

# --------------------------------------------------------------------------------

# Functions

# --------------------------------------------------------------------------------

def read\_table\_list(table\_list\_file):

"""

Reads the master CSV file that will help in creating Airflow tasks in

the DAG dynamically.

:param table\_list\_file: (String) The file location of the master file,

e.g. '/home/airflow/framework/master.csv'

:return master\_record\_all: (List) List of Python dictionaries containing

the information for a single row in master CSV file.

"""

master\_record\_all = []

logger.info('Reading table\_list\_file from : %s' % str(table\_list\_file))

try:

with open(table\_list\_file, 'rb') as csv\_file:

csv\_reader = csv.reader(csv\_file)

next(csv\_reader) # skip the headers

for row in csv\_reader:

logger.info(row)

master\_record = {

'table\_source': row[0],

'table\_dest': row[1]

}

master\_record\_all.append(master\_record)

return master\_record\_all

except IOError as e:

logger.error('Error opening table\_list\_file %s: ' % str(

table\_list\_file), e)

* The name of the DAG is bq\_copy\_us\_to\_eu\_01, and the DAG is not scheduled by default so needs to be triggered manually.

default\_args = {

'owner': 'airflow',

'start\_date': datetime.today(),

'depends\_on\_past': False,

'email': [''],

'email\_on\_failure': False,

'email\_on\_retry': False,

'retries': 1,

'retry\_delay': timedelta(minutes=5),

}

# DAG object.

with models.DAG('bq\_copy\_us\_to\_eu\_01',

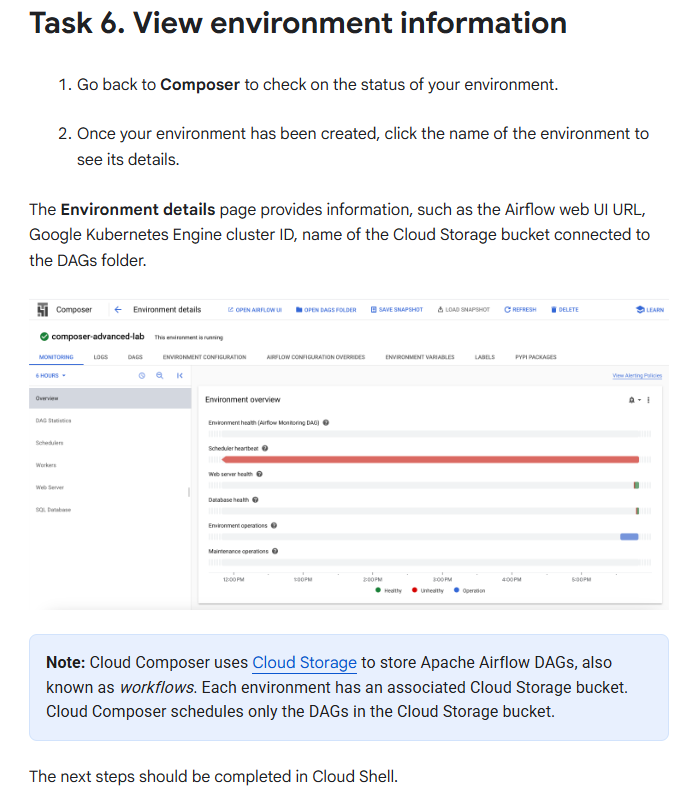
default\_args=default\_args,

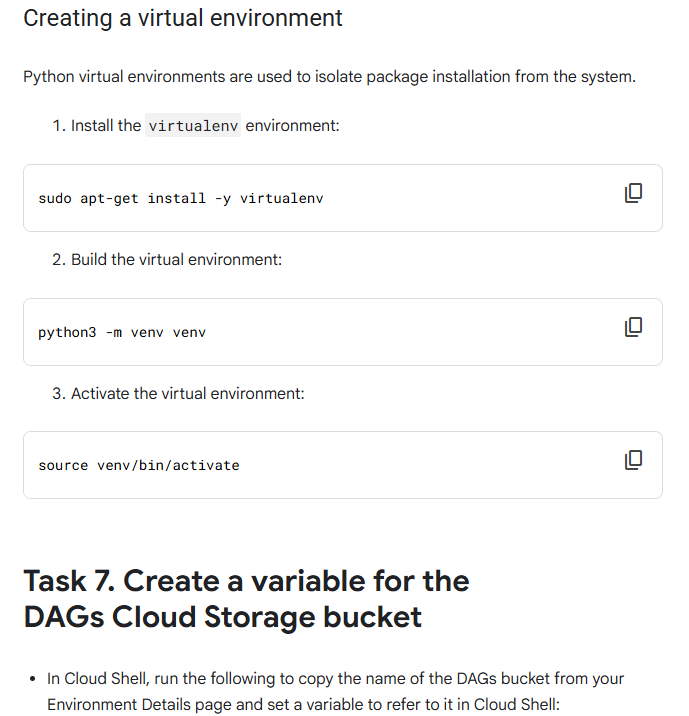
schedule\_interval=None) as dag:

