



Tot plugin AWS S3

Control de versiones	2
Modelo de integración	3
Extracción de metadatos	3
Muestreo de datos	4
Gobierno activo de estructuras	4
Gobierno activo de accesos	4
Edición de objetos	4
Credenciales requeridas	4
Extracción de metadatos	4
Muestreo de datos	5
Gobierno activo de estructuras	6
Gobierno activo de accesos	6
Edición de objetos	6
Limitación en los nombres de ficheros	6
Configuración	7

Control de versiones

Versión	Fecha de modificación	Responsable	Aprobador	Resumen de cambios
1.0	01/12/2023	Anjana Producto	Anjana Producto	Creación del documento
1.1	04/03/2024	Anjana Producto	Anjana Producto	Incluida la edición de objetos en la Integración y en Credenciales

Modelo de integración

Extracción de metadatos

Para listar las estructuras disponibles, se recupera la lista de *buckets* accesibles y todos los objetos con el objetivo de retornar un mapa de contenido. El comportamiento es un poco diferente si se configura un *bucket* en el fichero YML de configuración ya que, en tal caso, se limita el listado al contenido de ese *bucket*.

El valor retornado en el listado de estructuras es un objeto con el nombre de la estructura (que es un concatenado de la infraestructura, la tecnología y la zona) y una lista de objetos que en ella se encuentran. En este caso la estructura se corresponde con el *bucket* y el objeto con el objeto en Amazon S3.

Para la extracción de metadata de un objeto se utilizan las mismas herramientas, se lee el contenido del objeto para extraer su metadato y se devuelve el resultado.

Después de ejecutar la extracción del metadato se obtiene un objeto con un nombre (*elementName*), con una lista de atributos clave valor (como la infraestructura, la tecnología y la zona) y además una lista de campos (*fields*) asociados al mismo, también con su lista de atributos clave valor.

Para una exitosa extracción de metadato, los atributos deben llamarse igual en la tabla *attribute_definition* para el campo *name* con el objetivo de que aparezcan en pantalla:

- **physicalName** con el nombre con el mismo valor que el objeto
- **path** con la concatenación de los valores del objeto en Amazon S3
- **infrastructure** con el valor seleccionado
- **technology** con el valor seleccionado
- **zone** con el valor seleccionado

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

Nombre de atributo	Tipo de atributo
physicalName	INPUT_TEXT
path	INPUT_TEXT
infrastructure	SELECT
technology	SELECT
zone	SELECT
name	INPUT_TEXT

También enviará atributos relativos a los campos del recurso pedido, siempre dependiendo del contenido y tipo del recurso. Para más información leer el documento de [Extracción de metadata de ficheros](#).

Muestreo de datos

Para el muestreo de datos, se localiza el objeto a muestrear (hasta el número máximo de resultados configurados), se lee el contenido de sus ficheros utilizando librerías de Apache según la tipología de estos y se devuelven los resultados.

El valor retornado en el muestreo de datos es un objeto que contiene cabeceras (*headers*) y valores (*values*).

En las cabeceras (*headers*) se incluyen los nombres de los atributos de los objetos que se deben retornar tras el muestreo de los datos.

En los valores (*values*) se incluyen la lista de valores de cada cabecera para cada objeto que se desea retornar. Esto permitirá que el plugin pueda devolver, aparte de atributos como el nombre del fichero o el contenido del fichero, el resto de datos y metadatos.

Gobierno activo de estructuras

En el protocolo S3 las rutas son emuladas con lo cual no es posible pre-provisionar dichos elementos.

Gobierno activo de accesos

La gestión de accesos sobre esta tecnología se realiza directamente en AWS IAM por lo cual se delega este tipo de actuación en el plugin para dicha tecnología, siendo imprescindible la presencia de este plugin, por tanto, para disponer de la funcionalidad.

Edición de objetos

La edición de objetos sobre esta tecnología se realiza directamente en AWS IAM por lo cual se delega este tipo de actuación en el plugin para dicha tecnología, siendo imprescindible la presencia de este plugin, por tanto, para disponer de la funcionalidad.

