To launch the **roof ingest** job, run the following command:

*./build\_and\_run\_ingest\_roof.sh*

This will build the docker container and submit the spark job to run **roof\_ingest.py**.

This PySpark job will finish in around 20 seconds and achieve the following:

* Check each JSON file within **roof\_data** is valid, fixes minor issues if applicable, and stages a copy under the new folder, **roof\_data\_fixed**
* Read in each JSON file based on the expected schema

For reference, relevant logging messages will be written to **ingest\_roof.log**.