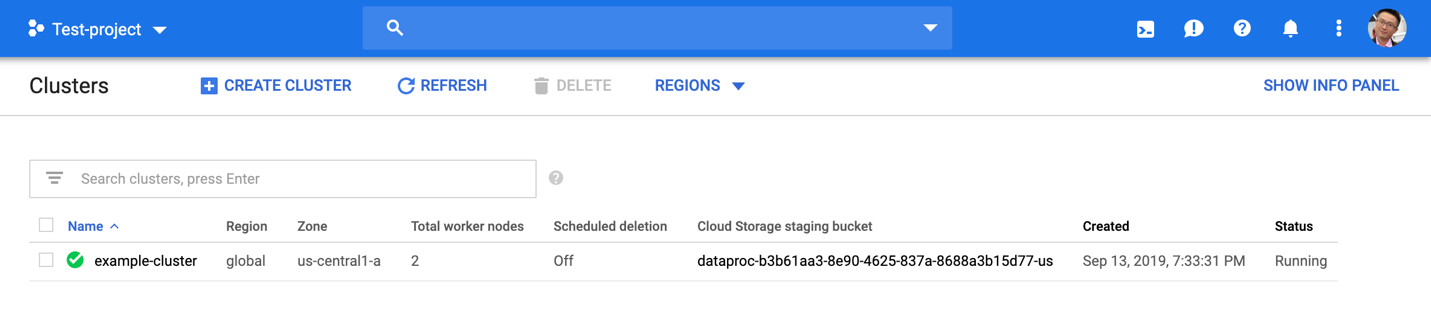
Homework0

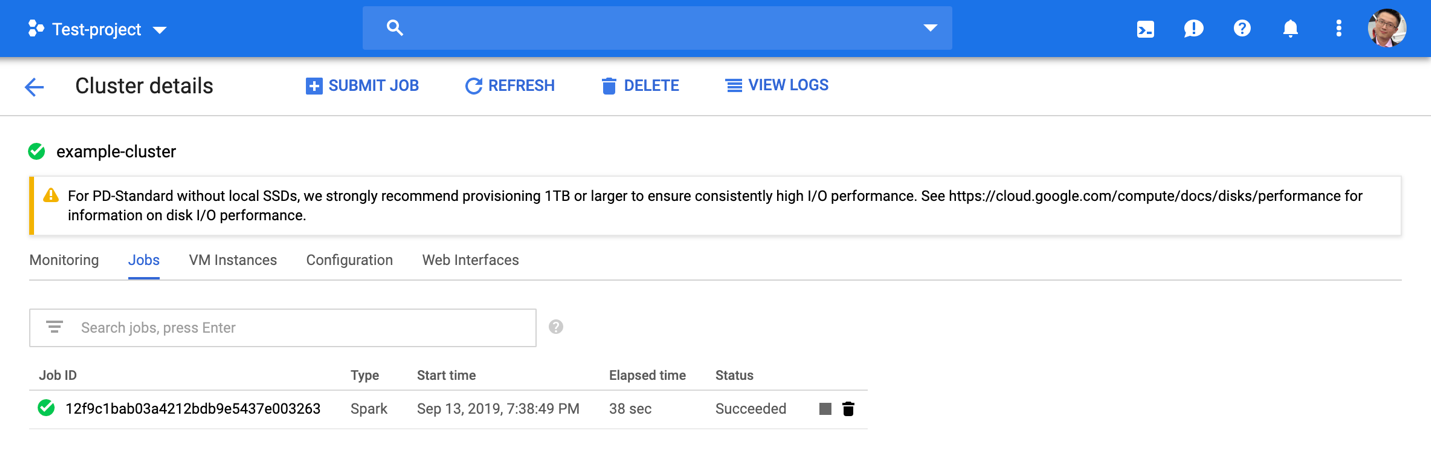
1. Warmup

(1) pi calculation

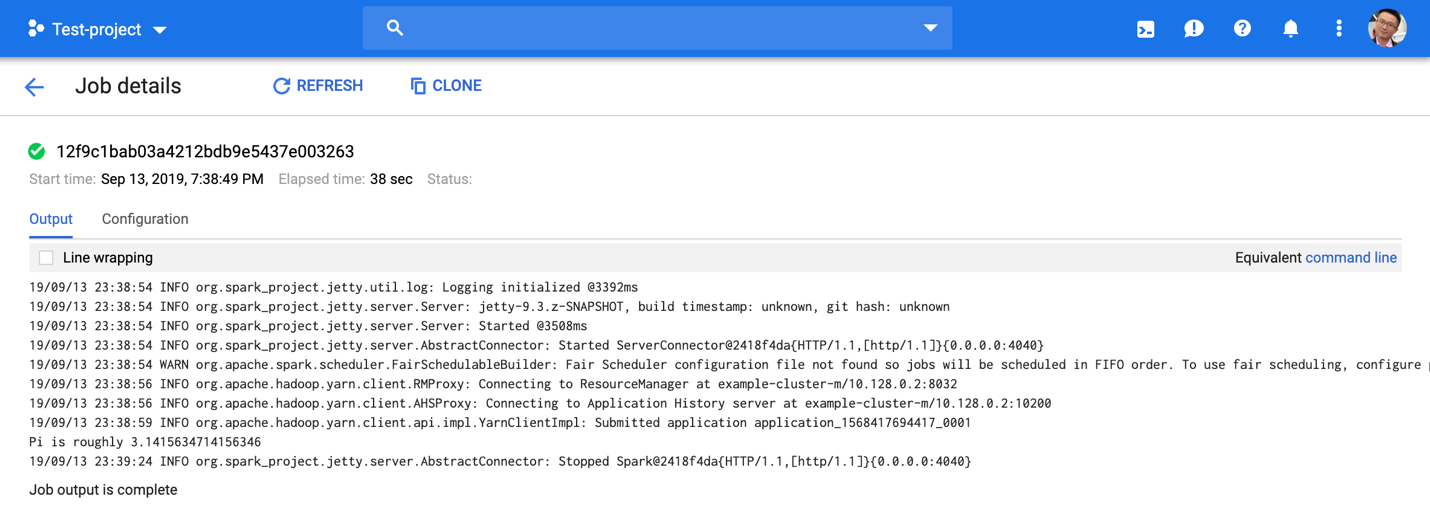
