MSCK Repair Command

...

IMPORTANT

Copyright Infringement and Illegal Content Sharing Notice

What is the purpose of msck repair?

The MSCK REPAIR TABLE command was designed to manually add partitions that are added to the HDFS file system, but are not present in the metastore.

Let us first create an external table

hadoop fs -mkdir /data (create data directory in hdfs)

Create an external table partitioned on state.

```
create external table orders_w_partition(
  id string,
  customer_id string,
  product_id string,
  quantity int,
  amount double,
 zipcode char(5))
partitioned by (state char(2))
row format delimited fields terminated by ','
location '/data';
```

Screenshot of the commands

```
[cloudera@quickstart Downloads]$ hadoop fs -mkdir /data
[cloudera@quickstart Downloads]$ hadoop fs -ls /data
[cloudera@quickstart Downloads]$
```

```
hive> use trendytech;
OK
Time taken: 0.011 seconds
hive> create external table orders_w_partition(
    >   id string,
    >   customer_id string,
    >   product_id string,
    >   quantity int,
    >   amount double,
    >  zipcode char(5))
    > partitioned by (state char(2))
    > row format delimited fields terminated by ','
    > location '/user/cloudera/data';
OK
Time taken: 0.04 seconds
hive> show partitions orders_w_partition
    > ;
OK
Time taken: 0.036 seconds
hive> select * from orders_w_partition;
OK
Time taken: 0.045 seconds
hive>
```

The table does not have any partitions

Manually create partition directories in hdfs

```
[cloudera@quickstart ~]$ hadoop fs -mkdir /data/orders=ca
[cloudera@quickstart ~]$ hadoop fs -mkdir /data/orders=ct
```

hadoop fs -mkdir /data/orders=ca
hadoop fs -mkdir /data/orders=ct

Now try to display the partitions in hive table.

show partitions orders_w_partition; (this will still not show any partitions as hive does not have metadata for the 2 partitions you manually added in hdfs)

Use msck repair command

msck repair table orders_w_partition;

show partitions orders_w_partition;

```
hive> show partitions orders_w_partition;
OK
Time taken: 0.029 seconds
hive> msck repair table orders_w_partition;
OK
Partitions not in metastore:    orders_w_partition:state=ca     orders_w_partition:
state=ct
Repair: Added partition to metastore orders_w_partition:state=ca
Repair: Added partition to metastore orders_w_partition:state=ct
Time taken: 0.088 seconds, Fetched: 3 row(s)
hive> show partitions orders_w_partition;
OK
state=ca
state=ct
Time taken: 0.032 seconds, Fetched: 2 row(s)
hive>
```

Using msck repair command we are able to add the hive metadata which was missing. We can see that.

We have learnt msck repair command

Happy Learning!!!

Follow US

Trainer	Mr. Sumit Mittal
Phone	9108179578
Email	trendytech.sumit@gmail.com
Website	https://trendytech.in/courses/big-data-online-training/
LinkedIn	https://www.linkedin.com/in/bigdatabysumit/
Twitter	@BigdataBySumit
Instagram	bigdatabysumit
Facebook	https://www.facebook.com/trendytech.in/
Youtube	https://www.youtube.com/channel/UCbTggJVf0NDTfWX-C_gUGSg