



```
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: true
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.hive.common.StringInternUtils (file:/usr/lib/hive/lib/hive-common-3.1.3.jar)
to field java.net.URI.string
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.hive.common.StringInternUtils
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
Hive Session ID = 36a922ed-98d4-451c-a83f-af36009916e7
hive>
> CREATE TABLE IF NOT EXISTS brewery_data (
>   Batch_ID INT,
>   Brew_Date STRING,
>   Beer_Style STRING,
>   SKU STRING,
>   Location STRING,
>   Fermentation_Time INT,
>   Temperature DOUBLE,
>   pH_Level DOUBLE,
>   Gravity DOUBLE,
>   Alcohol_Content DOUBLE,
>   Bitterness INT,
>   Color INT,
>   Ingredient_Ratio STRING,
>   Volume_Produced INT,
>   Total_Sales DOUBLE,
>   Quality_Score DOUBLE,
>   Brewhouse_Efficiency DOUBLE,
>   Loss_During_Brewing DOUBLE,
>   Loss_During_Fermentation DOUBLE,
>   Loss_During_Bottling_Kegging DOUBLE
> );
OK
Time taken: 1.114 seconds
hive>
```



```
> CREATE TABLE IF NOT EXISTS brewery_data (  
>   Batch_ID INT,  
>   Brew_Date STRING,  
>   Beer_Style STRING,  
>   SKU STRING,  
>   Location STRING,  
>   Fermentation_Time INT,  
>   Temperature DOUBLE,  
>   pH_Level DOUBLE,  
>   Gravity DOUBLE,  
>   Alcohol_Content DOUBLE,  
>   Bitterness INT,  
>   Color INT,  
>   Ingredient_Ratio STRING,  
>   Volume_Produced INT,  
>   Total_Sales DOUBLE,  
>   Quality_Score DOUBLE,  
>   Brewhouse_Efficiency DOUBLE,  
>   Loss_During_Brewing DOUBLE,  
>   Loss_During_Fermentation DOUBLE,  
>   Loss_During_Bottling_Kegging DOUBLE  
> );
```

OK

Time taken: 1.114 seconds

```
hive> LOAD DATA INPATH 'gs://final_project_bucket2/brewery_data.csv' INTO TABLE brewery_data;
```

FAILED: SemanticException Line 1:17 Invalid path 'gs://final_project_bucket2/brewery_data.csv': No files matching path gs://final_p
roject_bucket2/brewery_data.csv

```
hive>
```

```
>
```

```
> LOAD DATA INPATH 'gs://final_project_bucket2/brewery_data_raw.csv' INTO TABLE brewery_data;
```

Loading data to table default.brewery_data

OK

Time taken: 12.56 seconds

```
hive> █
```





>

> LOAD DATA INPATH 'gs://final_project_bucket2/brewery_data_raw.csv' INTO TABLE brewery_data;

Loading data to table default.brewery_data

OK

Time taken: 12.56 seconds

hive> SELECT * FROM brewery_data LIMIT 5;

OK

track_id	track_name	track_artist	track_popularity	track_album_id	NULL	NULL	NULL	NULL	NULL	NULL	NULL
6f807x0ima9a1j3VPbc7VN	I Don't Care (with Justin Bieber) - Loud Luxury	Remix	Ed Sheeran	66	2oCs0DGTsRO98Gh5ZS12Cx	NULL	NULL	NULL	NULL	NULL	NULL
0r7CVbZTWZgbTCYdfa2P31	Memories - Dillon Francis Remix	Maroon 5	67	63rPSO264uRjW1X5E6cWv6	NULL	NULL	NULL	NULL	NULL	NULL	NULL
1z1Hg7Vb0AhHDE791	All the Time - Don Diablo Remix	Zara Larsson	70	1HoSmj2eLcsrR0vE9gThr4	NULL	NULL	NULL	NULL	NULL	NULL	NULL
75FpbthrwQmzHlBJLuGdC7	Call You Mine - Keanu Silva Remix	The Chainsmokers	60	1nqYsOeflyKKuGOVchbsk6	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Time taken: 3.884 seconds, Fetched: 5 row(s)

hive>

>

>

>

> set hive.cli.print.header;

hive.cli.print.header=false

hive> set hive.cli.print.header = true;

hive> set hive.cli.print.header;

hive.cli.print.header=true

hive> SELECT * FROM brewery_data LIMIT 5;

