
Questions to Ask a Data Warehouse Appliance Vendor

Guidance for Evaluating Benchmarks and Interpreting Market Messaging

Sponsored by:

TERADATA
Raising Intelligence

By Neil Raden

Hired Brains Research

July 2006



Table of Contents

Executive Summary	3
The Proprietary/Non-Proprietary Debate	3
Cost Savings	3
Questions to Ask an Appliance Vendor	5
1. Do you cite benchmarks?	5
2. Are your benchmarks verified by a third party?	5
3. Do you compare against other databases generically?	6
4. When were all benchmarks performed?	6
5. Where were these benchmarks performed?	6
6. Who conducted the tests?	6
7. Were install and setup times tested?	6
8. Do you completely identify configuration?	7
9. Do you identify configuration of your competitor(s)?	7
10. Were other vendors aware of the benchmarking?	7
11. Did benchmarks violate license agreements?	7
12. Compare workloads of systems at time of benchmark	7
13. What did the benchmark measure?	8
14. Did benchmark consider fault tolerance, update load or scalability?	8
15. Did you test and report concurrent, in-flight queries?	8
16. Benchmark average age of your hardware and competitor's	8
17. Are competitors' benchmarks conducted by experts?	9
18. Describe tuning in your benchmark	9
19. Describe tuning in other systems	9
20. Describe projected load on system when in production	9
21. Who created the benchmark?	9
22. Define the term "other databases"	10
23. Do you use any third-party benchmarks?	10
24. Do you recommend offloading from the EDW with your appliance?	10
Conclusion	11
About the Author	11

Executive Summary

While there is no firm, official definition for a data warehouse appliance, the market is gradually converging on one – a combination of database software, optimized for data warehousing, with its own physical platform, including operating system, server (CPU and RAM), and integrated mass storage. While also leaning towards an open source and non-proprietary flavor, current offerings are a mix of proprietary and non-proprietary physical components and databases that may be derived from open source, but have been extensively modified. The more well-known appliance vendors are Netezza and Datallegro. There are lesser-known ones, too, and undoubtedly more will appear over time. Though some consider Teradata Corporation to be the original appliance vendor, the market perceives it to be the competitive target of the new entrants to this market. Certainly Teradata is the incumbent with the enviable client list, especially the largest, most vital, and most advanced data warehouses.

The Proprietary/Non-Proprietary Debate

The appliance vendors begin with open source relational database systems, but they all modify the systems and extend them with enveloping extensions that cannot be ported to other hardware platforms. The concept of a relational database designed and tuned for one and only one physical platform is not new. In fact, historically it has been more the rule than the exception. Teradata is now available for Microsoft Windows operating systems and will soon be available for Linux, but until a few years ago, it was wed to a single hardware/OS combination. Other successful examples (at the time) of this approach were Tandem, Digital's RDB, and IBM's DB2 for mainframes and DB/400 for the AS/400 environment. There is no inherent advantage or anything new about a database system wed to a single platform. The appeal of data warehouse appliances must lie elsewhere.

Appliance vendors cite a price differential and point to their use of non-proprietary hardware. In fact, no hardware is non-proprietary, in the strict sense. Using generic 7200 rpm disk drives, Intel Xeon chips or Altera FPGAs (Field Programmable Gate Arrays) implies a license arrangement with those manufacturers. Assembling the components in an appliance-designed enclosure and backplane, with interconnect hardware and software, can hardly be considered non-proprietary.

Implementing an open source database, such as PostgreSQL or

Ingres, could be considered a cost savings, but in some cases these databases have been massively modified and enhanced. And in most cases, the vendors may not be participating in the open source community, so they must bear all of the subsequent development costs themselves, and cannot avail themselves of the advances provided by the open source community, in the same way that

a purchased application customized by a customer is not upgradeable. In fairness, certain enhancements from the open source community on the periphery of the product might still be usable, such as ODBC or JDBC interfaces.

