

# Databricks on AWS

Last updated: December 1, 2023

## Migration guide for cluster-scoped init scripts from DBFS to S3

For Databricks Runtime 11.2 and below, move init scripts from DBFS to S3.

Note: In cases where your init scripts are “self-contained,” i.e., DO NOT reference other files such as libraries, configuration files, or shell scripts. We highly recommend storing init scripts as [workspace files](#).

In general, you must do the following:

- 1) [Configure the clusters to authenticate](#) to the S3.
- 2) Copy all your init scripts and files referenced by the init scripts from DBFS to S3.
- 3) [If applicable] Update the init scripts to reference files on S3.
- 4) Update cluster configuration and [cluster policies](#) to reference the init scripts on S3.

- 1) Configure the clusters to authenticate to the S3 using instance profiles

Follow the instructions in our [documentation](#) to configure your cluster with an instance profile to access S3.

- 2) Copy all your init scripts and files referenced by the init scripts from DBFS to S3

Identify all objects (clusters, jobs, DLT pipelines, etc.) that use init scripts on DBFS. [Use the script detection notebook](#) to prepare a list of individual init scripts. For init scripts that reference files on DBFS (e.g. libraries, configuration files or other files), prepare the list of referenced files.

1. Create a new S3 bucket (or use an existing one) to move all the init scripts and referenced files from DBFS to S3.
2. Check the bucket permissions (IAM) and confirm that the cluster’s instance profile can access this new S3 bucket.
3. In your Databricks Notebook, use the [dbutils.fs.cp](#) command to copy files from DBFS to AWS S3 like this:

```
dbutils.fs.cp("dbfs:<path>", "s3://<bucket>/<path>")
```

### 3) Update the init scripts to point to S3 locations for files referenced in the init script

You need to modify init scripts that use referenced files to use S3 objects. Change references inside the init script from `/dbfs/path` to `s3://bucket/path`.

With the instance profile attached to the cluster, you can directly use the [AWS CLI tool](#) to work with files on AWS S3. The AWS CLI can be installed using the `pip` command. Use `aws s3 cp` command ([doc](#)) to copy files from AWS S3 to the local disk as follows:

```
#!/bin/bash

pip install awscli

aws s3 ls "s3://<bucket>/<path>"
aws s3 cp "s3://<bucket>/<path>/some-file" some-file
```

### 4) Update cluster configuration and cluster policies to reference the init scripts on S3

Change the [init script paths](#) in affected clusters, jobs, Delta Live Tables pipelines, and cluster policies to point to S3 instead of DBFS. If you use the [init scripts detection notebook](#), click links in the generated HTML tables:

- **Clusters:** edit your existing clusters' configuration: remove init scripts that use files on DBFS, and add init scripts with source "S3", providing the path to the file on AWS S3, e.g. `s3://<bucket>/path/to/init-script` ([more information](#)).

Note: If you are using [cluster policies](#), you must update both the cluster policy and the configurations of any other clusters using the policy. Cluster policy changes do not propagate to other clusters using that policy.

- **Jobs:** Edit the cluster configuration for each task that uses a dedicated job cluster and each shared job cluster. Update init script location from DBFS to AWS S3, as described above.
- **If using Cluster policies:** Open and edit the [cluster policy](#). Search for blocks like the following, where N is the item number

```
{
  "init_scripts.N.dbfs.destination": {
    "type": "fixed",
    "value": "dbfs:/FileStore/init-scripts/empty_init_script.sh"
  }
}
```

and replace the value, adjusting the file path from DBFS to S3:

```
{
  "init_scripts.N.s3.destination": {
    "type": "fixed",
    "value": "s3://<bucket>/empty_init_script.sh"
  }
}
```

- **Delta Live Tables pipelines:** Open the [pipeline settings](#), select the "JSON" tab, and in the cluster definition(s), change entries in the `init_scripts` array from DBFS to S3;

```
{
  "dbfs": {
    "destination": "dbfs:/FileStore/init-scripts/empty_init_script.sh"
  }
}
```

```
{
  "s3": {
    "destination": "s3://<bucket>/empty_init_script.sh"
  }
}
```