pepperdata

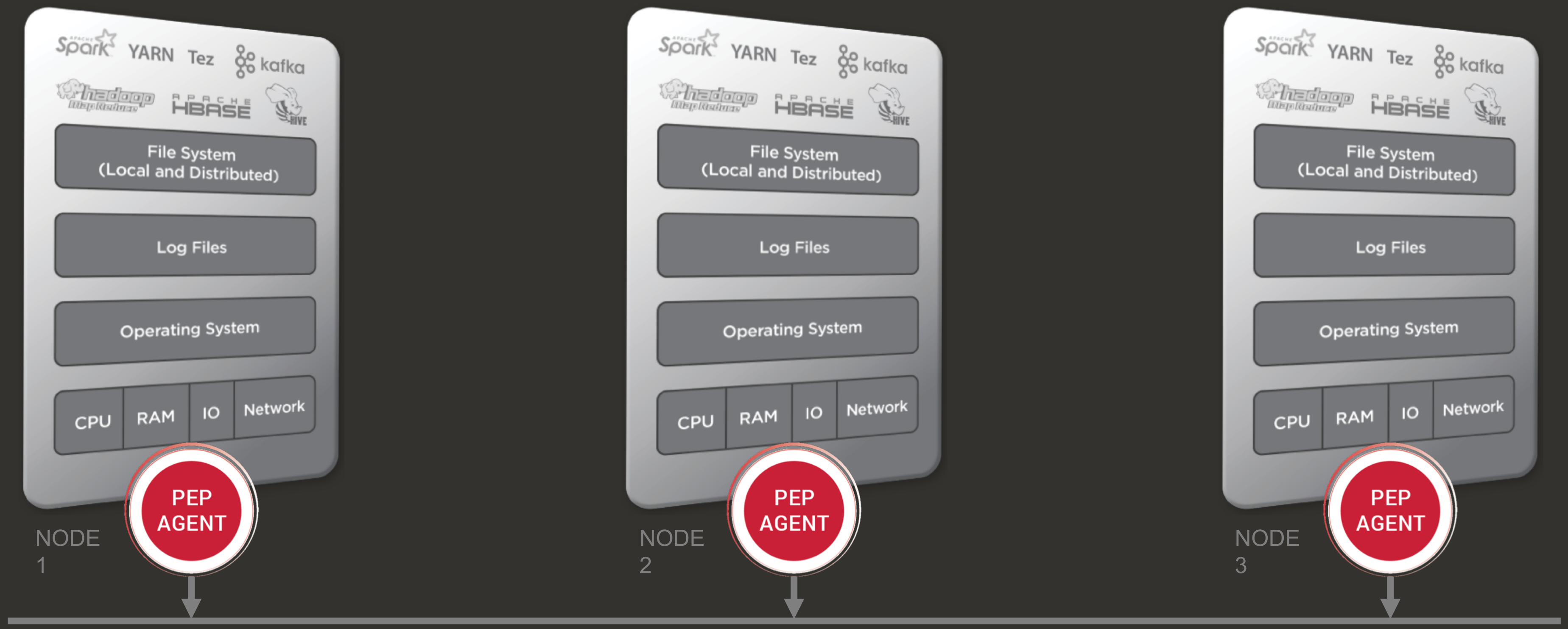
Optimize

Your Big Data Infrastructure, Your Apps, and Your Time

at Scale

On-Premise, Cloud, or Hybrid

Pep Agent | Get ALL of the metrics



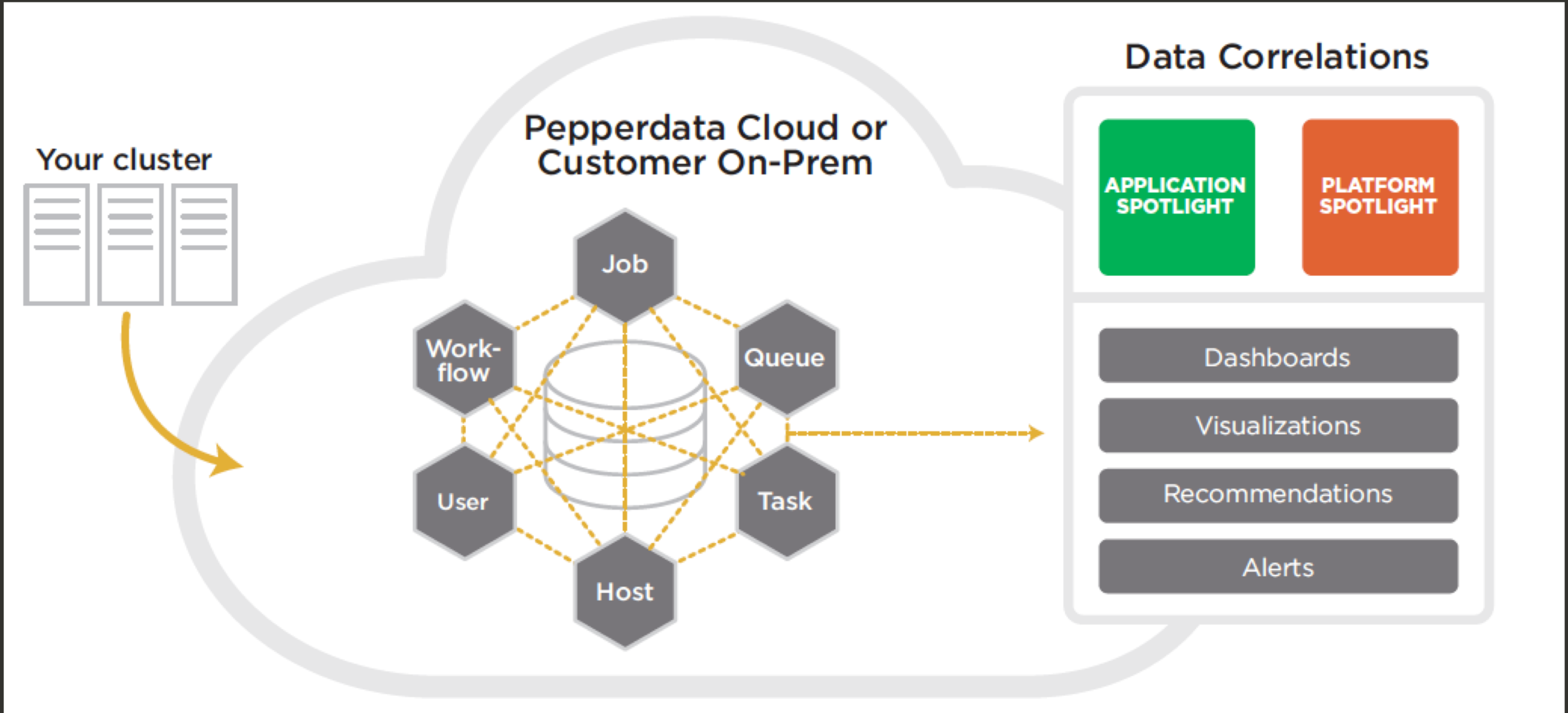
A Pep Agent on every node Continuously collects hundreds of metrics in real-time across the entire stack.

This unique real-time data is Correlated and powers Platform Spotlight and Application Spotlight.

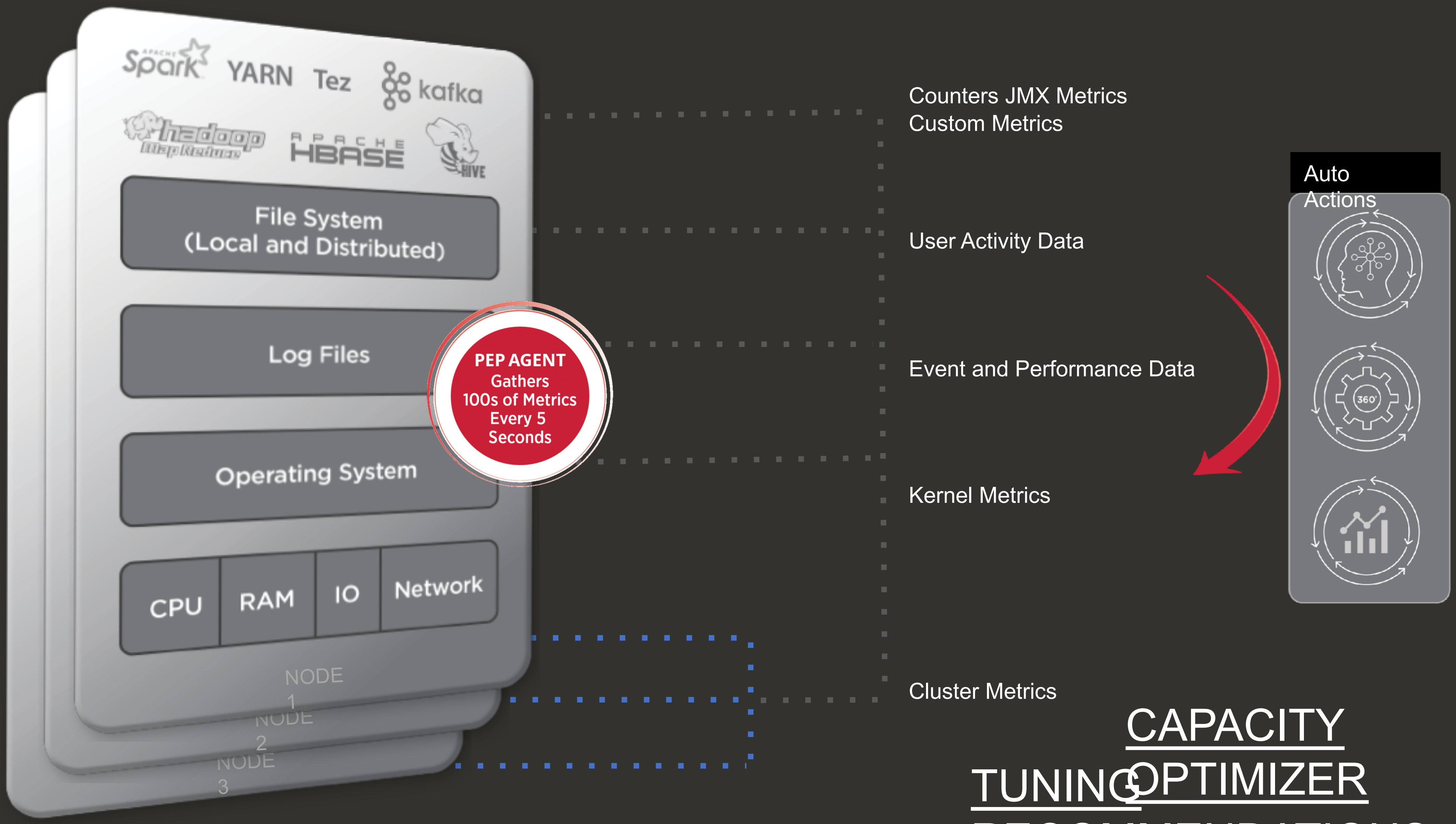
This unique real-time data also powers auto-tuning for capacity and application performance optimization

