

Workload Analysis - Hadoop

Vishnu Vardhan

7/12/2018

What was tested

S3 Operations from a single Hive query

Test Description: from: Promise

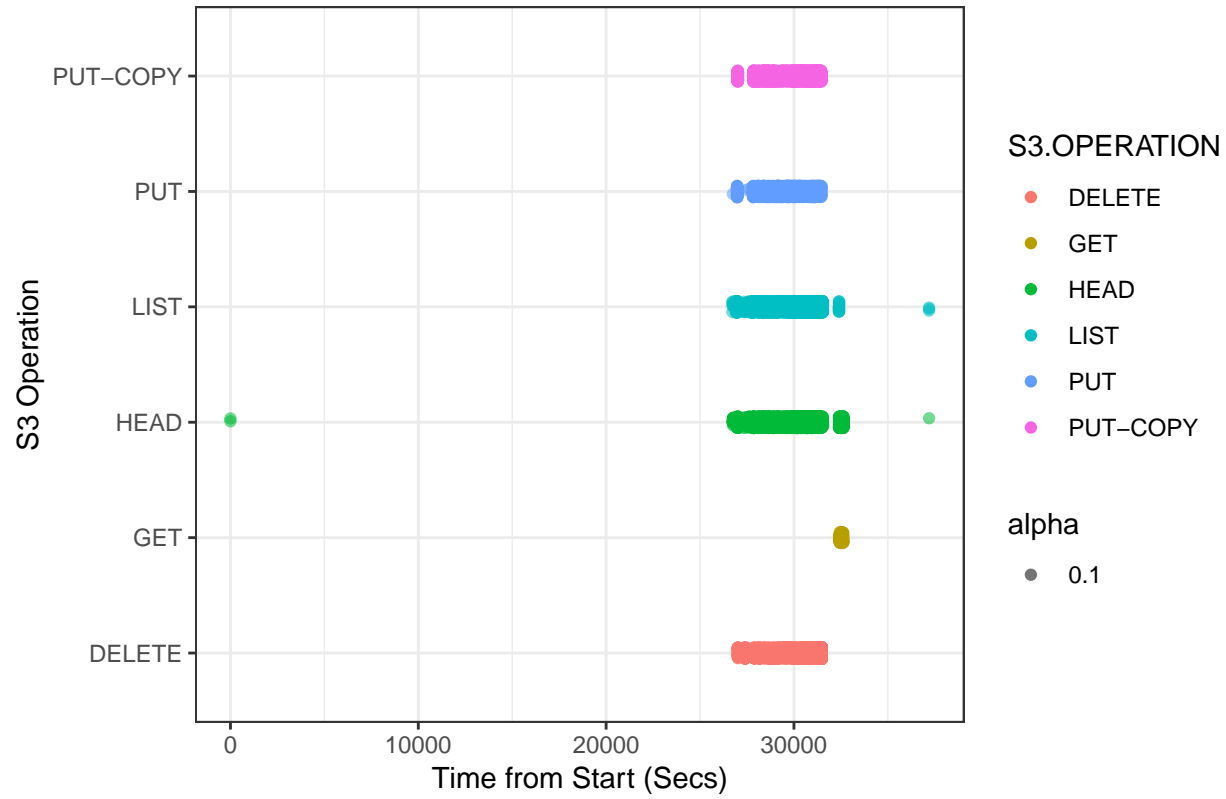
fs.s3a.multipart.size is now 50MB storage: Extrenal to hive data set: 100GB