Cost Savings

This begs the question, what are the sources of the cost savings that the vendors can pass along that they claim is as much as 50%? There are a number of answers. In some cases, these appliances are stripped down versions of fully functional relational databases for data warehousing. They lack much of the functionality that has evolved over 25 years of data warehousing. They are Moore's Law pure plays, substituting less expensive, less performant hardware (particularly disk drives) for intellectual capital and taking advantage of the technology cost curve that is now so evident. Close examination shows that they're largely feature poor, not only in

... there is no firm, official definition for a data warehouse appliance.

database function, but in vitally essential capabilities, such as system management, load balancing, and mixed load management.

In other cases, they have turned to less expensive hardware components with less stringent performance and reliability specifications, such as the chosen CPU. But the more dramatic cost savings may come through the use of less expensive disk drives with a lower Mean Time Between Failure (MTBF). In essence, they're trading off availability for lower price.

Does this imply that these appliances have no value? Of course not. But the purpose of this paper is to provide some guidance when comparing these tools head-to-head to the more mature, more capable, more "hardened" data warehouse-enabled database systems.

Much of the market messaging of these emerging vendors focuses on price/performance. On the surface, this makes for easy, attention-getting slogans, but it can also be extremely misleading. While performance in terms of query speed in a contrived benchmark may seem compelling early in the evaluation cycle, performance gains in one area may hide performance weaknesses in other areas. The myriad workload possibilities in a true production environment and the variety of query types presented, in parallel, can easily confound a system designed to perform in one mode only. This is precisely why the TPC develops benchmark *suites*, to measure the performance of a system across a wide range of probable situations. And query resolution speed is only one measure of performance. A focus on query performance may not expose a glaring lack of advanced database and data warehouse management features. These shortcomings typically manifest themselves as more difficult and expensive production environments, if not outright failure of the project.

Questions to Ask an Appliance Vendor

Like any new technology that enters the market, it's useful to separate the wheat from the chaff in the marketing messages. Some of the appliance vendors marry their sparse functionality and abundant hardware to contrive testing scenarios that show their products in the most favorable light. There is nothing wrong with this; any organization in a competitive market would be remiss if they didn't do the same. But it does highlight the need for the market to be vigilant and ensure that the marketing messages don't wander into the realm of misleading or outright false claims.

To assist you in evaluating these products, we've assembled a list of annotated questions to ask an appliance vendor about their benchmark claims. Because of the strong tendency of these vendors to exploit supposed "benchmark results," these questions are limited strictly to benchmarks and don't cover all of the other considerations in deciding on a platform vendor, such as market position, tenure, client references, depth and breadth of functionality, proven reliability and availability, professional services, or size of the user group.

1. Do you cite benchmarks?

Do you cite benchmarks in your material or in your sales presentations?

The use of performance benchmarks is an indication that an appliance vendor chooses to highlight certain kinds of performance as a competitive advantage. This may be instead of, or often in addition to, other measures, such as TCO (Total Cost of Ownership), breadth and depth of functionality and startup costs. The key point to keep in mind is that if performance benchmarks are prominent in a vendor's value proposition, it is vitally important to understand how the benchmarks were performed and even more importantly, what sort of conclusions you can draw from them.

For example, I might offer an automotive technology that can deliver 100 miles per gallon, but if it requires replacement of an expensive part (such as a battery) every 10,000 miles, the cost

savings vanish in short order. If a benchmark can show that a database can load new records ten times faster than a competitor's, it might come as a surprise that database reorganization is required frequently because of fragmentation or skew, which requires considerable downtime.

Remember, there isn't a single metric in the world that is useful unless its methodology is exposed, and it can be compared to something else.

2. Are your benchmarks verified by a third party?

Are your claims in your benchmarks about your performance or your competitors verified by a third party?

