

```
The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
Ankurs-MacBook-Pro:~ ankursugandhi$ ssh -
i/Users/ankursugandhi/Downloads/demo_key_pair.pem hadoop@ec2-35-175-240-
209.compute-1.amazonaws.com
The authenticity of host 'ec2-35-175-240-209.compute-1.amazonaws.com
(35.175.240.209)' can't be established.
ECDSA key fingerprint is
SHA256:HsHROXjGPWQ9x7lGztnshbZgydAATtgMyCKuqSXjr9Y.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-35-175-240-209.compute-
1.amazonaws.com,35.175.240.209' (ECDSA) to the list of known hosts.
Last login: Thu Jan 14 10:19:57 2021
```

```

_ _ | _ _ | _ )
_ | ( _ _ /
_ _ | \ _ _ | _ _ |

```

Amazon Linux AMI

```
https://aws.amazon.com/amazon-linux-ami/2018.03-release-notes/
43 package(s) needed for security, out of 74 available
Run "sudo yum update" to apply all updates.
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No
such file or directory
```

EEEEEEEEEEEEEEEEEEEE	MMMMMMMM	MMMMMMMM	RRRRRRRRRRRRRRRR			
E::::::::::::::::::::E	M:::::::::M	M:::::::::M	R:::::::::::::::::R			
EE::::::::EEEEEEEEEE::E	M:::::::::M	M:::::::::M	R:::::RRRRRR:::::R			
E:::::E	EEEE	M:::::::::M	M:::::::::M	RR::::R	R:::::R	
E:::::E		M:::::::::M:::M	M:::M:::::M	R:::R	R:::::R	
E:::::EEEEEEEEEEEE	M:::::M	M:::M	M:::M	M:::::M	R:::RRRRRR:::::R	
E::::::::::::::::::::E	M:::::M	M:::M:::M	M:::::M		R:::::::::::::::::RR	
E:::::EEEEEEEEEEEE	M:::::M	M:::::M	M:::::M		R:::RRRRRR:::::R	
E:::::E	M:::::M	M:::M	M:::::M	R:::R	R:::::R	
E:::::E	EEEE	M:::::M	MMM	M:::::M	R:::R	R:::::R
EE::::::::EEEEEEEEEE:::E	M:::::M		M:::::M	R:::R	R:::::R	
E::::::::::::::::::::::::::E	M:::::M		M:::::M	RR:::R	R:::::R	
EEEEEEEEEEEEEEEEEEEE	MMMMMMMM	MMMMMMMM	RRRRRRR	RRRRRR		

```
[hadoop@ip-172-31-34-126 ~]$ hadoop fs -mkdir /user/hive/online_sales
[hadoop@ip-172-31-34-126 ~]$ aws s3 ls e-commerce-events-ml
2020-03-17 11:47:09 545839412 2019-Nov.csv
2020-03-17 11:37:31 482542278 2019-Oct.csv
[hadoop@ip-172-31-34-126 ~]$ hadoop fs -ls /user/hive
```

```
Found 2 items
drwxr-xr-x    - hadoop hadoop          0 2021-01-14 10:25
/user/hive/online_sales
drwxrwxrwt    - hdfs    hadoop          0 2021-01-14 09:45
/user/hive/warehouse
[hadoop@ip-172-31-34-126 ~]$ hadoop distcp 's3://e-commerce-events-ml/*'
'/user/hive/online_sales/'
```

```
21/01/14 10:26:13 INFO tools.DistCp: Input Options:
DistCpOptions{atomicCommit=false, syncFolder=false, deleteMissing=false,
ignoreFailures=false, overwrite=false, skipCRC=false, blocking=true,
numListstatusThreads=0, maxMaps=20, mapBandwidth=100,
sslConfigurationFile='null', copyStrategy='uniformsize',
preserveStatus=[], preserveRawXattrs=false, atomicWorkPath=null,
logPath=null, sourceFileListing=null, sourcePaths=[s3://e-commerce-
events-ml/*], targetPath=/user/hive/online_sales, targetPathExists=true,
filtersFile='null'}
21/01/14 10:26:13 INFO client.RMProxy: Connecting to ResourceManager at
ip-172-31-34-126.ec2.internal/172.31.34.126:8032
21/01/14 10:26:20 INFO tools.SimpleCopyListing: Paths (files+dirs) cnt =
2; dirCnt = 0
21/01/14 10:26:20 INFO tools.SimpleCopyListing: Build file listing
completed.