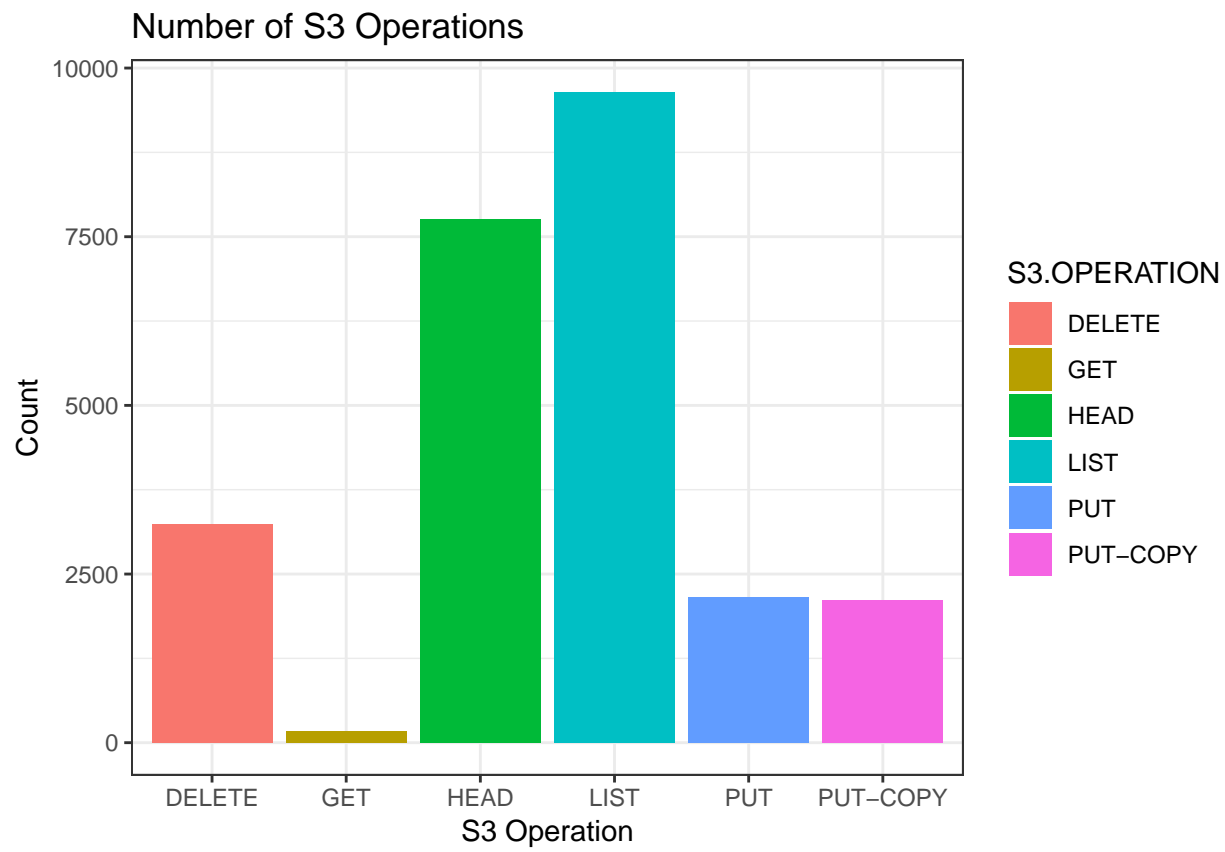
The data is external to hive, hive-bench creates external tables using the s3a connector i.e. ./tpcds-setup.sh 10 s3a://datalake/

