápido.
ADLS Gen2	Azure Data Lake Storage Gen2: almacenamiento jerárquico del lago.
LRS (Hot / Cool)	Locally-Redundant Storage; Hot (acceso frecuente), Cool (poco frecuente, más barato).
Egress	Transferencia de datos de salida desde Azure hacia fuera de la región (puede facturarse).
Holgura / contingencia	Margen porcentual sobre el costo base para cubrir incertidumbre y variaciones.

OPEX / TCO	Gasto operativo anual / costo total de propiedad acumulado (aquí, 12 años).
ROM	Rough Order of Magnitude: estimación de planeación, no definitiva.

Tabla 33. Glosario de términos de volúmenes y costos.

Los términos de arquitectura (Bronce/Plata/Oro, SHIR, pipeline, etc.) se definen en el Anexo E. La tabla siguiente concentra solo los términos propios de volúmenes y costos.

10.15.2 Catálogo de SKU y metodología

Catálogo de cómputo de Azure PostgreSQL Flexible Server (West US 2, Consumption, 1-06-2026). Costo mensual de cómputo = USD/hora × 730 h.

SKU	Serie	vCore	RAM	USD/h	USD/mes
B1ms	Burstable	1	2 GB	0,017	12,4
B2s	Burstable	2	4 GB	0,068	49,6
B2ms	Burstable	2	8 GB	0,136	99,3
D2ds_v5	General Purpose	2	8 GB	0,178	129,9
D4ds_v5	General Purpose	4	16 GB	0,356	259,9
E2ds_v5	Memory Optimized	2	16 GB	0,250	182,5

Tabla 34. Catálogo de SKU PostgreSQL y tarifas.

Metodología de cálculo:

Costo BD por año = (USD/hora × 730 × 12 × (1+HA)) + (GB aprovisionado × 0,115 × 12 × (1+HA)). Para E4, la HA zona-redundante duplica cómputo y almacenamiento; el respaldo queda dentro del tier gratuito. El total a 12 años suma los componentes ETL no-BD (~839 USD), invariantes. En la simulación de palancas, la reserva aplica 51 % al cómputo y el apagado 8×5 usa 173 h/mes; el almacenamiento no varía.

Almacenamiento: PostgreSQL 0,115 USD/GB-mes; Data Lake Hot LRS 0,0184; Cool LRS 0,010 USD/GB-mes. Backup LRS 0,095 USD/GB-mes (no incluido en el base). Spark (Memory Optimized) 0,138 USD/vCore-h (~12 vCores, ~1,1 h/corrida); Data flow (Basic) 0,257 USD/vCore-h (~8 vCores, 20 min/corrida).

Nota sobre el almacenamiento del lago: el Hot LRS tiene escalonamiento por volumen (0,0184 USD/GB-mes para el primer tramo; 0,0177 y 0,0169 en tramos superiores). Por el volumen de este proyecto (decenas a centenas de GB) aplica el primer tramo (0,0184), que es el usado en todos los cálculos.

11. Conclusiones

La arquitectura descrita corresponde al despliegue verificado de los procesos ETL de carga de avances del POT Medellín: conecta el servidor institucional on-premise con el entorno Azure del grupo de recursos QTE-DMDE-DAPCMP02 y cubre los flujos de variables y de capas, con política de retención, reglas de ejecución automática y estimación de volúmenes y costos en un solo documento.