Each Pep Agent is 1% of a single core and roughly 300 MB of RAM per node

Pepperdata **Metrics Platform**



Pepperdata Auto-Actions



File System (Local and Distributed)

Log Files

Operating System

CPU | RAM | IO | Network

NODE 1

NODE 2

NODE 3

PEP AGENT
Gathers 100s of Metrics Every 5 Seconds

Counters JMX Metrics
Custom Metrics

User Activity Data

Event and Performance Data

Kernel Metrics

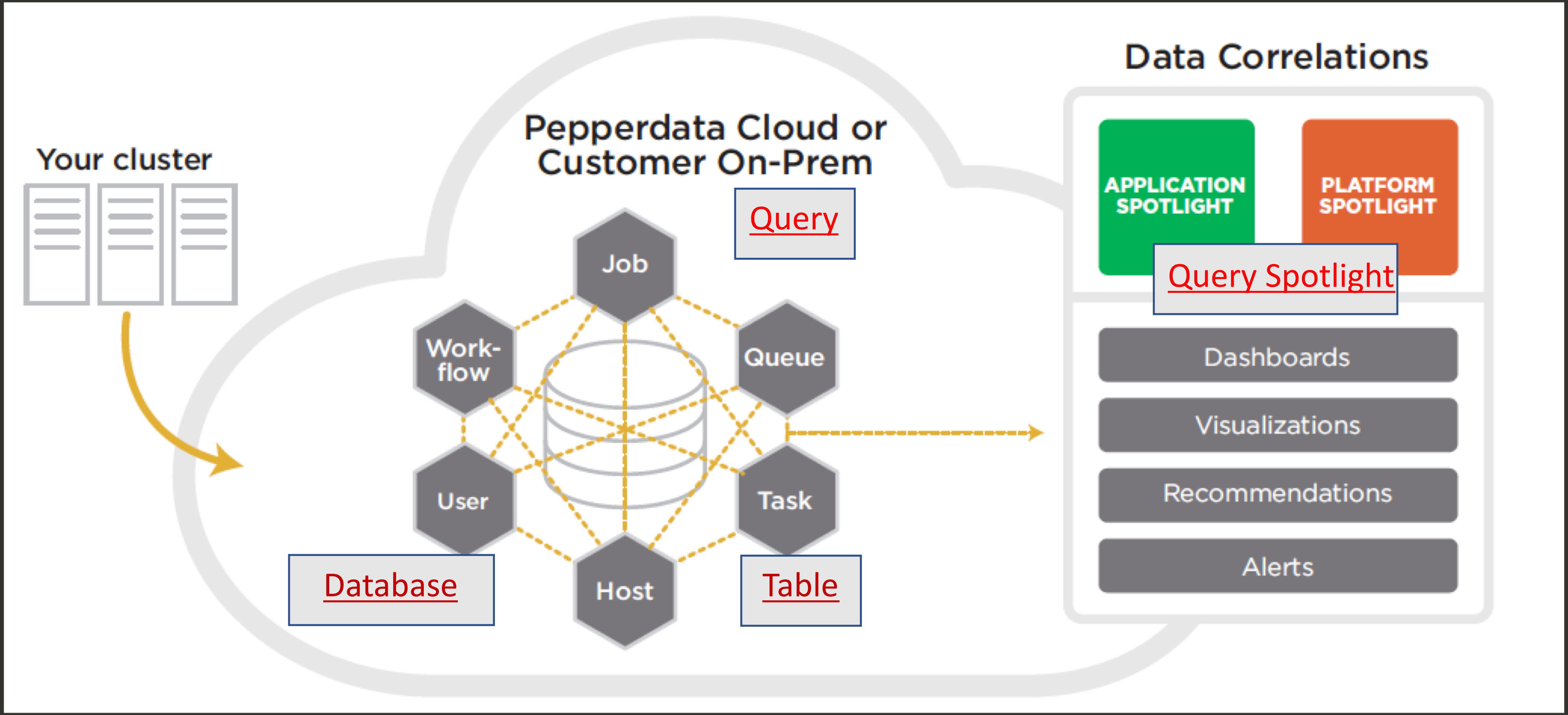
Cluster Metrics

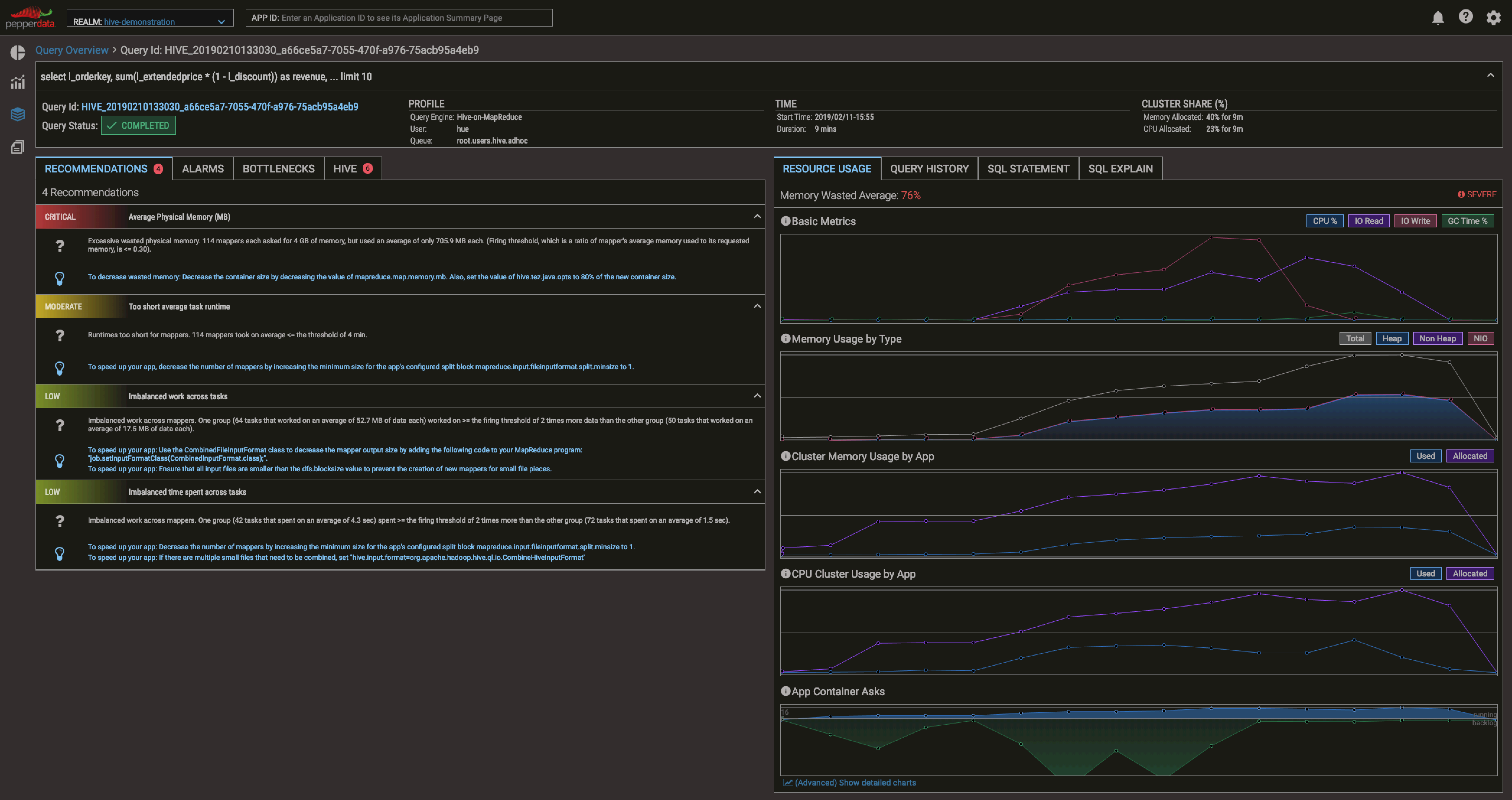
Auto Actions

CAPACITY OPTIMIZER

TUNING RECOMMENDATIONS

Pepperdata **QUERY SPOTLIGHT**




pepperdata
REALM: hive-demonstration
APP ID: Enter an Application ID to see its Application Summary Page

Query Overview > Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9

select l_orderkey, sum(l_extendedprice * (1 - l_discount)) as revenue, ... limit 10

Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9
Query Status: ✓ COMPLETED

PROFILE
Query Engine: Hive-on-MapReduce
User: hue
Queue: root.users.hive.adhoc

TIME
Start Time: 2019/02/11-15:55
Duration: 9 mins

CLUSTER SHARE (%)
Memory Allocated: 40% for 9m
CPU Allocated: 23% for 9m

4 Recommendations

CRITICAL — Average Physical Memory (MB)

? Excessive wasted physical memory. 114 mappers each asked for 4 GB of memory, but used an average of only 705.9 MB each. (Firing threshold, which is a ratio of mapper's average memory used to its requested memory, is <= 0.30).