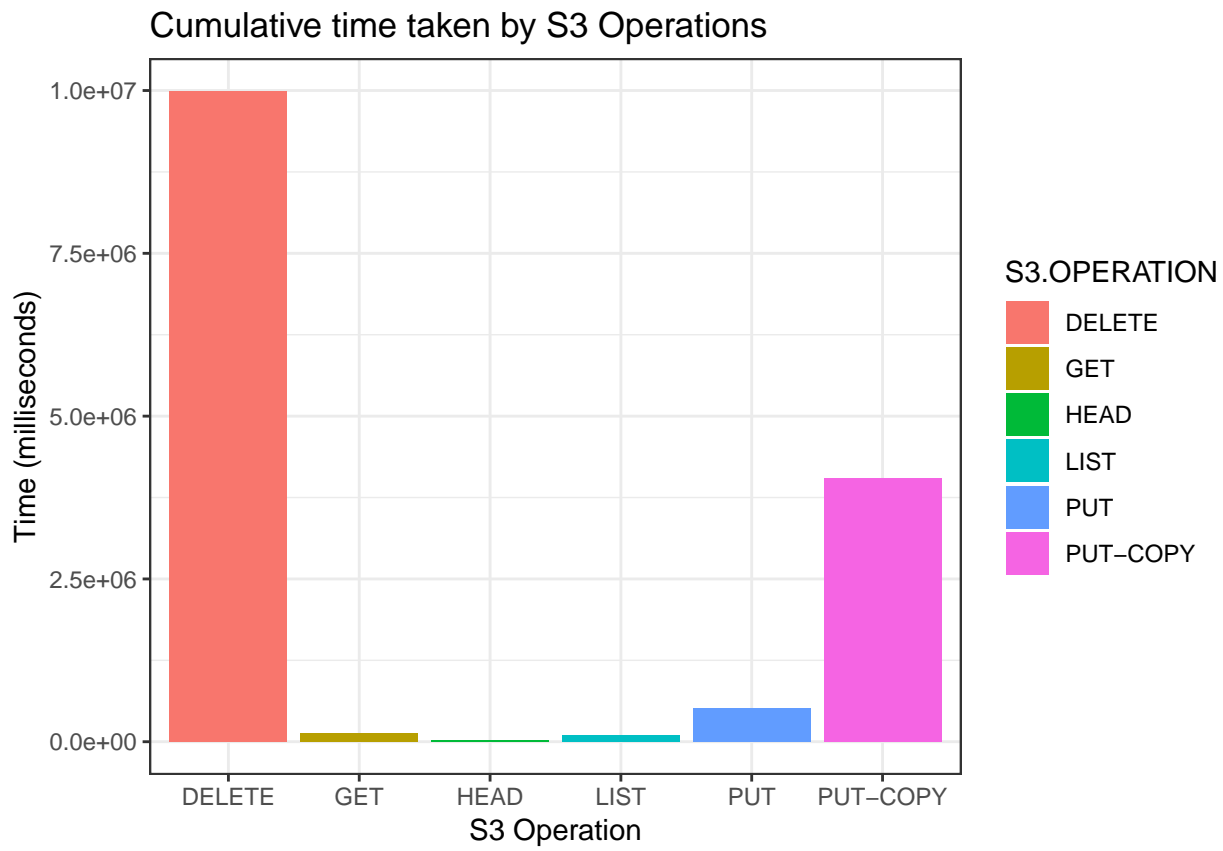
```
## [1] "Total test time from the first S3 operation to the last: "
```