• create cluster



• submit job

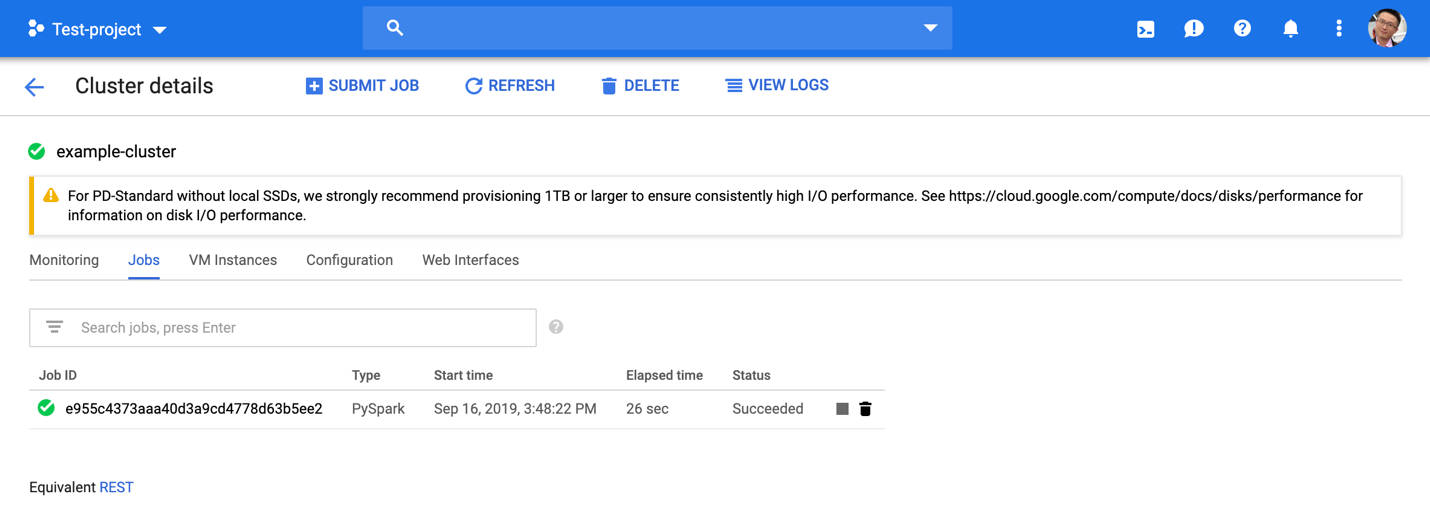


• get result

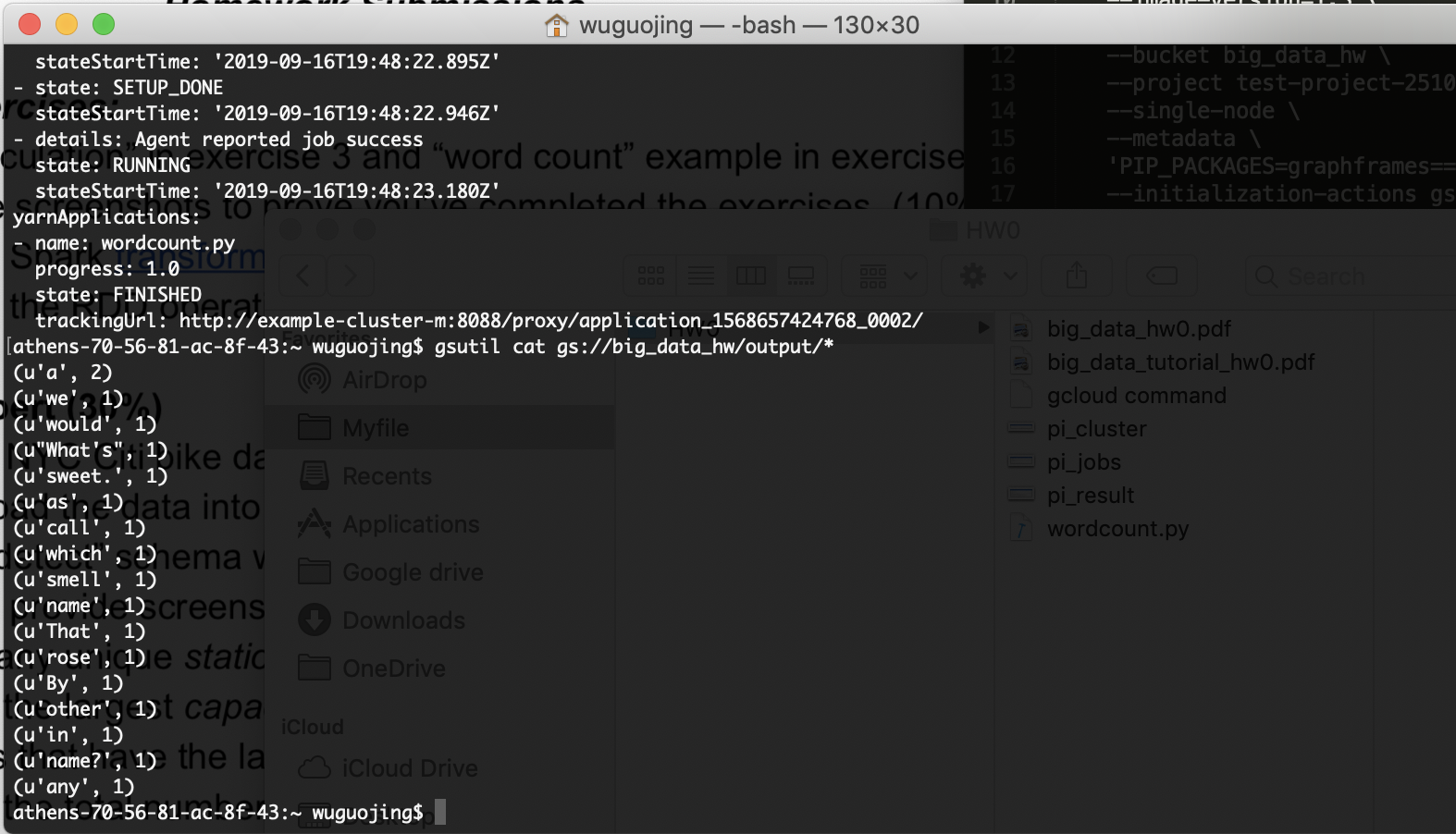


Word count