To decrease wasted memory: Decrease the container size by decreasing the value of mapreduce.map.memory.mb. Also, set the value of hive.tez.java.opts to 80% of the new container size.

MODERATE — Too short average task runtime

? Runtimes too short for mappers. 114 mappers took on average <= the threshold of 4 min.

To speed up your app, decrease the number of mappers by increasing the minimum size for the app's configured split block mapreduce.input.fileinputformat.split.minsize to 1.

LOW — Imbalanced work across tasks

? Imbalanced work across mappers. One group (64 tasks that worked on an average of 52.7 MB of data each) worked on >= the firing threshold of 2 times more data than the other group (50 tasks that worked on an average of 17.5 MB of data each).

To speed up your app: Use the CombinedFileInputFormat class to decrease the mapper output size by adding the following code to your MapReduce program: "job.setInputFormatClass(CombinedInputFormat.class);".
To speed up your app: Ensure that all input files are smaller than the dfs.blocksize value to prevent the creation of new mappers for small file pieces.

LOW — Imbalanced time spent across tasks

? Imbalanced work across mappers. One group (42 tasks that spent on an average of 4.3 sec) spent >= the firing threshold of 2 times more than the other group (72 tasks that spent on an average of 1.5 sec).

To speed up your app: Decrease the number of mappers by increasing the minimum size for the app's configured split block mapreduce.input.fileinputformat.split.minsize to 1.
To speed up your app: If there are multiple small files that need to be combined, set "hive.input.format=org.apache.hadoop.hive.ql.io.CombineHiveInputFormat"

Memory Wasted Average: 76% — SEVERE

Basic Metrics

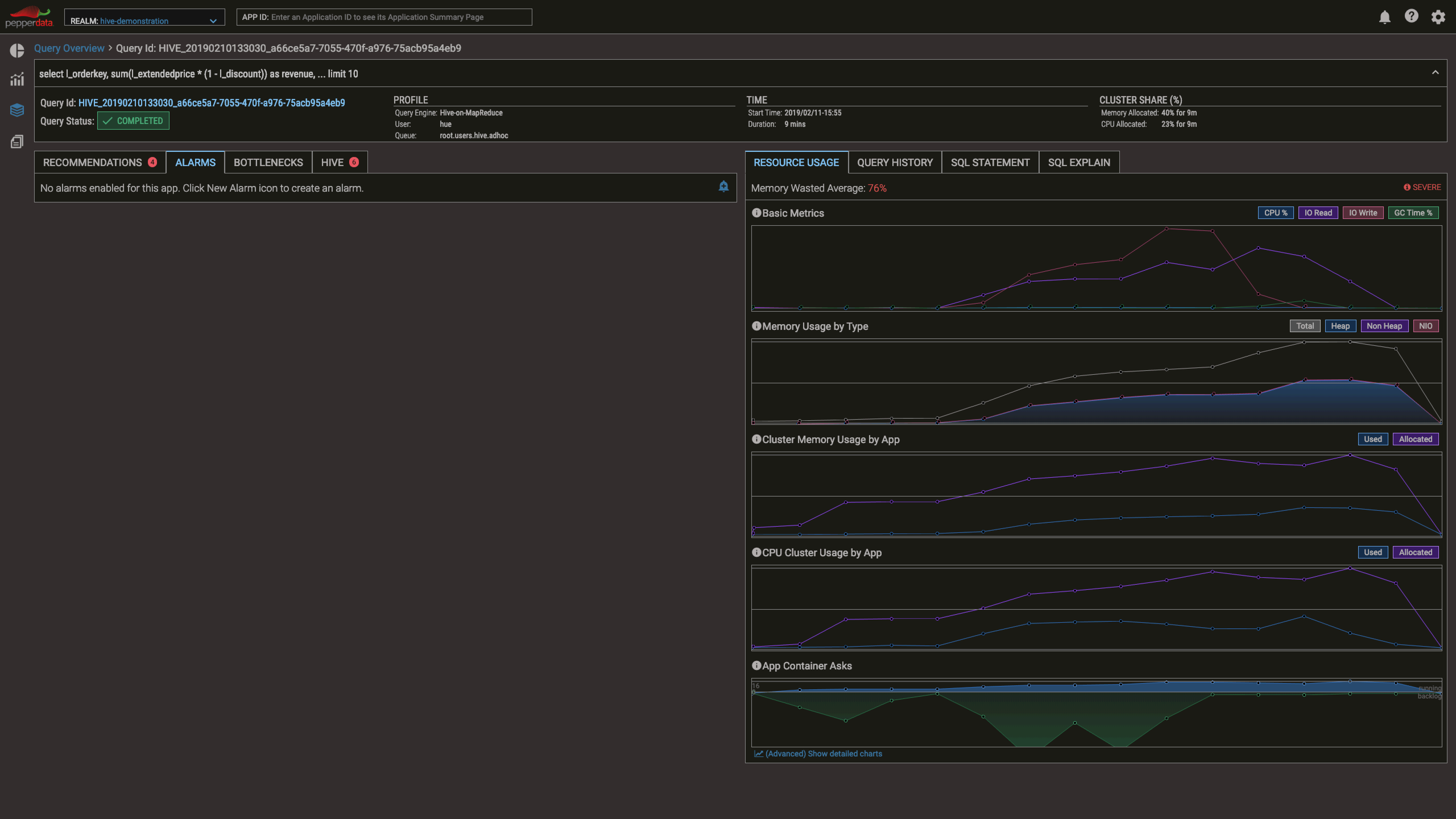
Memory Usage by Type

Cluster Memory Usage by App

CPU Cluster Usage by App

App Container Asks

(Advanced) Show detailed charts



pepperdata
REALM: hive-demonstration
APP ID: Enter an Application ID to see its Application Summary Page

Query Overview > Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9

select l_orderkey, sum(l_extendedprice * (1 - l_discount)) as revenue, ... limit 10

Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9
Query Status: ✓ COMPLETED

PROFILE
Query Engine: Hive-on-MapReduce
User: hue
Queue: root.users.hive.adhoc

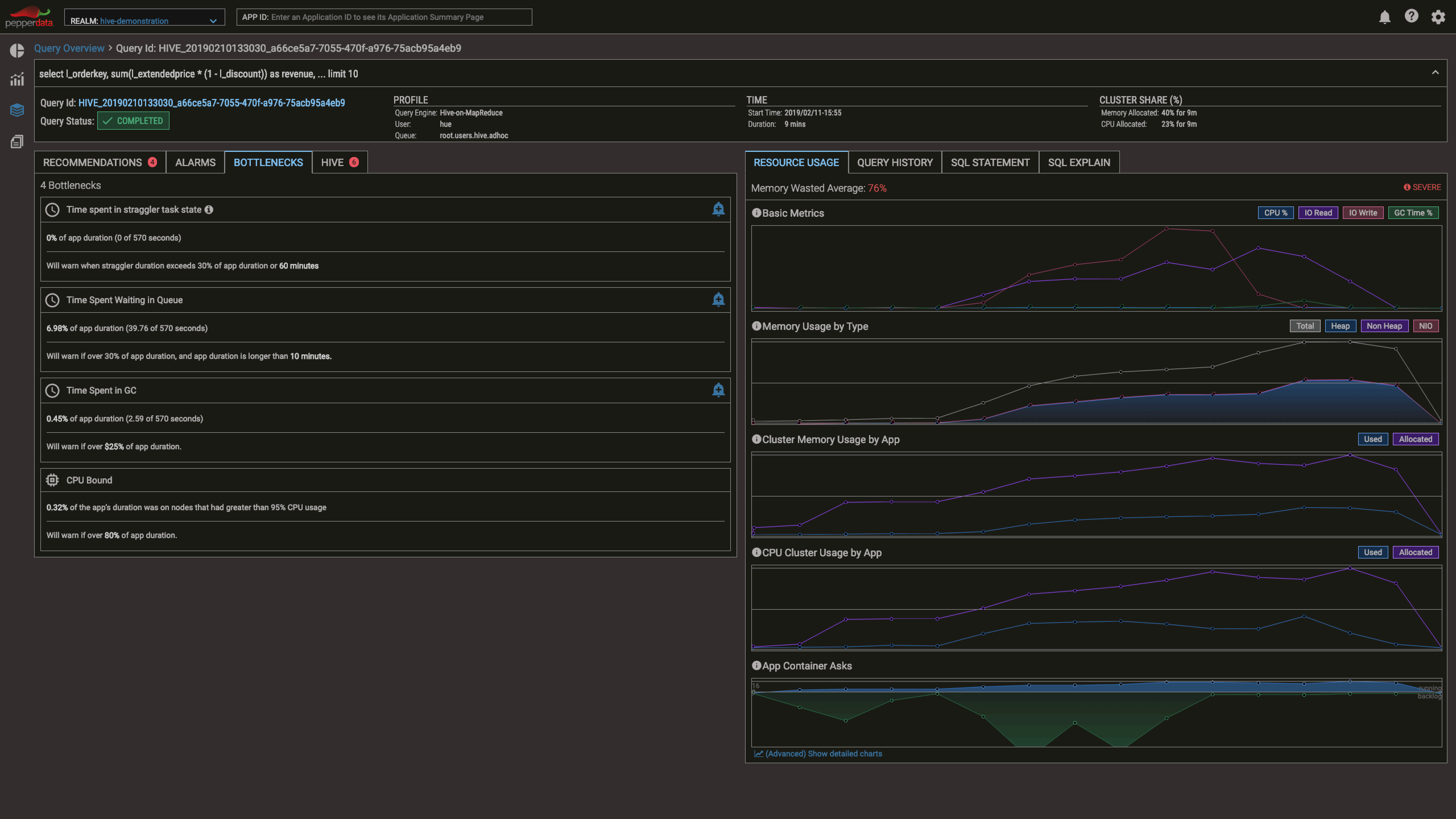
TIME
Start Time: 2019/02/11-15:55
Duration: 9 mins

CLUSTER SHARE (%)
Memory Allocated: 40% for 9m
CPU Allocated: 23% for 9m

No alarms enabled for this app. Click New Alarm icon to create an alarm.

