Tot plugin GCP BigQuery

Control de versiones

Versión	Fecha de modificación	Responsable	Aprobador	Resumen de cambios
1.0	20/12/2023	Anjana Producto	Anjana Producto	Creación del documento

Modelo de integración

Extracción de metadatos

Para la extracción de metadata se utiliza una conexión BigQuery mediante la cual se accede a la definición de estructuras.

El plugin extrae los siguientes atributos que deben llamarse igual en la tabla attribute_definition, campo name para que aparezcan en la plantilla:

- **catalog** con el valor de catalog en la base de datos
- **schema** con el valor de schema en la base de datos
- **physicalName** y **name** con el mismo valor, el nombre de la tabla
- **path** con la concatenación de los valores de catalog, schema and table
- **infrastructure** con el valor seleccionado
- **technology** con el valor seleccionado
- **zone** con el valor seleccionado
- **tags** son las etiquetas a nivel de vista que tienen las tablas.

También envía los siguientes atributos relativos a los campos del recurso pedido:

- **name** y **physicalName** con el valor del campo
- **defaultValue** con el valor por defecto definido para el campo
- **fieldDataType** con el tipo de dat
- o definido para el campo
- **length** con el tamaño del campo
- **incrementalField** indicando si es un campo incremental
- **position** posición que ocupa el campo
- **precision** con el valor de la precisión del campo
- **nullable** indicando si el campo es nullable
- **pk** indicando si el campo es una pk
- **description** con el valor para el campo
- **tags** son las etiquetas a nivel de columna que tienen las tablas.

Los atributos a crear en Anjana deben de tener los siguientes tipos:

Nombre de atributo	**Tipo de atributo**
catalog	INPUT_TEXT
schema	INPUT_TEXT
physicalName	INPUT_TEXT
path	INPUT_TEXT
infrastructure	SELECT

technology	SELECT
zone	SELECT
tags	ARRAY_ALPHANUMERICAL
name	INPUT_TEXT
defaultValue	INPUT_TEXT
fieldDataType	INPUT_TEXT
length	INPUT_NUMBER
incrementalField	INPUT_CHECKBOX
position	INPUT_NUMBER
precision	INPUT_NUMBER
nullable	INPUT_CHECKBOX
pk	INPUT_CHECKBOX
description	ENRICHED_TEXT_AREA_INTERNATIONAL

Muestreo de datos

Utilizando una conexión BigQuery con la credencial configurada se ejecuta una query con límite de registros sobre los campos inventariados en Anjana Data en la que, adicionalmente, se sustituyen los valores de los campos sensibles por asteriscos.

Aquellos campos que se modifiquen después de crear el objeto en Anjana (es decir, que estén definidos en el metadato pero no se hayan incorporado en la estructura física) aparecerán como no disponibles en el muestreo.

Gobierno activo

La gestión de acceso requiere el plugin “Tot plugin GCP IAM” para que genere los roles (funciones) custom que representan a los DSA.

El presente plugin asociará dichos roles custom con usuarios y condiciones de acceso a nivel tabla siguiendo la recomendación del fabricante:
https://cloud.google.com/bigquery/docs/table-access-controls#api

Edición de objetos

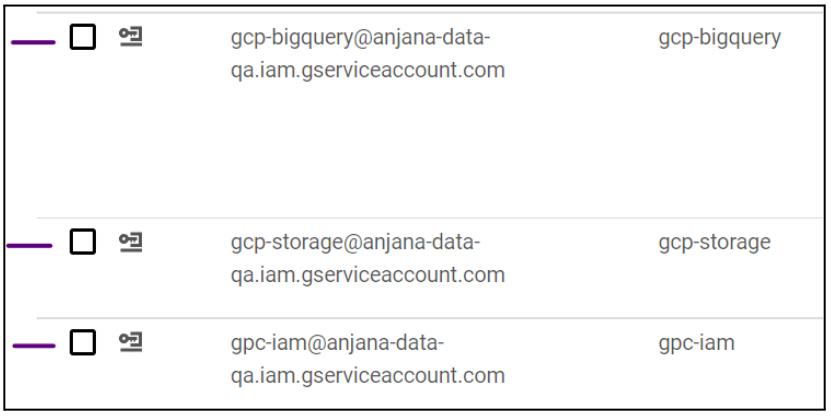
El plugin permite gestionar la activación o desactivación de entidades no nativas incluidas en DSAs, de modo que cuando una entidad no nativa se active se darán los permisos correspondientes en las tablas y cuando se desactive se eliminarán los permisos.

