

# Airflow DAG Automation Setup

## Repository Directory Structure

```
<root>/
  airflow_jobs/
    <team; eg: beardo>/
      <env; stage|prod>/
        <dag python files>
  scala_jobs/
    src/
    build.sbt
    config.txt
  python_jobs/
    <python_module>/
      __init__.py needs to be present in all sub directories(empty file)
      <sub dir>/
    setup.cfg
    pyproject.toml
    config.txt
```

## Configuration

### scala\_jobs build.sbt

- Ensure no local references to jars via file:// is made in the config. If so ensure the dependent jars are built and pushed to Nexus Artifactory beforehand.
- Ensure below configuration is present in build.sbt else scala version specific jars get built which is not what we want. **Note:** After this change the built jar will be present directly inside target/ and not inside target/scala\_\*/

```
crossPaths := false
```

- Ensure when mentioning Argoid internal dependency jars double percentage(%%) is not used after groupId only single percentage(%)
  - **Wrong** - libraryDependencies += "com.argoid" %% "argoid-common" % "1.0.0"
  - **Right** - libraryDependencies += "com.argoid" % "argoid-common" % "1.0.0"

### scala\_jobs config.txt

deployment\_path=<relative deployment path; beardo/jars> Path relative to airflow jar deployment directory. Normally /data/1

### python\_jobs setup.cfg / pyproject.toml

Refer: [Python Packaging Process](#)

### python\_jobs config.txt

venv=airflow\_venv Python virtual environment where the package is to be deployed. For packages which are only used as python library dependency (ex:python-config-client), this is not required.

yum\_dependencies=python3-devel space separated list of yum packages which needs to be installed as part of installing this package and its dependencies. This is python package requiring a yum dependency for it run in code NOT dependencies which are required before building the python packages.

## Pipeline Setup

Use below pipelines as references to create new DAG automation pipelines

<https://1.infra.argoid.com:8443/view/Argoid-SaaS-Stage1/job/be-stage-argoid-saas-processing/>

<https://1.infra.argoid.com:8443/view/Argoid-SaaS-Stage1/job/be-stage-argoid-saas-models/>

The variables at start of pipeline definition need to be modified for each env and repository.