

**PROCEDIMIENTO PARA CONFIGURAR METADATA DRIVEN**

**PROGRAMA DE GOBIERNO DEL DATO**

22/08/2025

**Contenido**

[1. OBJETIVO 2](#_Toc1517683388)

[2. ALCANCE 3](#_Toc641395272)

[3. ROLES Y RESPONSABILIDADES 3](#_Toc1690149433)

[4. CONCEPTOS CLAVE 3](#_Toc1565647853)

[5. LINEAMIENTOS 4](#_Toc371313935)

[6. PROCEDIMIENTO 4](#_Toc1229096711)

[6.1 Flujo para configurar Metadata Driven 4](#_Toc1068927204)

[6.2 Etapas del procedimiento 5](#_Toc1643626332)

[6.3 Estructura de plantillas del Metadata Driven 6](#_Toc1575311440)

[6.4 Flujo para cambios de parámetros en Metadata Driven 9](#_Toc1541465514)

[6.5 Procedimiento 9](#_Toc1522187631)

[7. REFERENCIA 10](#_Toc635739672)

[8. CONTROL DE VERSIONES 10](#_Toc20860104)

[9. APROBACIÓN 10](#_Toc1296128586)

[10. VIGENCIA 10](#_Toc1151006536)

1. **OBJETIVO**

Definir el procedimiento para configurar la Metadata Driven de cada iniciativa de analítica donde se definan los parámetros para establecer la conexión con las fuentes de información.

1. **ALCANCE**

Este procedimiento aplica a todas las ingestas de datos que aterrizan en el Datalake.

1. **ROLES Y RESPONSABILIDADES**

|  |  |  |
| --- | --- | --- |
| **N°** | **ROL** | **RESPONSABILIDADES** |
| 1 | Gestor Data TI | Tiene el alcance de las fuentes de información basado en el Artefacto de Arquitectura.  Solicita el llenado del formato de metadata driven. |
| 2 | Analista de Diseño del Dato | Apoya y valida el llenado del formato de metadata driven de acuerdo a las fuentes relevadas en la solución.  Crear nuevas carpetas en el site de metadata driven, en caso se mapee una nueva fuente de información.  En entornos QAS y PRD, es el encargado del llenado del formato de metadata driven.  Coordina con Arquitectura de Soluciones los cambios y ejecuciones del notebook de metadata driven. |
| 3 | Data Engineer / Arquitecto del Dato \*  (\*) Para desarrollos in-house: Data Engineer.  (\*) Para desarrollos externos: Arquitecto del Dato. | En entornos DEV, llena el formato de metadata driven.  En entornos QAS y PRD, apoya en el llenado del formato de metadata driven.  Es el encargado de desarrollar el proceso de ingesta de datos de su solución. |
| 4 | Arquitectura de Soluciones | Aprueba el Artefacto de Arquitectura.  Aprueba y realiza cambios en el notebook del Datafactory Analítica en caso se incluyan nuevas fuentes.  Ejecuta pipeline manual que despliegue el notebook del Datafactor Analítica en caso se incluyan nuevas fuentes. |

1. **CONCEPTOS CLAVE**

* **Metadata Driven:**

Estrategia que permite definir procesos de integración y control de accesos basados en metadatos estructurados, asegurando escalabilidad y estandarización.

Dentro de MINSUR, el procesamiento de la Metadata Driven es un proceso clave para la correcta ingesta al Datalake, además de brindar visibilidad de los datos que se van a ingestar, establecen parámetros para la conexión con las fuentes de información y asegurar el tipo de objeto que se va a cargar, la periodicidad, el tipo de carga, entre otros. Esta metadata garantiza que el equipo de desarrollo pueda tener claridad sobre las fuentes a consumir, los datos a extraer y, enriquece la gestión del proyecto al asegurar un plan de trabajo más robusto y aterrizado para los recursos que se van a consumir.

Actualmente, se tiene mapeado 3 fuentes de información: SAP, Pi System y Sharepoint. Estas fuentes tienen una plantilla de Metadata Driven por ambiente (Desarrollo, Pruebas, Producción) donde se llena la información referente a la metadata que se va a consumir de cada fuente.

