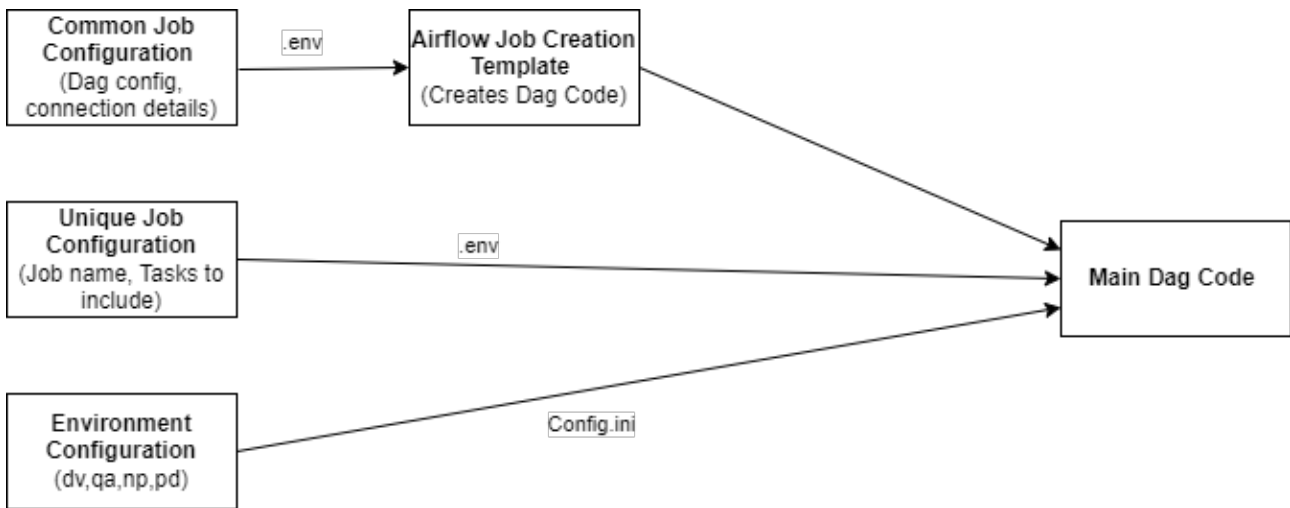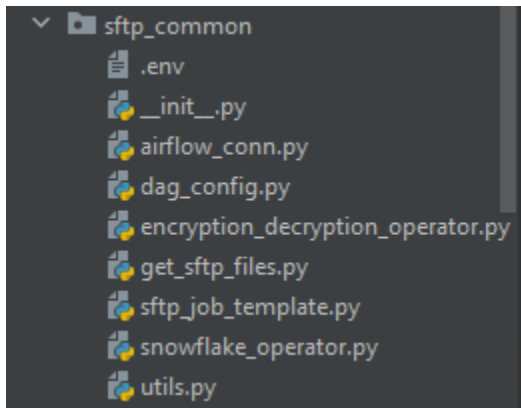# Airflow SFTP Job Common Template
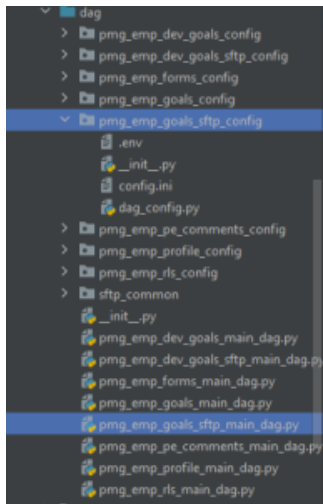
Airflow generic job Architecture

File structure of Airflow job creation template in Git

File structure of Main Dag

Naming convention:

Airflow job creation template:- sftp_common/sftp_job_template.py

Airflow job creation template configuration:- sftp_common/.env

Main Dag File:- pmg_emp_goals_main_dag.py

Main Dag configuration:- pmg_emp_goals_config/.env

Environment configuration:- pmg_emp_goals_config/config.ini

This is **Airflow job creation template** for creating new job for sftp pipeline

Github link: https://github.com/hris-data/HR-Data-Platform/tree/STRY0146446_V1

Process:-

1. **Update Airflow job creation template configuration with relevant details example given below**