• submit job



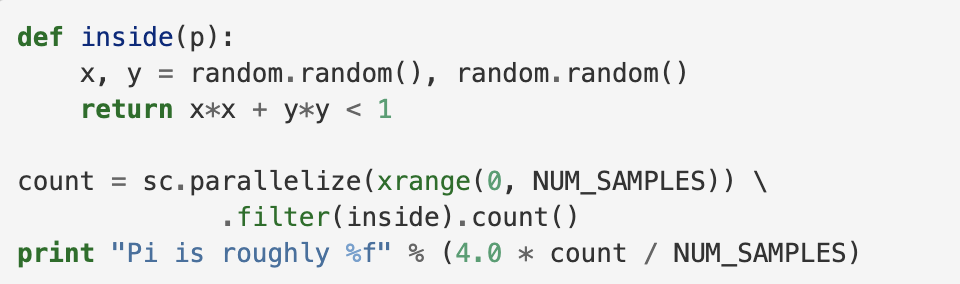
• output result



(2) pi calculation

• transformation: filter()

• action: count()



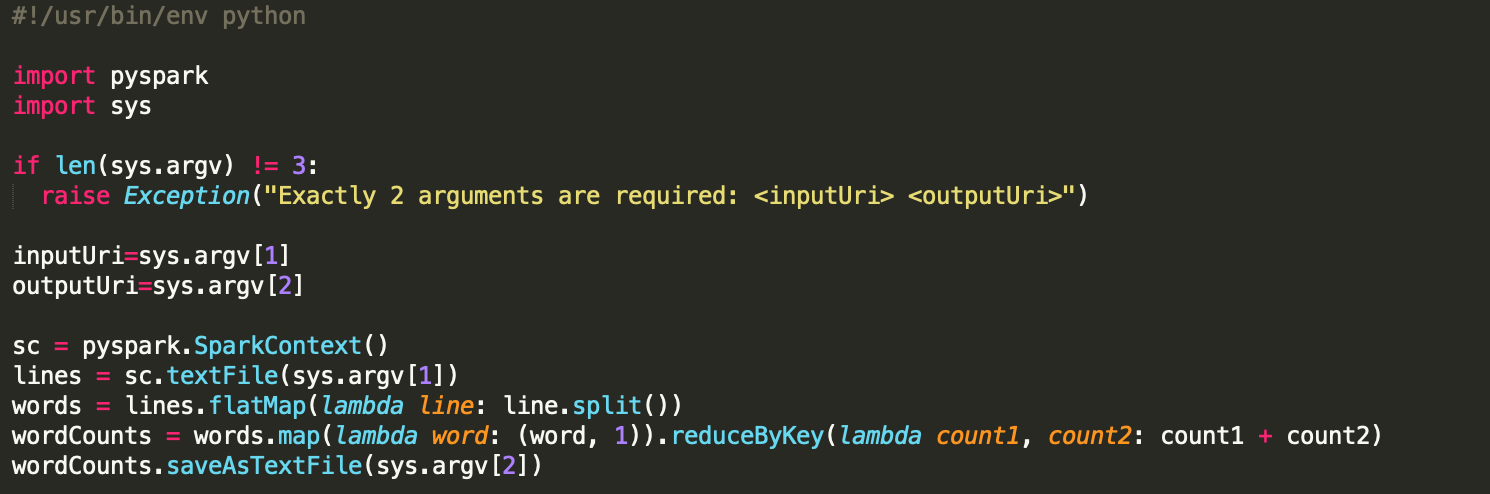
Action

transformation

Word count:

• transformation: map(), flatMap(), reduceByKey()

• action: saveAsTextFile()