OK

brewery_data.batch_id	brewery_data.brew_date	brewery_data.beer_style	brewery_data.sku	brewery_data.location	brewery_data.fermentation_time	brewery_data.temperature	brewery_data.ph_level	brewery_data.gravity	brewery_data.alcohol_content	brewery_data.bitterness	brewery_data.color	brewery_data.ingredient_ratio	brewery_data.volume_produced	brewery_data.total_sales	brewery_data.quality_score	brewery_data.brewhouse_efficiency	brewery_data.loss_during_brewing	brewery_data.loss_during_fermentation	brewery_data.loss_during_bottling_kegging
-----------------------	------------------------	-------------------------	------------------	-----------------------	--------------------------------	--------------------------	-----------------------	----------------------	------------------------------	-------------------------	--------------------	-------------------------------	------------------------------	--------------------------	----------------------------	-----------------------------------	----------------------------------	---------------------------------------	-------------------------------------------


```
>
> CREATE TABLE IF NOT EXISTS cleaned_brewery_data AS
> SELECT
>   COALESCE(Batch_ID, '0') AS Batch_ID,
>   COALESCE(Brew_Date, 'Unknown') AS Brew_Date,
>   COALESCE(Beer_Style, 'Unknown') AS Beer_Style,
>   COALESCE(SKU, 'Unknown') AS SKU,
>   COALESCE(Location, 'Unknown') AS Location,
>   COALESCE(Fermentation_Time, 0) AS Fermentation_Time,
>   COALESCE(Temperature, 0.0) AS Temperature,
>   COALESCE(pH_Level, 0.0) AS pH_Level,
>   COALESCE(Gravity, 0.0) AS Gravity,
>   COALESCE(Alcohol_Content, 0.0) AS Alcohol_Content,
>   COALESCE(Bitterness, 0) AS Bitterness,
>   COALESCE(Color, 0) AS Color,
>   COALESCE(Ingredient_Ratio, 'Unknown') AS Ingredient_Ratio,
>   COALESCE(Volume_Produced, 0) AS Volume_Produced,
>   COALESCE(Total_Sales, 0.0) AS Total_Sales,
>   COALESCE(Quality_Score, 0.0) AS Quality_Score,
>   COALESCE(Brewhouse_Efficiency, 0.0) AS Brewhouse_Efficiency,
>   COALESCE(Loss_During_Brewing, 0.0) AS Loss_During_Brewing,
>   COALESCE(Loss_During_Fermentation, 0.0) AS Loss_During_Fermentation,
>   COALESCE(Loss_During_Bottling_Kegging, 0.0) AS Loss_During_Bottling_Kegging
> FROM brewery_data;
```

Query ID = jaygames0143_20231212211913_cb94a295-5e78-4ad4-bdff-8afda7539e4a
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1702254398360_0012)

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
----------	------	--------	-------	-----------	---------	---------	--------	--------

```
> COALESCE(Alcohol_Content, 0.0) AS Alcohol_Content,
> COALESCE(Bitterness, 0) AS Bitterness,
> COALESCE(Color, 0) AS Color,
> COALESCE(Ingredient_Ratio, 'Unknown') AS Ingredient_Ratio,
> COALESCE(Volume_Produced, 0) AS Volume_Produced,
> COALESCE(Total_Sales, 0.0) AS Total_Sales,
> COALESCE(Quality_Score, 0.0) AS Quality_Score,
> COALESCE(Brewhouse_Efficiency, 0.0) AS Brewhouse_Efficiency,
> COALESCE(Loss_During_Brewing, 0.0) AS Loss_During_Brewing,
> COALESCE(Loss_During_Fermentation, 0.0) AS Loss_During_Fermentation,
> COALESCE(Loss_During_Bottling_Kegging, 0.0) AS Loss_During_Bottling_Kegging
> FROM brewery_data;
Query ID = jaygames0143_20231212211913_cb94a295-5e78-4ad4-bdff-8afda7539e4a
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1702254398360_0012)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED      4          4          0          0          0          0
-----
VERTICES: 01/01 [=====>>] 100%  ELAPSED TIME: 232.63 s
-----
Moving data to directory hdfs://finalprojectcluster2-m/user/hive/warehouse/cleaned_brewery_data
OK
batch_id      brew_date      beer_style      sku      location      fermentation_time      temperature      ph_level      gravi
ty      alcohol_content      bitterness      color      ingredient_ratio      volume_produced      total_sales      quality_score      brewhouse_eff
iciency      loss_during_brewing      loss_during_fermentation      loss_during_bottling_kegging
Time taken: 257.824 seconds
hive>
```

```
ty      alcohol_content bitterness      color  ingredient_ratio      volume_produced total_sales      quality_score  brewhouse_eff
iciency loss_during_brewing      loss_during_fermentation      loss_during_bottling_kegging
Time taken: 257.824 seconds
hive>
>
> INSERT OVERWRITE DIRECTORY 'gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv'
> SELECT * FROM
Display all 634 possibilities? (y or n)
> SELECT * FROM cleaned_brewery_data;
Query ID = jaygames0143_20231212213150_68eafb63-ec0e-4446-b4af-14ea6144bd97
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1702254398360_0013)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	4	4	0	0	0	0

