

Execution Instructions

CS 6350 Big data Management And Analytics

Homework 1

Execution Instructions

Introduction:

The following document briefly describes the set of instructions that needs to be followed for deploying the Hadoop Map reduce Jar files in the UTD cluster.

Commands:

- Question 1:

Jar : **EntityCount.jar**

Input arguments: **Input path(all yelp data), Output path**

command: **nohup hadoop jar \$HOME/bigdata/HW1/qn1/EntityCount.jar EntityCount /allyelpdata /\$USER/run1 &**

Output capture

```
axs143632@cs6360:~[07:11:54] $ time nohup hadoop jar $HOME/bigdata/HW1/qn1/EntityCount.jar EntityCount /allyelpdata /$USER/run1
nohup: ignoring input and appending output to 'nohup.out'

real    0m22.365s
user    0m5.926s
sys     0m0.287s
axs143632@cs6360:~[07:12:24] $ hdfs dfs -cat /$USER/run1/p*
15/06/26 19:13:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
business      78
review    132319
user      196309
-
```

- Question 2:

Jar : **FilterCity.jar**

Input arguments: **Input path(all yelp data), Output path**

command: **nohup hadoop jar \$HOME/bigdata/HW1/qn2/FilterCity.jar FilterCity /allyelpdata /\$USER/run2 &**

Output capture

```
axs143632@cs6360:~[07:13:01] $ time nohup hadoop jar $HOME/bigdata/HW1/qn2/FilterCity.jar FilterCity /allyelpdata /$USER/run2
nohup: ignoring input and appending output to 'nohup.out'

real    0m21.941s
user    0m7.096s
sys     0m0.295s
axs143632@cs6360:~[07:14:00] $ hdfs dfs -cat /$USER/run2/p*
15/06/26 19:14:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
7R0qrKQbwkBRQNFgc5tDcQ
S506FIQdswKA0anIlQSF_g
-
```

Execution Instructions

- Question 3:

Jar : **TopTenBusiness.jar**

Input arguments: **Input path(all yelp data), Output path**

command: **nohup hadoop jar \$HOME/bigdata/HW1/qn3/TopTenBusiness.jar TopTenBusiness /allyelpdata /\$USER/run3 &**

Output capture

```
axs143632@cs6360:~[07:14:08] $ time nohup hadoop jar $HOME/bigdata/HW1/qn3/TopTenBusiness.jar TopTenBusiness /allyelpdata /$USER/run3
nohup: ignoring input and appending output to `nohup.out'

real    0m21.484s
user    0m6.129s
sys     0m0.305s
axs143632@cs6360:~[07:15:40] $ hdfs dfs -cat /$USER/run3/p*
15/06/26 19:16:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Business ID
-HmR20K3fPL-MgVepuq1Q
-GEsoFE1yIavBPiJWdDRAA
-8MhEiQ0CXXQGIBC4EYpww
-LHk6aRcgF-Q0JGIau2uXQ
-6fMAgRynU3tL_ZbQV3eLw
-KedUrm_YGt2s4Q9p_Q6Mg
-J5wUuLJdr0s28InPxGkmw
-H10Dd00MVHL843ErsV6VA
-7r5-GdmuXsgRhkBnN0nqA
-Fef5c2ZcJ_YKh963XoaQw
```

- Question 4:

Jar : **TopTenBusinessDetails.jar**

Input arguments: **-D business.data=<path to business data>,-D review.data=<path to review data>, Output path**

command: **nohup hadoop jar \$HOME/bigdata/HW1/qn4/TopTenBusinessDetails.jar TopTenBusinessDetails -D business.data=/yelpbusiness/business.csv -D review.data=/yelpreview/review.csv /\$USER/run4 &**