Memory Wasted Average: 76%

SEVERE



pepperdata
REALM: hive-demonstration
APP ID: Enter an Application ID to see its Application Summary Page

Query Overview > Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9

select l_orderkey, sum(l_extendedprice * (1 - l_discount)) as revenue, ... limit 10

Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9

Query Status: ✓ COMPLETED

PROFILE

Query Engine: Hive-on-MapReduce

User: hue

Queue: root.users.hive.adhoc

TIME

Start Time: 2019/02/11-15:55

Duration: 9 mins

CLUSTER SHARE (%)

Memory Allocated: 40% for 9m

CPU Allocated: 23% for 9m

4 Bottlenecks

Time spent in straggler task state

0% of app duration (0 of 570 seconds)

Will warn when straggler duration exceeds 30% of app duration or 60 minutes

Time Spent Waiting in Queue

6.98% of app duration (39.76 of 570 seconds)

Will warn if over 30% of app duration, and app duration is longer than 10 minutes.

Time Spent in GC

0.45% of app duration (2.59 of 570 seconds)

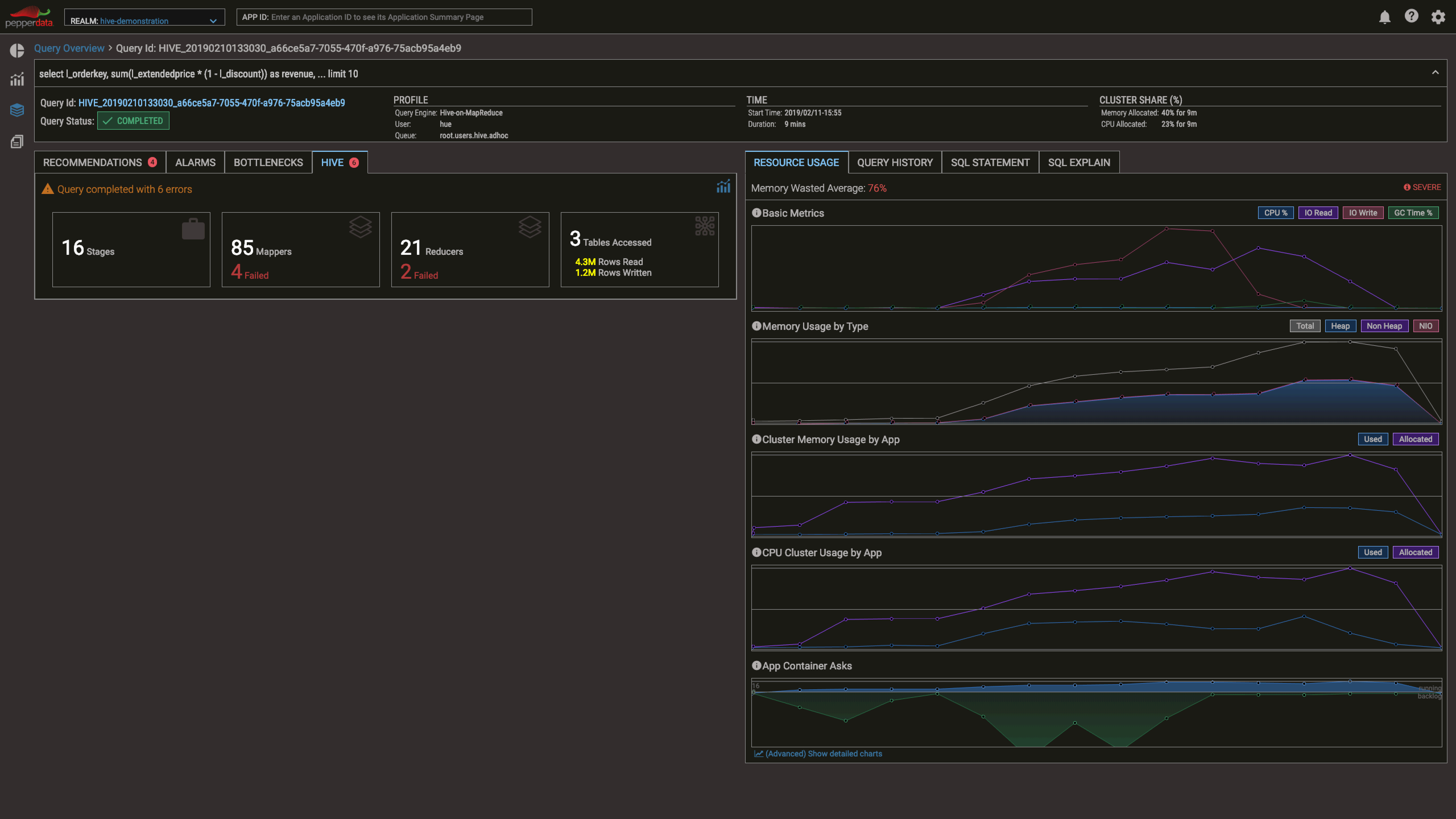
Will warn if over $25% of app duration.

CPU Bound

0.32% of the app's duration was on nodes that had greater than 95% CPU usage

Will warn if over 80% of app duration.

Memory Wasted Average: 76%



pepperdata
REALM: hive-demonstration
APP ID: Enter an Application ID to see its Application Summary Page

Query Overview > Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9

select l_orderkey, sum(l_extendedprice * (1 - l_discount)) as revenue, ... limit 10

Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9

Query Status: ✓ COMPLETED

PROFILE

Query Engine:	Hive-on-MapReduce
User:	hue
Queue:	root.users.hive.adhoc

TIME

Start Time:	2019/02/11-15:55
Duration:	9 mins

CLUSTER SHARE (%)

Memory Allocated:	40% for 9m
CPU Allocated:	23% for 9m

⚠ Query completed with 6 errors

16 Stages

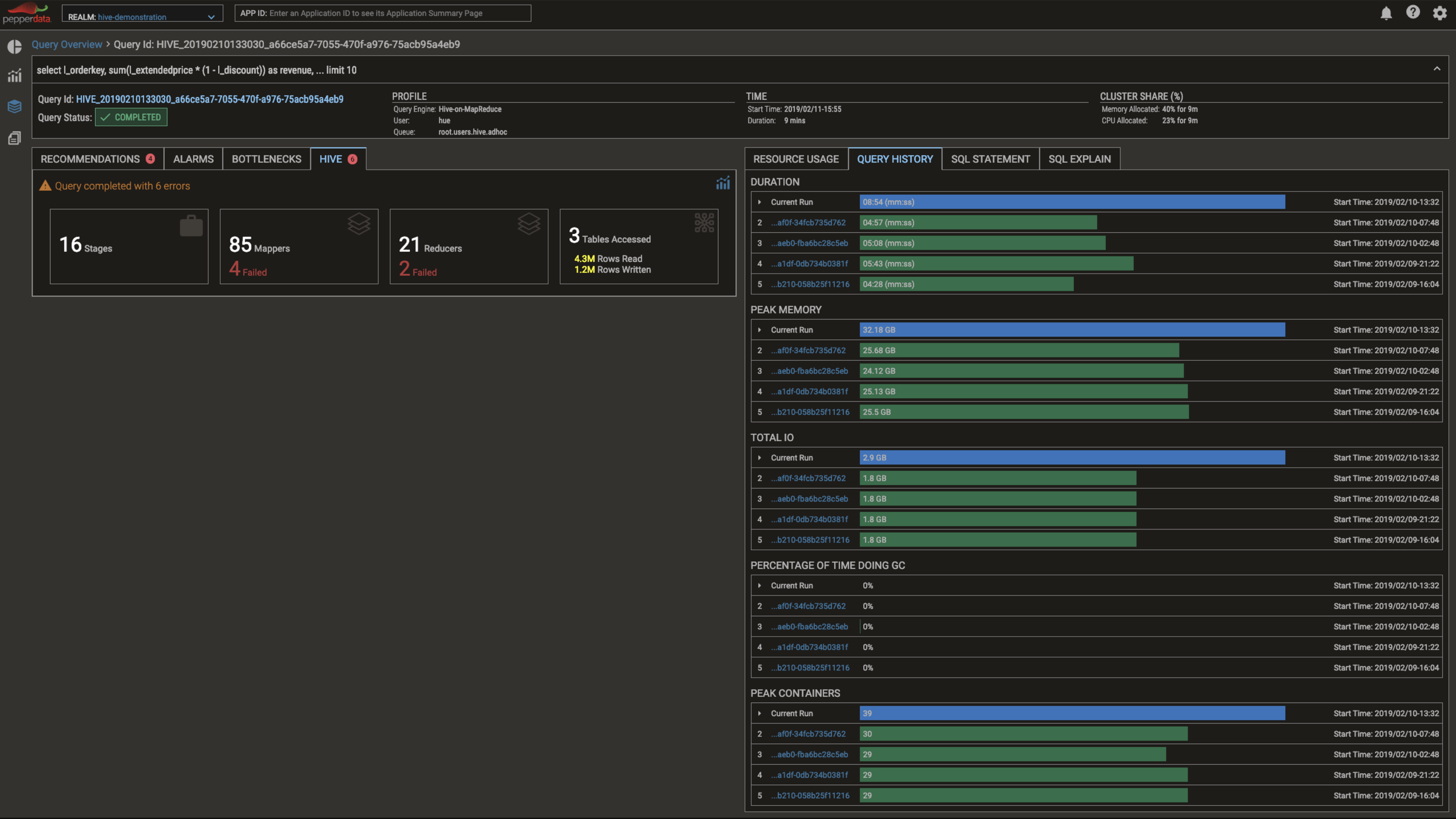
85 Mappers
4 Failed

21 Reducers
2 Failed

3 Tables Accessed
4.3M Rows Read
1.2M Rows Written

Memory Wasted Average: 76%

SEVERE



Query Overview > Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9

select l_orderkey, sum(l_extendedprice * (1 - l_discount)) as revenue, ... limit 10

Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9

Query Status: ✓ COMPLETED

PROFILE

Query Engine:	Hive-on-MapReduce
User:	hue
Queue:	root.users.hive.adhoc

TIME

Start Time:	2019/02/11-15:55
Duration:	9 mins

CLUSTER SHARE (%)

Memory Allocated:	40% for 9m
CPU Allocated:	23% for 9m

⚠ Query completed with 6 errors

16 Stages

85 Mappers
4 Failed

21 Reducers
2 Failed

3 Tables Accessed
4.3M Rows Read
1.2M Rows Written

DURATION

▸	Current Run	08:54 (mm:ss)	Start Time: 2019/02/10-13:32
2	...af0f-34fcb735d762	04:57 (mm:ss)	Start Time: 2019/02/10-07:48
3	...aeb0-fba6bc28c5eb	05:08 (mm:ss)	Start Time: 2019/02/10-02:48
4	...a1df-0db734b0381f	05:43 (mm:ss)	Start Time: 2019/02/09-21:22
5	...b210-058b25f11216	04:28 (mm:ss)	Start Time: 2019/02/09-16:04

PEAK MEMORY

▸	Current Run	32.18 GB	Start Time: 2019/02/10-13:32
2	...af0f-34fcb735d762	25.68 GB	Start Time: 2019/02/10-07:48
3	...aeb0-fba6bc28c5eb	24.12 GB	Start Time: 2019/02/10-02:48
4	...a1df-0db734b0381f	25.13 GB	Start Time: 2019/02/09-21:22
5	...b210-058b25f11216	25.5 GB	Start Time: 2019/02/09-16:04

TOTAL IO

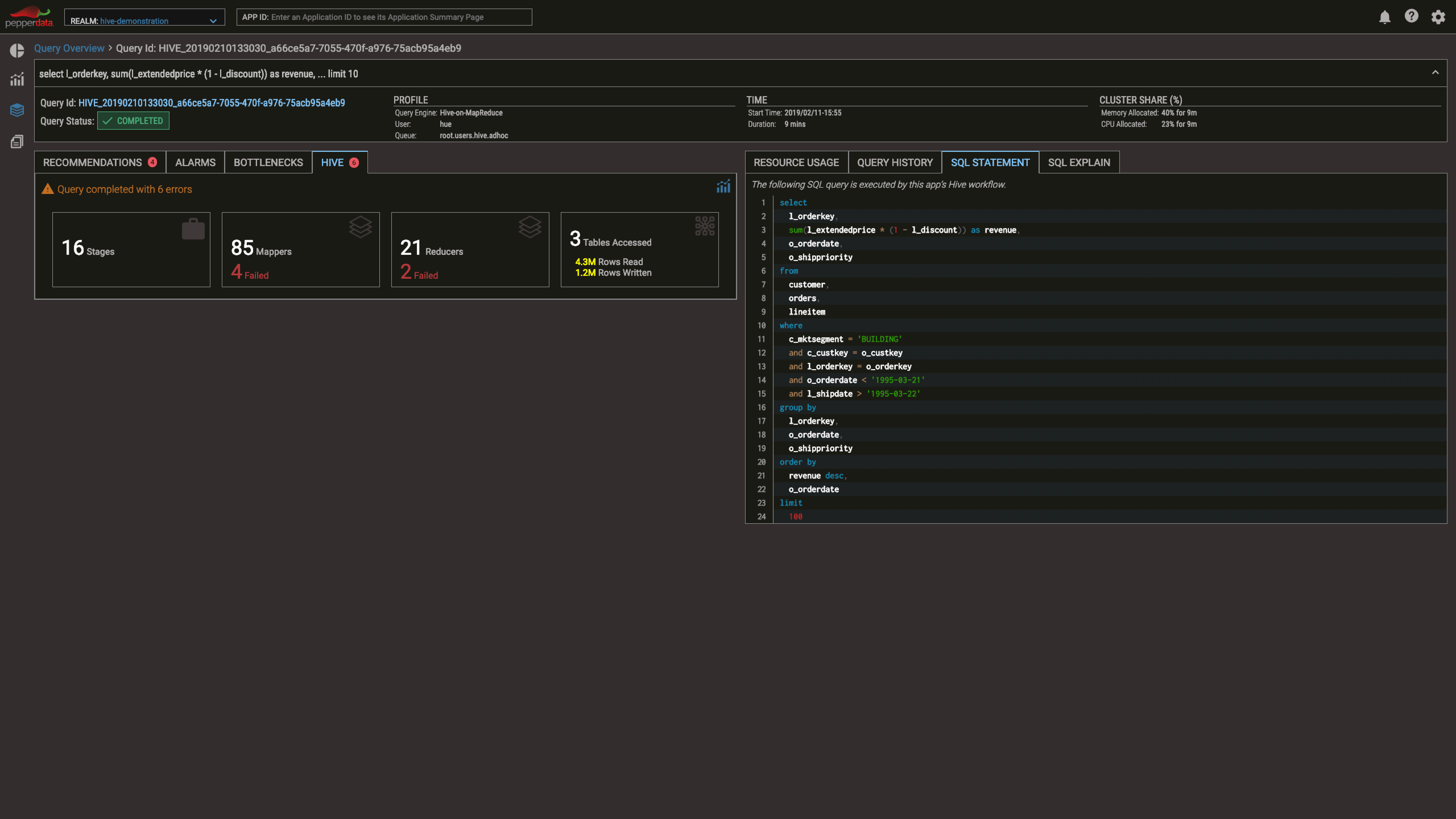
▸	Current Run	2.9 GB	Start Time: 2019/02/10-13:32
2	...af0f-34fcb735d762	1.8 GB	Start Time: 2019/02/10-07:48
3	...aeb0-fba6bc28c5eb	1.8 GB	Start Time: 2019/02/10-02:48
4	...a1df-0db734b0381f	1.8 GB	Start Time: 2019/02/09-21:22
5	...b210-058b25f11216	1.8 GB	Start Time: 2019/02/09-16:04

PERCENTAGE OF TIME DOING GC

▸	Current Run	0%	Start Time: 2019/02/10-13:32
2	...af0f-34fcb735d762	0%	Start Time: 2019/02/10-07:48
3	...aeb0-fba6bc28c5eb	0%	Start Time: 2019/02/10-02:48
4	...a1df-0db734b0381f	0%	Start Time: 2019/02/09-21:22
5	...b210-058b25f11216	0%	Start Time: 2019/02/09-16:04

PEAK CONTAINERS

▸	Current Run	39	Start Time: 2019/02/10-13:32
2	...af0f-34fcb735d762	30	Start Time: 2019/02/10-07:48
3	...aeb0-fba6bc28c5eb	29	Start Time: 2019/02/10-02:48
4	...a1df-0db734b0381f	29	Start Time: 2019/02/09-21:22
5	...b210-058b25f11216	29	Start Time: 2019/02/09-16:04



pepperdata
REALM: hive-demonstration
APP ID: Enter an Application ID to see its Application Summary Page

Query Overview > Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9

select l_orderkey, sum(l_extendedprice * (1 - l_discount)) as revenue, ... limit 10

Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9

Query Status: ✓ COMPLETED

PROFILE

Query Engine: Hive-on-MapReduce

User: hue

Queue: root.users.hive.adhoc

TIME

Start Time: 2019/02/11-15:55

Duration: 9 mins

CLUSTER SHARE (%)