   Airflow job creation template configuration (sftp_common/.env):
   ```
   #Project Configurations
   project_name='<project id>'
   location='<region>'
   zone = '<zone>'

   #Airflow job configurations
   depends_on_past = <True or False> # if this job is depends on last run
   start_date='<job start date>'
   retries= <number of retries when job get failed>
   retry_delay= <retry delay>
   execution_timeout = <execution timeout for task in minutes>
   dagrun_timeout = <total dagrun time in minutes>
   max_active_runs = <max active run for same job>
   concurrency = <concurrency of job>
   catchup = <True or False> # run job for history


   # GCS Bucket, location & Folder names
   DESTINATION_BUCKET = "<bucket name to store data>"
   DESTINATION_PATH = "<folder path for data in bucket>"
   ZIP_FILE_PATH = "<zip file path>"
   PROCESSED_FILE_PATH = "<modified File path>"

   # Source Directory
   DIR = ""
   ```

```
# Encryption configuration
ENCRYPTION_SECRET = "<encryption key id from secret>"
LMS_SCP_SECRET = "<sftp id from secret>"
LMS_SNOWFLAKE_SECRET = "<snowflake connection id from secret>"

DBT_CLOUD_CONN_ID = "<DBT connection id name>"
DBT_CLOUD_CONN_SECRET = "<dbt secret id>"
```

2. **Update Main Dag configuration**

```
#Airflow job configurations
job_name ='<job name>'
description = "<job description>"
tags = <list of tags>   #["pmgm"]

# task to be executed
IS_UNZIP = <True is run unzip task>  #False
IS_LOAD_SNOWFLAKE = <True is run task to load data in snowflake>   #True
IS_ENCRYPT_FILE = <Task to encrypt file>     #False
IS_DBT_LOAD = <Run DBT task>   #True
IS_EXCEL_CONVERT = <Convert data into CSV if excel report available>    #True

# SFTP Config

PREVIOUS_DATE = "<run job for previous day>" # initially empty
FILE_LIST = '<list of file>'  # initially empty
#Example for PREVIOUS_DATE = "2022-05-29"
#Example for FILE_LIST = '["LOREALSA01_D_20220520000600.zip", "LOREALSA01_D_20220521000533.zip",
"LOREALSA01_D_20220519000508.zip"]'


# Snowflake configuration details
FILE_FORMAT = "<snowflake file format>" #"CSV_FF"
FILE_DETAILS = <list of folder name, database name, schema name and table name>  #'[("PROD_ExportPerfGoalsActivities", "HRDP_LND_%
ENV%_DB", "PMG_LND_SCH", "PERF_GOAL_REPORT"),]'


# Encryption configuration

LMS_SNOWFLAKE_CONN = "<snowflake conn id>"    #"snowflake-conn-%env%"
LMS_SNOWFLAKE_SECRET = "<snowflake secret id>"   #"snowflake_conn_%env%"
```

3. **Update Environmental configuration**

```
[DV]
DESTINATION_BUCKET = <destination bucket name>
SNOWFLAKE_STAGE_NAME=GCS_INT_STAGE_DV
DBT_CLOUD_ACCOUNT_ID = 69742
DBT_CLOUD_JOB_ID = 130544
schedule_interval = 0 1 * * *
LMS_SCP_CONN = lms-sftp-conn
FILE_PATTERN = T3_ExportPerfGoalsActivities
SOURCE_PATH = /DBDUMP/T3_PMGM/

[QA]
DESTINATION_BUCKET = hrdp_gcs_data_eu_qa
SNOWFLAKE_STAGE_NAME=GCS_INT_STAGE_QA
DBT_CLOUD_ACCOUNT_ID = 69742
DBT_CLOUD_JOB_ID = 131204
schedule_interval = 0 1 * * *
LMS_SCP_CONN = pmg-sftp-conn
FILE_PATTERN = T3_ExportPerfGoalsActivities
SOURCE_PATH = /DBDUMP/T3_PMGM/

[NP]
DESTINATION_BUCKET = hrdp_gcs_data_eu_np
SNOWFLAKE_STAGE_NAME=GCS_INT_STAGE_NP
DBT_CLOUD_ACCOUNT_ID = 69742
DBT_CLOUD_JOB_ID = 131211
schedule_interval = 0 1 * * *
LMS_SCP_CONN = lms-sftp-conn-np
FILE_PATTERN = PROD_ExportPerfGoalsActivities
SOURCE_PATH = /DBDUMP/PMGM/
```

```
[PD]
DESTINATION_BUCKET = hrdp_gcs_data_eu_pd
SNOWFLAKE_STAGE_NAME=GCS_INT_STAGE_PD
DBT_CLOUD_ACCOUNT_ID = 69742
DBT_CLOUD_JOB_ID = 132899
schedule_interval = 0 1 * * *
LMS_SCP_CONN = lms-sftp-conn-pd
FILE_PATTERN = PROD_ExportPerfGoalsActivities
SOURCE_PATH = /DBDUMP/PMGM/
```