Backend Coding Task

Context

AWS maintains documentation about each of its services — the actions those services support, the resources and conditions that are applicable for those actions etc. At Symmetry, we use this information to validate and check if a particular policy is well formed etc.

However, this information is not provided by the AWS in any structured manner other than the basic HTML pages and tables in the documentation liked above.

Problem Statement

Your task is to write a Python script that can crawl all the service pages in the documentation (links to each service are listed at root of documentation) and extract the data from those pages to produce json files for each service. A sample JSON produced for *airflow* service is shown in the appendix at the end of this document.

The elements to be parsed from the page are:

1. prefix — AWS service prefix. (In the Fig. 2 below, it is “airflow” mentioned in the first line of the service documentation.)
2. link — link of the documentation page
3. actions — an array of actions supported by this service. This data comes from the actions table as seen in the first table in Fig. 3 below. Each row in the action becomes one struct entry of this array. See the sample json at the end.
4. resources — an array of resources supported by this service. This data comes from the resources table as seen in the second table in Fig. 3 below. See the sample json at the end.
5. conditions — an array of conditions supported by this service. This data comes from the conditions table as in the third table in Fig. 3 below. See the sample json at the end.

Please feel free to ask any question in case of a doubt.

Please see the Appendix on the next page.

Appendix.

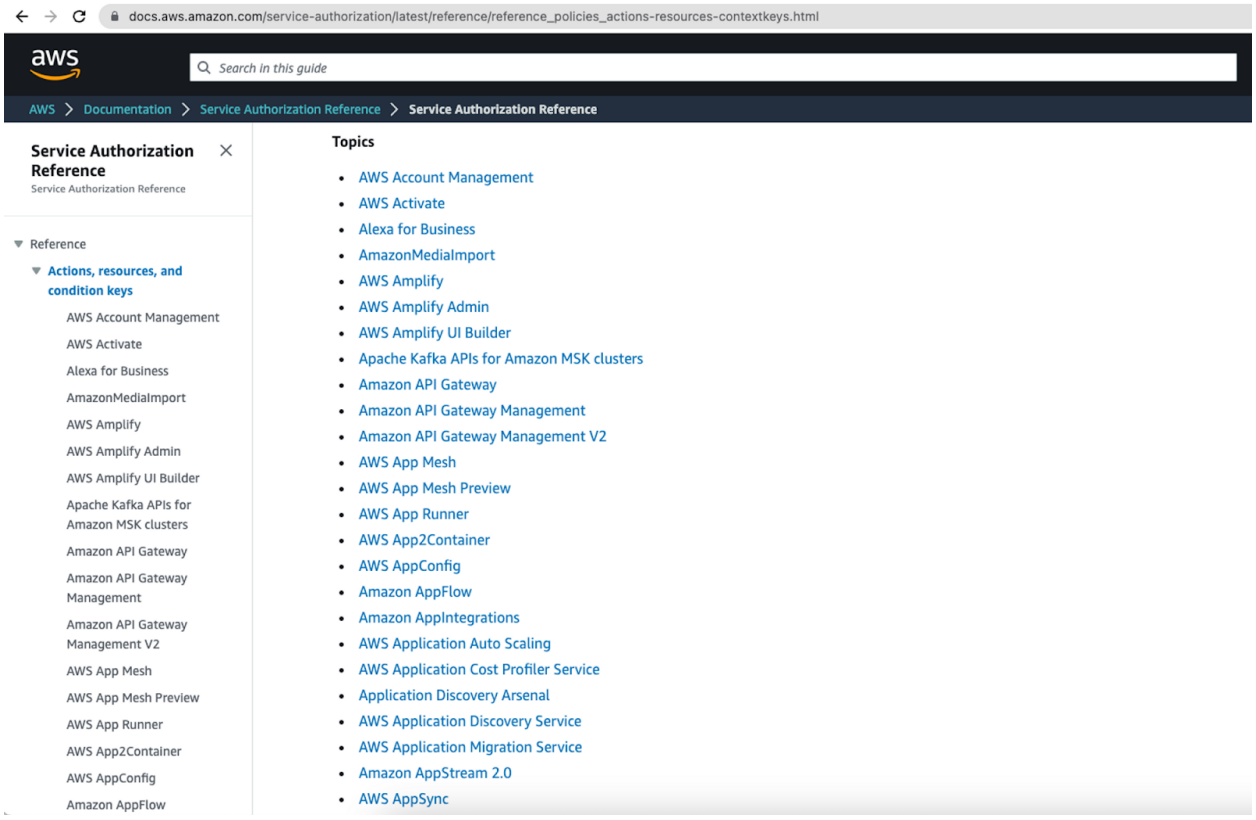


Fig. 1 List of services at the root of documentation along with their links

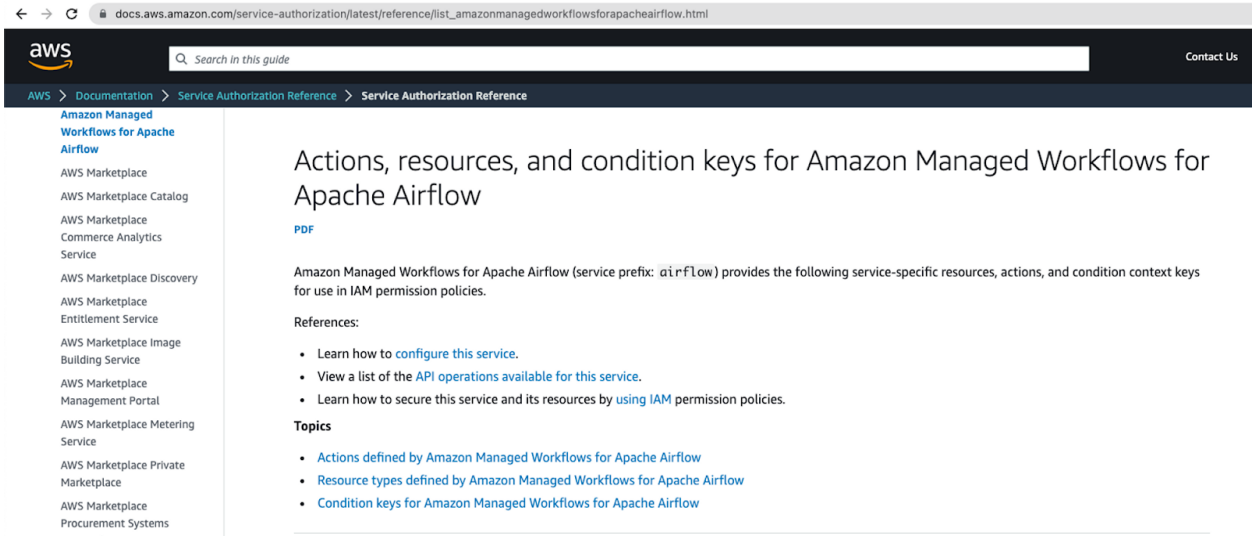
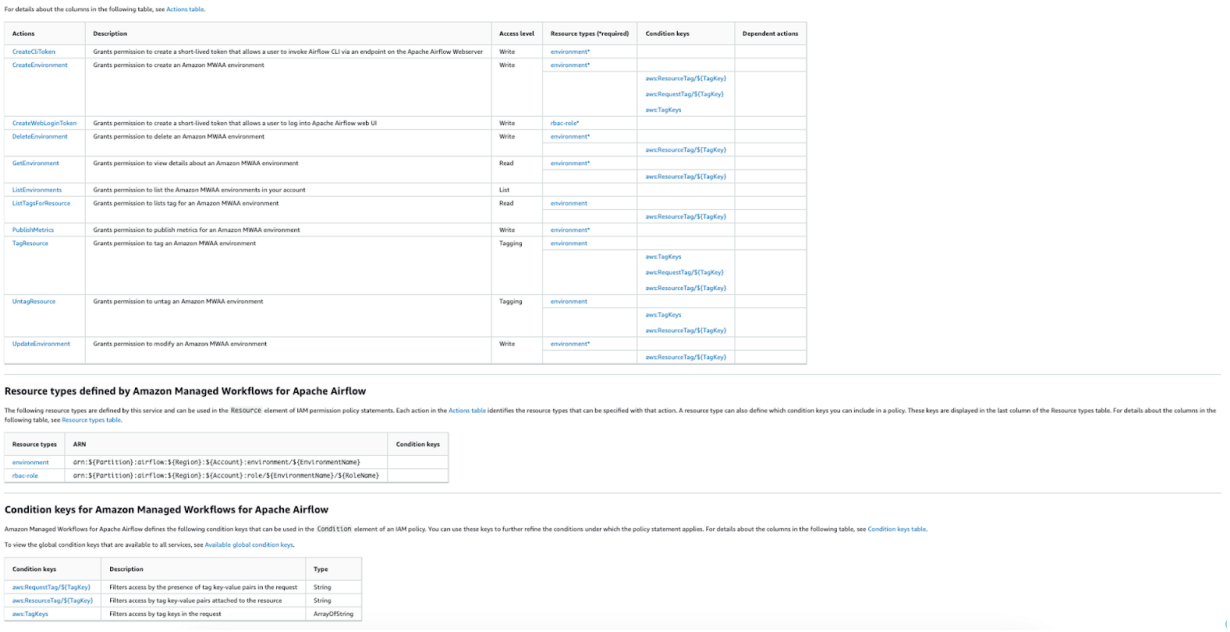


Fig. 2 Airflow service specific page after clicking link in the root document



For details about the columns in the following table, see Actions table.