Memory Allocated: 40% for 9m

CPU Allocated: 23% for 9m

⚠ Query completed with 6 errors

16 Stages

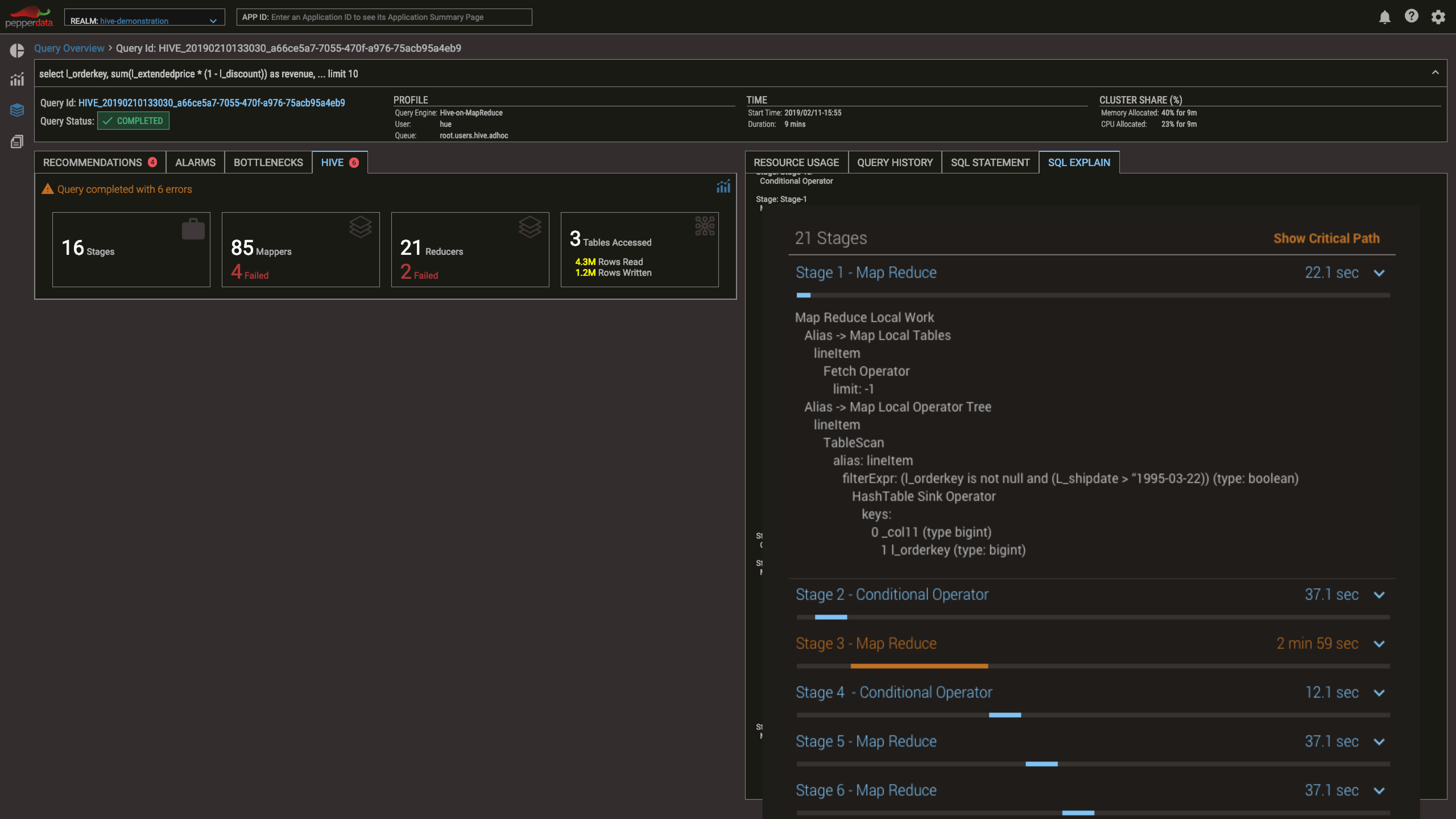
85 Mappers
4 Failed

21 Reducers
2 Failed

3 Tables Accessed
4.3M Rows Read
1.2M Rows Written

The following SQL query is executed by this app's Hive workflow.

```
select
  l_orderkey,
  sum(l_extendedprice * (1 - l_discount)) as revenue,
  o_orderdate,
  o_shippriority
from
  customer,
  orders,
  lineitem
where
  c_mktsegment = 'BUILDING'
  and c_custkey = o_custkey
  and l_orderkey = o_orderkey
  and o_orderdate < '1995-03-21'
  and l_shipdate > '1995-03-22'
group by
  l_orderkey,
  o_orderdate,
  o_shippriority
order by
  revenue desc,
  o_orderdate
limit
  100
```



pepperdata
REALM: hive-demonstration
APP ID: Enter an Application ID to see its Application Summary Page

Query Overview > Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9

select l_orderkey, sum(l_extendedprice * (1 - l_discount)) as revenue, ... limit 10

Query Id: HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9
Query Status: ✓ COMPLETED

PROFILE
Query Engine: Hive-on-MapReduce
User: hue
Queue: root.users.hive.adhoc

TIME
Start Time: 2019/02/11-15:55
Duration: 9 mins

CLUSTER SHARE (%)
Memory Allocated: 40% for 9m
CPU Allocated: 23% for 9m

Query completed with 6 errors

16 Stages

85 Mappers
4 Failed

21 Reducers
2 Failed

3 Tables Accessed
4.3M Rows Read
1.2M Rows Written

Conditional Operator

Stage: Stage-1

21 Stages
Show Critical Path

Stage 1 - Map Reduce 22.1 sec

```
Map Reduce Local Work
  Alias -> Map Local Tables
    lineItem
      Fetch Operator
        limit: -1
  Alias -> Map Local Operator Tree
    lineItem
      TableScan
        alias: lineItem
          filterExpr: (l_orderkey is not null and (L_shipdate > "1995-03-22)) (type: boolean)
            HashTable Sink Operator
              keys:
                0 _col11 (type bigint)
                  1 l_orderkey (type: bigint)
```

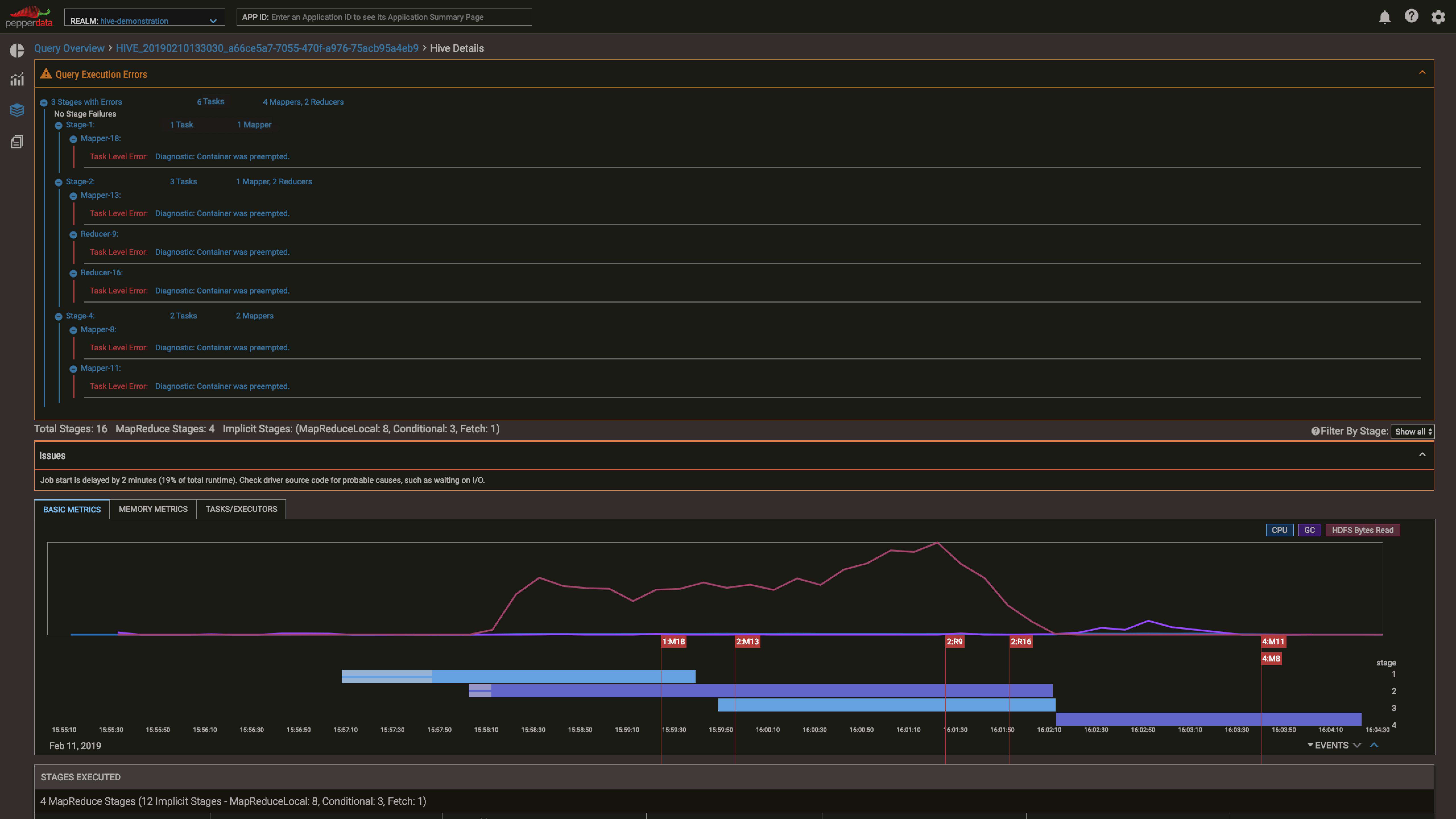
Stage 2 - Conditional Operator 37.1 sec

Stage 3 - Map Reduce 2 min 59 sec

Stage 4 - Conditional Operator 12.1 sec

Stage 5 - Map Reduce 37.1 sec

Stage 6 - Map Reduce 37.1 sec


pepperdata
REALM: hive-demonstration
APP ID: Enter an Application ID to see its Application Summary Page

Query Overview > HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9 > Hive Details

Query Execution Errors

3 Stages with Errors 6 Tasks 4 Mappers, 2 Reducers

No Stage Failures

Stage-1: 1 Task 1 Mapper

Mapper-18:

Task Level Error: Diagnostic: Container was preempted.

Stage-2: 3 Tasks 1 Mapper, 2 Reducers

Mapper-13:

Task Level Error: Diagnostic: Container was preempted.

Reducer-9:

Task Level Error: Diagnostic: Container was preempted.

Reducer-16:

Task Level Error: Diagnostic: Container was preempted.

Stage-4: 2 Tasks 2 Mappers

Mapper-8:

Task Level Error: Diagnostic: Container was preempted.

Mapper-11:

Task Level Error: Diagnostic: Container was preempted.