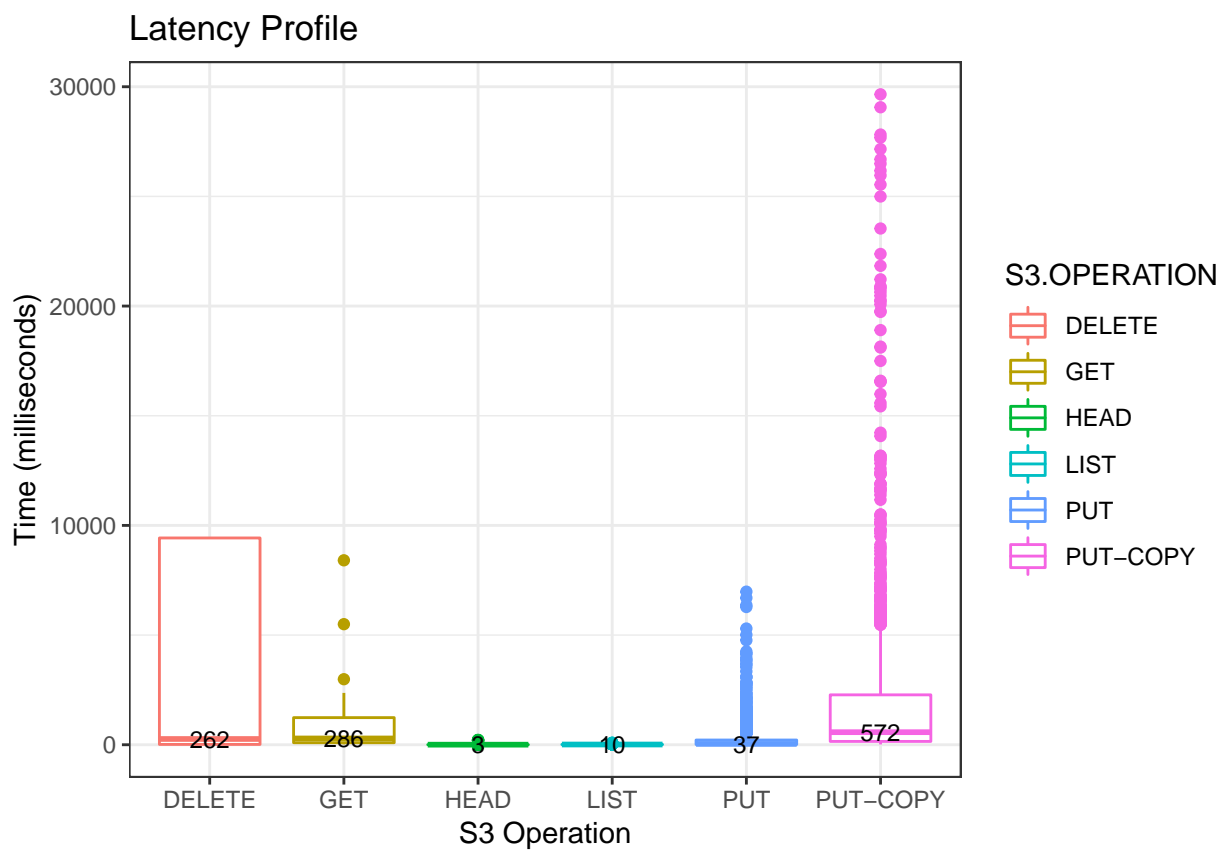
```
## Time difference of 10.33232 hours
```