Actions	Description	Access level	Resource types (*required)	Condition keys	Dependent actions
CreateCliToken	Grants permission to create a short-lived token that allows a user to invoke Airflow CLI via an endpoint on the Apache Airflow Webserver	Write	environment*		
CreateEnvironment	Grants permission to create an Amazon MWAA environment	Write	environment*		
				aws:ResourceTag/${TagKey} aws:RequestTag/${TagKey} aws:TagKeys	
CreateWebLoginToken	Grants permission to create a short-lived token that allows a user to log into Apache Airflow web UI	Write	rbac-role*		
DeleteEnvironment	Grants permission to delete an Amazon MWAA environment	Write	environment*		
				aws:ResourceTag/${TagKey}	
GetEnvironment	Grants permission to view details about an Amazon MWAA environment	Read	environment*		
				aws:ResourceTag/${TagKey}	
ListEnvironments	Grants permission to list the Amazon MWAA environments in your account	List			
ListTagsForResource	Grants permission to lists tag for an Amazon MWAA environment	Read	environment		
				aws:ResourceTag/${TagKey}	
PublishMetrics	Grants permission to publish metrics for an Amazon MWAA environment	Write	environment*		
TagResource	Grants permission to tag an Amazon MWAA environment	Tagging	environment		
				aws:TagKeys aws:RequestTag/${TagKey} aws:ResourceTag/${TagKey}	
UntagResource	Grants permission to untag an Amazon MWAA environment	Tagging	environment		
				aws:TagKeys aws:ResourceTag/${TagKey}	
UpdateEnvironment	Grants permission to modify an Amazon MWAA environment	Write	environment*		
				aws:ResourceTag/${TagKey}	

Resource types defined by Amazon Managed Workflows for Apache Airflow

The following resource types are defined by this service and can be used in the `Resource` element of IAM permission policy statements. Each action in the Actions table identifies the resource types that can be specified with that action. A resource type can also define which condition keys you can include in a policy. These keys are displayed in the last column of the Resource types table. For details about the columns in the following table, see Resource types table.

Resource types	ARN	Condition keys
environment	arn:${Partition}:airflow:${Region}:${Account}:environment/${EnvironmentName}	
rbac-role	arn:${Partition}:airflow:${Region}:${Account}:role/${EnvironmentName}/${RoleName}	

Condition keys for Amazon Managed Workflows for Apache Airflow

Amazon Managed Workflows for Apache Airflow defines the following condition keys that can be used in the `Condition` element of an IAM policy. You can use these keys to further refine the conditions under which the policy statement applies. For details about the columns in the following table, see Condition keys table.

To view the global condition keys that are available to all services, see Available global condition keys.

Condition keys	Description	Type
aws:RequestTag/${TagKey}	Filters access by the presence of tag key-value pairs in the request	String
aws:ResourceTag/${TagKey}	Filters access by tag key-value pairs attached to the resource	String
aws:TagKeys	Filters access by tag keys in the request	ArrayOfString

Fig 3. Tabular data on the service page that needs to be read and parsed. Actions, resources, and conditions data in tabular format show here. This is the data that needs to be parsed and stored as json.

