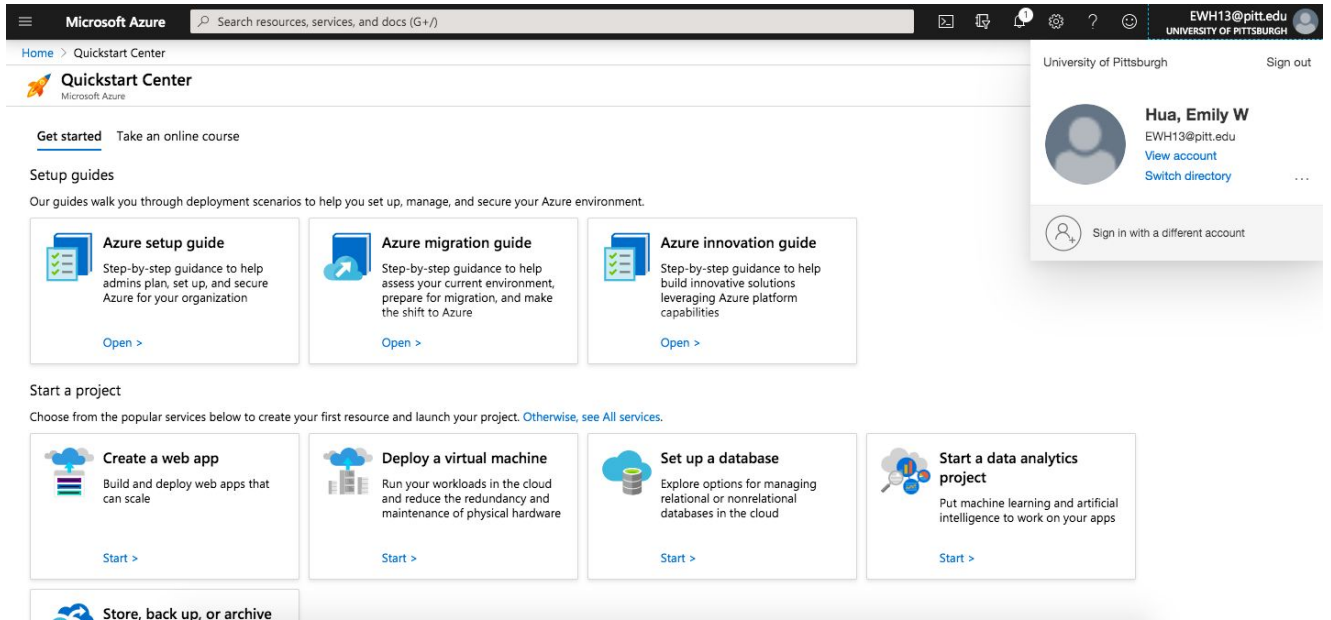


Emily Hua

Homework 3: Running WordCount MapReduce on Microsoft Azure

- Screenshot of your Azure Account showing your username (25%)



- Screenshot of your implementation of Word Count on Azure (25%)

```

package org.apache.hadoop.examples;

import java.io.IOException;
import java.util.StringTokenizer;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.util.GenericOptionsParser;

public class WordCount {

    public static class TokenizerMapper
        extends Mapper<Object, Text, Text, IntWritable>{

        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();

        public void map(Object key, Text value, Context context
            ) throws IOException, InterruptedException {
            StringTokenizer itr = new StringTokenizer(value.toString());
            while (itr.hasMoreTokens()) {
                word.set(itr.nextToken());
                context.write(word, one);
            }
        }
    }

    public static class IntSumReducer
        extends Reducer<Text, IntWritable, Text, IntWritable> {
        private IntWritable result = new IntWritable();

        public void reduce(Text key, Iterable<IntWritable> values,
            Context context
            ) throws IOException, InterruptedException {
            int sum = 0;
            for (IntWritable val : values) {
                sum += val.get();
            }
            result.set(sum);
            context.write(key, result);
        }
    }
}