21/01/14 10:26:20 INFO Configuration.deprecation: io.sort.mb is
deprecated. Instead, use mapreduce.task.io.sort.mb
21/01/14 10:26:20 INFO Configuration.deprecation: io.sort.factor is
deprecated. Instead, use mapreduce.task.io.sort.factor
21/01/14 10:26:20 INFO tools.DistCp: Number of paths in the copy list: 2
21/01/14 10:26:20 INFO tools.DistCp: Number of paths in the copy list: 2
21/01/14 10:26:20 INFO client.RMProxy: Connecting to ResourceManager at
ip-172-31-34-126.ec2.internal/172.31.34.126:8032
21/01/14 10:26:20 INFO mapreduce.JobSubmitter: number of splits:2
21/01/14 10:26:20 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1610617611610_0001
21/01/14 10:26:21 INFO impl.YarnClientImpl: Submitted application
application_1610617611610_0001
21/01/14 10:26:22 INFO mapreduce.Job: The url to track the job:
http://ip-172-31-34-
126.ec2.internal:20888/proxy/application_1610617611610_0001/
21/01/14 10:26:22 INFO tools.DistCp: DistCp job-id:
job_1610617611610_0001
21/01/14 10:26:22 INFO mapreduce.Job: Running job: job_1610617611610_0001
21/01/14 10:26:31 INFO mapreduce.Job: Job job_1610617611610_0001 running
in uber mode : false
21/01/14 10:26:31 INFO mapreduce.Job: map 0% reduce 0%
21/01/14 10:26:52 INFO mapreduce.Job: map 100% reduce 0%
21/01/14 10:27:06 INFO mapreduce.Job: Job job_1610617611610_0001
completed successfully
21/01/14 10:27:06 INFO mapreduce.Job: Counters: 38
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=345684
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=898
    HDFS: Number of bytes written=1028381690
    HDFS: Number of read operations=26
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=8
    S3: Number of bytes read=1028381690
    S3: Number of bytes written=0
```

```
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0
Job Counters
  Launched map tasks=2
  Other local map tasks=2
  Total time spent by all maps in occupied slots (ms)=2009856
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=62808
  Total vcore-milliseconds taken by all map tasks=62808
  Total megabyte-milliseconds taken by all map tasks=64315392
Map-Reduce Framework
  Map input records=2
  Map output records=0
  Input split bytes=272
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=925
  CPU time spent (ms)=40370
  Physical memory (bytes) snapshot=957206528
  Virtual memory (bytes) snapshot=6588588032
  Total committed heap usage (bytes)=763363328
File Input Format Counters
  Bytes Read=626
File Output Format Counters
  Bytes Written=0
DistCp Counters
  Bytes Copied=1028381690
  Bytes Expected=1028381690
  Files Copied=2
[hadoop@ip-172-31-34-126 ~]$ hadoop fs -cat /user/hive/online_sales/2019-
Nov.csv | head
event_time,event_type,product_id,category_id,category_code,brand,price,us
er_id,user_session
2019-11-01 00:00:02
UTC,view,5802432,1487580009286598681,,,0.32,562076640,09fafd6c-6c99-46b1-
834f-33527f4de241
2019-11-01 00:00:09
UTC,cart,5844397,1487580006317032337,,,2.38,553329724,2067216c-31b5-455d-
alcc-af0575a34ffb
2019-11-01 00:00:10
UTC,view,5837166,1783999064103190764,,pnb,22.22,556138645,57ed222e-a54a-
4907-9944-5a875c2d7f4f
2019-11-01 00:00:11
UTC,cart,5876812,1487580010100293687,,jessnail,3.16,564506666,186c1951-
8052-4b37-adce-dd9644b1d5f7
2019-11-01 00:00:24
UTC,remove_from_cart,5826182,1487580007483048900,,,3.33,553329724,2067216
c-31b5-455d-alcc-af0575a34ffb
2019-11-01 00:00:24
UTC,remove_from_cart,5826182,1487580007483048900,,,3.33,553329724,2067216
c-31b5-455d-alcc-af0575a34ffb
```

```

2019-11-01 00:00:25
UTC,view,5856189,1487580009026551821,,runail,15.71,562076640,09fafd6c-
6c99-46b1-834f-33527f4de241
2019-11-01 00:00:32
UTC,view,5837835,1933472286753424063,,,3.49,514649199,432a4e95-375c-4b40-
bd36-0fc039e77580
2019-11-01 00:00:34
UTC,remove_from_cart,5870838,1487580007675986893,,milv,0.79,429913900,2f0
bff3c-252f-4fe6-afcd-5d8a6a92839a
cat: Unable to write to output stream.