JavaScript

```
{
  "prefix": "airflow",
  "link": [
"https://docs.aws.amazon.com/service-authorization/latest/referen
ce/list_amazonmanagedworkflowsforapacheairflow.html"
  ],
  "actions": [
    {
      "action": "CreateCliToken",
      "access": "Write",
      "resources": [
        "environment*"
      ],
      "conditionKeys": [],
      "dependentActions": [],
```

```
      "desc": "Grants permission to create a short-lived token
that allows a user to invoke Airflow CLI via an endpoint on the
Apache Airflow Webserver"
    },
    {
      "action": "CreateEnvironment",
      "access": "Write",
      "resources": [
        "environment*"
      ],
      "conditionKeys": [],
      "dependentActions": [],
      "desc": "Grants permission to create an Amazon MWAA
environment"
    },
    {
      "action": "CreateEnvironment",
      "access": "Write",
      "resources": [],
      "conditionKeys": [
        "aws:ResourceTag/${TagKey}",
        "aws:RequestTag/${TagKey}",
        "aws:TagKeys"
      ],
      "dependentActions": [],
      "desc": "Grants permission to create an Amazon MWAA
environment"
    },
    {
      "action": "CreateWebLoginToken",
      "access": "Write",
      "resources": [
        "rbac-role*"
      ],
      "conditionKeys": [],
      "dependentActions": [],
```

```
      "desc": "Grants permission to create a short-lived token
that allows a user to log into Apache Airflow web UI"
    },
    {
      "action": "DeleteEnvironment",
      "access": "Write",
      "resources": [
        "environment*"
      ],
      "conditionKeys": [],
      "dependentActions": [],
      "desc": "Grants permission to delete an Amazon MWAA
environment"
    },
    {
      "action": "DeleteEnvironment",
      "access": "Write",
      "resources": [],
      "conditionKeys": [
        "aws:ResourceTag/${TagKey}"
      ],
      "dependentActions": [],
      "desc": "Grants permission to delete an Amazon MWAA
environment"
    },
    {
      "action": "GetEnvironment",
      "access": "Read",
      "resources": [
        "environment*"
      ],
      "conditionKeys": [],
      "dependentActions": [],
      "desc": "Grants permission to view details about an Amazon
MWAA environment"
    },
```

```
    {
      "action": "GetEnvironment",
      "access": "Read",
      "resources": [],
      "conditionKeys": [
        "aws:ResourceTag/${TagKey}"
      ],
      "dependentActions": [],
      "desc": "Grants permission to view details about an Amazon
MWAA environment"
    },
    {
      "action": "ListEnvironments",
      "access": "List",
      "resources": [],
      "conditionKeys": [],
      "dependentActions": [],
      "desc": "Grants permission to list the Amazon MWAA
environments in your account"
    },
    {
      "action": "ListTagsForResource",
      "access": "Read",
      "resources": [
        "environment"
      ],
      "conditionKeys": [],
      "dependentActions": [],
      "desc": "Grants permission to lists tag for an Amazon MWAA
environment"
    },
    {
      "action": "ListTagsForResource",
      "access": "Read",
      "resources": [],
      "conditionKeys": [
```

```
        "aws:ResourceTag/${TagKey}"
      ],
      "dependentActions": [],
      "desc": "Grants permission to lists tag for an Amazon MWAA
environment"
    },
    {
      "action": "PublishMetrics",
      "access": "Write",
      "resources": [
        "environment*"
      ],
      "conditionKeys": [],
      "dependentActions": [],
      "desc": "Grants permission to publish metrics for an Amazon
MWAA environment"
    },
    {
      "action": "TagResource",
      "access": "Tagging",
      "resources": [
        "environment"
      ],
      "conditionKeys": [],
      "dependentActions": [],
      "desc": "Grants permission to tag an Amazon MWAA
environment"
    },
    {
      "action": "TagResource",
      "access": "Tagging",
      "resources": [],
      "conditionKeys": [
        "aws:TagKeys",
        "aws:RequestTag/${TagKey}",
        "aws:ResourceTag/${TagKey}"
```

```
      ],
      "dependentActions": [],
      "desc": "Grants permission to tag an Amazon MWAA
environment"
    },
    {
      "action": "UntagResource",
      "access": "Tagging",
      "resources": [
        "environment"
      ],
      "conditionKeys": [],
      "dependentActions": [],
      "desc": "Grants permission to untag an Amazon MWAA
environment"
    },
    {
      "action": "UntagResource",
      "access": "Tagging",
      "resources": [],
      "conditionKeys": [
        "aws:TagKeys",
        "aws:ResourceTag/${TagKey}"
      ],
      "dependentActions": [],
      "desc": "Grants permission to untag an Amazon MWAA
environment"
    },
    {
      "action": "UpdateEnvironment",
      "access": "Write",
      "resources": [
        "environment*"
      ],
      "conditionKeys": [],
      "dependentActions": [],
```

```
      "desc": "Grants permission to modify an Amazon MWAA
environment"
    },
    {
      "action": "UpdateEnvironment",
      "access": "Write",
      "resources": [],
      "conditionKeys": [
        "aws:ResourceTag/${TagKey}"
      ],
      "dependentActions": [],
      "desc": "Grants permission to modify an Amazon MWAA
environment"
    }
  ],
  "resources": [
    {
      "resourceType": "environment",
      "arn":
"arn:${Partition}:airflow:${Region}:${Account}:environment/${Envi
ronmentName}",
      "conditionKeys": []
    },
    {
      "resourceType": "rbac-role",
      "arn":
"arn:${Partition}:airflow:${Region}:${Account}:role/${Environment
Name}/${RoleName}",
      "conditionKeys": []
    }
  ],
  "conditions": [
    {
      "conditionKey": "aws:RequestTag/${TagKey}",
      "desc": "Filters access by the presence of tag key-value
pairs in the request",
```

```
      "typ": "String"
    },
    {
      "conditionKey": "aws:ResourceTag/${TagKey}",
      "desc": "Filters access by tag key-value pairs attached to
the resource",
      "typ": "String"
    },
    {
      "conditionKey": "aws:TagKeys",
      "desc": "Filters access by tag keys in the request",
      "typ": "ArrayOfString"
    }
  ]
}
```


Fig 4. Sample JSON file generated by parsing the data present in tables on the airflow service page.