Total Stages: 16 MapReduce Stages: 4 Implicit Stages: (MapReduceLocal: 8, Conditional: 3, Fetch: 1)

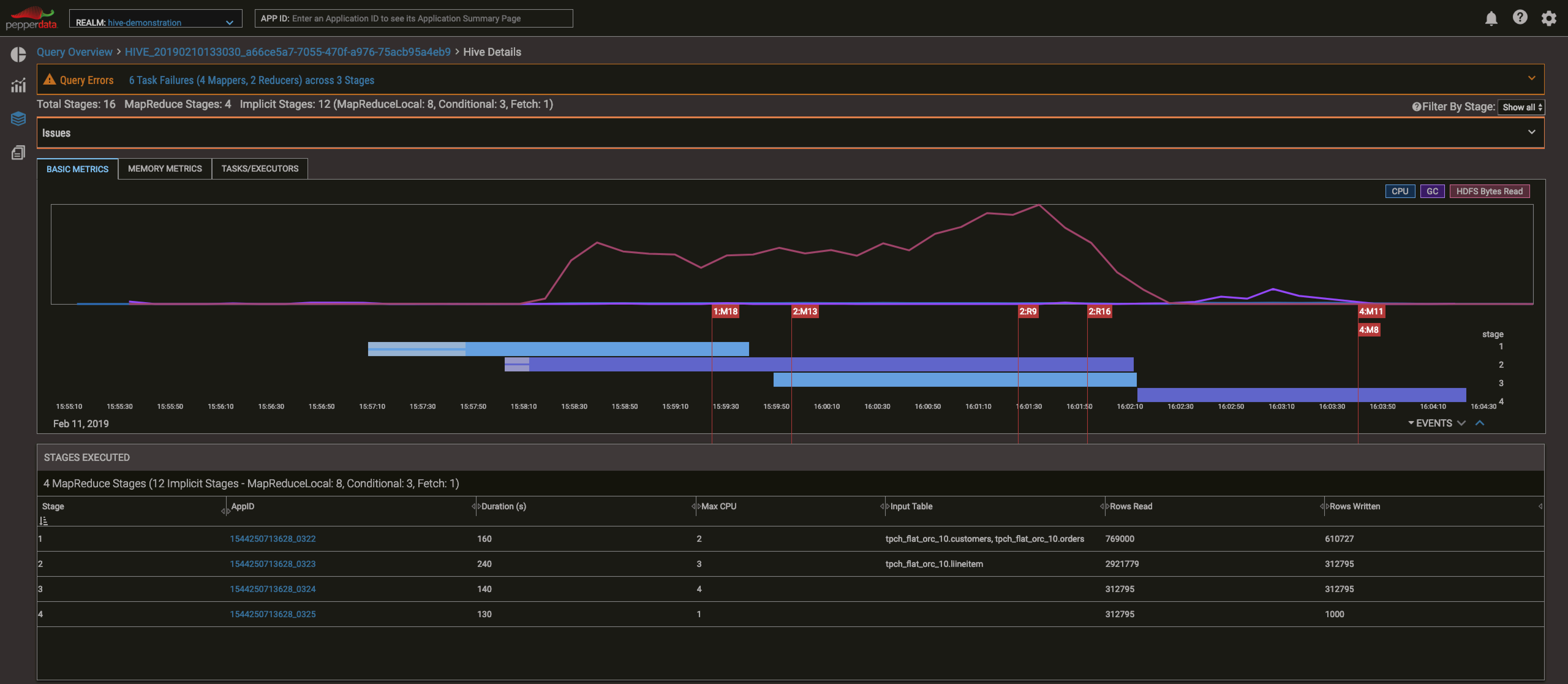
?Filter By Stage: Show all

Issues

Job start is delayed by 2 minutes (19% of total runtime). Check driver source code for probable causes, such as waiting on I/O.

STAGES EXECUTED

4 MapReduce Stages (12 Implicit Stages - MapReduceLocal: 8, Conditional: 3, Fetch: 1)



Query Overview > HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9 > Hive Details

Query Errors 6 Task Failures (4 Mappers, 2 Reducers) across 3 Stages

Total Stages: 16 MapReduce Stages: 4 Implicit Stages: 12 (MapReduceLocal: 8, Conditional: 3, Fetch: 1)

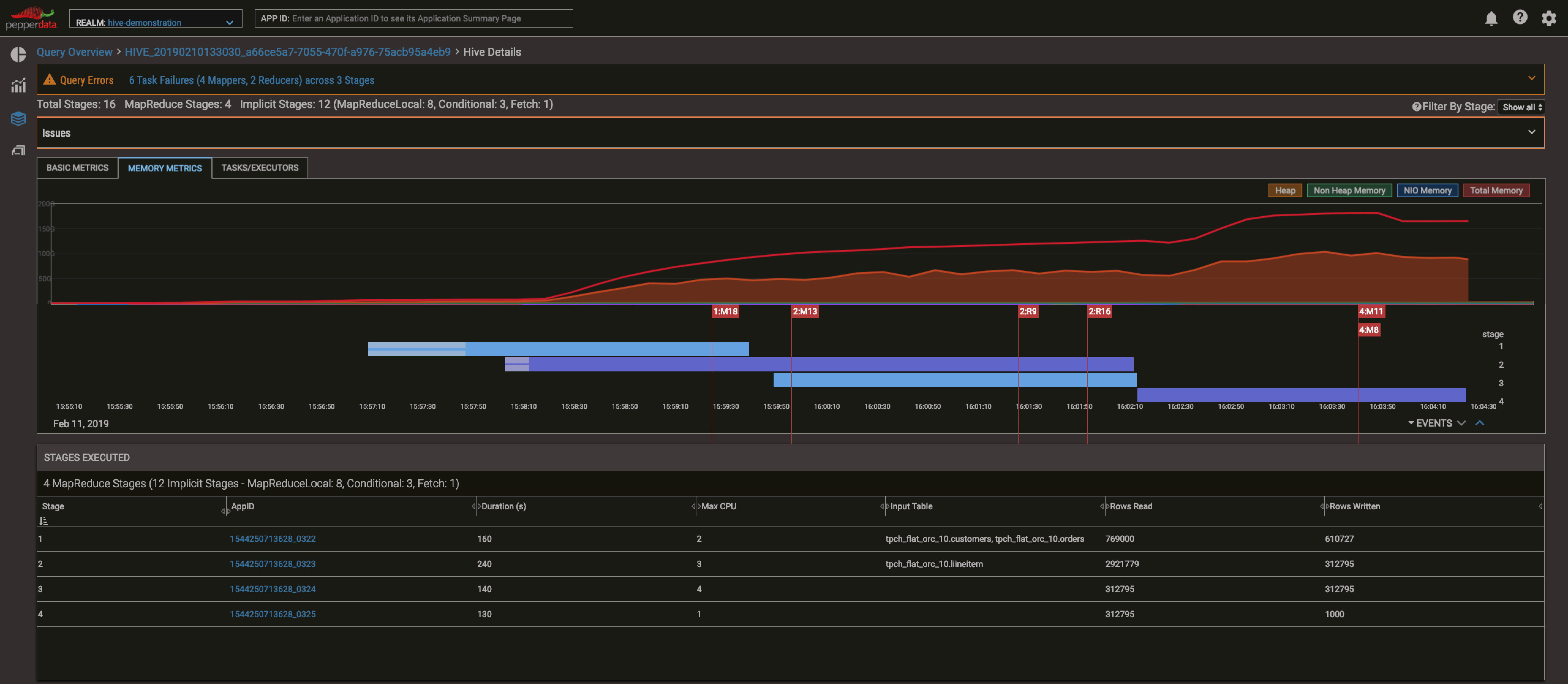
Filter By Stage: Show all

Issues

STAGES EXECUTED

4 MapReduce Stages (12 Implicit Stages - MapReduceLocal: 8, Conditional: 3, Fetch: 1)

Stage	AppID	Duration (s)	Max CPU	Input Table	Rows Read	Rows Written
1	1544250713628_0322	160	2	tpch_flat_orc_10.customers, tpch_flat_orc_10.orders	769000	610727
2	1544250713628_0323	240	3	tpch_flat_orc_10.lineitem	2921779	312795
3	1544250713628_0324	140	4		312795	312795
4	1544250713628_0325	130	1		312795	1000



Query Overview > HIVE_20190210133030_a66ce5a7-7055-470f-a976-75acb95a4eb9 > Hive Details

⚠ Query Errors 6 Task Failures (4 Mappers, 2 Reducers) across 3 Stages

Total Stages: 16 MapReduce Stages: 4 Implicit Stages: 12 (MapReduceLocal: 8, Conditional: 3, Fetch: 1)