Credenciales requeridas

Extracción de metadatos

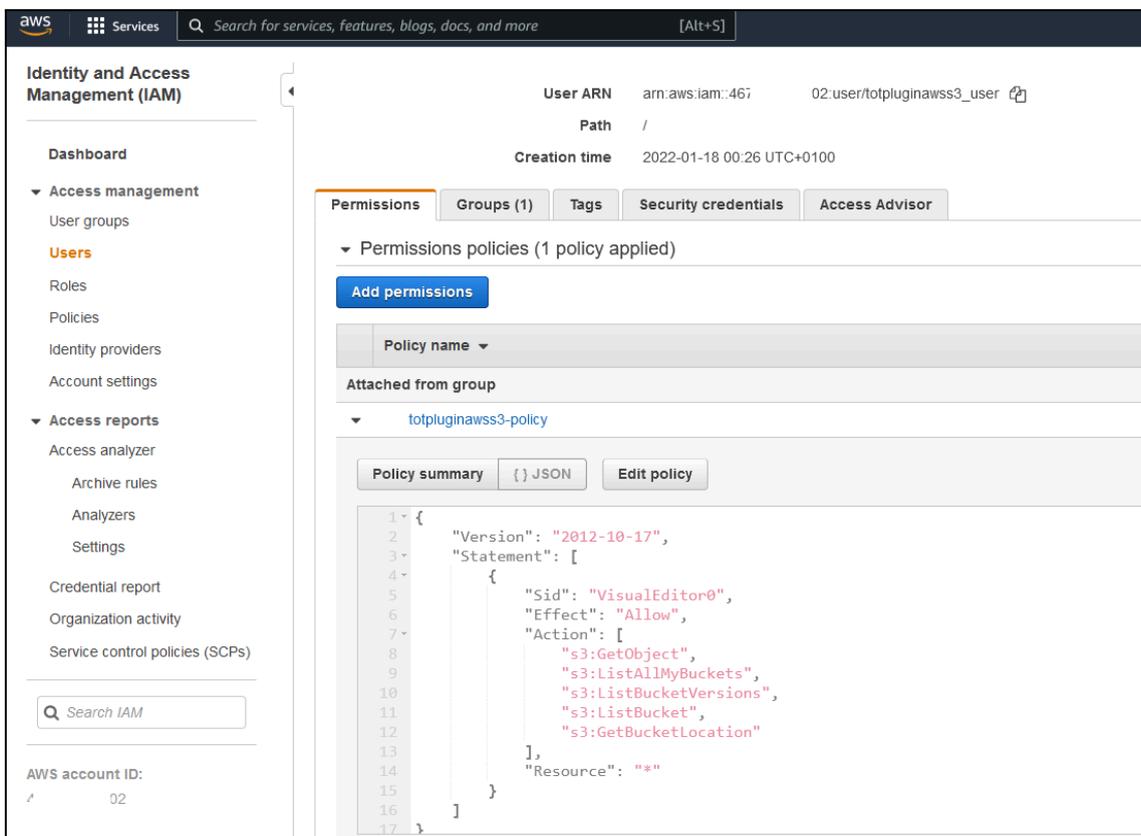
Si se desea realizar acciones de extracción de metadato se requiere una conexión con Amazon S3. Para establecer dicha conexión son necesarias una *accessKey* y una *secretKey* de la cuenta que se proporcione a Anjana para manejar el gobierno del dato. De manera opcional, se requiere un proxy si el usuario no quiere que se conecte directamente a Amazon S3.

De los datos de Amazon S3 se necesita la región en la que se encuentra y de manera opcional, si sólo se quiere gobernar un único *bucket*, el *bucket* deseado (si no se incluye un *bucket*, se gobiernan todos los *buckets* en los que la cuenta tenga permisos).

Por su parte, Amazon S3 define un conjunto de acciones¹ que se pueden especificar en una política y que ayudarán a que el usuario con acceso a la tecnología pueda obtener la información deseada.

Para este plugin resultan interesantes los siguientes acciones:

- **s3:ListAllMyBuckets** para listar todos los *buckets* del usuario autenticado.
- **s3:ListBucket** para listar el contenido de un *bucket*.
- **s3:GetBucketLocation** para retornar la región en la que reside el *bucket*.
- **s3:GetObject** para retornar objetos de Amazon S3. Para poder leer el objeto se deberá, además, tener permisos de lectura sobre el mismo.