```
VERTICES: 01/01 [=====>>] 100% ELAPSED TIME: 250.11 s
Moving data to directory gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv
OK
cleaned_brewery_data.batch_id      cleaned_brewery_data.brew_date      cleaned_brewery_data.beer_style      cleaned_brewery_data.sku      clean
ed_brewery_data.location      cleaned_brewery_data.fermentation_time      cleaned_brewery_data.temperature      cleaned_brewery_data.
ph_level      cleaned_brewery_data.gravity      cleaned_brewery_data.alcohol_content      cleaned_brewery_data.bitterness      cleaned_brew
ry_data.color      cleaned_brewery_data.ingredient_ratio      cleaned_brewery_data.volume_produced      cleaned_brewery_data.total_sales      c
leaned_brewery_data.quality_score      cleaned_brewery_data.brewhouse_efficiency      cleaned_brewery_data.loss_during_brewing      c
leaned_brewery_data.loss_during_fermentation      cleaned_brewery_data.loss_during_bottling_kegging
Time taken: 278.854 seconds
hive> █
```



```
leaned_brewery_data.quality_score      cleaned_brewery_data.brewhouse_efficiency      cleaned_brewery_data.loss_during_brewing      c
leaned_brewery_data.loss_during_fermentation      cleaned_brewery_data.loss_during_bottling_kegging
Time taken: 278.854 seconds
hive> exit
> ;
jaygames0143@finalprojectcluster2-m:~$
jaygames0143@finalprojectcluster2-m:~$ hadoop fs -ls 'gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv'
Dec 12, 2023 10:31:30 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_get_file_status. latencyMs=606; previousMaxLatencyMs=0; operationCount=1; context=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv
Dec 12, 2023 10:31:31 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_glob_status. latencyMs=783; previousMaxLatencyMs=0; operationCount=1; context=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv; pattern=com.google.cloud.hadoop.fs.gcs.GoogleHadoopFileSystemBase$Lambda$75/0x000000010021f040@2459319c
Found 4 items
-rwx----- 3 jaygames0143 jaygames0143 805306551 2023-12-12 21:36 gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000000_0
-rwx----- 3 jaygames0143 jaygames0143 805306559 2023-12-12 21:36 gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000001_0
-rwx----- 3 jaygames0143 jaygames0143 799153565 2023-12-12 21:36 gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000002_0
-rwx----- 3 jaygames0143 jaygames0143 207433992 2023-12-12 21:36 gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000003_0
jaygames0143@finalprojectcluster2-m:~$ hadoop fs -mv 'gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000000_0' 'gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_1';
Dec 12, 2023 10:32:51 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_get_file_status. latencyMs=546; previousMaxLatencyMs=0; operationCount=1; context=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_1
Dec 12, 2023 10:32:51 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_glob_status. latencyMs=703; previousMaxLatencyMs=0; operationCount=1; context=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_1; pattern=com.google.cloud.hadoop.fs.gcs.GoogleHadoopFileSystemBase$Lambda$74/0x000000010021fc40@4aedaf61
Dec 12, 2023 10:32:52 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_rename. latencyMs=474; previousMaxLatencyMs=0; operationCount=1; context=renam
```