Third-party organizations involved in setting up and monitoring benchmarks achieve only modest results because of the large number of variables in database performance. These standards organizations attempt to level the playing field by having all participants perform, in a verifiable way, a standard and repeatable set of tests, but even these benchmarks have serious drawbacks. Vendors may configure their systems in a way that favors the benchmark, but customers apply the technology to other kinds of applications that cannot maintain the same performance without extensive tuning. There are many factors that can lead to large differences in benchmark results.

When a benchmark by a vendor is not monitored by a third party, the objectivity of the results has to be an issue for examination. In one recent case, an appliance vendor ran a side-by-side benchmark with an incumbent vendor. The most current versions were used for the appliance, while results from the incumbent were based on a functionality test, not a performance test, which had been done four-years ago. In addition, the functionality test used technology that was at least two generations old in both the hardware and the software. Needless to say, the appliance product produced much better numbers and the presented comparisons made no mention of the disparity in time and product generations.

3. Do you compare against other databases generically?

Do you distinguish in your statements between “other databases” and those that are specifically designed for large-scale data warehousing?

When it comes to data warehousing, there are two kinds of relational databases – those that are general purpose, or “merchant” databases, such as Oracle or DB2, and those that have been designed from the ground up for analytical processing, such as Teradata® Database. For an appliance vendor to state that they perform 10X – 100X better than “other” databases is extremely misleading. In one particularly egregious case, a vendor claimed a 1000X improvement over an unspecified competitor. In fact, it was a ten-year-old internal benchmark that was being compared to their own older database product.

4. When were all benchmarks performed?

When were your benchmark tests performed? When were your competitor’s benchmark tests performed?

Timing of benchmarks is crucial because the progression of technology is a major factor. A four-year-old system may still be perfectly usable in an organization, but its raw performance simply cannot compare with something state-of-the-art. Features added to the products over that four year span may yield significant performance improvements on a wide range of queries without the need to change technology.

5. Where were these benchmarks performed?

Where were these benchmarks performed?

When an appliance vendor produces a competitive benchmark, it implies that the vendor conducted a benchmark on a competitor’s system. Under what circumstances can that occur? Does the appliance vendor own a fully configured environment of the competitor(s)’ system? That would be very unlikely. Was the appliance vendor in a competitive situation where the incumbent

was an unwilling or unknowing participant in a competitive benchmark? If there was an incumbent, how exactly was there time to take the system out of production to set up and run an apples-to-apples benchmark? Was this production system using the latest hardware and software versions?

6. Who conducted the tests?

Who conducted your benchmark tests? Who conducted your competitor’s tests? Who prepared the statistics (those you are using in your sales and marketing material) in each case?

It may seem too obvious to ask, but knowing who actually performed the benchmark, especially if that party is available for independent confirmation, is quite useful. A few metrics rarely tell the whole story. Response may range from, “We ran it seven times before we got it to the number we were looking for,” to, “We could have gotten a much better time, but we only had one afternoon, and the DBA got called away.”

Like any test, the setup and running of the test is one sort of discipline, measuring and documenting what happened is another.

Like any test, the setup and running of the test is one sort of discipline, measuring and documenting what happened is another.

7. Were install and setup times tested?

Do you compare install and set up times, as well as query performance?

Because an appliance is a complete bundle of server, storage, and database software, appliance vendors often claim that the system can be running out of the box in a few hours. The key point is that it doesn’t matter if you can unpack an appliance and have it running in an hour. Once the system is running, engineers should run extensive diagnostic tests for at least a day to be sure there is no damage to components from shipping, such as loose boards or damaged disk drives.

In addition, the most time consuming part of a data warehouse project is data management, including finding, profiling, cleansing,

extracting, transforming, and loading the data. It takes a lot longer to actually design and build a warehouse than it does to configure the platform. If an appliance can truly be running in an hour compared to another system that takes a day or two, those days are just layered into a six-month or 12-month or 18-month project plan. It isn't going to happen repeatedly and wouldn't slow anything down.