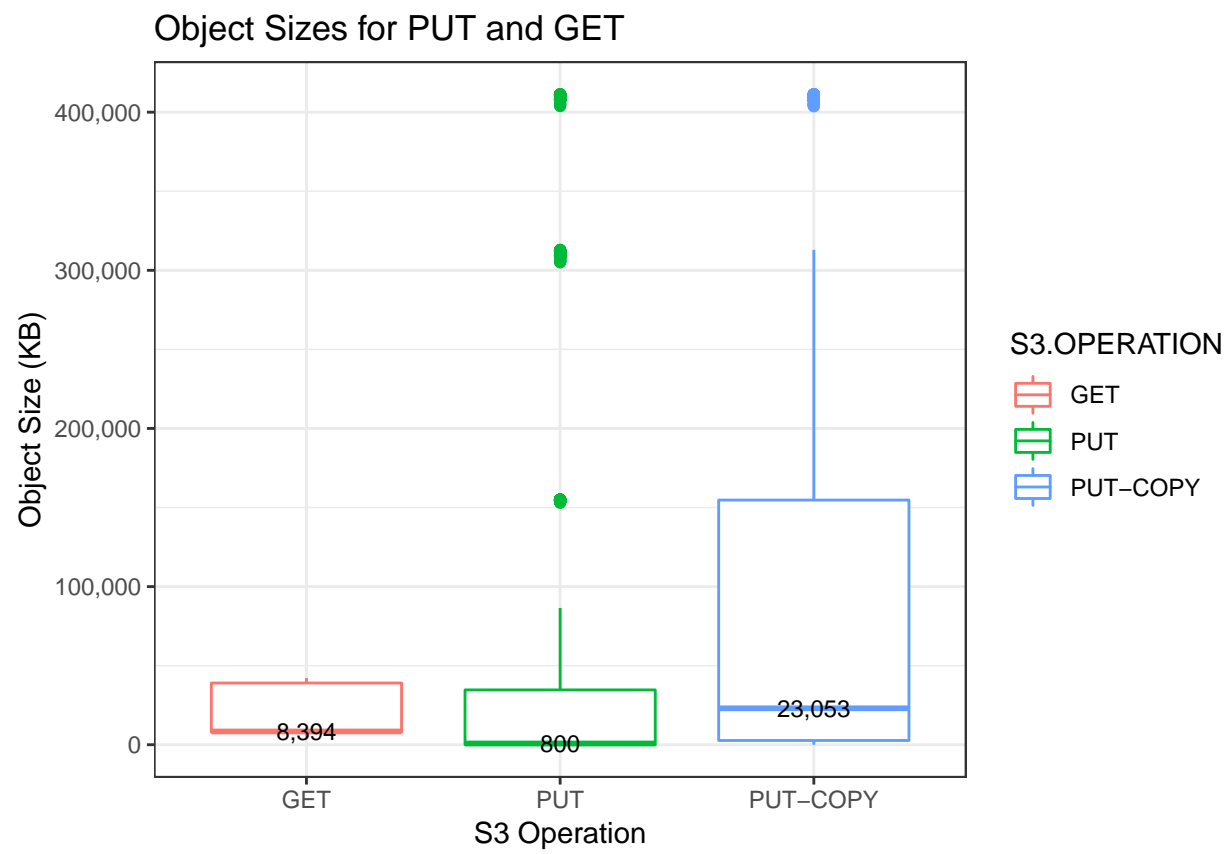
Time when operations happen from the earliest S3 Operation





