**DMSTaskCopy Script Read Me Document**

1. Setup Instructions
   1. Install Python version 3.6 or above
   2. Verify python is installed
      1. python3 –version (or) python --version
   3. Extract the DMSTaskAutomation.zip file to your desired working directory.
   4. Modify the configuration file – dms\_config.json with the source and destination account details.
   5. Setup a virtual environment to run the tool.
      1. python -m venv venv – use python or python3 as installed on your system
      2. source venv/bin/activate.sh – for Mac or Linux
      3. venv/Scripts/activate.bat – for windows
   6. Set the service account password as environment variable for AD authentication. Please note that is optional and only required if the ‘ad\_authentication’ flag is set to true.

set PASSWORD=<password>

* 1. Run the script from within the service account as follows:

python DMSTaskCopy.py -m <option> [-src\_ep\_passwd <password>] [-tgt\_ep\_passwd <password>]

(or)

Python DMSTaskCopy.py --mode <option> [--src\_ep\_passwd <password>] –[tgt\_ep\_passwd <password>]

where option can be any one of the following:-

* import = use the downloaded the tablemappings json and replication task settings json files from the ‘data’ sub directory
* export = download the tablemappings json and replication task settings json files from the source account
* The src\_ep\_passwd and tgt\_ep\_passwd arguments are optional and only required if the config file cannot be used to store them.

The script will prompt the user if a cleanup of resources is required. It is assumed that the script is run for a second time, it will cleanup the resources used by it based on the replication tasks in the target account.

Sample Prompts from Script:

*Do you want to perform cleanup of resources before starting? (y/n) : n*

The script also prompts if you want to start the replication task. If ‘n’ is entered, it does not start the replication task which remains in ready state.

*Do you want to start the replication task - fullload, (y/n)?y*

1. Python Module dependencies
   1. pip install boto3
   2. pip install requests

pip install -r requirements.txt should install all the required dependencies.

Note: It was observed that for some users the pip install commands fail with CERTIFICATE VERIFY errors. In that case, please use the below commands:

$ pip install --trusted-host pypi.org --trusted-host files.pythonhosted.org pip boto3

$ pip install --trusted-host pypi.org --trusted-host files.pythonhosted.org pip requests

$ pip install --trusted-host pypi.org --trusted-host files.pythonhosted.org –upgrade pip

1. Configuration Settings

The available configuration file options are shown in the below tables. You can also refer to the Appendix section which presents the options in a JSON format (as available on the dms\_config.json file) with embedded descriptions/comments explaining each option in detail. Please do not copy the Appendix section as such into your dms\_config.json file as JSON does not support comments.

**Source Account Settings**

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Value | Optional | Description |
| src\_account\_id | AWS account ID | No | Source account ID from which AWS DMS  tasks must be copied |
| src\_account\_role | AWS account role | No | Predefined IAM user role such as admin or devops |
| src\_account\_region | AWS Region | No | The region where the resources are hosted. |
| rep\_task\_tag\_filters | Tag | No | AWS tag defined on AWS resources to be copied in the source account |

**Target Account Settings**

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Value | Optional | Description |
| dest\_account\_id | AWS account ID | No | The destination account ID |
| dest\_account\_role | AWS account role | No | The destination account role |
| dest\_account\_region | AWS Region | No | The region where the resources are hosted. |
| dest\_environment | Alphanumeric | No | Any unique identifier for the target environment to which AWS DMS tasks are copied or promoted such as qa, uat, or prod |
| dms\_task\_import\_export\_subdir | Folder name | No | A subfolder that the script creates to download the AWS DMS task table mapping and task setting json files |

