

In this demonstration we upgrade our demonstration DAG to work on logically partitioned data. The data that we use in this lesson has been pre-partitioned in Amazon Web Services (AWS) S3 by creation date. The partition follows the format: `<year>/<month>/<day>/<file>.csv`.

In practice, it is often best to have Airflow process pre-partitioned data. If your upstream data sources cannot partition data, it is possible to write an Airflow DAG to partition the data. However, it is worth keeping in mind memory limitations on your Airflow workers. If the size of the data to be partitioned exceeds the amount of memory available on your worker, the DAG will not successfully execute.