```
path=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_1; pattern=com.google.cloud.hadoop.fs.gcs.GoogleHadoopFileSy
stemBase$Lambda$74/0x000000010021fc40@4aedaf61
Dec 12, 2023 10:32:52 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_rename. latencyMs=474; previousMaxLatencyMs=0; operationCount=1; context=renam
e(gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000000_0 -> gs://final_project_bucket2/brewery_data_cleaned_using_hi
ve.csv/Part_1)
jaygames0143@finalprojectcluster2-m:~$ hadoop fs -mv 'gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000001_0' 'gs://
final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_2';
Dec 12, 2023 10:33:35 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_get_file_status. latencyMs=629; previousMaxLatencyMs=0; operationCount=1; cont
ext=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_2
Dec 12, 2023 10:33:35 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_glob_status. latencyMs=800; previousMaxLatencyMs=0; operationCount=1; context=
path=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_2; pattern=com.google.cloud.hadoop.fs.gcs.GoogleHadoopFileSy
stemBase$Lambda$74/0x000000010021fc40@3ed03652
Dec 12, 2023 10:33:36 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_rename. latencyMs=490; previousMaxLatencyMs=0; operationCount=1; context=renam
e(gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000001_0 -> gs://final_project_bucket2/brewery_data_cleaned_using
ve.csv/Part_2)
jaygames0143@finalprojectcluster2-m:~$ hadoop fs -mv 'gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000002_0' 'gs://
final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_3';
Dec 12, 2023 10:33:51 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_get_file_status. latencyMs=610; previousMaxLatencyMs=0; operationCount=1; cont
ext=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_3
Dec 12, 2023 10:33:51 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_glob_status. latencyMs=762; previousMaxLatencyMs=0; operationCount=1; context=
path=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_3; pattern=com.google.cloud.hadoop.fs.gcs.GoogleHadoopFileSy
stemBase$Lambda$74/0x000000010021fc40@3ed03652
Dec 12, 2023 10:33:51 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_rename. latencyMs=507; previousMaxLatencyMs=0; operationCount=1; context=renam
e(gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000002_0 -> gs://final_project_bucket2/brewery_data_cleaned_using_hi
ve.csv/Part_3)
jaygames0143@finalprojectcluster2-m:~$ hadoop fs -mv 'gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000003_0' 'gs://
final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_4';
```




```
INFO: Detected potential high latency for operation op_glob_status. latencyMs=800; previousMaxLatencyMs=0; operationCount=1; context=
path=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_2; pattern=com.google.cloud.hadoop.fs.gcs.GoogleHadoopFileSy
stemBase$Lambda$74/0x000000010021fc40@3ed03652
Dec 12, 2023 10:33:36 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_rename. latencyMs=490; previousMaxLatencyMs=0; operationCount=1; context=rename
(gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000001_0 -> gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_2)
jaygames0143@finalprojectcluster2-m:~$ hadoop fs -mv 'gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000002_0' 'gs://
final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_3';
Dec 12, 2023 10:33:51 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_get_file_status. latencyMs=610; previousMaxLatencyMs=0; operationCount=1; context=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_3
Dec 12, 2023 10:33:51 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_glob_status. latencyMs=762; previousMaxLatencyMs=0; operationCount=1; context=
path=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_3; pattern=com.google.cloud.hadoop.fs.gcs.GoogleHadoopFileSy
stemBase$Lambda$74/0x000000010021fc40@3ed03652
Dec 12, 2023 10:33:51 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_rename. latencyMs=507; previousMaxLatencyMs=0; operationCount=1; context=rename
(gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000002_0 -> gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_3)
jaygames0143@finalprojectcluster2-m:~$ hadoop fs -mv 'gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000003_0' 'gs://
final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_4';
Dec 12, 2023 10:34:05 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_get_file_status. latencyMs=508; previousMaxLatencyMs=0; operationCount=1; context=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_4
Dec 12, 2023 10:34:05 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_glob_status. latencyMs=708; previousMaxLatencyMs=0; operationCount=1; context=
path=gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_4; pattern=com.google.cloud.hadoop.fs.gcs.GoogleHadoopFileSy
stemBase$Lambda$74/0x000000010021fc40@3ed03652
Dec 12, 2023 10:34:06 PM com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics updateMinMaxStats
INFO: Detected potential high latency for operation op_rename. latencyMs=467; previousMaxLatencyMs=0; operationCount=1; context=rename
(gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/000003_0 -> gs://final_project_bucket2/brewery_data_cleaned_using_hive.csv/Part_4)
jaygames0143@finalprojectcluster2-m:~$
```