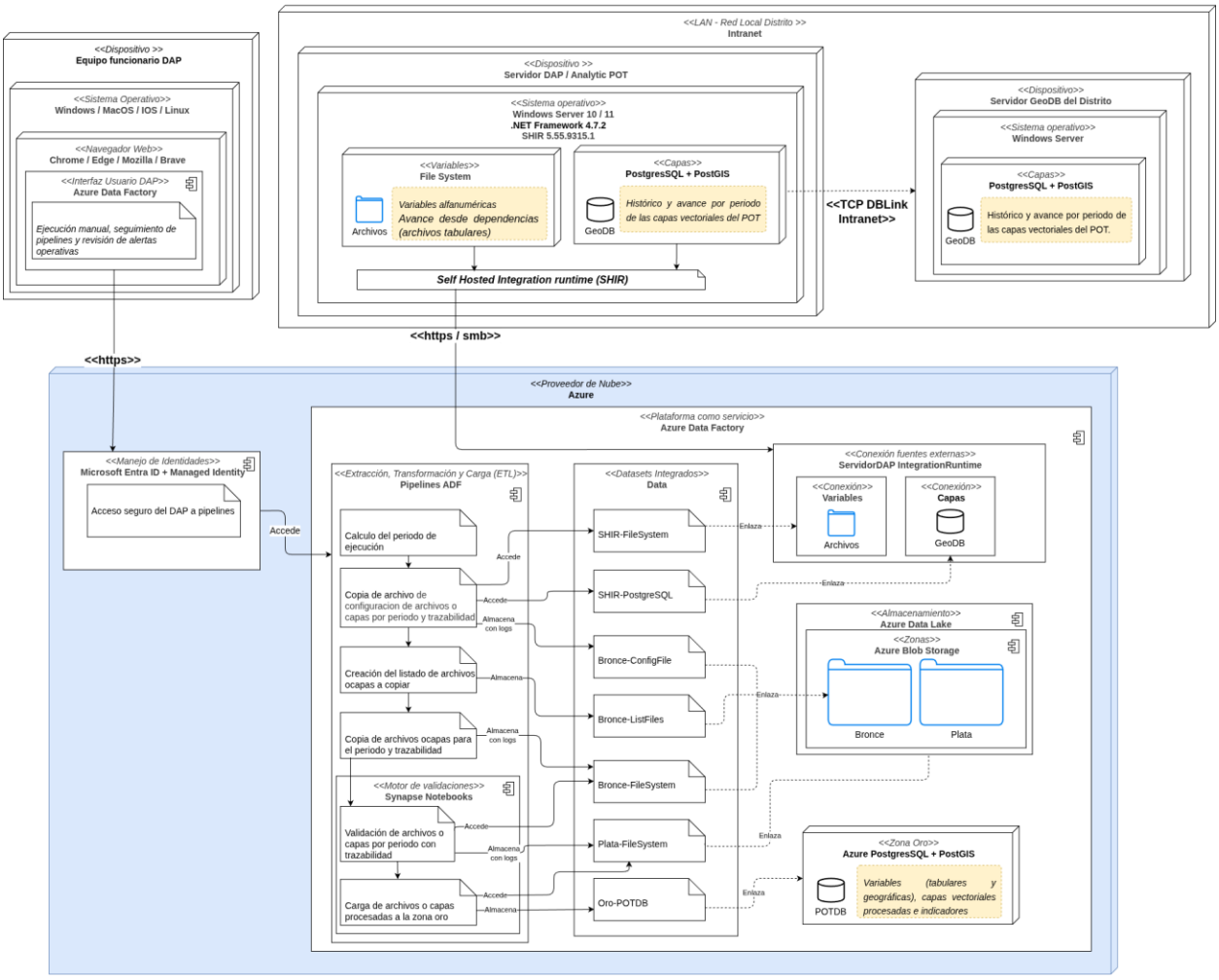
Azure Synapse Analytics orquesta la extracción, validación, transformación y publicación mediante los maestros ETL-CARGA-VARIABLES y ETL-CARGA-CAPAS (tres sub-pipelines cada uno). El agente DAP-IntegrationRuntime1 habilita la integración privada con el recurso compartido del Servidor DAP (variables y config.json) y con el Servidor GeoDB (extracción dinámica de capas). El Data Lake organiza los insumos en zonas Bronce y Plata, y postgresql-dap concentra la Zona Oro. El modelo de seguridad se basa en mínimo privilegio, Identidad Administrada y Key Vault.

La separación entre el lago de datos y la capa de aplicación permite incorporar después herramientas de visualización o consulta (Power BI, Tableau, aplicaciones web o SQL) sin rediseñar la plataforma ETL. El Servidor GeoDB del Distrito es la fuente activa de capas vectoriales: el ETL las extrae dinámicamente mediante SHIRCapasDB a partir de la matrícula en config.json; el File System aporta la configuración y los archivos de variables, no las geometrías de capas.
Con el repositorio dap-pipelines y este documento, un equipo técnico dispone de la referencia para operar la solución: arquitectura, roles, retención, triggers y dimensionamiento económico, aplicando el endurecimiento de red recomendado al pasar de piloto a producción.