Esto no limita la creación de plantillas para nuevas fuentes que se detecten en una iniciativa, por el contrario, el presente procedimiento, contempla actividades que involucran las posibles creaciones de nuevas plantillas para nuevas fuentes.

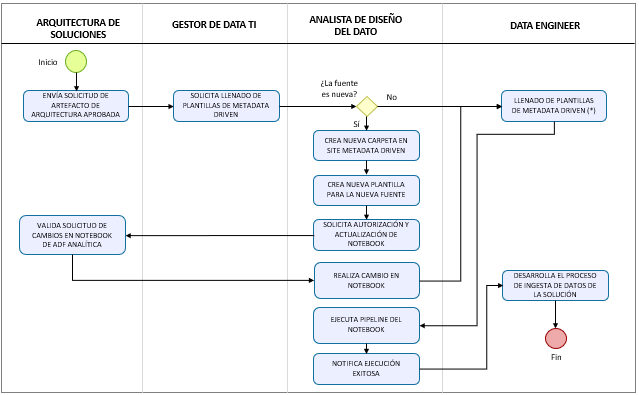
* **Gobierno del Dato:**

Conjunto de políticas, procesos y roles que aseguran la calidad, seguridad y trazabilidad del dato en toda la organización.

1. **LINEAMIENTOS**

* Toda metadata se debe registrar en las plantillas oficiales ubicadas en el site de Metadata Driven en el siguinte enlace de SharePoint: <https://mineriabreca.sharepoint.com/sites/Gobierno-del-Dato/MetaData_Driven/Forms/AllItems.aspx?viewid=b32d28d6%2D1fce%2D4e40%2D9207%2D936b7b0eddc7>
* Todas las plantillas deben seguir la misma estructura de campos definida en el procedimiento.
* Se deberán crear plantillas específicas cuando se detecten nuevas fuentes.
* La única herramienta corporativa y oficial para procesar las plantillas de metadata driven es el Azure Data Factory de Analítica, herramienta que custodia Arquitectura de Soluciones.
* El pipeline que ejecuta el notebook que procesa las plantillas de metadata driven del Azure Data Factory de Analítica se ejecuta a demanda, de manera manual, previa coordinación entre el Analista de Diseño del Dato y el equipo de Infraestructura.
* Se deberá actualizar el presente procedimiento en caso se detecten nuevas fuentes y se creen sus respectivas plantillas.
* Arquitectura de Soluciones será el responsable de aprobar, dentro de la programación del notebook, el procesamiento de las nuevas plantillas de metadata driven de las nuevas fuentes detectadas.
* Toda modificación en parámetros debe quedar registrada y aprobada por el Analista de Diseño del Dato.
* Todo proceso debe cumplir con las políticas de Gobierno del Dato y de Arquitectura y Metadata Analítica.

1. **PROCEDIMIENTO** 
   1. **Flujo para configurar Metadata Driven**



* 1. **Etapas del procedimiento**

|  |  |  |  |
| --- | --- | --- | --- |
| **N° PASO** | **ACTIVIDAD** | **RESPONSABLE** | **HERRAMIENTA** |
| 1 | Enviar Artefacto de Arquitectura aprobado. | Arquitectura de Soluciones | Correo electrónico |
| 2 | Solicitar llenado de plantillas de metadata driven. | Gestor Data TI | Correo electrónico |
| 3 | Evaluar si existe una nueva fuente ajena a Pi System, SAP o Sharepoint. | Analista de Diseño del Dato |  |
| 4.1 | En caso no exista nueva fuente: Derivar con el Data Engineer. | Analista de Diseño del Dato | Correo electrónico |
| 4.2 | En caso sí exista nueva fuente: Crear nueva carpeta en site Sharepoint de Metadata Driven. | Analista de Diseño del Dato | Sharepoint |
| 5 | Crear nuevas plantillas para la nueva fuente. | Analista de Diseño del Dato | Sharepoint |
| 6 | Solicitar autorización y actualización en la programación del notebook del Azure Datafactory Analítico para incluir los parámetros de la nueva fuente. | Analista de Diseño del Dato | Correo electrónico |
| 7 | Validar solicitud de cambios en la programación del notebook del Azure Datafactory Analítico. | Arquitectura de Soluciones | Correo electrónico |
| 8 | Realizar cambios en la programación del notebook. | Analista de Diseño del Dato | Azure |
| 10 (\*) | Llenar formatos de metadata driven de acuerdo con el tipo de fuente. | **En entorno DEV:**  Data Engineer  **En entorno QAS/PRD:**  Analista de Diseño del Dato | Sharepoint |
| 11 | Ejecutar el pipeline del notebook modificado. | Analista de Diseño del Dato | Azure |
| 12 | Notificar al Data Engineer la ejecución exitosa del pipeline. | Analista de Diseño del Dato | Correo electrónico |
| 13 | Desarrollar el proceso de ingesta de datos de la solución. | Data Engineer | Azure |