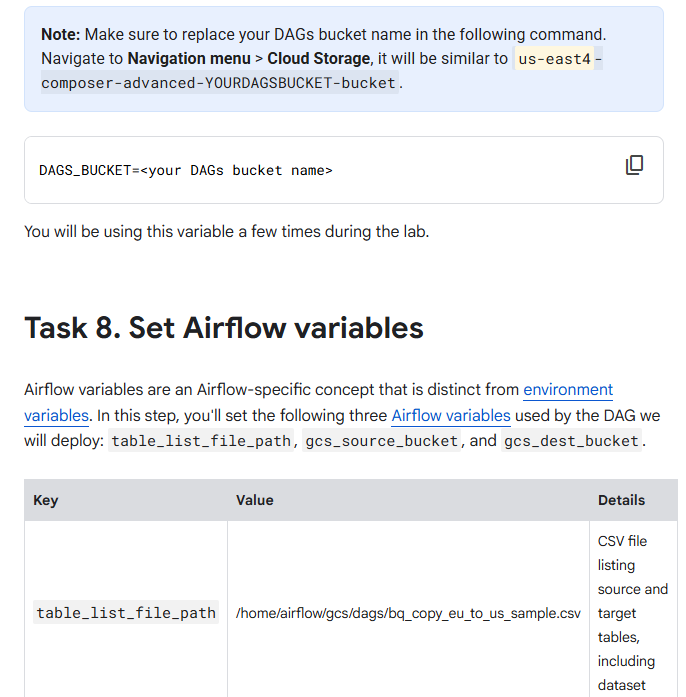
* To define the Cloud Storage plugin, the class Cloud StoragePlugin(AirflowPlugin) is defined, mapping the hook and operator downloaded from the Airflow 1.10-stable branch.

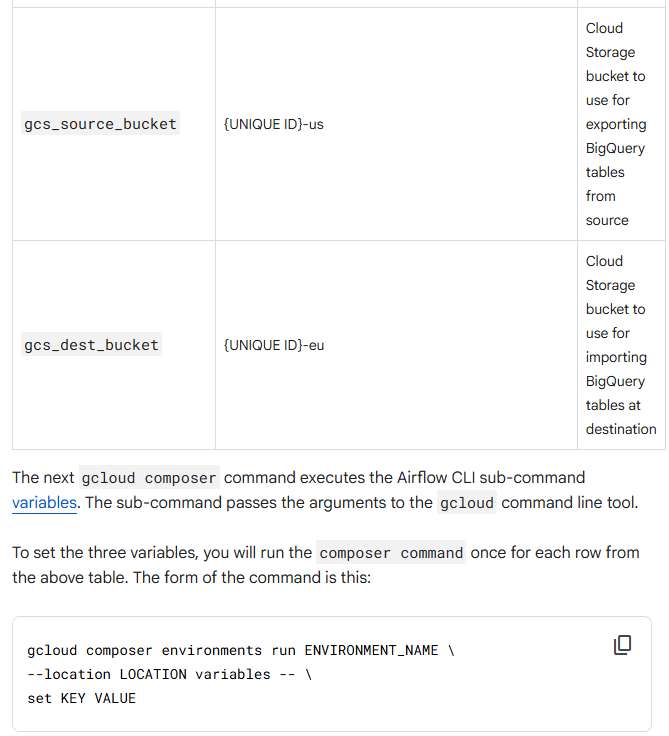
# Import operator from plugins

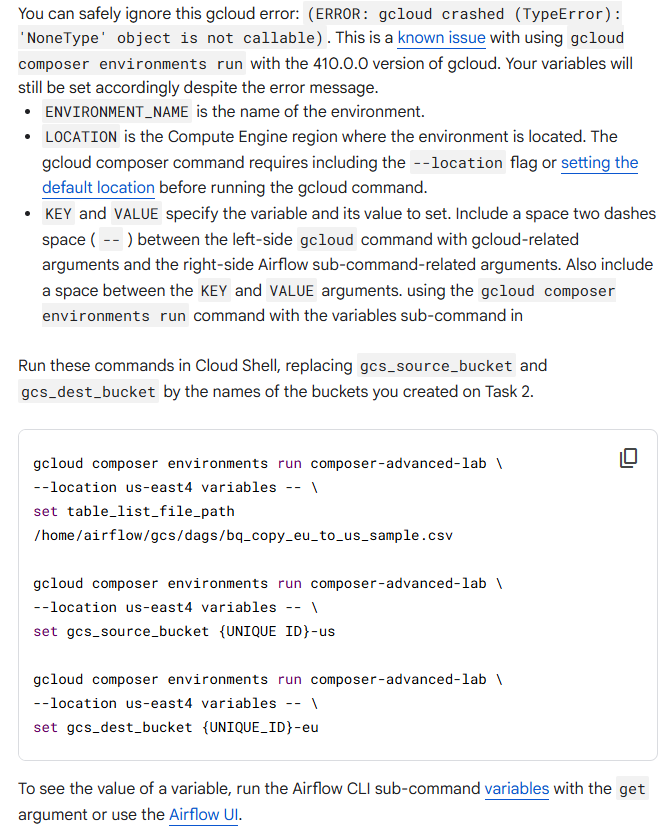
from gcs\_plugin.operators import gcs\_to\_gcs











For example, run the following:

gcloud composer environments run composer-advanced-lab \

--location us-east4 variables -- \

get gcs\_source\_bucket

**Note**: Make sure to set all three Airflow variables used by the DAG.