**Generic Settings**

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Value | Optional | Description |
| ad\_username | AD login | No | The Active Directory user name. The password must be defined as an environment variable. |
| dms\_task\_import\_export\_subdir | Folder name | No | A subfolder that the script creates to download the AWS DMS task table mapping and setting text files. |
| promt\_start\_dms\_tasks | True/False | No | Setting to true causes the script to prompt and wait for user input and confirmation before starting each AWS DMS task. |
| auto\_start\_dms\_tasks\_on\_creation | True/False | No | Setting to true automatically starts all AWS DMS tasks after creation without waiting for user confirmation. |
| endpoints\_separate\_elem\_approach | False | No | This feature is still in testing and requires further work. Please ignore for now. |
| ad\_authentication | True/False | No | Toggles between AD-based authentication or IAM. The default is False. If set to True, the identity\_service\_url must be set to point to the AD-based authentication service REST endpoint. |
| identity\_service\_url | URL | Yes | This is optional if ‘ad\_authentication’ is set to False. Otherwise, it’s a mandatory option and needs to be pointed to the REST URL endpoint for the identity service based on AD Authentication. |
| cert\_verify\_flag | True/False | No | This is mandatory if the ‘identity\_service\_url’ REST API requires certificate verification during POST requests. |
| sts\_src\_role\_arn | ARN | Yes | IAM role ARN from the source account. Required only when ad\_authentication is set to False. |
| sts\_tgt\_role\_arn | ARN | Yes | IAM role ARN from the target account. Required only when ad\_authentication is set to False. |

**Endpoints Settings**

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Value** | **Optional** | **Description** |
| endpoints | JSON Array | No | This array relates to endpoints configuration. |
| endpoints.endpoint\_arn | ARN | Yes | Target AWS account endpoint ARN if its already created. This requires the endpoint\_identifier field to be defined at a minimum. |
| endpoints.endpoint\_identifier | String | Yes | The endpoint\_identifier must exactly match the source AWS account endpoint name when defined. Note: This is not the endpoint identifier name in the target AWS account. The endpoint name will be constructed from the database name with a suffix containing the dest\_environment value and a numerical value in the format shown here: <databasename>-<dest\_environment>-1 |
| endpoints.endpoint\_identifier\_new | String | Yes | This value will be used as the endpoint name when creating it on the target AWS account. This field requires the endpoint\_identifier to be defined. |
| endpoints.endpoint\_identifier\_special\_chars | Chars | No | This field can hold special characters that are not allowed in endpoint names but are part of the endpoint database name. Eg: '\_' or '$'. These values if found in the database name will be replaced with '-' and used for creating the endpoint identifier name. |
| endpoints.endpoint\_type | String | No | ‘source' or 'target' |
| endpoints.engine\_name | String | Yes | aurora-postgresql, sqlserver, etc. All values defined in the AWS documentation for engine names are acceptable. |
| endpoints.user\_name | String | Yes | The username for the endpoint database. |
| endpoints.password | String | Yes | The password for the endpoint database. This is optional and can be supplied as a command line argument. |
| endpoints.server\_name | String | Yes | The host name for the endpoint database. |
| endpoints.database\_port | Integer | Yes | The port number for the endpoint database. |
| endpoints.database\_name | String | Yes | The database name for the endpoint. |
| endpoints.extra\_conn\_attr | String | Yes | The extra connection attributes for endpoint. All values as defined in AWS documentation are acceptable. |

**Replication Instance Settings**

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Value** | **Optional** | **Description** |
| rep\_instances | JSON Array | No | This array relates to replication instance configuration. Each replication instance whose class or other attributes needs to be overridden in the target AWS account can be defined in this section. |
| rep\_instances.rep\_instance\_arn | String | Yes | The target AWS account replication instance arn. |
| rep\_instances.rep\_instance\_identifier | String | Yes | The source account replication instance identifier whose attributes must be overrridden in target AWS account. |
| rep\_instanes.rep\_instance\_identifier\_new | String | Yes | The replication instance name to be used for instance creation. |
| rep\_instances.allocated\_storage | String | Yes | The replication instance allocated storage in GB. |
| rep\_instances.rep\_instance\_class | String | Yes | All values as defined in AWS documentation are allowed. |
| rep\_instances.rep\_engine\_version | String | Yes | All values as defined in AWS documentation are allowed. |
| rep\_instances.vpc\_security\_group\_ids | String | Yes | All values as defined in AWS documentation are allowed. |
| rep\_instances.availability\_zone | String | Yes | All values as defined in AWS documentation are allowed. |
| rep\_instances.rep\_subnet\_grp\_identifier | String | Yes | All values as defined in AWS documentation are allowed. |
| rep\_instances.preferred\_maint\_window | String | Yes | All values as defined in AWS documentation are allowed. |
| rep\_instances.multi\_az\_bool | Boolean | Yes | All values as defined in AWS documentation are allowed. |
| rep\_instances.publicly\_accessible\_bool | Boolean | Yes | All values as defined in AWS documentation are allowed. |

**APPENDIX:**

The configuration file follows the json format and has the following structure defined. The description for each attribute and its value is explained in detail and embedded in the below json layout wherever appropriate:

Note: You can refer to the Appendix section which has all these options in an tabular column format for quick reference if required.

