



Tot plugin AWS Glue

Control de versiones	2
Modelo de integración	2
Extracción de metadatos	3
Credenciales requeridas	4
Extracción de metadatos	4
Despliegue	5
Configuración necesaria	5

Control de versiones

Versión	Fecha de modificación	Responsable	Aprobador	Resumen de cambios
1.0	28/10/2022	Anjana Producto	Anjana Producto	Creación del documento

Modelo de integración

Extracción de metadatos

Se utilizan los métodos que ofrece la API de Glue para acceder a la definición de bases de datos y tablas.

Extrae los siguientes atributos que deben llamarse igual en la tabla `attribute_definition`, campo `name` para que aparezcan en la plantilla.

- **dataBase** con el valor del nombre de la base de datos
- **physicalName** y **name** con el mismo valor, el nombre de la tabla
- **description** con el valor de la descripción 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
- **data_format** con el valor seleccionado
- **compressed** indicando si es comprimido
- **partitionKeys** con la concatenación de las diferentes particiones

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

- **name** con el valor del campo correspondiente
- **physicalName** con el valor del campo correspondiente
- **description** con el valor del campo correspondiente al comentario
- **fieldDataType** con el tipo de dato definido para el campo correspondiente
- **position** posición que ocupa el campo correspondiente
- **pk** indicando si el campo es una pk

El plugin es capaz de realizar la extracción de metadatos de los siguientes tipos de elementos:

- Tablas de las diferentes base de datos que estén alojadas en el servicio; siempre en el mismo formato que aparezca en Glue, nunca en el formato de la localización de los ficheros generados (S3).

Credenciales requeridas

Extracción de metadatos

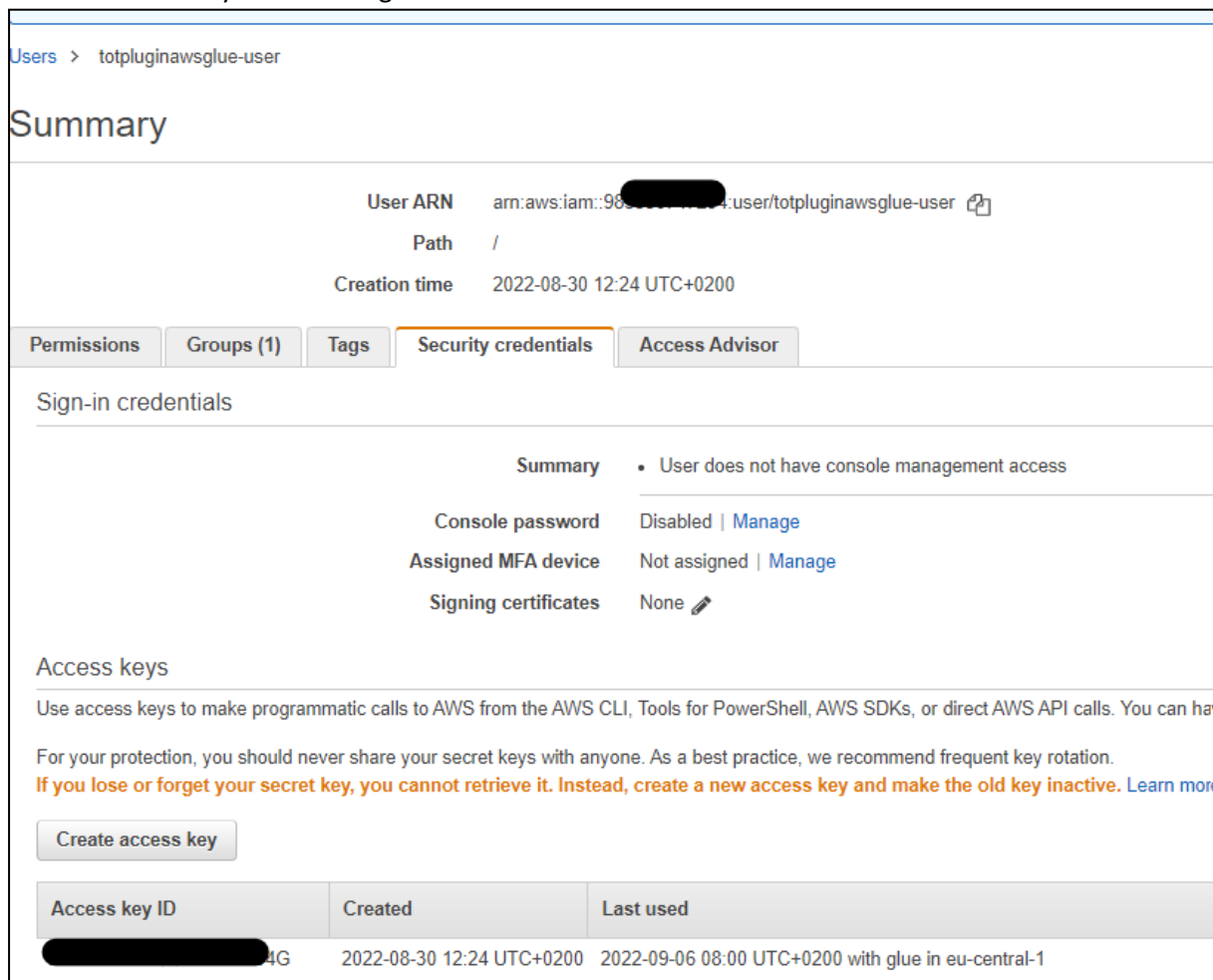
Es necesario un usuario con permisos de lectura sobre el servicio de Glue. Entre los permisos* de lectura son necesarios los siguientes:

- glue:SearchTables
- glue:GetConnection
- glue:GetTable
- glue:GetTables
- glue:GetDatabases
- glue:GetPartitions

*Nota: Estos son los permisos mínimos para un funcionamiento correcto de Amazon Glue. Cada tecnología por la que pase Glue para obtener los datos puede necesitar sus propios permisos en Glue.

Si se desea realizar extracción de metadato se requiere de una conexión con Amazon Glue. Para establecer dicha conexión se requiere un accessKey y un secretKey de la cuenta AWS IAM. Además se requiere la región en la que se encuentra.

Para poder obtener el acceskey y secretkey necesario para la instalación, se debe crear un usuario en AWS IAM y, en la pestaña de credenciales de seguridad, crear unas claves de acceso, las cuales se introducirán en el yaml de configuración.



The screenshot shows the AWS IAM console page for a user named 'totpluginawsglue-user'. The 'Summary' section displays the User ARN, Path, and Creation time. Below this, there are tabs for 'Permissions', 'Groups (1)', 'Tags', 'Security credentials', and 'Access Advisor'. The 'Security credentials' tab is active, showing 'Sign-in credentials' with a summary that the user does not have console management access. It also lists 'Console password' (Disabled), 'Assigned MFA device' (Not assigned), and 'Signing certificates' (None). The 'Access keys' section provides instructions on how to use access keys and includes a 'Create access key' button. Below this, a table lists the created access keys.

Access key ID	Created	Last used
[Redacted]	2022-08-30 12:24 UTC+0200	2022-09-06 08:00 UTC+0200 with glue in eu-central-1

Se debe crear una política de AWS y asignarla al grupo donde se encuentre el usuario.
Debe ser así:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "VisualEditor0",
      "Effect": "Allow",
      "Action": [
        "glue:SearchTables",
        "glue:GetTable",
        "glue:GetConnection",
        "glue:GetTables",
        "glue:GetPartitions",
        "glue:GetPartition",
        "glue:GetDatabases"
      ],
      "Resource": "*"
    }
  ]
}
```

Despliegue

Se ha de seguir el manual genérico del despliegue de plugins.
Doc: Anjana Data x.x - DOC - Tot despliegue de plugins

Configuración necesaria

```
server:
  port: 15019

totplugin:
  location: http://totpluginawsglueserver:15019/plugin/aws/glue/api/v1
  server:
    url: http://totserver:15000/tot/
  connection:
    accessKey: ""
    secretKey: ""
    idAccount: ""
    region: ""
  aris:
    - ari: "anja:totplugin:extract:/AWS/GLUE/DEV/"
    - ari: "anja:totplugin:im:/AWS/GLUE/DEV/"
    imAri: "anja:totplugin:im:/AWS/GLUE/IAM/"
```

Server:

- port: El puerto en el que se va a desplegar el plugin.

TotPlugin (Apartado con la configuración específica del plugin):

- Location: URL del plugin cuando está desplegado (lo que se debe modificar es el host y el puerto, la ruta de entrada no debe modificarse)

- Server:
 - Url: URL de tot
- Aris:
 - ari: ARI usada para registrarse en Tot y poder ser referido y llamado según eventos en Anjana.
- Connection (apartado con la configuración relativa a las credenciales de conexión con Amazon S3):
 - accessKey: Clave de acceso de la cuenta de Anjana generada para el plugin.
 - secretKey: La contraseña de acceso de la cuenta de Anjana generada para el plugin.
 - region: La región donde se encuentra registrado Glue.