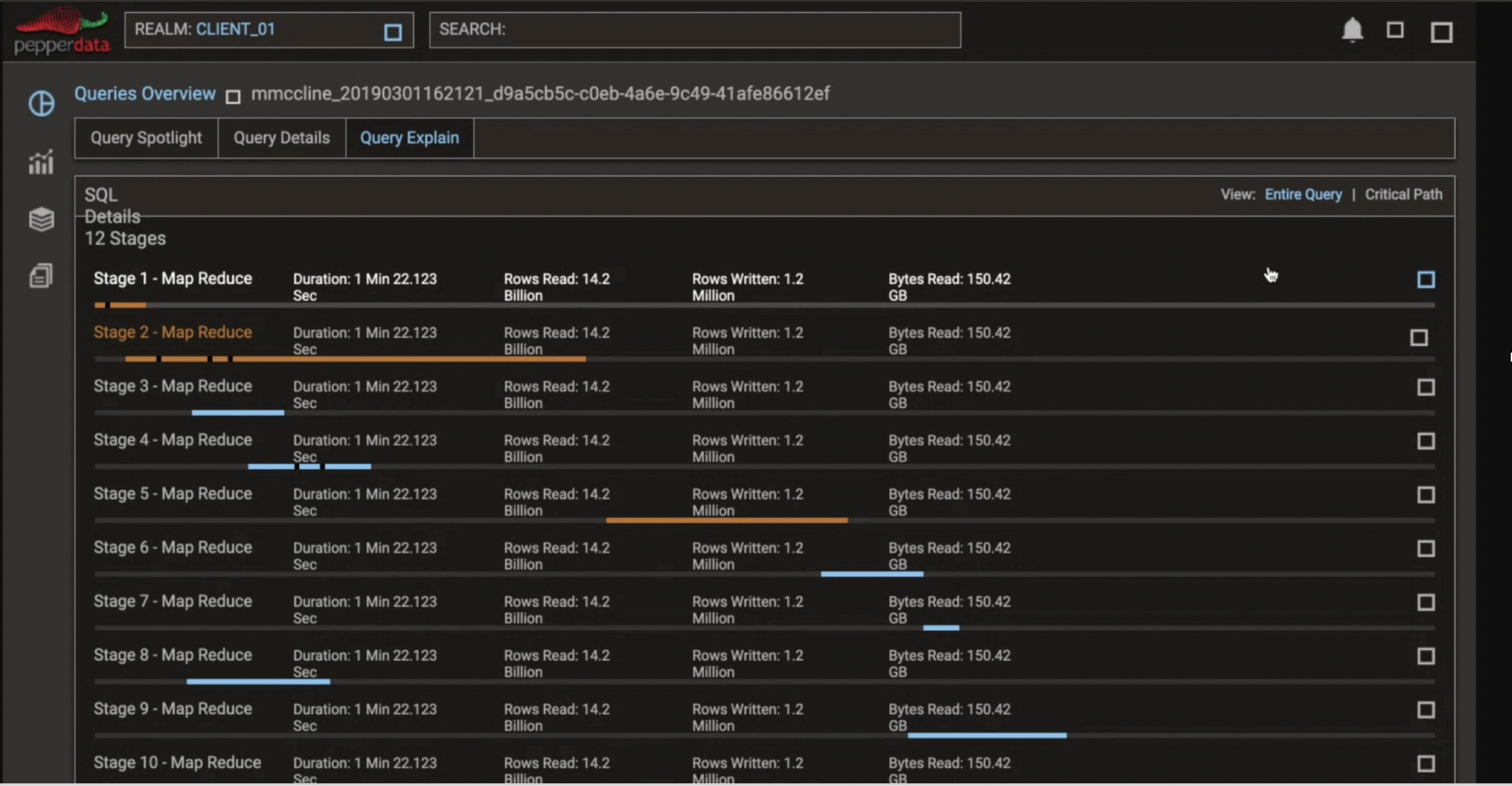
Filter By Stage: Show all

Issues

STAGES EXECUTED

4 MapReduce Stages (12 Implicit Stages - MapReduceLocal: 8, Conditional: 3, Fetch: 1)

Stage	AppID	Duration (s)	Max CPU	Input Table	Rows Read	Rows Written
1	1544250713628_0322	160	2	tpch_flat_orc_10.customers, tpch_flat_orc_10.orders	769000	610727
2	1544250713628_0323	240	3	tpch_flat_orc_10.lineitem	2921779	312795
3	1544250713628_0324	140	4		312795	312795
4	1544250713628_0325	130	1		312795	1000

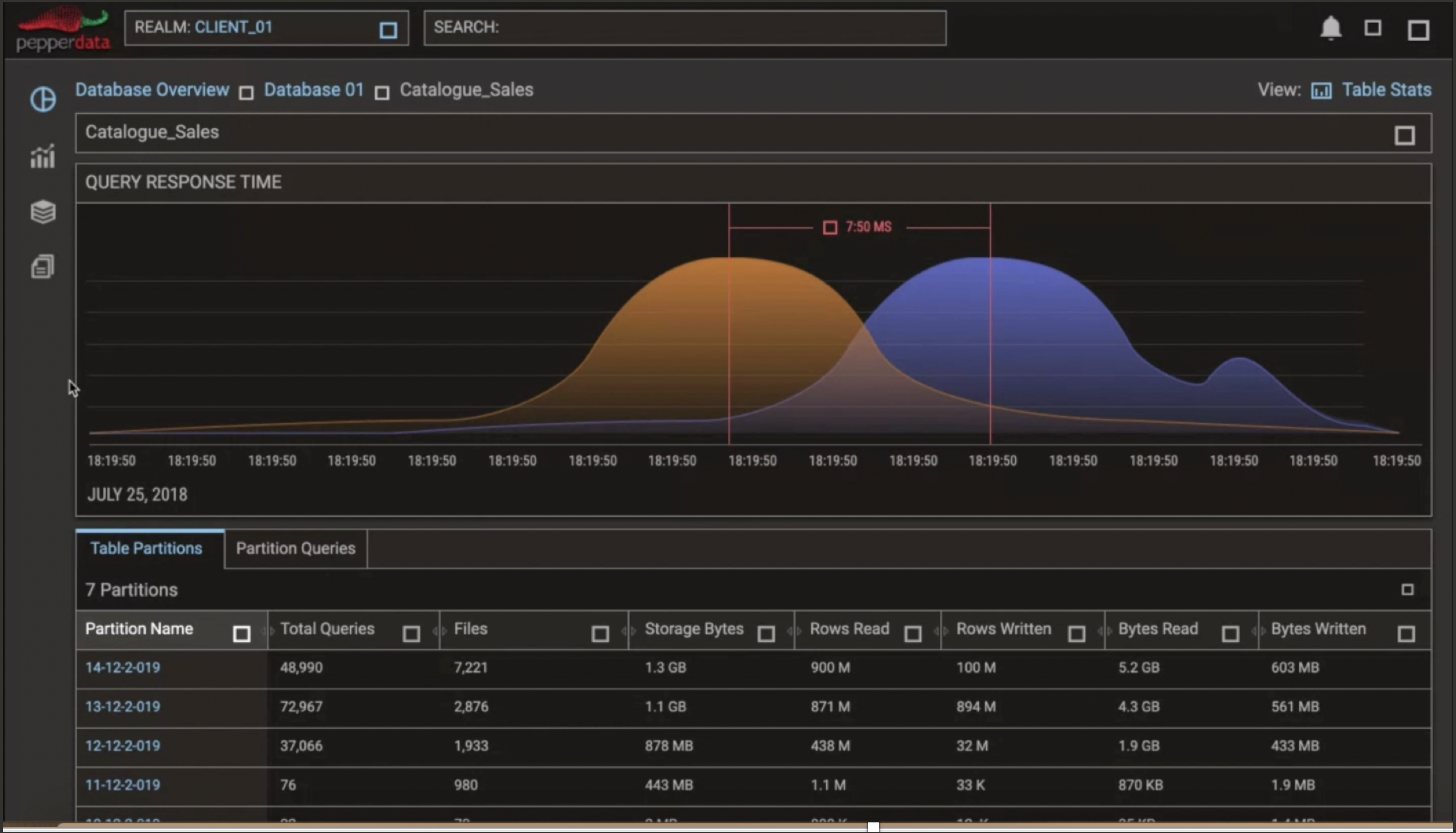


pepperdata
REALM: CLIENT_01 □
SEARCH:
Queries Overview □ mmccline_20190301162121_d9a5cb5c-c0eb-4a6e-9c49-41afe86612ef
Query Spotlight
Query Details
Query Explain
View: Entire Query | Critical Path

SQL Details

12 Stages

Stage 1 - Map Reduce	Duration: 1 Min 22.123 Sec	Rows Read: 14.2 Billion	Rows Written: 1.2 Million	Bytes Read: 150.42 GB	□
Stage 2 - Map Reduce	Duration: 1 Min 22.123 Sec	Rows Read: 14.2 Billion	Rows Written: 1.2 Million	Bytes Read: 150.42 GB	□
Stage 3 - Map Reduce	Duration: 1 Min 22.123 Sec	Rows Read: 14.2 Billion	Rows Written: 1.2 Million	Bytes Read: 150.42 GB	□
Stage 4 - Map Reduce	Duration: 1 Min 22.123 Sec	Rows Read: 14.2 Billion	Rows Written: 1.2 Million	Bytes Read: 150.42 GB	□
Stage 5 - Map Reduce	Duration: 1 Min 22.123 Sec	Rows Read: 14.2 Billion	Rows Written: 1.2 Million	Bytes Read: 150.42 GB	□
Stage 6 - Map Reduce	Duration: 1 Min 22.123 Sec	Rows Read: 14.2 Billion	Rows Written: 1.2 Million	Bytes Read: 150.42 GB	□
Stage 7 - Map Reduce	Duration: 1 Min 22.123 Sec	Rows Read: 14.2 Billion	Rows Written: 1.2 Million	Bytes Read: 150.42 GB	□
Stage 8 - Map Reduce	Duration: 1 Min 22.123 Sec	Rows Read: 14.2 Billion	Rows Written: 1.2 Million	Bytes Read: 150.42 GB	□
Stage 9 - Map Reduce	Duration: 1 Min 22.123 Sec	Rows Read: 14.2 Billion	Rows Written: 1.2 Million	Bytes Read: 150.42 GB	□
Stage 10 - Map Reduce	Duration: 1 Min 22.123 Sec	Rows Read: 14.2 Billion	Rows Written: 1.2 Million	Bytes Read: 150.42 GB	□



7 Partitions

Partition Name	Total Queries	Files	Storage Bytes	Rows Read	Rows Written	Bytes Read	Bytes Written
14-12-2-019	48,990	7,221	1.3 GB	900 M	100 M	5.2 GB	603 MB
13-12-2-019	72,967	2,876	1.1 GB	871 M	894 M	4.3 GB	561 MB
12-12-2-019	37,066	1,933	878 MB	438 M	32 M	1.9 GB	433 MB
11-12-2-019	76	980	443 MB	1.1 M	33 K	870 KB	1.9 MB

pepperdata

Query Spotlight

1. FULLY Visualize Query Performance
2. Provide ACTIONABLE insights to Developers
3. Automatically Optimize Query Performance

Get early access to Query Spotlight

Email: Dan@Pepperdata.com