Anexos

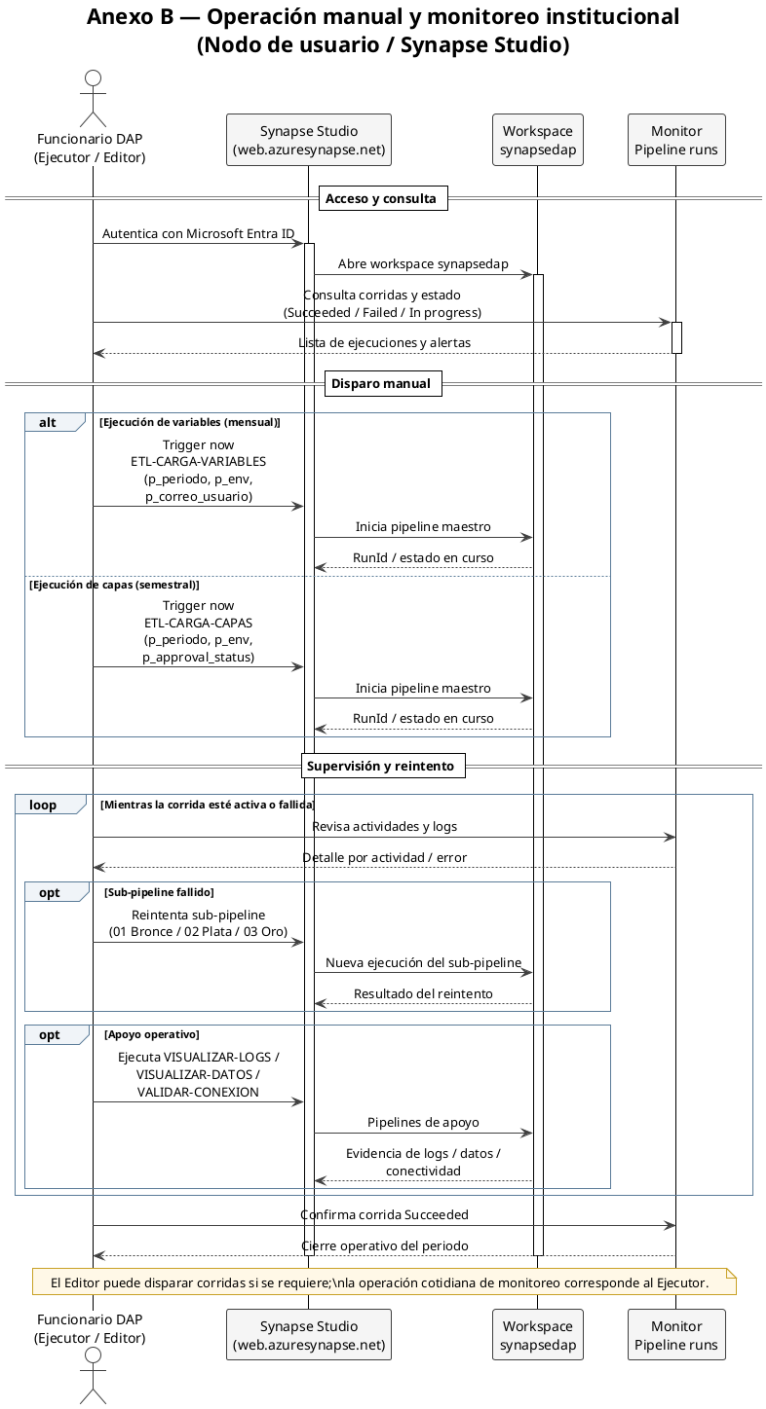
Anexo A - Diagrama de despliegue implementado

Arquitectura de despliegue, procesos ETL para carga de información de avances de POT



