

Hive Queries Fail Automatically when Disk Space Limits Reached

Learn the cause and how to resolve the following error message.

```
Diagnostic Messages for this Task:
Error: Task exceeded the limits:
    org.apache.hadoop.mapred.Task$TaskReporter$TaskLimitException:
        too much data in local scratch dir=/mnt4/hadoop/yarn/cache/yarn/nm-local-dir/usercache/1/appcache
/application_1522879910596_701143.
        current size is 322234851654 the limit is 322122547200
Error: java.lang.RuntimeException:
    org.apache.hadoop.hive.ql.metadata.HiveException:
    org.apache.hadoop.hdfs.protocol.DSQuotaExceededException:
    The DiskSpace quota of /mnt/hive/hive-1/0 is exceeded:
    quota = 8246337208320 B = 7.50 TB but diskspace consumed = 8246728346553 B = 7.50 TB
```

Description

Hive processing will at times move data from S3 into local storage or HDFS storage on the Hadoop cluster nodes. Hive jobs fail when disk space is exhausted on Hadoop cluster nodes-- either when one of the Hadoop nodes runs out of local space or when a job reaches its overall storage limit on HDFS.

In the past, under these conditions, a Treasure Data operator would be alerted to manually kill the job, and support would manually notify the customer and follow up. We are changing our handling of these situations to be more predictable and automated, and more in line with other job failures.

The new behavior is, such jobs will fail, and the job Output Log will contain a diagnostic message about the failure.

If the disk space is full on one Hadoop worker, the output log contains:

```
Diagnostic Messages for this Task:
Error: Task exceeded the limits:
    org.apache.hadoop.mapred.Task$TaskReporter$TaskLimitException:
        too much data in local scratch dir=/mnt4/hadoop/yarn/cache/yarn/nm-local-dir/usercache/1/appcache
/application_1522879910596_701143.
        current size is 322234851654 the limit is 322122547200
```

If the query exceeds the limit of HDFS storage, the output log contains:

```
Error: java.lang.RuntimeException:
    org.apache.hadoop.hive.ql.metadata.HiveException:
    org.apache.hadoop.hdfs.protocol.DSQuotaExceededException:
    The DiskSpace quota of /mnt/hive/hive-1/0 is exceeded:
    quota = 8246337208320 B = 7.50 TB but diskspace consumed = 8246728346553 B = 7.50 TB
```

Resolution

To reduce disk usage by Hadoop jobs, limit the amount of input data, such as by applying `TD_TIME_RANGE` to restrict the time period of any subqueries that scan data, or consider applying more restrictive conditions to JOINS. See [Performance Tuning for Hive](#) for more suggestions.