greenpixie

GPX Data Onboarding

For AWS Cost & Usage Report (CUR) Enrichment with Emissions Data

Greenpixie's **"CUR Enrichment"** process is designed to integrate emissions data into AWS Cost & Usage Reports (CUR), offering a more holistic view of cloud resource utilisation. Clients begin by uploading their CUR files to an Amazon S3 bucket and granting Greenpixie the necessary read and write permissions for these buckets.

Setup Requirements

- Client will need to create, or have pre-existing, a Cost and Usage Report writing to S3. Instructions for creating a Cost and Usage Report are provided by AWS.
- Greenpixie recommends **including Resource IDs** to better understand emissions sources, **hourly time granularity** for time-of-use data, and **parquet format** for output (also available for gzip). Please inform Greenpixie if the report settings differ from these recommendations.
- Client will also need to create, or have pre-existing, an S3 bucket for Greenpixie to output enriched CUR files.
- **If required, client can use a single bucket with separate prefixes for inputs and outputs**

For ongoing enrichment, Client must set up the process for CUR files to be added to the specified bucket when they are available.

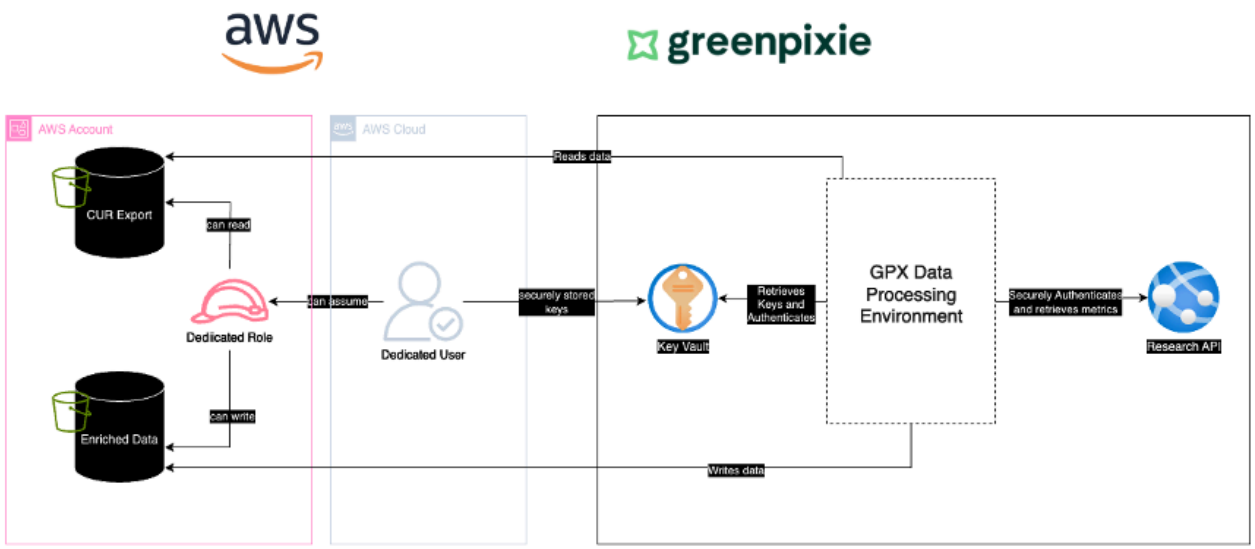
Both S3 buckets should have ACLs disabled to ensure object ownership reverts to the bucket owner, even if Greenpixie or an AWS billing entity wrote the objects.

By disabling ACLs, the IAM policies should be enforced rather than any policies written by the original object owner. This should be applied to both the read and write buckets.

greenpixie

Setting up IAM Roles / Policies

GPX Data accesses the input and output buckets using AWS IAM Users and Roles. An IAM Role with read-access to the input bucket and an IAM Role with write-access to the output bucket is created in the Client tenant. Permissions are given for an IAM User, which sits in the Greenpixie tenant, to assume each role. This assume role policy also uses an External ID string to increase security. This is shown in the diagram below:



IAM Role in Client Account

Input – This policy grants read permissions and should be applied to the input bucket. Client will need to specify the S3 bucket where Greenpixie picks up CUR data (<*SOURCE_S3*>), as well as any folder restrictions (<*OPTIONAL_FOLDER*>*)* to limit Greenpixie's access to non-CUR data in the same bucket.

Output – This policy also grants read and write permissions and should be applied to the output bucket for enriched CUR files. This permission requires an <*DESTINATION_S3*> and optional folder (<*OPTIONAL_FOLDER*>) specifying where enriched CUR files should be written.

These policies can either be created separately or combined into one policy / role. Overleaf you can see an example of a combined policy for a single role.

