Benchmarking Cloud Storage/Database Systems

The most commonly used cloud database benchmark is YCSB — the Yahoo! Cloud Storage Benchmark. It is a simple multi-threaded randomized request generator with an extensible backend adaptor for linking to specific storage/database systems. It is open source.

https://github.com/brianfrankcooper/YCSB

A few years ago a group of CMU students built a set of extensions to YCSB that they called YCSB++. It also produced open source, but the YCSB maintainers were not willing to take the much to pervasive patch. So the widely used YCSB benchmarks still does not have features YCSB++ developed.

http://www.pdl.cmu.edu/ycsb++/index.shtml

The primary goal of this project is to redevelop YCSB++ features in a manner that is more likely to be accepted by the YCSB maintainers, so the features can be widely used. This means researching the status of YCSB and its maintainers as well as their expectations for patches (i.e.., small and simple) and patch tests. It means reviewing YCSB++ features, subletting the most useful, improving on their implementation and utility, and packaging in bite size, testable patches.

Likely YCSB++ features to focus on include: multi-client coordination using Zookeeper; multi-phase tests (coordinated across clients); read-after-write measurements using Zookeeper; test profiles and metrics for deducing the performance impact of delayed background work like SSTable compaction; white-box monitoring and reporting of system-under-test internal metrics (OTUS). Additionally, one painful feature of YCSB and standard databases is the performance sensitivity to offered load (number of client threads) — a tool js needed to automatically repeat experiments at different performance levels to construct a graph showing the best performance and performance degradation at higher load levels.