How to create a new staging switch (and subsequent MDM runs)

Prerequisites:

- Conf files
  - related data files need to be post-ETL
  - target conf files should be in `<BUCKET>/MDM/conf/<SOURCE>/incoming/`

Steps to create a new stg switch:

1. Run ETL
   a. Prepare the files
      i. For Salesvision, put the appropriate data files in the SV sftp site.
      ii. Run the job that will pick up these data files. You can use either:
         1. For now, FM works (python sv.py in ~edmfilemgr/edm-fm/). The rest of this page assumes the use of FM.
         2. Opswise jobs will also work if the date in the filename is either today (for LQE files) or yesterday (for FOP files). A date override feature is available but I don't know yet how to use it when having Opswise jobs kicked off by the ops team. (@ Vikram Kondadasula can we use the date override feature in the file loader yet?)
         3. When this job is done, the conf files should show in <BUCKET>/ETL/conf/incoming/
   b. Execute ETL:
      i. In test, you can run the following from in a test EMR: /usr/bin/spark-submit --num-executors 2 --jars s3://lazard-test-client-master/code/jars/postgresql-42.2.5.jar --master yarn --conf spark.pyspark.python=/usr/bin/python3.6 --py-files s3://lazard-test-client-master/code/commons/master/edm-commons.zip,s3://lazard-test-client-master/code/dpl/master/edm-dpl.zip s3://lazard-test-client-master/code/dpl/master/etl_driver.py -c s3://lazard-test-client-master/ETL/conf/incoming/<YOURCONFFILENAME> -m false -e
      ii. In production, ETL gets triggered automatically when using FM because a lambda is triggered whenever a file is placed in ETL/incoming/.
      iii. Within 5 minutes of starting the ETL job, ensure that the files moved from the 'incoming' to the 'active' folder in s3.
      iv. Also in s3, ensure that the conf files have been copied to the MDM incoming folder which is <BUCKET>/MDM/conf/SV/incoming/ . Once the new files are in the MDM incoming directory in s3, we are ready to run MDM.
2. Point to the new database switch
   a. Update `edm_params` to reflect the target data set by running the below query in Aqua Data: `update edm_params set value1 = '<desired_stg_switch>' where key = 'active_mdm_stage'`
3. Run SV MDM
   a. Launch MDM with parameter SV via your preferred entry point:
      i. Jenkins SNS push
      ii. Opswise job
   b. In test, can also run with the following command in a test EMR: aws sns publish --topic-arn arn:aws:sns:us-east-2:430815409173:mdm_lambda --message "SV" --region=us-east-2
   c. Ensure the conf files move from <BUCKET>/MDM/conf/SV/incoming/ to <BUCKET>/MDM/conf/SV/done/ . Once the files are in 'done', MDM has completed.
4. If needed, run SF MDM
   a. Request the SF entity file. This can be done by pinging the CMSF Slack channel and requesting the entity file. Please note the target environment for the SF export (either "test" or "prod").
   b. After confirmation, wait for 15 minutes and email the Ops team to run the following workflow in production: amb_edmcm_SalesForce_input. This workflow will trigger ETL for the Salesforce entity file. @ Adam Perez how can we run SF MDM in test? would this command work? aws sns publish --topic-arn arn:aws:sns:us-east-2:430815409173:mdm_lambda --message "SF" --region=us-east-2
   c. @ Adam Perez Does production workflow amb_edmcm_SalesForce_input run the full SF file data pipeline, through MDM SF, or only the SF ETL job?