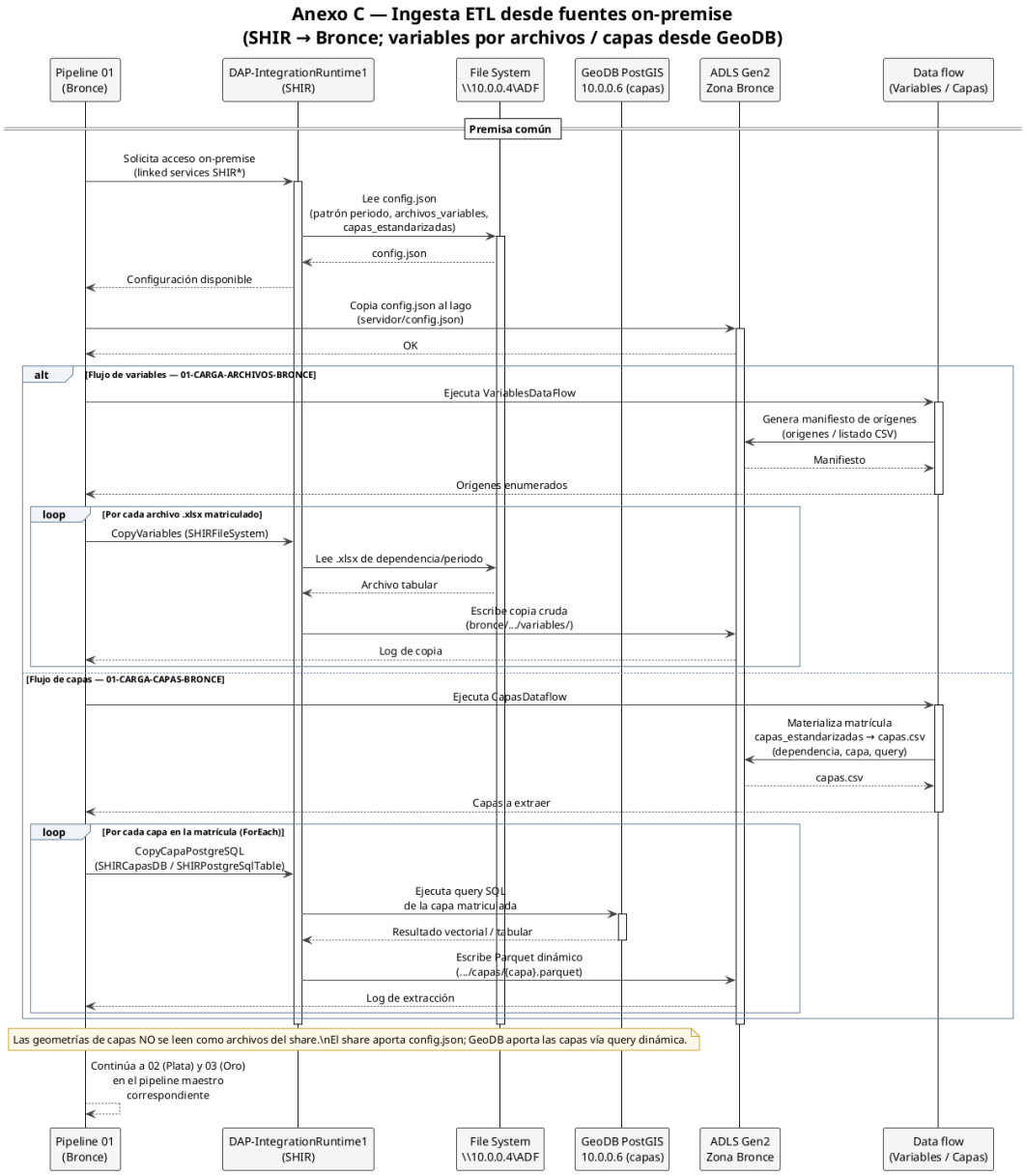
Representa los tres nodos, las zonas Bronce/Plata/Oro, el agente de integración, los notebooks y la base PostgreSQL de publicación.