Credenciales requeridas

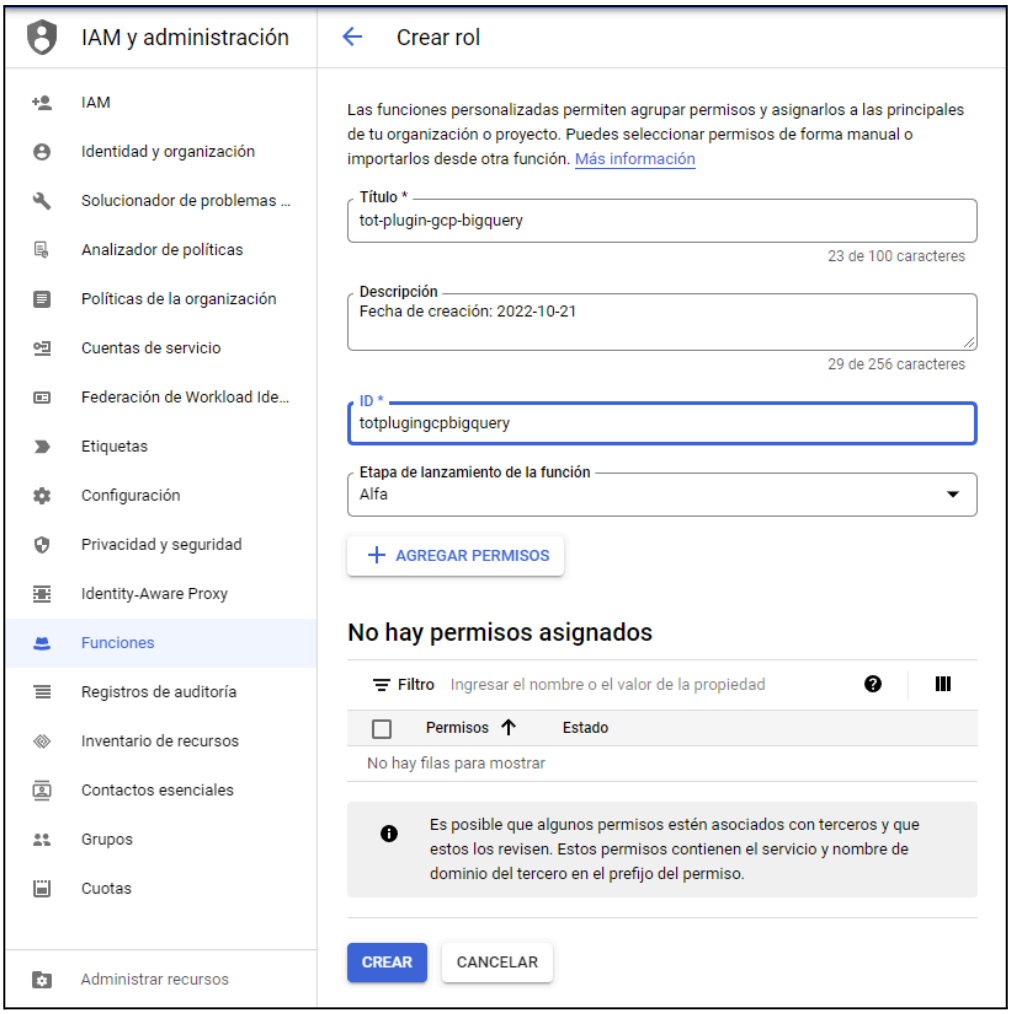
Las credenciales requeridas se deben configurar en el fichero yaml en la propiedad "totplugin.connection.credentialsContent".