```

```

public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    String[] otherArgs = new GenericOptionsParser(conf, args).getRemainingArgs();
    if (otherArgs.length != 2) {
        System.err.println("Usage: wordcount <in> <out>");
        System.exit(2);
    }
    Job job = new Job(conf, "word count");
    job.setJarByClass(WordCount.class);
    job.setMapperClass(TokenizerMapper.class);
    job.setCombinerClass(IntSumReducer.class);
    job.setReducerClass(IntSumReducer.class);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    FileInputFormat.addInputPath(job, new Path(otherArgs[0]));
    FileOutputFormat.setOutputPath(job, new Path(otherArgs[1]));
    System.exit(job.waitForCompletion(true) ? 0 : 1);
}

```

- https://pitt-my.sharepoint.com/:t/g/personal/ewh13_pitt_edu/EQTSN86KW3IDuYRiw483-SEBUgpyvTWmsze2SWvwB67UtA?e=8sSr7O

- ```

wordcountjava — sshuser@hn0-cs1660: ~ — ssh sshuser@cs1660-homework3-ssh.azurehdinsight.net — 201x53
sshuser@hn0-cs1660:~$ hdfs dfs -cat /example/data/wordcountoutput/*
111
"Anna, 3
"Come 4
"Dieu 4
"Dio 4
"EH, 3
"Emen-Hétani!" 3
"Flock"; 4
"From 4
"Grant 4
"Have 4
"I 26
"It's 3
"Joineet 3
"Marriage 4
"Monsieur 8
"No 7
"Nor," 4
"Nothing," 3
"Now 4
"Phoebus!" 3
"Phoebus",--what 3
"Phoebus," 3
"Pen 3
"Russia 4
"S 3
"Sergey 4
"The 11
"They 4
"Tit 145
>To 7
Told 4
"Twat 6
"Wash 4
>We 4
"what 8
"Whither 4
"With 4
"Ye," 4
"You 12
"ANARKH," 3
"Tit 96
"twat 3
"---," 4
"--And 4
"--Dier." 4
"--General 4
"--May 3
"--Our 4
"--The 4
"--What 4
"--With 4
Julie 164
Julie's 40
Julie, 28
Julie," 8
Julie--and 4
Julie--whose 4
Julie, 28
Julie," 4
Julie? 4
Juliet 39
Juliet! 9
Juliet's 21
Juliet, 54
Juliet, 25
Juliet: 3
Juliet: 3
Juliet? 12
Julietta 3
Julietta's 3
Julio 3
Julius 48
Julius! 3
Julius, 3
Julner, 4
July 111
July's 3
July, 84
July, 33
Julys: 3
July? 4
Jumiéges, 3
Jumping 3
June 145
June, 232
June, 15
June: 8
Junes 3
Jungfrau 4
Jungfrau, 4
Junior: 4
Junius 9
Juno 31
Juno! 3
Juno's 12
Juno, 15
Juno-like, 3
Juno; 3
Junot's 24
Jupiter 95
Jupiter! 21
Jupiter!" 3
Jupiter's 4
Jupiter, 67
Jupiter,-- 3

```

```
wordcountjava — sshuser@hn0-cs1660: ~ — ssh sshuser@cs1660-homework3-ssh.azurehdinsight.net — 204x54

zed! 3
zen 4
zenith 26
zenith, 12
zenith, 4
zephyrod!" 3
zephyrs 3
zere's 4
zero, 8
zero, 4
zest 11
zi 8
zides, 4
zig-zag 12
zig-zag, 4
zigzag 10
zigzag, 3
zigzags 3
zigzags, 10
zinc 12
zinc, 4
zinc, 4
zingara, 3
zir, 3
ziri, 3
ziruph 3
zis 4
zo 3
zodiac 7
zodiac, 4
zodiacs 3
zone 25
zone*, 3
zone, 9
zone, 3
zones 3
zones, 6
zoology 7
zoology, 8
zoology, 3
zounds, 3
zu 11
zaggered 3
{EDITOR'S 4
| 2001
à 3
évêchés, 3
ø 24
øa 3
øa-, 3
øcoles," 3
øtre 3
øtø 3
sshuser@hn0-cs1660:~$
```

\*File too large to take a full screenshot of the output, here are screenshots of the beginning, middle, and end of the file.