

Benchmark Deception

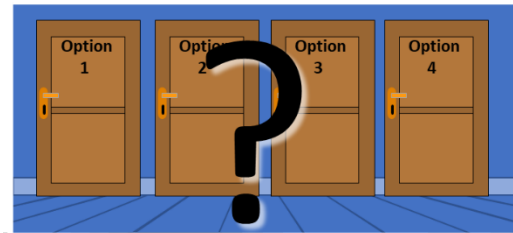
By Douglas H Ebel, Benchmark Trainer and Mentor

As a manager or an agent of change, you are considering the future of your company's analytic capabilities, motivated by:

- A desire to extend your capabilities to ingest data from suppliers or publish to customers
- Leverage new analytic capabilities and tools
- Move to the cloud to eliminate on-premises expense
- Address performance or capacity issues of your current platform
- Replace current technology that has reached end-of-life.

There are many options which may include:

- A new version of your existing software
- A new platform (newer hardware or the cloud)
- A new DBMS



You are being called on by sales personnel from multiple vendors that all have in common:

- Promises
- References of their customer's successes and money savings

How do you choose? Every platform will probably work. You know this because each sales team told you so. (They look trustworthy). But any change will require expense, effort by your technical staff, deferral of other analytic development work, retraining of users and conversion of user's reports. After all of that, will it be better or worse than today? Will it really save money? Will your users need to make compromises? Will you be rewarded or looking for a new job?

So, you narrow down your choices to a small number of likely candidate technologies, and you sponsor comparative testing. New capabilities are tested with Proof of Concept (POC) tests. For those, you conjure up small experiments with one or more new features.

But most likely you will need to move your existing workload to the new analytic platform as well and the comparison of performance between two or more platforms is a benchmark.

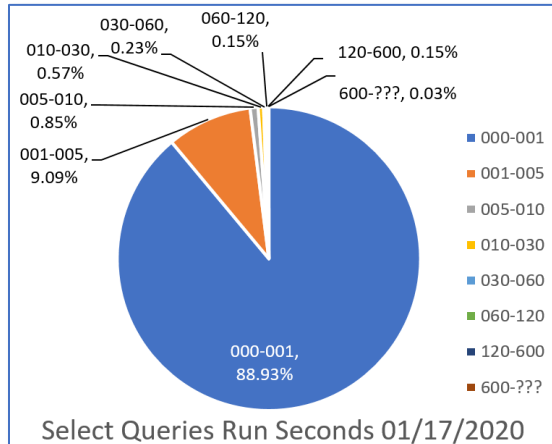
... and the deception begins.

"Give us your 5 worst queries and we will beat the pants off your current platform" says one vendor. (That is obviously a strategy their salespeople have been successful with or they wouldn't be proposing it). The question you should consider: Does the fact they can run 5 queries faster than on my current platform prove they can run all of my workload faster?

... another vendor uses deception cut from a different cloth.

“We recommend you run an industry standard benchmark because it is free of vendor bias.” Oh really? The question you should consider, “is that because you believe it matches my workload needs or because you believe you can win with it?” Again, no salesperson is going to recommend tests where they think they will lose.

Here is an actual workload profile for a typical enterprise data warehouse customer:



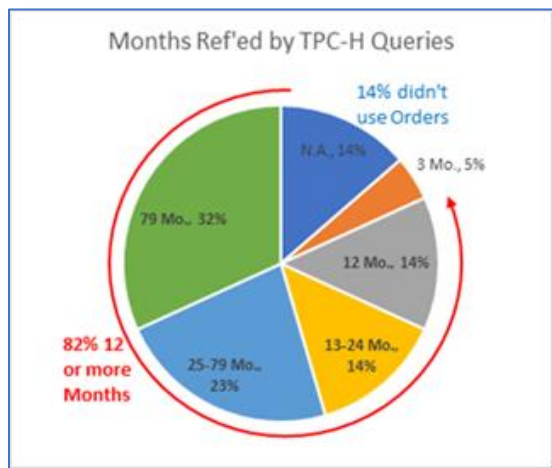
I’ve studied workload profiles across multiple industries and companies. Surprisingly 60-90% of the queries complete in less than 1 second and 5-10% in the 1-5 second range.

So what?

If you went along with a vendor suggested test of your worst queries and they performed better, then all you’ve proved is that 0.1% of your workload will run faster. But that isn’t what most of your users are doing, so you’ll have some explaining if the majority of your workload runs worse.

Oops.

If you went along with the vendor suggesting the “Industry Standard Benchmark”, notice the pattern of query references by the TPC-H benchmark below:



Where are the queries by the order clerk about an order, or a sales rep about their customer's business or the marketing analyst about the new item introduction or the sales district manager about month-to-date vs quota or the logistics manager about parts on order?

Oops again.

Not all deception is externally driven.

In one benchmark at a US insurance company, the customer's DBA constructed a test where 500 simple queries were released at the same instant, and another test with 1000 queries released the same way. The tests lasted less than 1 minute each. The winner of the contest was faster due to less query optimization. Six months later, the company was living with recurring crashes when attempting to process a normal workload. Apparently, all users didn't get the memo to only run simple queries all at the same time. Who knew?

In a benchmark at a large mortgage institution, they planned to run 56 queries in each concurrent session ranging from 1 to 500 sessions in 9 increments. Given that the test for 5 concurrent sessions ran for 1 hour, they had to reduce the scope to avoid taking several hundred hours testing on each of the three vendors. They selected 8 queries that consumed 402,813 seconds of CPU and produced an average of 199,212,977 rows of output. (Seems like a lot of reading material for the average user). To avoid filling up the server simulating the users, they modified the queries to return the count of rows that would be created. With the change, the DBMS recognized it didn't need to produce the rows to say how many rows would be produced. The (non-representative) 8 queries only used 15 seconds of CPU with the flawed testing. It would make equal sense to have all sessions execute "SELECT 2 +2;".

Did they do their due diligence in selecting platforms? Uh, No.

The answer: In over a decade of executing and assisting other in performing benchmarks, I've come up with a simple recipe that gives a good basis of selection without boiling the ocean:

1. Profile the current workload (by application, query type, duration, tables, etc.)
2. Select a representative workload matching that profile and adapt it for a benchmark
3. Take a snapshot of your production data (protecting sensitive and PII data)
4. Test the selected workload at various concurrencies using a query driver designed for data warehouse simulations.
5. Test with some data maintenance along with the queries to make it "real world"

It should only take a couple of weeks to prepare a fair, representative benchmark test. It is a small investment in design to avoid a mistake in a huge investment when you choose a new analytic platform. For those who want to know more, I've laid it out in detail in "Benchmark Deception...and How to Avoid It."

For those who don't want to follow the simple recipe above, may I suggest a simplified technology selection approach? Flip a coin. It's quick. And you should have enough time leftover to freshen up that resume.