Five Characteristics of a Successful Benchmark

Andrew Bond
Principal Software Engineer
Red Hat, Inc.

A wide variety of of benchmarks exist in the computer industry. Small benchmarks targeted at a single hardware subsystem or software component and large complex benchmarks that use whole rooms of equipment. Benchmarks developed by industry standard organizations and benchmarks developed by individuals. However, most successful benchmarks have five common characteristics: a well-defined target audience, an easy to understand single metric, ease of use, low cost of entry, and regular updates.

The reason benchmarks are created is to provide performance information to a particular audience. However, the success of a benchmark actually involves two separate audiences, those that run the benchmark and those that are the intended recipient of the data. Sometimes, these two audiences end up being the same set of individuals, but many times they are different. A successful benchmark will have very well-defined and understood target audiences, which means the data the benchmark is providing is valuable to them. As an example, let's look at the CPU2006 benchmark developed by the industry consortium Standard Performance Evaluation Corporation (SPEC). This benchmark has a well-defined audience of hardware manufacturers that want to run the benchmark to provide CPU performance data on their hardware. However, these hardware manufacturers are not the intended recipients of the data. Instead the benchmark data is intended for potential customers that are looking for ways to compare the CPU capabilities of hardware solutions from various vendors. Therefore, this benchmark has a well-defined target audience wanting to run the benchmark, and a well-defined target audience wanting to use the benchmark data.

A benchmark can have a great target audience, but if the data the benchmark is producing is not easy to understand and use for comparison, then the benchmark will not become successful. Therefore, a successful benchmark should have an easy to understand single main metric. The benchmark can be comprised of various workloads and have various scores associated with those workloads, but a methodology for combining these scores into a single metric is extremely important. If a benchmark metric has too many numbers or is too complex a metric, then the benchmark's target audience will quickly lose interest in trying to use the benchmark data. As an example, the industry consortia SPEC and the Transaction Processing Council (TPC) both understand how important a single metric is to the success of a benchmark. Therefore, all the benchmarks developed by those two organizations have a single main metric associated with them.

Another characteristic of a successful benchmark is that it is easy to use. Simple to run. However, simple to run does not mean the benchmark has to be a simple workload, just that a majority of the complexity should not be exposed to those running the benchmark. If a benchmark is comprised of twenty separate workloads that require manual intervention between each workload as well as some hardware changes during the benchmark, then the benchmark is not well-designed. So the design plays a big role in the ease of use. Another aspect of the ease of use is what type kit or harness is provided to run the benchmark. If the infrastructure to run the benchmark and process the results is already provided, then the ease of use increases dramatically.

Related to the ease of use is the cost of entry of the benchmark. How much does it cost to setup and run the benchmark for someone starting from scratch? If the benchmark is just made up of a specification document that describes the benchmark and no infrastructure to run the benchmark is provided, then the cost of entry of that benchmark can be very high. Time and engineering resources would be necessary to build the infrastructure (code, scripts, software, etc.) necessary to run the benchmark, and the cost of these resources can be very high. Therefore, a successful benchmark provides some or all of the environment necessary to run the benchmark to create a low cost of entry. The other aspect of a low cost of entry is how much hardware and software is required to run the benchmark. If the minimum configuration to run the benchmark would be very expensive, then only a small number of companies would be capable of running the benchmark. Low cost of entry means a wider potential adoption of the benchmark and its results.

A buggy or outdated benchmark will typically die a quick death. Therefore, a successful benchmark will have regular updates. These updates take the forms of bug fix maintenance releases, as well as major updates to the benchmark to add functionality, change the workloads, and bring the benchmark in line with current standards. Regular updates also show a commitment by the benchmark creator to the benchmark, which in turn makes it easier for users of the benchmark to commit the necessary time and money investments in running the benchmark. Again as an example, let's use SPEC CPU. This benchmark has gone through four major releases since its first release in 1995 and each release has had multiple bug fix releases. This level of commitment by SPEC to the benchmark makes it easier for hardware vendors to make time and money commitments and investments in running the benchmark.

Benchmark users are like flowing water. They always flow to the paths of least resistance. Therefore a successful benchmark should create an environment that provides the least resistance. If the benchmark has the five characteristics of a well-defined target audience, an easy to understand single metric, ease of use, low cost of entry, and regular updates, then that benchmark will be a path of least resistance. Having benchmark users flowing to your benchmark is always a good thing.