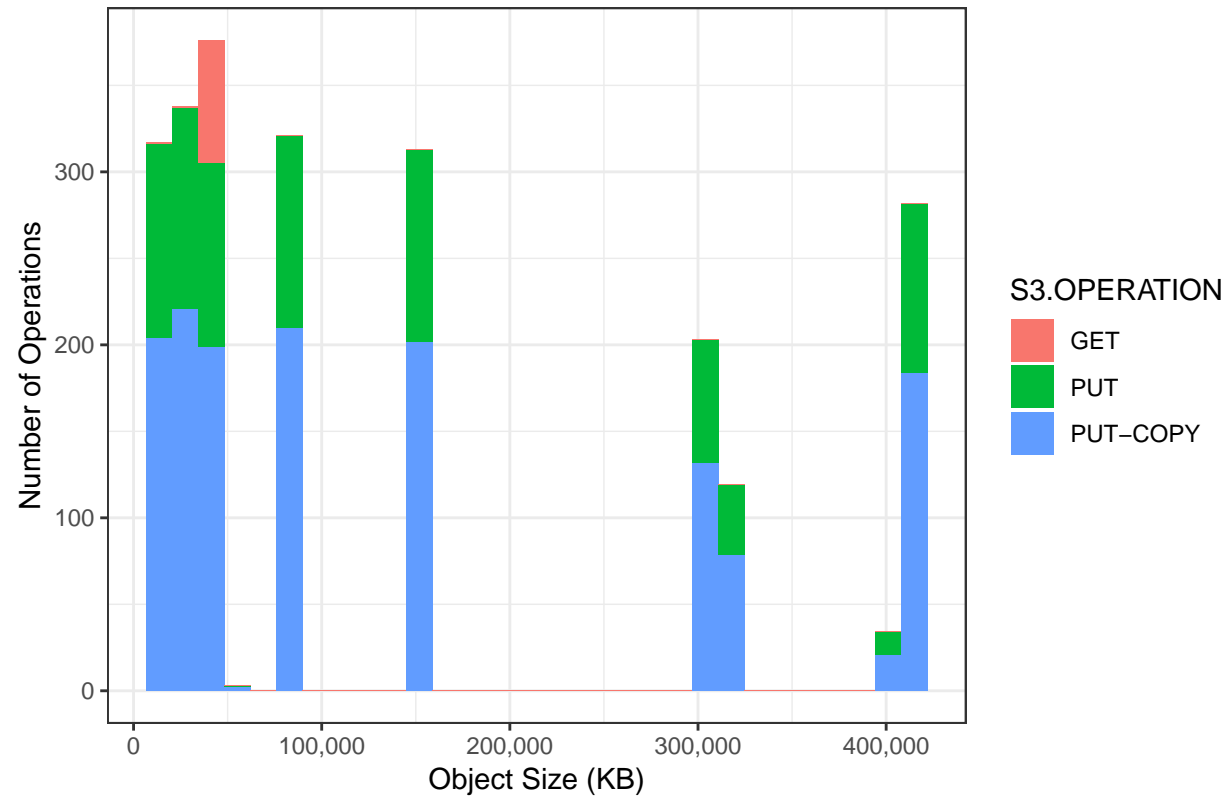






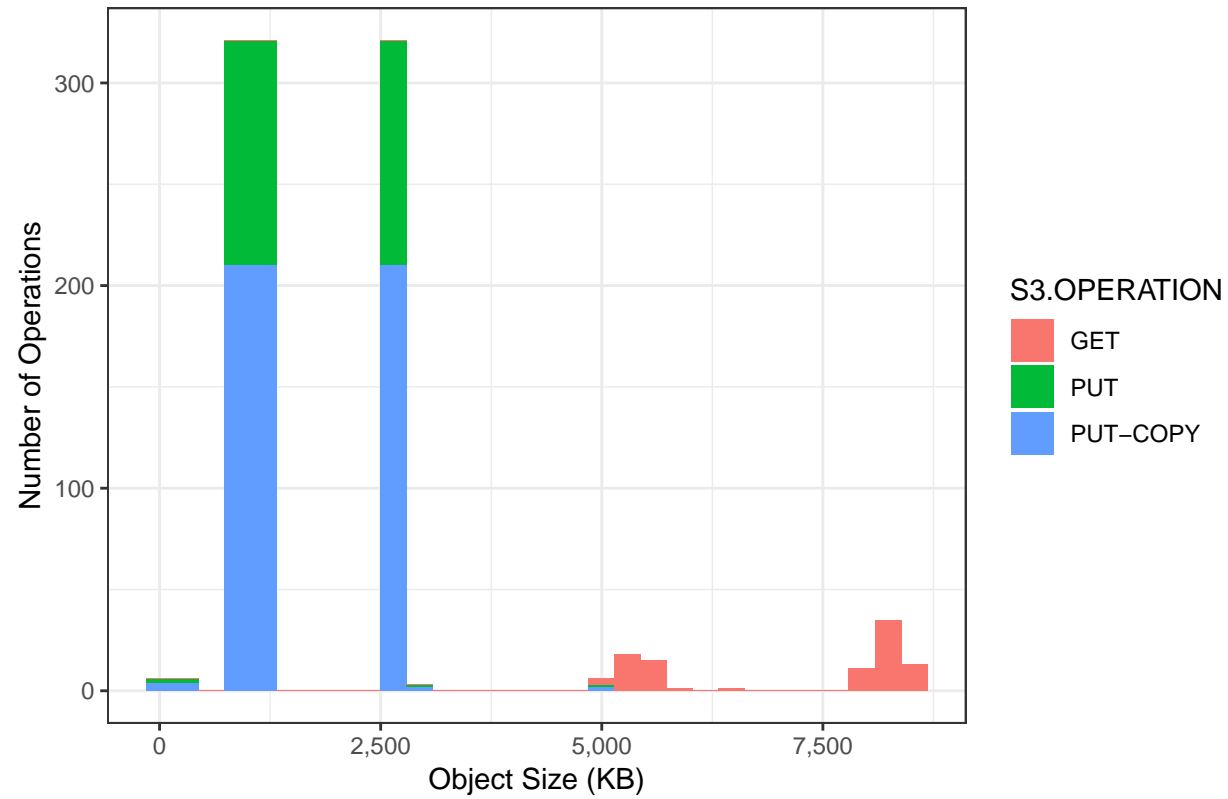
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Object Size Histogram for objects greater than 10MB



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Object Size Histogram for objects greater than 100KB, <10MB




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Object Size Histogram for objects less than 100KB