* 1. **Estructura de plantillas del Metadata Driven**

De acuerdo con el procedimiento presentado en las secciones anteriores, se ha trabajo una plantilla de Metadata Driven para registrar los parámetros de conexión con las fuentes de la solución. Estas plantillas se encuentran en el site de Metadata Driven dentro del repositorio Sharepoint de Gobierno del Dato. Actualmente, se cuentan con 3 plantillas paras las siguientes fuentes de información (ruta del repositorio donde se encuentran las plantillas de Metadata Driven):

* Link de Metadata Driven – Shrarepoint: [Metadata Driven - Sharepoint.xlsx](https://mineriabreca.sharepoint.com/:x:/r/sites/Gobierno-del-Dato/MetaData_Driven/Sharepoint/Metadata%20Driven%20-%20Sharepoint.xlsx?d=we1dd6e54fb7641e9a8fffb443736e473&csf=1&web=1&e=MrtscA)
* Link de Metadata Driven – SAP: [Metadata Driven - SAP.xlsx](https://mineriabreca.sharepoint.com/:x:/r/sites/Gobierno-del-Dato/MetaData_Driven/SAP/DEV/Metadata%20Driven%20-%20SAP.xlsx?d=w8a3986039d754a11bb0b0b51cd76d06d&csf=1&web=1&e=USNNmy)
* Link de Metadata Driven – Pi System: [Metadata Driven - PiSystem.xlsx](https://mineriabreca.sharepoint.com/:x:/r/sites/Gobierno-del-Dato/MetaData_Driven/PiSystem/Metadata%20Driven%20-%20PiSystem.xlsx?d=w0bc40865c07d48a1911362d233361c47&csf=1&web=1&e=IhS1lt)

En caso se detecten nuevas fuentes de información, se debe realizar la creación de las carpetas y las plantillas de Metadata Driven. Adicionalmente, se debe actualizar el presente Procedimiento y en particular, la sección 5.5 donde se agregue: Link de Metadata Driven – [Nombre de la fuente]: [Link de la plantilla]

La estructura de toda plantilla de Metadata Driven creada o por crear, debe contar con la siguiente estructura:

|  |  |  |
| --- | --- | --- |
| **N°** | **CAMPO** | **DESCRIPCIÓN** |
| 1 | Id | Identificador, correlativo por cada fuente que se va a consumir. |
| 2 | UM | Unidad minera (ejemplo: Fundición Pisco). |
| 3 | Dominio | Dominio asignado para la solución. Relevado en el Artefacto de Arquitectura (ejemplo: Logística) |
| 4 | Subdominio | Subdominio asignado para la solución. Relevado en el Artefacto de Arquitectura (ejemplo: Compras) |
| 5 | SourceType | Nombre del tipo de recurso (ejemplo: Pisco) |
| 6 | Type | Tipo de fuente (ejemplo: Sharepoint) |
| 7 | SourceSettings | Configuración de fuente con parámetros como “host”, “instance”, “userName”, “secretName”, “databaseName”, “schema”, “table”, “tableSchema”, “sourceContainer”, “sourcePath”, “sourceFileName”, “periodType”, “shetName”, “extension”. |
| 8 | CopyQueries | Configuración de fuente con parámetros como “sqlReaderFull” y “sqlReaderDelta”. |
| 9 | CopySettings | Configuración de fuente con parámetros como “PrimaryKey”. |
| 10 | SinkSetting | Configuración de fuente con parámetros como “fileName”, “folderPath” y “fileSystem”. |
| 11 | DataLoadingBehavior | Tipo de carga: Full o Incremental. |
| 12 | EntityType | Tipo de tabla que se carga: Maestra o Transaccional. |
| 13 | CopyEnable | Condición de tabla si va a ser ingestado: 1=sí o 0=no. |
| 14 | LastIncrementalUpdate | Última actualización incremental: 1=sí o 0=no. |
| 15 | SchemaADFSettings |  |
| 16 | DataCatalog | Indicar nombre del catálogo de datos bajo siguiente nomenclatura: domain:concept:región |
| 17 | UserDefinedFunction |  |

