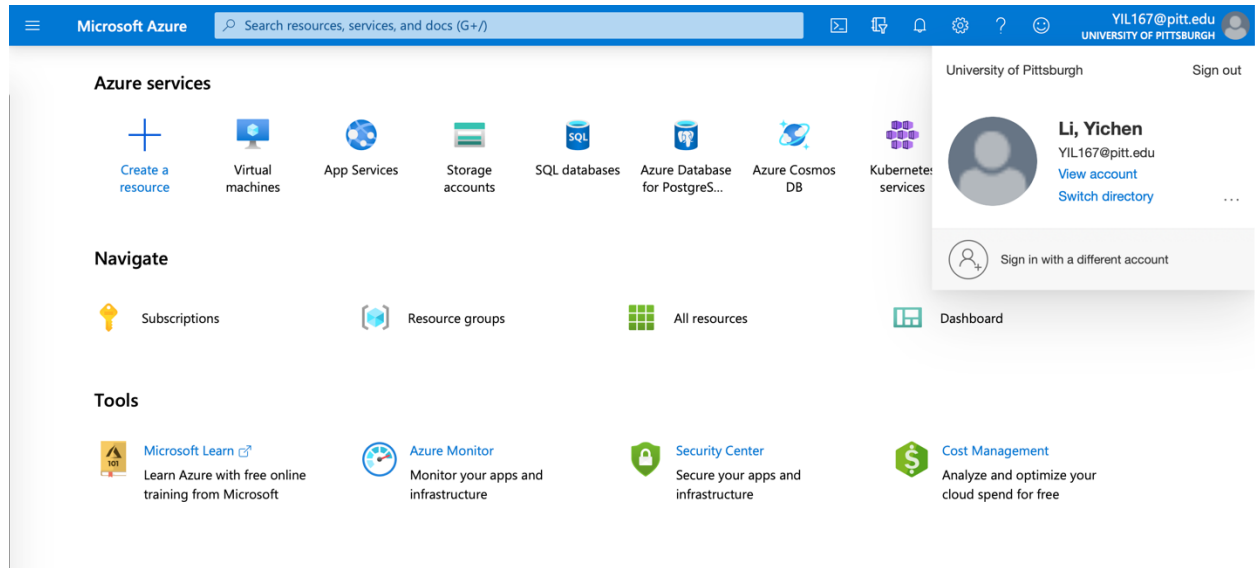


CS1660 – Cloud Computing  
HW#3  
Yichen Li



Lin to the input data file:

[https://pitt-my.sharepoint.com/:t:/g/personal/yil167\\_pitt\\_edu/Ecs6fbPyYmNGq2-\\_cy4x3lgBIL-a9atFyyGUW9BBcMfcQ?e=pgrdZw](https://pitt-my.sharepoint.com/:t:/g/personal/yil167_pitt_edu/Ecs6fbPyYmNGq2-_cy4x3lgBIL-a9atFyyGUW9BBcMfcQ?e=pgrdZw)

[https://pitt-my.sharepoint.com/:t:/g/personal/yil167\\_pitt\\_edu/Ecs6fbPyYmNGq2-\\_cy4x3lgBILL-a9atFyyGUW9BBcMfcQ?e=oqpugG](https://pitt-my.sharepoint.com/:t:/g/personal/yil167_pitt_edu/Ecs6fbPyYmNGq2-_cy4x3lgBILL-a9atFyyGUW9BBcMfcQ?e=oqpugG)

```
sshuser@hn0-hadoop:~$ yarn jar wordcountjava-1.0-SNAPSHOT.jar org.apache.hadoop.  
examples.WordCount /example/data/gutenberg/davinci.txt /example/data/wordcountou  
t  
20/04/15 14:15:55 INFO client.AHSPProxy: Connecting to Application History server  
at headnodehost/10.0.0.16:10200  
20/04/15 14:15:55 INFO client.RequestHedgingRMFailoverProxyProvider: Looking for  
the active RM in [rm1, rm2]...  
20/04/15 14:15:55 INFO client.RequestHedgingRMFailoverProxyProvider: Found activ  
e RM [rm2]  
20/04/15 14:15:56 INFO input.FileInputFormat: Total input paths to process : 1  
20/04/15 14:15:56 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library  
20/04/15 14:15:56 INFO lzo.LzoCodec: Successfully loaded & initialized native-lz  
o library [hadoop-lzo rev b5efb3e531bc1558201462b8ab15bb412ffa6b89]  
20/04/15 14:15:56 INFO mapreduce.JobSubmitter: number of splits:1  
20/04/15 14:15:56 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_15  
86959735905_0002  
20/04/15 14:15:57 INFO impl.YarnClientImpl: Submitted application application_15  
86959735905_0002  
20/04/15 14:15:57 INFO mapreduce.Job: The url to track the job: http://hn1-hadoo  
p.wx0p1khwt0ve3mcqugtncuzpc.bx.internal.cloudapp.net:8088/proxy/application_158  
6959735905_0002/  
20/04/15 14:15:57 INFO mapreduce.Job: Running job: job_1586959735905_0002
```

Screenshot of mapreduce job execution on HDinsight

```
youth, 3
youth--devoted 1
youth. 2
youth.] 1
youth; 1
youthful 3
youwant 1
z 1
z_ 2
z_. 2
z_; 1
zeal 1
zelus 1
zenith 2
zerfielen. 1
zero; 1
zum 1
zur 1
zwanzig 1
zweite 1
❖ 4
❖ 1
❖crit_ 1
❖pieza; 1
sshuser@hn0-hadoop:~$
```

Screenshot of output

```
WordCount.java — wordcountjava
EXPLORER
OPEN EDITORS
WordCount.java 1
pom.xml 1
WORDCOUNTJAVA
.settings
.vscode
settings.json
src
main
java
org
apache
hadoop
examples
WordC... 1
test
target
.classpath
.project
pom.xml 1
OUTLINE
JAVA DEPENDENCIES
MAVEN PROJECTS

WordCount.java
1 package org.apache.hadoop.examples;
2
3 import java.io.IOException;
4 import java.util.StringTokenizer;
5 import org.apache.hadoop.conf.Configuration;
6 import org.apache.hadoop.fs.Path;
7 import org.apache.hadoop.io.IntWritable;
8 import org.apache.hadoop.io.Text;
9 import org.apache.hadoop.mapreduce.Job;
10 import org.apache.hadoop.mapreduce.Mapper;
11 import org.apache.hadoop.mapreduce.Reducer;
12 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
13 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
14 import org.apache.hadoop.util.GenericOptionsParser;
15
16 public class WordCount {
17
18     public static class TokenizerMapper
19         extends Mapper<Object, Text, Text, IntWritable>{
20
21         private final static IntWritable one = new IntWritable(1);
22         private Text word = new Text();
23
24         public void map(Object key, Text value, Context context
25             ) throws IOException, InterruptedException {
26             StringTokenizer itr = new StringTokenizer(value.toString());
27             while (itr.hasMoreTokens()) {
28                 word.set(itr.nextToken());
29                 context.write(word, one);
30             }
31         }
32     }
33
34     public static class IntSumReducer
35         extends Reducer<Text, IntWritable>{
36         private IntWritable result = new I
37
Could not create temporary directory: Permission denied

sshuser@hn0-hadoop:~$ ls
wordcountjava-1.0-SNAPSHOT.jar
sshuser@hn0-hadoop:~$
```

Screenshot of mapreduce implementation on azure