Creación de la cuenta de servicio

Para GCP es necesario crear una cuenta de servicio en IAM para cada plugin de forma individual y, tras eso, asignarle los permisos necesarios para la ejecución de las tareas específicas de cada plugin.



Para personalizar los permisos de forma adecuada es necesaria la creación de roles personalizados en los cuáles se engloban los permisos que luego son asociados a las cuentas de servicio.



Extracción de metadatos

Los permisos utilizados son los siguientes:

- bigquery.datasets.get
- bigquery.tables.get
- bigquery.tables.list

Muestreo de datos

Los permisos utilizados son los siguientes:

- bigquery.datasets.get
- bigquery.tables.get
- bigquery.tables.getData
- bigquery.tables.list
- bigquery.jobs.create

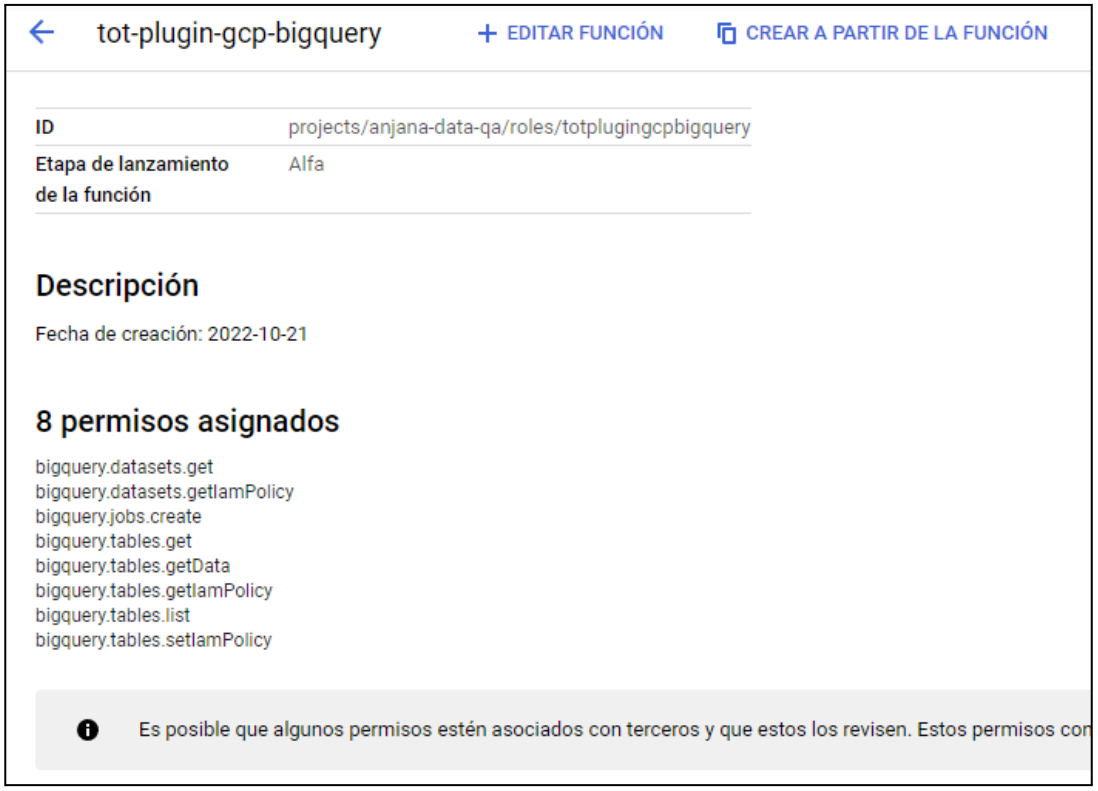
Gobierno activo

La gestión de acceso requiere el plugin “Tot plugin GCP IAM” para que genere los roles (funciones) custom que representan a los DSA. Los permisos que necesita este plugin para poder llevar a cabo el gobierno activo son los siguientes:

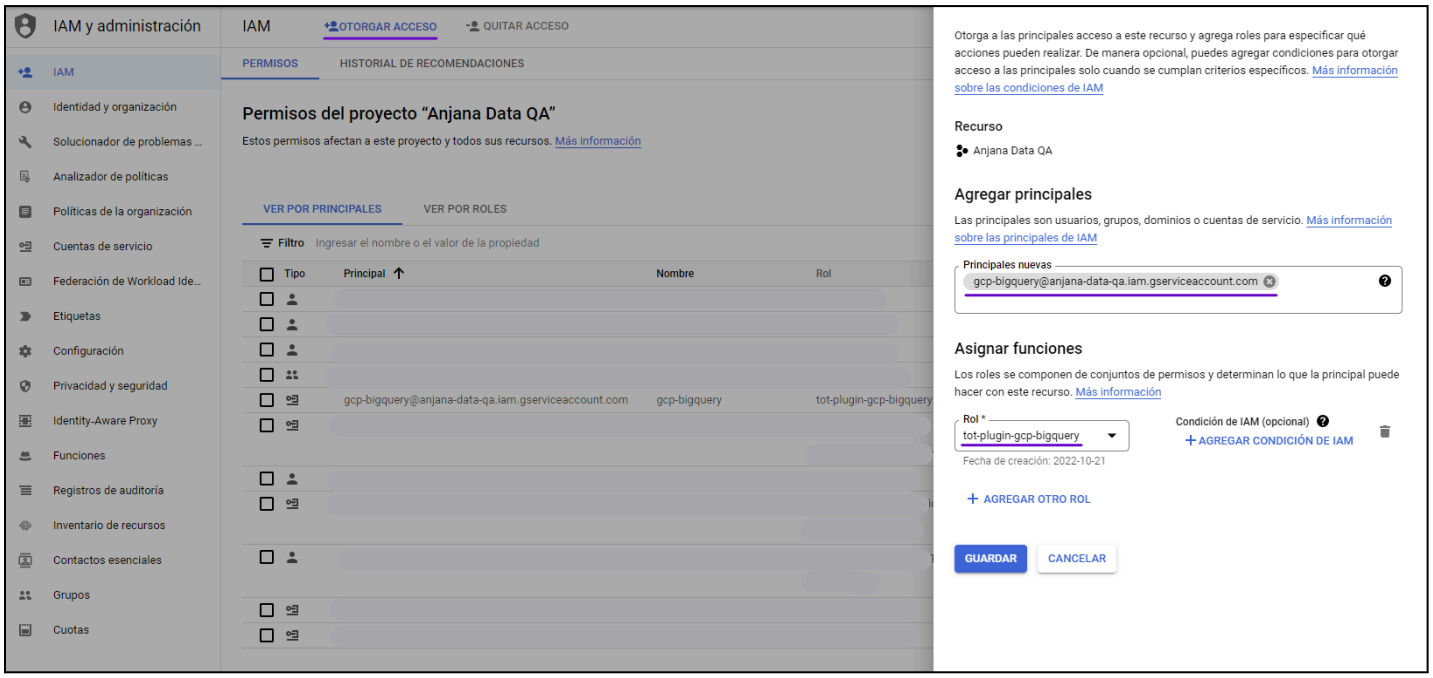
- bigquery.datasets.get
- bigquery.tables.get
- bigquery.tables.getIamPolicy
- bigquery.tables.setIamPolicy

En resumen, los permisos utilizados para el rol personalizado son los siguientes:

- bigquery.datasets.get
- bigquery.tables.get
- bigquery.tables.getData
- bigquery.tables.list
- bigquery.jobs.create
- bigquery.tables.getIamPolicy
- bigquery.tables.setIamPolicy



Para asignar los permisos a la cuenta de servicio de BigQuery es necesario asignar el rol con los permisos al usuario:



Hay que tener en cuenta que Anjana sólo se encarga de dar acceso a los recursos concretos que se gobiernen en Anjana. Para poder realizar consultas o queries en los mismos, se necesita que los usuarios normales tengan una serie de permisos previamente:

- bigquery.jobs.create
- bigquery.datasets.get
- bigquery.jobs.list
- bigquery.models.list
- bigquery.tables.list
- resourcemanager.projects.get

Edición de objetos

Los permisos que necesita este plugin para poder llevar a cabo la activación o desactivación de una entidad no nativa son los siguientes:

- bigquery.datasets.get
- bigquery.tables.get
- bigquery.tables.getIamPolicy
- bigquery.tables.setIamPolicy

Para poder realizar consultas o queries en los recursos, los usuarios normales tienen que tener previamente los siguientes permisos:

- bigquery.jobs.create
- bigquery.datasets.get
- bigquery.jobs.list
- bigquery.models.list
- bigquery.tables.list
- resourcemanager.projects.get

Limitaciones BigQuery

El número máximo de bindings para usuarios en una tabla es 1500, lo que quiere decir es que, como máximo, Anjana podrá tener 1500 usuarios entre propietarios y adheridos en los DSAs que contengan una tabla en particular.

Ejemplo de configuración

Aquí se incluye el detalle de la configuración específica del plugin, para revisar la configuración común, mirar el documento Anjana Data 23.1 - DS - Guía de configuración técnica.

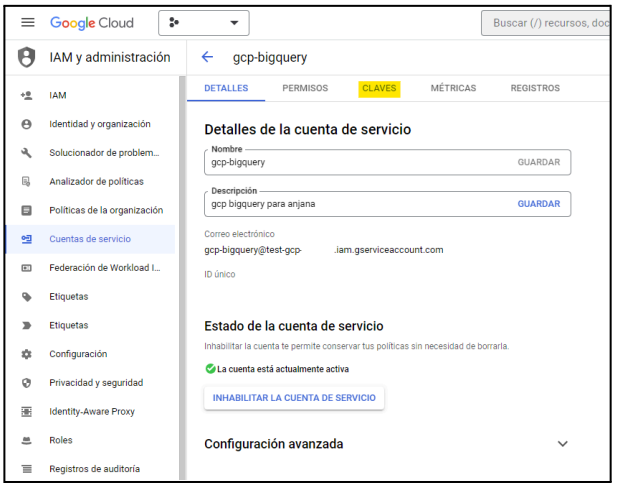
```
server:
  port: 15002

totplugin:
 server:
 urls:
   - http://totserver:15000/tot/
  aris:
    - ari: "anja:totplugin:extract:/Google/gcpBigQuery/devQA/"
    - ari: "anja:totplugin:sample:/Google/gcpBigQuery/devQA/"
    - ari: "anja:totplugin:edit:/Google/gcpBigQuery/devQA/"
    - ari: "anja:totplugin:im:/Google/gcpBigQuery/devQA/"
      imAri: "anja:totplugin:im:/Google/gcpIam/devQA/"
  connection:
    credentialsContent: |
      {
        "type": "service_account",
        "project_id": "anjana-data-qa",
        "private_key_id": "**********************",
        "private_key": "-----BEGIN PRIVATE KEY-----\n\n-----END
PRIVATE KEY-----\n",
        "client_email": "gcp-bigquery@*****.com",
        "client_id": "**********************",
        "auth_uri": "https://accounts.google.com/o/oauth2/auth",
        "token_uri": "https://oauth2.googleapis.com/token",
        "auth_provider_x509_cert_url":
"https://www.googleapis.com/oauth2/v1/certs",
        "client_x509_cert_url":
"https://www.googleapis.com/robot/v1/metadata/x509/gcp-bigquery%40****
*.com",
        "universe_domain": "googleapis.com"
      }
    sample-rows: 15
    path-separator: "/"
```

```
eureka:
 client:
   serviceUrl:
     defaultZone: http://totserver:15000/tot/eureka
```

Configuraciones específicas:

- connection:
  - credentialsContent: Credenciales de acceso a GCP. Se obtienen en la pestaña claves de la cuenta de servicio



  - sample-rows: Tamaño del muestreo de datos.
  - path-separator: Separador que GCP usa en los roles