{

# The ad service account user name

"ad\_username": "",

# The source account identifier from which tasks will be copied over

"src\_account\_id": "",

# The source account user role

"src\_account\_role": "",

# The source account region

"src\_account\_region": "us-east-1",

# Whether to authenticate to AD or using IAM based cross account roles. When set to false, the ‘identity\_service\_url’ is ignored and ‘sts\_src\_role\_arn’, ‘sts\_tgt\_role\_arn’ are mandatory fields. Whereas, when this is set to true, the identity\_service\_url and cert\_verify\_flag are mandatory fields.

"ad\_authentication": false,

# The AD identity service exposed as a REST URL. Users authenticate using AD credentials and receive AWS STS tokens.

"identity\_service\_url": <https://api.aic-identity-ssvc.xyz.com/creds>

# The source account user role ARN where ‘111111111111’ must be replaced with AWS source account ID.

"sts\_src\_role\_arn": "arn:aws:iam::111111111111:role/MyAssumeRole"

# The target account user role ARN where ‘222222222222’ must be replaced with AWS source account ID.

"sts\_tgt\_role\_arn": "arn:aws:iam::222222222222:role/MyAssumeRole"

# The ‘identity\_service\_url’ verification flag which is passed into the AD REST API for authentication.

"cert\_verify\_flag": false

# The tag value used to filter the replication tasks and endpoints. The replication instances are mapped to tasks and will be automatically selected based on the selection of appropriate tasks. Note that the tag and its value are case sensitive.

"rep\_task\_tag\_filters": [

{

"Key": "DMS\_TASK\_ENV",

"Values": [

"DEV"

]

}

],

# The target account identifier where the tasks will be copied to

"dest\_account\_id": "",

# The target account user role

"dest\_account\_role": "",

# The target environment or staging region name. This will be appended to the endpoints created on target account like “<endpoint\_dbname>-qa”.

"dest\_environment": "",

#This field defines the folder name (within the current working directory of the script) where the ‘tablemappings’ or ‘replicationtasksettings’ json files shall be stored.

"dms\_task\_import\_export\_subdir": "",

#If this field is set to true, the user prompts for the replication tasks to be started shall be presented.

"promt\_start\_dms\_tasks": "false",

#If this field is set to true, all the created replication tasks will be auto started. This field requires the ‘prompt\_start\_dms-tasks’ set to false to auto start the replication tasks.

"auto\_start\_dms\_tasks\_on\_creation": "false",

#Endpoints if separate element is defined for every single available one from source AWS account. This is not fully developed and is still in testing phase. Please ignore this option for now.

"endpoints\_separate\_elem\_approach": "false",

# The endpoints array. Any number of endpoints can be defined here. Each endpoint defined here must tie up to a source account endpoint. The endpoint\_identifier defined must match with the corresponding source account endpoint identifier value. Also, the endpoint password is required for creation. If the endpoint\_arn is defined, then all the other attributes are optional except for the endpoint\_identifier.

"endpoints": [

{

# The endpoint\_arn of an existing endpoint in target account. Optional field. If left blank, the remaining fields are used to create a new endpoint on the target account.

"endpoint\_arn": "",

#The endpoint identifier from the source account. This field is mandatory. Note, the script will create a new endpoint with identifier name as “<endpoint\_database\_name>-<dest\_environment>”. The endpoint\_database\_name is defined in this block and the dest\_environment is defined in the above portion of this json configuration file. This field is optional.

"endpoint\_identifier": "",

#The actual endpoint identifier that shall be used for creating the endpoint. This optional field shall override the ‘endpoint\_identifier’ defined above.