*If the folder specified contains CUR partition folders such as /**year=2024/month=1/** or **/20240101-20240201/20240124T194625Z/**, then please take note of the structure of this.*

greenpixie

Example Read/Write Access Policy:

```
Unset
{
	"Version": "2012-10-17",
	"Statement": [
		{
			"Action": [
				"s3:GetBucketLocation",
				"s3:ListBucket",
				"s3:GetObject",
				"s3:GetObjectTagging",
				"s3:GetObjectVersionTagging"
			],
			"Resource": [
				"arn:aws:s3:::<SOURCE_S3>",
				"arn:aws:s3:::<SOURCE_S3>/<OPTIONAL_FOLDER>/*"
			],
			"Effect": "Allow",
			"Sid": "AccessSourceBucket"
		},
		{
			"Action": [
				"s3:PutObject",
				"s3:GetObject",
				"s3:ListBucket",
				"s3:DeleteObject",
				"s3:GetBucketLocation",

				"s3:PutObjectAcl"
			],
			"Resource": [
				"arn:aws:s3:::<DESTINATION_S3>",
				"arn:aws:s3:::<DESTINATION_S3>/<OPTIONAL_FOLDER>/*"
			],
			"Effect": "Allow",
			"Sid": "AccessDestBucket"
		}
	]
}
```

greenpixie

IAM Role Trust Relationship

The created role must allow the dedicated IAM User from the Greenpixie account to assume it. This is done through adding a Trust Relationship to the role as follows:

Here the <*CLIENT_NAME*> is a unique identifier agreed between Greenpixie and the client. This should be a string with no spaces or special characters.

We also require an increase in the default MaxSessionDuration to avoid processing timeouts.

Example IAM Role Trust Policy:

```
Unset
{
        "Version": "2012-10-17",
        "Statement": [
                {
                        "Effect": "Allow",
                        "Principal": {
                                "AWS":
                "arn:aws:iam::457748205563:user/gpx-data-<CLIENT_NAME>-user"
                        },
                        "Action": "sts:AssumeRole",
                        "Condition": {
                                "StringEquals": {
                                        "sts:ExternalId": "gpx-data-<CLIENT_NAME>-assume-id"
                                }
                        }
                }
        ],
        "MaxSessionDuration": 43200
}
```

greenpixie

Finalising the setup

Once the appropriate buckets and roles are created, please notify Greenpixie along with the following information:

- **Source Bucket ARN**
  (+ optional folders if applicable)
- **Destination Bucket ARN**
  (+ optional folders if applicable)
- **Input Role ARN**
- **Output Role ARN**
  (can be same as input)
- **External ID String**
  (example: “gpx-data-<CLIENT_NAME>-assume-id”)
- **Partition Folder Structure**
  (year=XXXX/month=XX, 20240101-20240201/20240124T194625Z/, etc)

The GPX team will then run a configuration test to make sure that everything is set up correctly, and we can get started on your data enrichment.

greenpixie

CloudFormation Setup

To simplify this process, we have created CloudFormation templates which automatically create the roles and policies listed above in the client environment. To run this navigate to ***CloudFormation*** in your AWS console and follow the instructions to setup a new CF Stack using the template below.

For more information on this visit:
https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/cfn-using-console.html

This can also be done using the CLI.

gpx-data-client-setup.yaml:

```
Unset
AWSTemplateFormatVersion: '2010-09-09'
Description: Setup IAM role and policies for GPX CUR enrichment bucket access with role and
bucket policy.

Parameters:
  ClientName:
        Type: String
        Description: The name of the client.

  SourceBucketName:
        Type: String
        Description: The name of the source S3 bucket to be accessed.

  DestBucketName:
        Type: String
        Description: The name of the destination S3 bucket to be accessed.

Resources:
  GPXDataBucketPolicy:
        Type: AWS::IAM::Policy
        Properties:
                PolicyName: !Sub 'gpx-data-${ClientName}-policy'
                PolicyDocument:
                        Version: "2012-10-17"
                Statement:
                - Sid: "AccessSourceBucket"
                        Effect: "Allow"
                        Action:
```

greenpixie

```
                        - "s3:GetObject"
                        - "s3:ListBucket"
                        - "s3:GetBucketLocation"
                    Resource:
                        - !Sub 'arn:aws:s3:::${SourceBucketName}'
                        - !Sub 'arn:aws:s3:::${SourceBucketName}/*'
            - Sid: "AccessDestBucket"
                    Effect: "Allow"
                    Action:
                        - "s3:PutObject"
                        - "s3:GetObject"
                        - "s3:ListBucket"
                        - "s3:DeleteObject"
                        - "s3:GetBucketLocation"
                        - "s3:PutObjectAcl"
                    Resource:
                        - !Sub 'arn:aws:s3:::${DestBucketName}'
                        - !Sub 'arn:aws:s3:::${DestBucketName}/*'
      Roles:
      - !Ref GPXDataRole

  GPXDataRole:
      Type: AWS::IAM::Role
      Properties:
            RoleName: !Sub 'gpx-data-${ClientName}-role'
            AssumeRolePolicyDocument:
                    Version: "2012-10-17"
                    Statement:
                    - Effect: "Allow"
                        Principal:
                            AWS: !Sub
            'arn:aws:iam::457748205563:user/gpx-data-${ClientName}-user'
                        Action:
                            - "sts:AssumeRole"
                        Condition:
                            StringEquals:
                                sts:ExternalId: !Sub
            'gpx-data-${ClientName}-assume-id'
            MaxSessionDuration: 43200
            Path: "/"

Outputs:
  RoleARN:
      Description: "ARN of the created IAM role"
      Value: !GetAtt GPXDataRole.Arn
```