Anexo B - Operación manual y monitoreo institucional



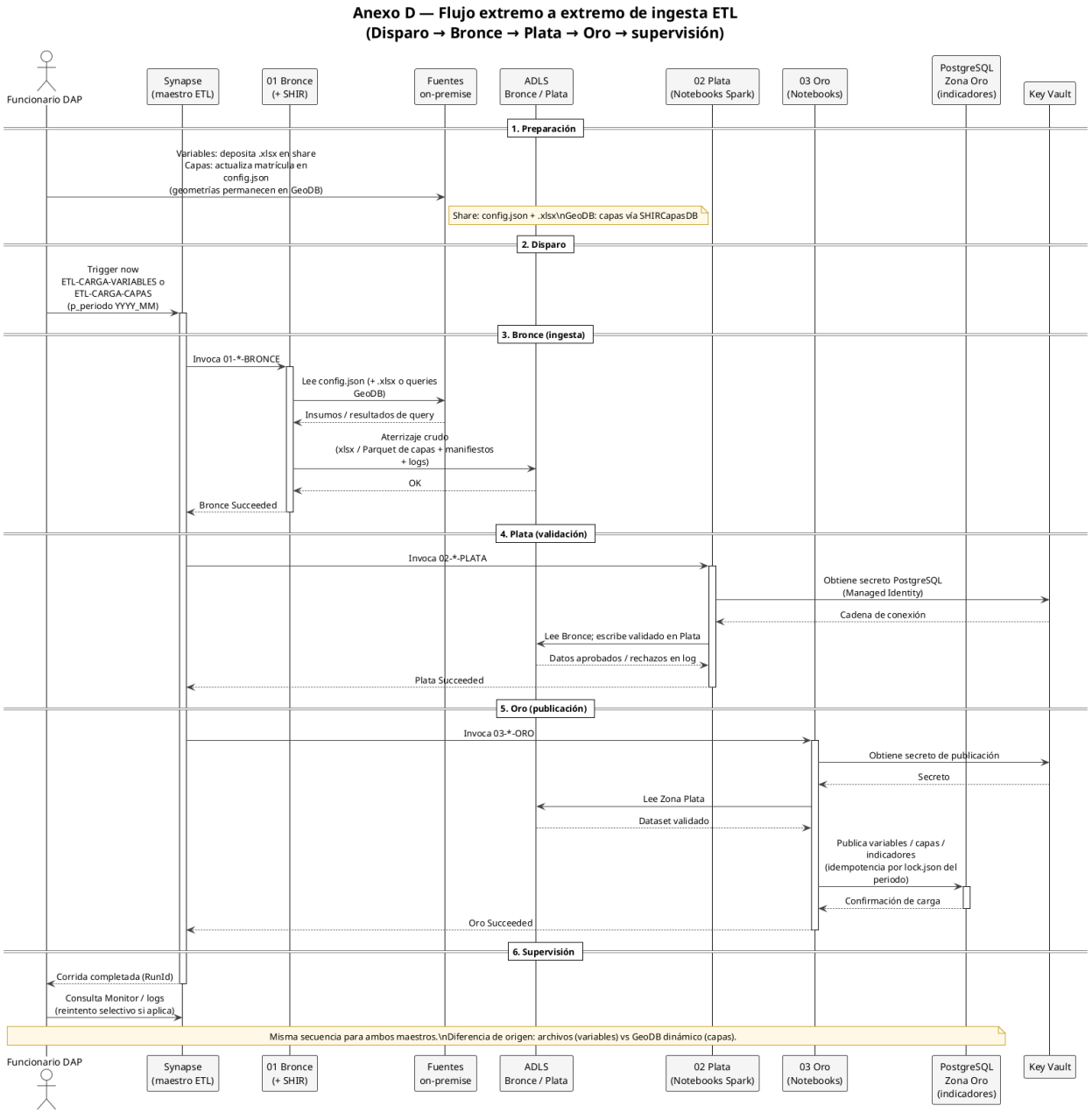
Esta secuencia representa el uso del sistema desde el nodo de usuario institucional: el Funcionario DAP accede a Synapse Studio, consulta corridas, revisa el estado del ETL y lanza ejecuciones manuales o reintenta sub-pipelines fallidos.

Anexo C - Ingesta ETL desde fuentes on-premise



Describe el acceso de Synapse a las fuentes institucionales mediante DAP-IntegrationRuntime1. En el flujo de variables, 01-CARGA-ARCHIVOS-BRONCE obtiene la configuración del periodo, genera el manifiesto de orígenes con VariablesDataFlow y copia los archivos .xlsx hacia la Zona Bronce. En el flujo de capas, 01-CARGA-CAPAS-BRONCE copia config.json, materializa la matrícula con CapasDataflow (capas_estandarizadas) y ejecuta la copia dinámica desde GeoDB (SHIRCapasDB / CopyCapaPostgreSQL) hacia Bronce en formato Parquet. Cada maestro continúa con su sub-pipeline 02 (validación a Plata) y 03 (publicación a Oro).

Anexo D - Flujo extremo a extremo de ingesta ETL



Consolida el flujo desde el disparo del proceso hasta la publicación en la base de indicadores de PostgreSQL y la supervisión por el Funcionario DAP. Participan: Funcionario DAP, workspace Synapse, Servidor DAP (share + SHIR), los sub-pipelines de cada maestro (variables o capas), las zonas Bronce → Plata del lago y la Zona Oro en PostgreSQL.

Anexo E - Glosario de arquitectura

- ADLS Gen2: Azure Data Lake Storage Gen2 (almacenamiento con *Hierarchical namespace*).
- SHIR: Self-Hosted Integration Runtime (agente de integración en red privada).
- MI: Managed Identity (identidad administrada, autenticación sin contraseñas).
- Zona Bronce/Plata/Oro: capas de maduración del dato (crudo / validado / publicado).
- p_periodo: parámetro de periodo en formato YYYY_MM.
- CAF: Cloud Adoption Framework (convención de nomenclatura de Azure).