"endpoint\_identifier\_new": "",

#The special characters list which shall be replaced with a ‘-‘ (hyphen) in the endpoint\_identifier field.

"endpoint\_identifier\_special\_chars": "\_$",

#The endpoint type. This field is **Mandatory**.

"endpoint\_type": "source",

#The endpoint engine name

"engine\_name": "sqlserver",

#The endpoint user name. This field is **Mandatory**.

"user\_name": "",

#The endpoint password. This is optional here but needs to be supplied as an argument at run time.

"password": "",

#The endpoint server\_name

"server\_name": "",

#The endpoint database\_port

"database\_port": "1433",

#The endpoint database\_name. This field is optional.

"database\_name": "",

#The endpoint extra connection attributes if needed. This is optional.

"extra\_conn\_attr": ""

},

{

# The endpoint\_arn of an existing endpoint in target account. Optional field. If left blank, the remaining fields are used to create a new endpoint on the target account.

"endpoint\_arn": "",

#The endpoint identifier from the source account. This field is mandatory. Note, the script will create a new endpoint with identifier name as “<endpoint\_database\_name>-<dest\_environment>”. The endpoint\_database\_name is defined in this block and the dest\_environment is defined in the above portion of this json configuration file. . This field is **optional**.

"endpoint\_identifier": "",

#The actual endpoint identifier that shall be used for creating the endpoint. This optional field shall override the ‘endpoint\_identifier’ defined above.

"endpoint\_identifier\_new": "",

#The special characters list which shall be replaced with a ‘-‘ (hyphen) in the endpoint\_identifier field.

"endpoint\_identifier\_special\_chars": "\_$",

#The endpoint type. This field is **Mandatory**.

"endpoint\_type": "target",

#The endpoint engine name

"engine\_name": "aurora-postgresql",

#The endpoint user name. This field is **Mandatory**.

"user\_name": "",

#The endpoint password. This is optional here but needs to be supplied as an argument at run time.

"password": "",

#The endpoint server\_name

"server\_name": "",

#The endpoint database port

"database\_port": "5432",

#The endpoint database name

"database\_name": "",

#The endpoint extra connection attributes if needed. This is optional.

"extra\_conn\_attr": ""

}

],

#This block describes the replication instances to be created on the target account. This is also an array and any number of instances can be defined. However, the rep\_instance\_identifier must refer to a valid instance from the source account. Again, the rep\_instance\_arn and rep\_instance\_identifier are sufficient and if defined will cause the remaining attributes for that instance to be ignored.

"rep\_instances": [

{

#The replication instance arn. This is an optional field. If the replication instance already exists on the target account, the arn can be provided. However, it requires the rep\_instance\_identifier to be provided which must match an existing instance on the source account. This is required to understand which replication task to tie up this instance with.

"rep\_instance\_arn": "",

#The replication instance identifier. This is an mandatory field.

"rep\_instance\_identifier": "",

#The replication instance identifier to be used for actually creating the replication instance. This optional field will override the ‘rep\_instance\_identifier’ field above..

"rep\_instance\_identifier\_new": "",

# The storage in GB for the replication instance

"allocated\_storage": "100",

#The replication instance class type.

"rep\_instance\_class": "dms.t3.medium",

#The replication instance engine version. This is optional.

"rep\_engine\_version": "",

#The replication instance security group. This is optional. The default value will be used.

"vpc\_security\_group\_ids": [

""

],

# The replication instance availability zone. This is optional. The default value will be used.

"availability\_zone": "",

# The replication instance vpc-id. This is optional. The default value will be used.

"rep\_subnet\_grp\_identifier": "",

# The replication instance preferred maintenance window. The instance will go down during this time period for applying patches. Must be atleast 30 mins in duration.

"preferred\_maint\_window": "Sat:04:10-Sat:05:00",

# The replication instance multi zone availability flag. The default value is false. You can't set the AvailabilityZone parameter if the Multi-AZ parameter is set to true .

"multi\_az\_bool": false,

# The replication instance public IP address. The default is false. If set to true, it will be assigned a public IP address.

"publicly\_accessible\_bool": false

}

]

}