Detalle de parámetros:

SourceSettings – Configuración de la fuente:

* **Host:** Dirección o nombre del servidor donde reside la fuente de datos.

Ejemplo: SVDCXDBPI0\\\\\\\\SQLEXPRESS

* **Instance:** Nombre o número de instancia de la base de datos o sistema de origen (aplica a motores como SAP, SQL Server, Oracle).

Ejemplo: PRD01

* **UserName:** Usuario técnico utilizado para la conexión a la fuente de datos.

Importante: se recomienda que no sea un usuario personal, sino uno de servicio.

* **SecretName:** Referencia al secreto almacenado en un gestor seguro de credenciales (ejemplo: Azure Key Vault). Evita exponer contraseñas en texto plano.
* **DatabaseName:** Nombre de la base de datos que contiene las tablas a extraer.

Ejemplo: DB\_LOGISTICA

* **Schema:** Esquema de base de datos al que pertenece la tabla de origen.

Ejemplo: dbo

* **Table:** Nombre de la tabla o entidad a consumir desde la fuente.

Ejemplo: Predictivo\_FPSC\_AF\_View

* **TableSchema:** Define la estructura de la tabla de origen (campos, tipos de dato, longitudes).

Usado para validaciones de consistencia antes de la ingesta.

* **SourceContainer:** Contenedor del almacenamiento (en Data Lake o Blob Storage) donde se encuentran los archivos fuente.

Ejemplo: raw-data

* **SourcePath:** Ruta dentro del contenedor donde se localiza el archivo o carpeta.

Ejemplo: /logistica/compras/2025/

* **SourceFileName:** Nombre del archivo fuente a consumir.

Ejemplo: ordenes\_compra\_202508.csv

* **PeriodType:** Frecuencia o granularidad del archivo a consumir.

Ejemplos: diario, mensual, anual.

* **SheetName:** Nombre de la hoja de cálculo (en caso de fuentes Excel).

Ejemplo: Hoja1.

* **Extensión:** Extensión del archivo de origen.

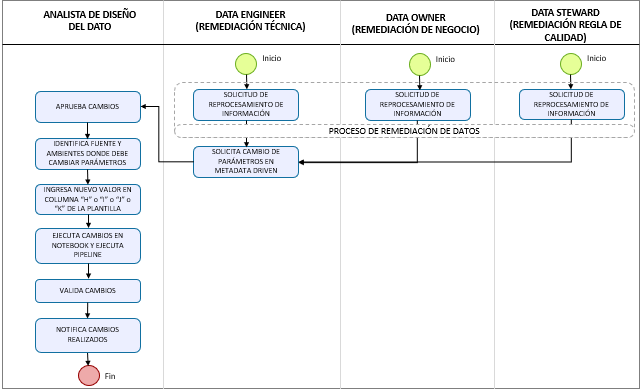
Ejemplo: .csv, .xlsx, .parquet.

