# Lab - 3: Task 3

## Objective:

Read a huge file from the bucket and copy the contents to another file using Dataflow.

## Instructions:

1. Install apache-beam[gcp] : pip install apache-beam[gcp].
2. Use the following code to copy the contents of a large file to another file

```
import apache_beam as beam
from apache_beam.io import ReadFromText
from apache_beam.io import WriteToText
from apache_beam.options.pipeline_options import PipelineOptions
from apache_beam.options.pipeline_options import GoogleCloudOptions
from apache_beam.options.pipeline_options import StandardOptions
options = PipelineOptions()
google_cloud_options = options.view_as(GoogleCloudOptions)
google_cloud_options.project = '<projectid>'
google_cloud_options.job_name = '<job name>'
Google_cloud_options.region = "us-central1"
google_cloud_options.temp_location="gs://bdl2022/tmp"
options.view_as(StandardOptions).runner = 'DataflowRunner'
with beam.Pipeline(options=options)as p:
    lines = p | 'Read' >> beam.io.ReadFromText( 'gs://bdl2022/out.txt' ) | 'Write' >>
beam.io.WriteToText( '<output_path>')
```