测试环境说明

Hadoop目录：/hadoop

Hive目录：/hadoop/hive

Hive-testbench目录：/hadoop/hive-testbench

Spark目录：/hadoop/spark

Emon目录：/root/Emon

Note: 测试脚本需要放置于pat执行目录下，本次测试DFSIO测试脚本位于/root/PAT，TPCDS位于/hadoop/pat-suite，不同测试请切换至相应文件夹；.cmd目录下包含一些测试用的工具，测试时将其放置于/root目录下

DFSIO

1. 测试前将PAT文件夹放置于/root目录下
2. 测试数据生成：数据生成脚本位于worload目录下，dfsio\_\*.sh，通过修改nrFiles和fileSize控制测试文件数量与大小。
3. Add cache pool: hdfs cacheadmin –addPool benchmark

Cache files: hdfs cacheadmin –addDirective –path /benchmarks/TestDFSIO/io\_data benchmark –replication 2

1. 通过执行PAT-collecting-data目录下的run.py进行DFSIO测试，run.py负责生成pat执行脚本，然后调用run.sh，pat结果放置于results目录下

TPCDS

1. 测试前将pat-suite文件夹放置于/hadoop目录下
2. 测试数据生成：/hadoop/hive-testbench/tpcds-setup.sh scale type dir

Scale为生成文件大小，单位为GB；type为文件类型，支持text、orc、parquet，orc和parquet需要在已经生成相同规模text数据时才能生成；dir为生成文件的目录，text文件位于hdfs://user/root/${dir}/${scale}，orc和parquet位于hdfs://hive/tpcds\_bin\_partitioned\_${type}\_${scale}.db

1. 通过执行pat-suite目录下的runall.sh进行TPCDS测试。

该脚本会调用auto.sh，执行workload目录下的测试脚本（tpcds-run\*.sh），需要执行不同的测试只需要将脚本放置于workload目录下，并在调用auto.sh时传入脚本名。

tpcds-run\*.sh脚本中 perl $HIVE\_TESTBENCH/runSuite.pl $QUERY\_DIR $run\_id spark orc 2000 最后三个参数控制执行引擎、文件类型以及文件大小

Pat结果放置于results目录下

Note：workload目录下的cache.sh脚本用于cache数据，AEP测试中cache监控使用cacheMonitor.sh，DRAM测试中cache进度监控使用cacheMonitor-DRAM.sh，不同测试环境下需要对cache.sh进行相应的修改

Tools

1. /root/.cmd/dropCache：清除缓存
2. /root/.cmd/cacheMonitor-DRAM.sh：监控cache进度，如果cache没有结束，该脚本一直阻塞
3. /root/.cmd/RCache-1.0.jar：递归缓存数据。DFSIO测试不需要使用，cacheadmin可以直接缓存；TPCDS测试数据有复杂的文件目录层次，请使用该工具缓存数据

示例：hadoop jar RCache-1.0.jar –d /user/root/tpcds/1000/ -p benchmark –r 2

-p指定cache pool；-r指定cache replication

4. /root/.cmd/qt.sh：用于收集TPCDS测试结果。将该脚本放置于results目录下执行。