CopySettings – Configuración de llaves y particiones

* **PrimaryKey:** claves primarias lógicas del dataset (tag, time).
* **watermarkColumnName:** columna que define el control de cambios (time).
* **watermarkColumnName\_2:** segunda columna de watermark (vacía).
* **watermarkColumnType:** tipo de dato de la columna watermark (datetime).
* **watermarkColumnStartValue:** valor inicial para empezar la carga (vacío, arranca desde el primer registro disponible).
* **partitionOption, partitionColumnName, partitionLowerBound, partitionUpperBound, partitionNames:** usados si se quiere paralelizar la carga en particiones de datos (vacíos, aquí no se usan).

SinkSetting:

* **FileName:** nombre del archivo de salida (Predictivo\_FPSC\_AF\_View).
* **FolderPath:** ruta dentro del Data Lake / Blob Storage (Minsur/Pisco/PISystem/Predictivo\_FPSC\_AF\_View).
* **FileSystem:** contenedor de almacenamiento (prelanding).
  1. **Flujo para cambios de parámetros en Metadata Driven**



* 1. **Procedimiento**

|  |  |  |  |
| --- | --- | --- | --- |
| **N° PASO** | **ACTIVIDAD** | **RESPONSABLE** | **HERRAMIENTA** |
| 1 | Solicitar reprocesamiento de información: esta actividad proviene por una actividad solicitada en el proceso operativo de remediación de datos. | Data Engineer | Correo electrónico |
| Data Owner |
| Data Steward |
| 2 | Solicitar cambio de parámetros en Metadata Driven | Data Engineer | Correo electrónico |
| 3 | Aprobar cambios de parámetros. | Analista de Diseño del Dato |  |
| 4 | Identificar fuentes y ambientes donde se deben aplicar los cambios de parámetros. | Analista de Diseño del Dato | Sharepoint |
| 5 | Ingresar nuevos valores de parámetros en las columnas H y/o I y/o J y/o K | Analista de Diseño del Dato | Sharepoint |
| 6 | Ejecuta cambios en la programación del notebook analítico y ejecuta pipeline que contienen al notebook | Analista de Diseño del Dato | Azure |
| 7 | Validar cambios realizados | Analista de Diseño del Dato |  |
| 8 | Notificar cambios realizados. | Analista de Diseño del Dato | Correo electrónico |

1. **REFERENCIA**

* DG-POL-PCGD-MINS-001-Política Corporativa de Gobierno de Datos.
* DG-POL-PCAMA-MINS-001 – Política Corporativa de Arquitectura y Metadata Analítica – Datalake, Deltalake, Unity Catalog y Metadata Driven.

1. **CONTROL DE VERSIONES**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Versión** | **Fecha** | **Responsable** | **Cambios Realizados** | **Aprobado por** |
| 1.0 | 22/08/2025 | Jean Cabrera | Versión inicial | Comité Directivo |

1. **APROBACIÓN**

Esta política ha sido revisada y aprobada por el Comité Directivo y es de cumplimiento obligatorio para toda la organización.

[Nombre] Comité de Gobierno de Datos [Fecha]

1. **VIGENCIA**

Este lineamiento entra en vigor a partir de su aprobación formal y permanecerá vigente hasta su reemplazo o derogación.

**Aprobado por:** [Nombre del responsable]  
**Cargo:** [Cargo]  
**Fecha de aprobación:** [Fecha]  
**Versión:** 1.0  
**Fecha de próxima revisión:** [Fecha]

**APROBACION DEL DOCUMENTO**

|  |  |  |  |
| --- | --- | --- | --- |
| **Cargo** | **Nombre** | **Firma** | **Fecha** |
| Supervisor De Sistemas De Información | Milagros Paico |  |  |
| Jefe De Arquitectura Y Desarrollo Ágil | Oscar Valencia |  |  |
| Jefe De Seguridad De La Información | Manuel Iturrizaga |  |  |
| Jefe de Analytics | Sthip Blas |  |  |
| Jefe Data Engineering | Luis Grados |  |  |
| Superintendente TI Marcobre | Marco Yacarini |  |  |
| Supervisor de Aplicación Marcobre | Jimmy Bendezú |  |  |
| Superintendente de TI & OT | Edwin Saldaña |  |  |
| Gerente de TI & OT | Dante Cárdenas |  |  |
| Gerente de TD | Piero Saravia |  |  |