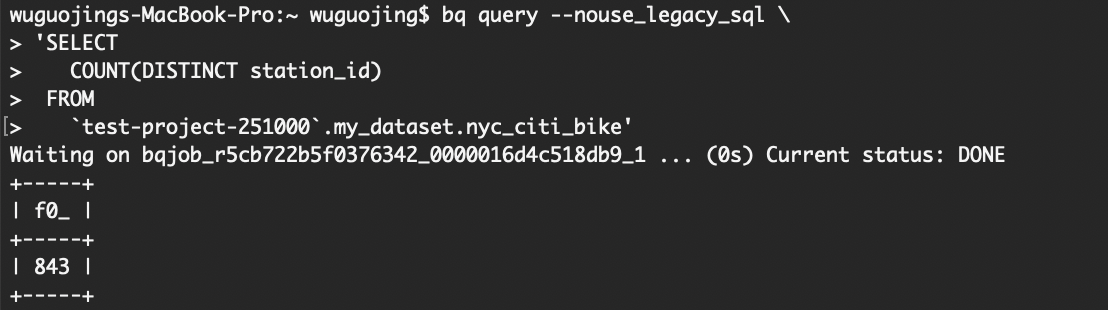
action

transformation

2. NYC Bike Expert

(1) unique station\_ids

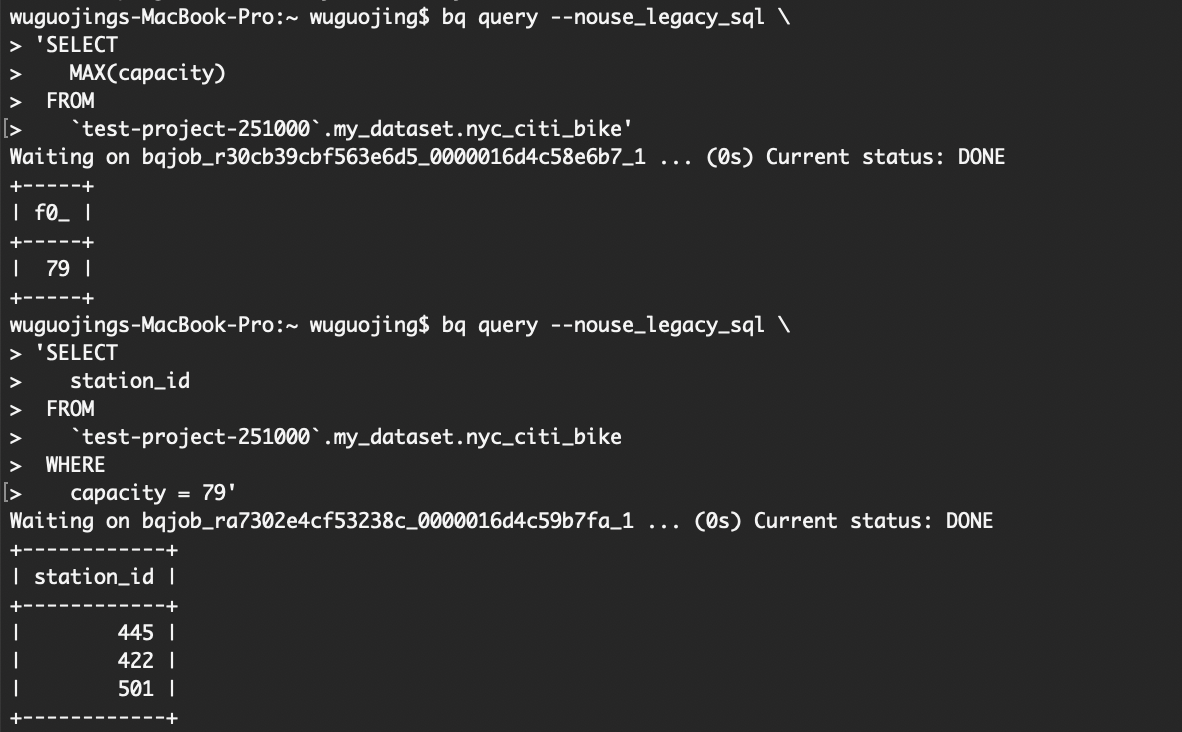
n = 843



(2) largest capacity and corresponding station\_id

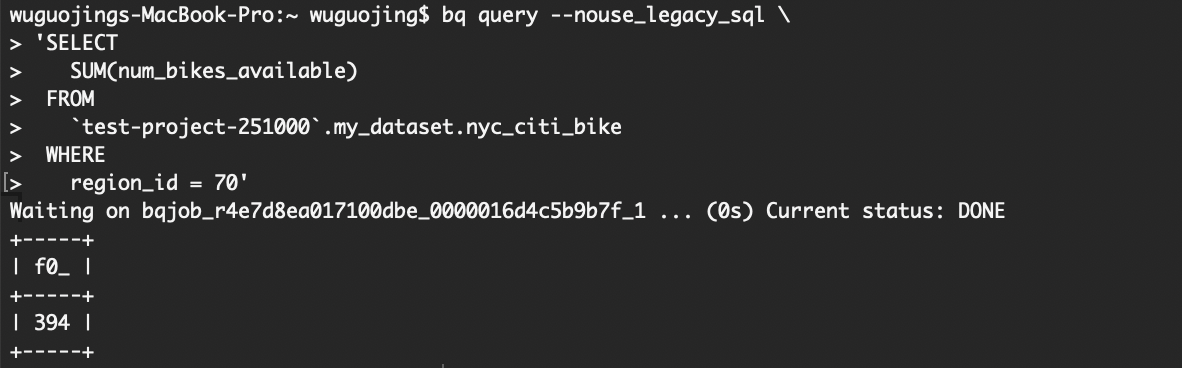
MAX = 79

Station\_id = {422, 445, 501}



