Hive HQL Hands-on: Logs & Errors Analysis

Create External Table for Log Data

```
cloudera@localhost:~
File  Edit  View  Search  Terminal  Help
[cloudera@localhost ~]$ hive
Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-comm
n-0.10.0-cdh4.4.0.jar!/hive-log4j.properties
Hive history file=/tmp/cloudera/hive_job_log_b8404494-8a73-4bc8-abf0-9b186be62e
0_248843411.txt
hive> CREATE EXTERNAL TABLE IF NOT EXISTS logs (
log_date STRING,
log_level STRING,
service_name STRING,
message STRING
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/hive/logs/';
OK
Time taken: 2.073 seconds
hive>
```

View All Logs

```
hive> select * from logs;
OK
2025-05-01 11:39:00     INFO    InventoryService        Order placed
2025-05-01 11:25:00     WARN    PaymentService  Address not verified
2025-05-01 12:49:00     ERROR   ShippingService Shipping error
2025-05-01 14:33:00     WARN    ShippingService Slow response time
2025-05-01 10:27:00     INFO    ShippingService Inventory updated
2025-05-01 14:11:00     INFO    PaymentService  Order placed
2025-05-01 12:28:00     ERROR   OrderService    Payment failed
2025-05-01 13:54:00     ERROR   OrderService    Payment failed
2025-05-01 14:54:00     ERROR   AuthService     Order placement failed
2025-05-01 10:54:00     WARN    OrderService    Address not verified
2025-05-01 14:38:00     ERROR   ShippingService Payment failed
2025-05-01 11:18:00     INFO    OrderService    Order placed
2025-05-01 12:44:00     INFO    PaymentService  Order placed
```

Count Total Logs

```
2025-05-17 06:43:56,167 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.66
c
2025-05-17 06:43:57,185 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.7
sec
2025-05-17 06:43:58,199 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.7
sec
2025-05-17 06:43:59,214 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.7
sec
2025-05-17 06:44:00,222 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.7
sec
2025-05-17 06:44:01,230 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.7
sec
MapReduce Total cumulative CPU time: 2 seconds 770 msec
Ended Job = job_202505170500_0001
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 2.77 sec   HDFS Read: 6357 HDFS Wri
: 4 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 770 msec
OK
100
Time taken: 26.837 seconds
hive>
```

Count Logs by Log Level

```
2025-05-17 00:47:20,291 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.
sec
MapReduce Total cumulative CPU time: 2 seconds 740 msec
Ended Job = job_202505170500_0002
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 2.74 sec   HDFS Read: 6357 HDFS Wr
: 25 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 740 msec
OK
ERROR   38
INFO    33
WARN    29
Time taken: 20.891 seconds
hive>
```

Count ERROR logs per service

```
Ended Job = job_202505170500_0003
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 1.85 sec   HDFS Read: 6357 HDFS Write
: 85 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 850 msec
OK
AuthService      4
InventoryService         5
OrderService     10
PaymentService  8
ShippingService 11
Time taken: 20.008 seconds
hive>
```

Filter Logs by Date

```
Total MapReduce CPU Time Spent: 920 msec
OK
2025-05-01 11:39:00     INFO    InventoryService        Order placed
2025-05-01 11:25:00     WARN    PaymentService  Address not verified
2025-05-01 12:49:00     ERROR   ShippingService Shipping error
2025-05-01 14:33:00     WARN    ShippingService Slow response time
2025-05-01 10:27:00     INFO    ShippingService Inventory updated
2025-05-01 14:11:00     INFO    PaymentService  Order placed
2025-05-01 12:28:00     ERROR   OrderService    Payment failed
2025-05-01 13:54:00     ERROR   OrderService    Payment failed
2025-05-01 14:54:00     ERROR   AuthService     Order placement failed
2025-05-01 10:54:00     WARN    OrderService    Address not verified
2025-05-01 14:38:00     ERROR   ShippingService Payment failed
2025-05-01 11:18:00     INFO    OrderService    Order placed
2025-05-01 12:44:00     INFO    PaymentService  Order placed
```

Create Partitioned Table by Date

```
hive> CREATE EXTERNAL TABLE logs_partitioned (
log_time STRING,
log_level STRING,
service_name STRING,
message STRING
)
PARTITIONED BY (log_date STRING)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE;
OK
Time taken: 0.839 seconds
```

Create a View for ERROR Logs

```
Total MapReduce CPU Time Spent: 630 msec
OK
2025-05-01 12:49:00     ERROR   ShippingService Shipping error
2025-05-01 12:28:00     ERROR   OrderService    Payment failed
2025-05-01 13:54:00     ERROR   OrderService    Payment failed
2025-05-01 14:54:00     ERROR   AuthService     Order placement failed
2025-05-01 14:38:00     ERROR   ShippingService Payment failed
2025-05-01 10:51:00     ERROR   ShippingService Order placement failed
2025-05-01 11:28:00     ERROR   PaymentService  Order placement failed
2025-05-01 12:04:00     ERROR   ShippingService User authentication failed for u
ser 105
2025-05-01 13:56:00     ERROR   OrderService    Order placement failed
2025-05-01 13:32:00     ERROR   InventoryService        Shipping error
2025-05-01 11:27:00     ERROR   InventoryService        Inventory sync error
2025-05-01 13:03:00     ERROR   ShippingService Shipping error
2025-05-01 10:28:00     ERROR   OrderService    Order placement failed
2025-05-01 10:51:00     ERROR   InventoryService        Inventory sync error
2025-05-01 12:41:00     ERROR   OrderService    Shipping error
2025-05-01 14:49:00     ERROR   ShippingService User authentication failed for u
ser 106
```