The screenshot shows the AWS IAM console interface. On the left is a navigation menu for Identity and Access Management (IAM). The main content area displays the details for a user named 'totpluginawss3_user'. The user's ARN is 'arn:aws:iam::46702:user:totpluginawss3_user'. Under the 'Permissions' tab, it shows one policy applied: 'totpluginawss3-policy'. The policy's JSON is displayed below, listing several S3 actions: 's3:GetObject', 's3:ListAllMyBuckets', 's3:ListBucketVersions', 's3:ListBucket', and 's3:GetBucketLocation'.

Muestreo de datos

Para desencadenar acciones relacionadas con el muestreo de datos se requiere de la misma configuración y credenciales que las mencionadas anteriormente para la [Extracción de metadatos](#).

¹ Acciones permitidas en Amazon S3: https://docs.aws.amazon.com/AmazonS3/latest/API/API_Operations_Amazon_Simple_Storage_Service.html

Gobierno activo de estructuras

En el protocolo S3 las rutas son emuladas con lo cual no es posible pre-provisionar dichos elementos.

Gobierno activo de accesos

La gestión de accesos sobre esta tecnología se realiza directamente en AWS IAM por lo cual se delega este tipo de actuación en el plugin para dicha tecnología, siendo imprescindible la presencia de este último por tanto para disponer de la funcionalidad.

Edición de objetos

La edición de objetos sobre esta tecnología se realiza directamente en AWS IAM por lo cual se delega este tipo de actuación en el plugin para dicha tecnología, siendo imprescindible la presencia de este último por tanto para disponer de la funcionalidad.

Limitación en los nombres de ficheros

Existen ciertas restricciones en AWS con los nombres de ficheros para que funcione todo correctamente. Los caracteres a usar que son seguros son los siguientes:

Alphanumeric characters	<ul style="list-style-type: none">• 0-9• a-z• A-Z
Special characters	<ul style="list-style-type: none">• Exclamation point (!)• Hyphen (-)• Underscore (_)• Period (.)• Asterisk (*)• Single quote (')• Open parenthesis (()• Close parenthesis ())

También existen caracteres que pueden necesitar un tratamiento especial, aunque se recomienda no usarlos para evitar problemas:

- Ampersand ("&")
- Dollar ("\$")
- ASCII character ranges 00–1F hex (0–31 decimal) and 7F (127 decimal)
- 'At' symbol ("@")
- Equals ("=")
- Semicolon (";")
- Forward slash ("/")
- Colon (":")
- Plus ("+")
- Space – Significant sequences of spaces might be lost in some uses (especially multiple spaces)
- Comma (",")
- Question mark ("?")

Y finalmente hay caracteres a evitar y que no son soportados por AWS <https://docs.aws.amazon.com/AmazonS3/latest/userguide/object-keys.html>

- Backslash ("\")
- Left curly brace ("{")
- Non-printable ASCII characters (128–255 decimal characters)
- Caret ("^")
- Right curly brace ("}")
- Percent character ("%")
- Grave accent / back tick ("`")
- Right square bracket ("]")
- Quotation marks
- 'Greater Than' symbol (">")
- Left square bracket ("[")
- Tilde ("~")
- 'Less Than' symbol ("<")
- 'Pound' character ("#")
- Vertical bar / pipe ("|")

Para más información consultar la documentación de AWS.

Configuración

Aquí se incluye el detalle de la configuración específica del plugin.
En la Guía de Configuración técnica se explica la configuración común.

```
server:
  port: 15007

totplugin:
  server:
    urls:
      - http://totserver:15000/tot/

  aris:
    - ari: "anja:totplugin:extract:/AWS/S3/DEV/"
    - ari: "anja:totplugin:sample:/AWS/S3/DEV/"
    - ari: "anja:totplugin:im:/AWS/S3/DEV/"
      imAri: "anja:totplugin:im:/AWS/S3/IAM/"

  connection:
    #proxy:
    accessKey: <accessKey>
    secretKey: <secretKey>
    bucket: <bucket>
    region: <region>

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

Server:

- port: El puerto en el que se va a desplegar el plugin.
- connection.proxy(opcional): La url del proxy si no se quiere que el plugin tenga conexión directa con S3
- connection.accessKey: Clave de acceso de la cuenta de Anjana generada para el plugin.
- connection.secretKey: La contraseña de acceso de la cuenta de Anjana generada para el plugin.
- connection.bucket(opcional): Si se define, limita a gobernar un bucket específico. En caso de que no se defina, se gobernarán todos los buckets accesible desde la cuenta.
- region: La región donde se encuentra registrado el S3.