

```
[cloudera@quickstart ~]$ # Create directory in HDFS
[cloudera@quickstart ~]$ hdfs dfs -mkdir -p /user/cloudera/logs

[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ # Copy local file into HDFS
[cloudera@quickstart ~]$ hdfs dfs -put logs.txt /user/cloudera/logs/

[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hdfs dfs -ls /user/cloudera/logs/
Found 1 items
-rw-r--r--    1 cloudera cloudera          192 2025-09-22 23:48
/user/cloudera/logs/logs.txt
[cloudera@quickstart ~]$ hive
```

```
Logging initialized using configuration in
file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> CREATE DATABASE IF NOT EXISTS logdb;
```

```
OK
```

```
Time taken: 1.399 seconds
```

```
hive> USE logdb;
```

```
OK
```

```
Time taken: 0.107 seconds
```

```
hive>
```

```
> CREATE EXTERNAL TABLE logs_structured (
>   log_time STRING,
>   log_level STRING,
>   user STRING,
>   action STRING
> )
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ' '
> LINES TERMINATED BY '\n'
> STORED AS TEXTFILE
> LOCATION '/user/cloudera/logs/';
```

```
OK
```

```
Time taken: 1.02 seconds
```

```
hive> -- View for error logs only
```

```
> CREATE VIEW error_logs AS
> SELECT *
> FROM logs_structured
```

```

    > WHERE log_level = 'ERROR';
OK
Time taken: 1.436 seconds
hive>
    > -- View for user login actions
    > CREATE VIEW login_logs AS
    > SELECT *
    > FROM logs_structured
    > WHERE action LIKE 'logged%';
OK
Time taken: 2.218 seconds
hive> -- Create an index on log_level for faster querying
    > CREATE INDEX idx_log_level
    > ON TABLE logs_structured (log_level)
    > AS 'COMPACT'
    > WITH DEFERRED REBUILD;
OK
Time taken: 1.864 seconds
hive>

hive>
    > -- Rebuild index
    > ALTER INDEX idx_log_level ON logs_structured REBUILD;
Query ID = cloudera_20250922235050_2bd3cc0d-378d-4eff-aa1d-86070b592d32
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
    set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
    set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
    set mapreduce.job.reduces=<number>
Starting Job = job_1758605977300_0017, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1758605977300_0017/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1758605977300_0017
Hadoop job information for Stage-1: number of mappers: 1; number of
reducers: 1
2025-09-22 23:51:06,479 Stage-1 map = 0%, reduce = 0%
2025-09-22 23:51:43,284 Stage-1 map = 100%, reduce = 0%, Cumulative CPU
5.27 sec
2025-09-22 23:52:04,474 Stage-1 map = 100%, reduce = 100%, Cumulative CPU
10.02 sec

```

```

MapReduce Total cumulative CPU time: 10 seconds 20 msec
Ended Job = job_1758605977300_0017
Loading data to table logdb.logdb__logs_structured_idx_log_level__
Table logdb.logdb__logs_structured_idx_log_level__ stats: [numFiles=1,
numRows=5, totalSize=355, rawDataSize=350]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 10.02 sec HDFS Read:
8594 HDFS Write: 456 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 20 msec
OK
Time taken: 85.186 seconds
hive> -- Select all logs
> SELECT * FROM logs_structured;
OK
2025-09-23 10:00:00 INFO User
2025-09-23 10:05:00 ERROR Failed
2025-09-23 10:10:00 INFO User
2025-09-23 10:15:00 INFO User
NULL NULL NULL
Time taken: 0.267 seconds, Fetched: 5 row(s)
hive>
> -- Select only INFO logs
> SELECT * FROM logs_structured WHERE log_level='INFO';
OK
Time taken: 0.178 seconds
hive>
hive>
> -- Count logs per user
> SELECT user, COUNT(*) as total_logs
> FROM logs_structured
> GROUP BY user;
Query ID = cloudera_20250922235252_42923d71-a66d-4b51-bb63-8dbe4e5e2947
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
    set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
    set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
    set mapreduce.job.reduces=<number>
Starting Job = job_1758605977300_0018, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1758605977300_0018/

```

```
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1758605977300_0018
Hadoop job information for Stage-1: number of mappers: 1; number of
reducers: 1
2025-09-22 23:52:33,373 Stage-1 map = 0%, reduce = 0%
2025-09-22 23:52:49,681 Stage-1 map = 100%, reduce = 0%, Cumulative CPU
4.44 sec
2025-09-22 23:53:07,566 Stage-1 map = 100%, reduce = 100%, Cumulative CPU
8.13 sec
MapReduce Total cumulative CPU time: 8 seconds 130 msec
Ended Job = job_1758605977300_0018
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.13 sec HDFS Read:
7232 HDFS Write: 20 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 130 msec
OK
NULL 1
ERROR 1
INFO 3
Time taken: 54.835 seconds, Fetched: 3 row(s)
hive>
    > -- Query using view
    > SELECT * FROM error_logs;
OK
Time taken: 0.343 seconds
hive>
```