[hadoop@ip-172-31-34-126 ~]$ hadoop fs -cat /user/hive/online_sales/2019-
Oct.csv | head
event_time,event_type,product_id,category_id,category_code,brand,price,us
er_id,user_session
2019-10-01 00:00:00
UTC,cart,5773203,1487580005134238553,,runail,2.62,463240011,26dd6e6e-
4dac-4778-8d2c-92e149dab885
2019-10-01 00:00:03
UTC,cart,5773353,1487580005134238553,,runail,2.62,463240011,26dd6e6e-
4dac-4778-8d2c-92e149dab885
2019-10-01 00:00:07
UTC,cart,5881589,2151191071051219817,,lovely,13.48,429681830,49e8d843-
adf3-428b-a2c3-fe8bc6a307c9
2019-10-01 00:00:07
UTC,cart,5723490,1487580005134238553,,runail,2.62,463240011,26dd6e6e-
4dac-4778-8d2c-92e149dab885
2019-10-01 00:00:15
UTC,cart,5881449,1487580013522845895,,lovely,0.56,429681830,49e8d843-
adf3-428b-a2c3-fe8bc6a307c9
2019-10-01 00:00:16
UTC,cart,5857269,1487580005134238553,,runail,2.62,430174032,73deale7-
664e-43f4-8b30-d32b9d5af04f
2019-10-01 00:00:19
UTC,cart,5739055,1487580008246412266,,kapous,4.75,377667011,81326ac6-
daa4-4f0a-b488-fd0956a78733
2019-10-01 00:00:24
UTC,cart,5825598,1487580009445982239,,,0.56,467916806,2f5b5546-b8cb-9ee7-
7ecd-84276f8ef486
2019-10-01 00:00:25
UTC,cart,5698989,1487580006317032337,,,1.27,385985999,d30965e8-1101-44ab-
b45d-cc1bb9fae694
cat: Unable to write to output stream.
[hadoop@ip-172-31-34-126 ~]$ hadoop fs -ls /user/hive/online_sales
Found 2 items
-rw-r--r--    1 hadoop hadoop   545839412 2021-01-14 10:27
/user/hive/online_sales/2019-Nov.csv
-rw-r--r--    1 hadoop hadoop   482542278 2021-01-14 10:27
/user/hive/online_sales/2019-Oct.csv
[hadoop@ip-172-31-34-126 ~]$ hadoop fs -ls /user/hive
Found 2 items
drwxr-xr-x    - hadoop hadoop              0 2021-01-14 10:27
/user/hive/online_sales
drwxrwxrwt    - hdfs   hadoop              0 2021-01-14 10:30
/user/hive/warehouse

```

```
[hadoop@ip-172-31-34-126 ~]$ hadoop fs -ls /user/hive/warehouse
Found 1 items
drwxrwxrwt   - hadoop hadoop           0 2021-01-14 10:32
/user/hive/warehouse/online_sales.db
[hadoop@ip-172-31-34-126 ~]$ hadoop fs -ls
/user/hive/warehouse/online_sales.db
Found 1 items
drwxrwxrwt   - hadoop hadoop           0 2021-01-14 10:34
/user/hive/warehouse/online_sales.db/sales_bucket
[hadoop@ip-172-31-34-126 ~]$ hadoop fs -ls
/user/hive/warehouse/online_sales.db/sales_bucket
Found 4 items
drwxrwxrwt   - hadoop hadoop           0 2021-01-14 10:34
/user/hive/warehouse/online_sales.db/sales_bucket/event_type=cart
drwxrwxrwt   - hadoop hadoop           0 2021-01-14 10:34
/user/hive/warehouse/online_sales.db/sales_bucket/event_type=purchase
drwxrwxrwt   - hadoop hadoop           0 2021-01-14 10:34
/user/hive/warehouse/online_sales.db/sales_bucket/event_type=remove_from_
cart
drwxrwxrwt   - hadoop hadoop           0 2021-01-14 10:34
/user/hive/warehouse/online_sales.db/sales_bucket/event_type=view
[hadoop@ip-172-31-34-126 ~]$
```