Here is a generic sample Python script file which takes an input sentence, splits the words out on individual lines and outputs the results to a defined storage account. This script, wordcount.py can be called from the Spark activity.

import sys

from operator import add

from pyspark.sql import SparkSession

def main():

spark = SparkSession\

.builder\

.appName("PythonWordCount")\

.getOrCreate()

lines = spark.read.text("wasbs://adftutorial@<storageaccountname>.blob.core.windows.net/spark/inputfiles/minecraftstory.txt").rdd.map(lambda r: r[0])

counts = lines.flatMap(lambda x: x.split(' ')) \

.map(lambda x: (x, 1)) \

.reduceByKey(add)

counts.saveAsTextFile("wasbs://adftutorial@<storageaccountname>.blob.core.windows.net/spark/outputfiles/wordcount")

spark.stop()

if \_\_name\_\_ == "\_\_main\_\_":

main()