(3) total number of bikes available in region\_id 70

n = 394

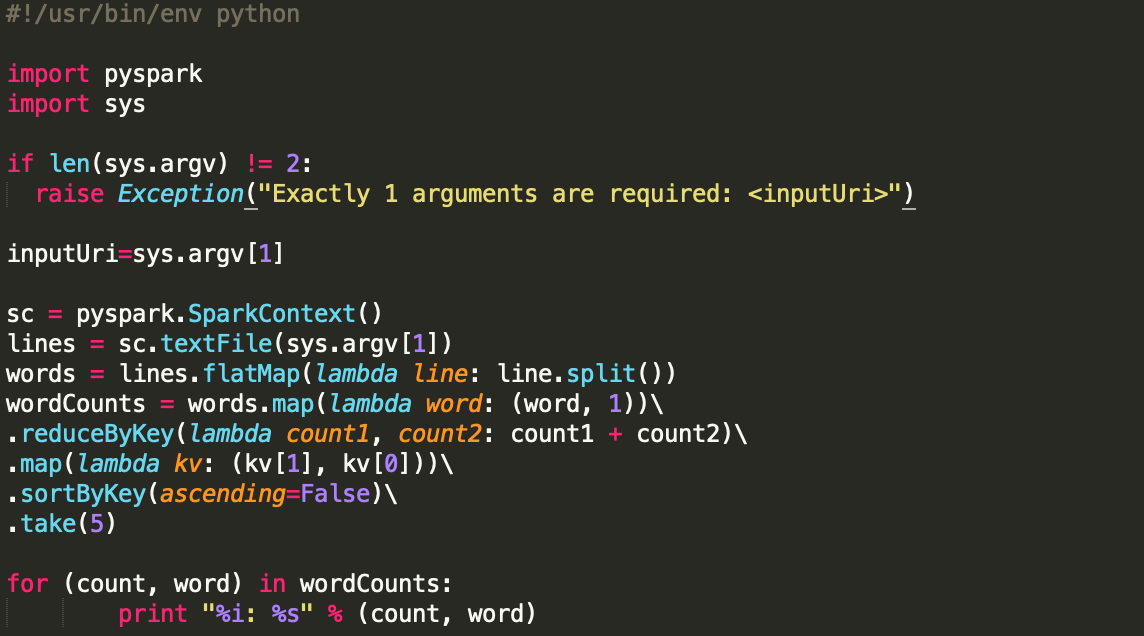


3. Understanding William Shakespeare

(1) without preprocess



Code:



(2) NLTK preprocessed

Here we removed stop words, punctuations and words like *‘s*, *‘d, ‘m*



Code:

