

Instances | EC2 | ap-south-1

EC2 Instance Connect | ap-south-1

ap-south-1.console.aws.amazon.com/ec2-instance-connect/ssh?region=ap-south-1&connType=standard&instanceId=i-02664183508b74bfb...

aws

Services

Search

[Alt+S]

Mumbai

Ashutosh

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

Last login: Wed Feb 28 16:43:16 2024 from 13.233.177.4
ubuntu@ip-172-31-8-12:~\$ cat mapper.py
#!/usr/bin/env python
"""mapper.py"""

import sys

input comes from STDIN (standard input)
for line in sys.stdin:
 # remove leading and trailing whitespace
 line = line.strip()
 # split the line into words
 words = line.split()
 # increase counters
 for word in words:
 # write the results to STDOUT (standard output);
 # what we output here will be the input for the
 # Reduce step, i.e. the input for reducer.py
 #
 # tab-delimited; the trivial word count is 1
 print('%s\t%s' % (word, 1))
ubuntu@ip-172-31-8-12:~\$

i-02664183508b74bfb (202318050 - Main)

PublicIPs: 3.110.177.119 PrivateIPs: 172.31.8.12

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Instances | EC2 | ap-south-1

EC2 Instance Connect | ap-south-1

ap-south-1.console.aws.amazon.com/ec2-instance-connect/ssh?region=ap-south-1&connType=standard&instanceId=i-02664183508b74bfb...

aws

Services

Search

[Alt+S]

Mumbai

Ashutosh

ubuntu@ip-172-31-8-12:~\$ cat reducer.py
#!/usr/bin/env python
"""reducer.py"""

from operator import itemgetter
import sys

current_word = None
current_count = 0
word = None

input comes from STDIN
for line in sys.stdin:
 # remove leading and trailing whitespace
 line = line.strip()

 # parse the input we got from mapper.py
 word, count = line.split('\t', 1)

 # convert count (currently a string) to int
 try:
 count = int(count)
 except ValueError:
 # count was not a number, so silently
 # ignore/discard this line
 continue

i-02664183508b74bfb (202318050 - Main)

PublicIPs: 3.110.177.119 PrivateIPs: 172.31.8.12

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

```
aws
Services Search [Alt+S]

# parse the input we got from mapper.py
word, count = line.split('\t', 1)

# convert count (currently a string) to int
try:
    count = int(count)
except ValueError:
    # count was not a number, so silently
    # ignore/discard this line
    continue

# this IF-switch only works because Hadoop sorts map output
# by key (here: word) before it is passed to the reducer
if current_word == word:
    current_count += count
else:
    if current_word:
        # write result to STDOUT
        print('%s\t%s' % (current_word, current_count))
    current_count = count
    current_word = word

# do not forget to output the last word if needed!
if current_word == word:
    print('%s\t%s' % (current_word, current_count))
ubuntu@ip-172-31-8-12:~$
```

i-02664183508b74bfb (202318050 - Main)
PublicIPs: 3.110.177.119 PrivateIPs: 172.31.8.12

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

```
aws
Services Search [Alt+S]

ubuntu@ip-172-31-8-12:~$ touch textfile.txt
ubuntu@ip-172-31-8-12:~$ nano textfile.txt
ubuntu@ip-172-31-8-12:~$ cat textfile.txt | python3 mapper.py | python3 reducer.py
We've 1
been 1
thrilled 1
to 1
see 1
participants' 1
enthusiasm 1
and 1
dedication 1
in 1
capturing 1
the 1
fest's 1
essence. 1
We 1
understand 1
that 1
several 1
of 1
you 1
have 1
reached 1
out 1

i-02664183508b74bfb (202318050 - Main)  
PublicIPs: 3.110.177.119 PrivateIPs: 172.31.8.12
```

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

```
number 1
of 1
requests, 1
we 1
are 1
pleased 1
to 1
announce 1
that 1
we 1
will 1
be 1
extending 1
the 1
deadline 1
for 1
submission 1
of 1
photos 1
for 1
the 1
Festive 1
Frames 1
Photography 1
Contest. 1
ubuntu@ip-172-31-8-12:~$
```

i-02664183508b74bfb (202318050 - Main)

PublicIPs: 3.110.177.119 PrivateIPs: 172.31.8.12

```
ubuntu@ip-172-31-8-12:~$ SHADOOP_HOME/sbin/start-dfs.sh
Starting namenodes on [ec2-3-110-177-119.ap-south-1.compute.amazonaws.com]
Starting datanodes
Starting secondary namenodes [ip-172-31-8-12]
ubuntu@ip-172-31-8-12:~$ SHADOOP_HOME/sbin/start-yarn.sh
Starting resourcemanager
Starting nodemanagers
ubuntu@ip-172-31-8-12:~$ hdfs dfsadmin -safemode leave
Safe mode is OFF
ubuntu@ip-172-31-8-12:~$ hadoop jar $HADOOP_HOME/share/hadoop/tools/lib/hadoop-*streaming*.jar -files mapper.py, reducer.py -input /home/ubuntu/input/* -o
utput /home/ubuntu/input-output
packageJobJar: [/tmp/hadoop-unjar1244906853132343741/] [] /tmp/streamjob1359769414052255239.jar tmpDir=null
2024-02-28 18:10:46,327 INFO client.RMProxy: Connecting to ResourceManager at ec2-3-110-177-119.ap-south-1.compute.amazonaws.com/172.31.8.12:8032
2024-02-28 18:10:46,771 INFO client.RMProxy: Connecting to ResourceManager at ec2-3-110-177-119.ap-south-1.compute.amazonaws.com/172.31.8.12:8032
2024-02-28 18:10:47,256 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/ubuntu/.staging/job_1709143828356_0001
2024-02-28 18:10:48,707 INFO mapred.FileInputFormat: Total input files to process : 1
2024-02-28 18:10:48,845 INFO mapreduce.JobSubmitter: number of splits:2
2024-02-28 18:10:49,252 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1709143828356_0001
2024-02-28 18:10:49,254 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-02-28 18:10:49,672 INFO conf.Configuration: resource-types.xml not found
2024-02-28 18:10:49,673 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-02-28 18:10:50,259 INFO impl.YarnClientImpl: Submitted application application_1709143828356_0001
2024-02-28 18:10:50,387 INFO mapreduce.Job: The url to track the job: http://ec2-3-110-177-119.ap-south-1.compute.amazonaws.com:8088/proxy/application_1709143828356_0001/
2024-02-28 18:10:50,390 INFO mapreduce.Job: Running job: job_1709143828356_0001
```

i-02664183508b74bfb (202318050 - Main)

PublicIPs: 3.110.177.119 PrivateIPs: 172.31.8.12