Execution Instructions

Output capture

```
axs143632@cs6360:~[07:16:11] $ time nohup hadoop jar $HOME/bigdata/HW1/qn4/TopTenBusinessDetails.jar TopTenBusinessDetails -D business.data=/yelpbusiness/business.csv -D review.data=/yelpreview/review.csv /$USER/run4
nohup: ignoring input and appending output to 'nohup.out'

real    0m21.504s
user    0m8.072s
sys      0m0.347s
axs143632@cs6360:~[07:17:19] $ hdfs dfs -cat /$USER/run4/p*
15/06/26 19:17:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Business ID      Full Address      Categories      Avg. Rating
D1_25rFl7k6vhTJc110PoQ  Amherst, MA 01002  ['Photographers', 'Event Planning & Services'] 5.0
2a-UIu0o_IpGcj_45IHBRa  West Lafayette, IN 47906  ['Local Services', 'Sewing & Alterations'] 5.0
ItB4MzEK8bZae3FdmqZ5sQ  Claremont, CA 91711  ['Local Services', 'Carpet Cleaning'] 5.0
GhViuGdDMbhIwJ2JzC707w  Pasadena, CA 91125  ['Photographers', 'Event Planning & Services'] 5.0
AbB1M0Le_gylcAoRZMe1UA  Mountain View, CA 94309  ['Photographers', 'Event Planning & Services'] 5.0
4T7yBnAHJZD-1vTeD9WlyQ  San Luis Obispo, CA 93405  ['Photographers', 'Event Planning & Services'] 5.0
Ao7lyljfdSzZTIvat0dryA  Ithaca, NY 14850  ['Shopping Centers', 'Shopping', 'Local Flavor'] 5.0
3J9sNtEs0NIL-F2tQHFTUG  San Luis Obispo, CA  ['Photographers', 'Videographers', 'Event Planning & Services'] 5.0
1YH_208GTETtLuJeFTqQvlg  San Luis Obispo, CA  ['Food', 'Bakeries'] 5.0
Du6SkVBGkb7dqtCpdz-_Kg  San Luis Obispo, CA  ['Wedding Planning', 'Event Planning & Services'] 5.0
```

- Question 5:

Jar : UserDetails.jar

Input arguments: -D business.data=<path to business data>,Input path(all yelp data),

Output path

command: nohup time hadoop jar UserDetails.jar UserDetails -D

business.data=/yelpdata/data.csv /yelpdata/data.csv /\$USER/run5 &

Output capture:

```
abinav@BIGDASH: /media/abinav/Conn/MSINCSINUS/US/Summer 15/CS 6358 Big Data Management and Analytics/HW 1/Solutions/qn5[07:36:50] $ cat /tmp/output/part-r-00000
0MaCdhR3aXb0G0ntwTMTGtG We hired Stanford's Bartender for a private movie screening party and will definitely use them again for all our events in the future.
Carlos showed up on time, set up a beautiful bar and was very professional the entire evening. It is so nice to work with someone that clearly takes pride in
what they do. I would highly recommend using Stanford's Bartender for your next event!
1UUqEV63CB6hR8j_UNjfxQ Stanford's Bartender is amazing! From planning, shopping list, prep, event and clean-up - everything was top notch/outstanding!! I ca
n't say enough good things about Carlos and his operation!! From the moment we spoke on the phone, I knew he was it - he knew his liquors and is not just a ba
rtender but a true mixologist!! His catprinhas, mojitos and margaritas were the best I ever had - he introduced me to rum/margarita that will now be the stand
ard here at our home => Not just great with drinks, but he's also good on gauging the crowd - it made for an energetic and fun party and all our guests loved
him!!! I highly recommend him! Thanks Carlos, you are outstanding!!!
2Nn-2YH6Jznfy7lfSpGB9g Thank you so much for making our Cinco de Mayo party such an awesome event! I was so impressed with the professionalism of both our ba
rtender and Carlos! We will definitely be booking again! Thank you thank you!
57WX12z2rEMt6S0i_QxdFQ Things could not have gone better for my husband's 40th birthday party. Jessica showed up right on time, and I felt so bad that our ba
r was outside in the 95 degree weather, but she didn't seem to mind at all. She was SO nice and so excited about the party. Everyone was blown away by her an
d the delicious cocktails! I noticed some people take cards, and I'm sure everyone will keep Stanford's in mind!! Also working with Carlos on the process was
a breeze- thanks again for everything!
```

NOTE:

When the code for the Mapjoin was deployed in the UTD cluster, I was facing some exceptions, hence I tested it successfully by deploying it in my local machine.