Also, quick setup times with minimal schema requirements can be a strategy to focus the benchmark on only those functions or queries that perform well with data structured that way. Non-indexed, large tables and queries that require full table scans will perform better on a bare-bones system that is optimized for large-scale throughput. Data warehouse environments are complex and are getting more complex with the emergence of operational BI and composite applications requiring a blend of OLTP and decision support functions.

8. Do you completely identify configuration?

Precisely and accurately identify the configuration of the system(s) on which you ran your tests, including

- a. Hardware
- b. Network configuration
- c. Other software running on the system(s)
- d. A thorough description of the benchmark
- e. Specific version of the Operating System
- f. Specific version of the Database engine
- g. Specific configuration settings for the Database engine

Unless a vendor can provide this information, for both its own configuration and any others compared in the benchmark, you should consider the information of questionable value.

9. Do you identify configuration of your competitor(s)?

Also provide the same information for the competitor's tests

It is particularly important to understand the conditions under

which the competitor's system was configured because the appliance vendor is typically not able to control this and can even exploit this. Be especially careful of sweeping assumptions like, "We'll run it at 4:00 a.m. when the database is quiet," or, "The normal workload won't interfere with this benchmark because it's only running at 20% right now."

10. Were other vendors aware of the benchmarking?

Were all of the vendors aware of these benchmarks of their systems?

It is far more likely that benchmark results will be skewed or even worthless if the vendors being benchmarked are not aware of the tests. Although client DBAs may be quite fluent and efficient with the normal operation of the system, benchmarking is a particularly fine skill and pitting in-house DBAs against the best the appliance vendor has to offer will yield misleading results.

11. Did benchmarks violate license agreements?

Did the licenses for these systems allow for these benchmark results to be published?

No vendor can prevent a customer from testing their system, or running their own internal benchmarks. However, most license agreements and Non-Disclosure Agreements preclude clients from publishing benchmarks without the consent of the vendor. In fact, if the benchmark was actually performed by the appliance vendor, it would also represent a violation of the license agreement, and its publication would be a clear violation of the Non-Disclosure Agreement.

12. Compare workloads of systems at time of benchmark

What was the load on the system(s) – yours and your competitors' – at the time that the benchmark was run?

An appliance arriving fresh from the factory for a benchmark can be configured anyway the vendor wishes. A system in production simply cannot be wiped clean and similarly configured to perform a particular benchmark.

13. What did the benchmark measure?

What was the goal of the benchmark – to benchmark query response time for a single query at a time, throughput of a large number of queries, table scans, joins, updates, etc.?

A high-performance relational database platform for data warehousing has to provide excellent performance under often extreme requirements, many of which may be unpredictable. While a benchmark may provide an indication of speed and throughput in a controlled environment, unless it is sophisticated enough to model an actual production environment, it represents only one small piece of information.

Large queries submitted one at a time may exhibit excellent response times, but how well does the system handle multiple large queries simultaneously?

Being able to scan large tables at blinding speed is clearly an advantage, but there may be queries that can be resolved far more efficiently with selective use of indexes or partitioning, good optimizer technology or even SQL query rewriting. After all, not every query submitted for execution is written in the optimal form. It hardly makes sense to have the fastest elapsed time to read a billion records if you only needed to read 100,000 in the first place. Optimizer technology is often the crown jewel of a database system, and it matures with time and experience. Appliance platforms that focus on raw horsepower have not invested in optimizer technology yet.

14. Did the benchmark consider fault tolerance, update load, or scalability?

Were other tests considered, such as fault recovery under a heavy query and update load or scalability/expansion tests?

Everything doesn't always go according to plan. If a system reaches an unstable state, the single most important benchmark metric may be the time to recover and restart production. Alternatively, can the system fail in one process or even one component and still elegantly continue by shifting the load? The system components that will fail most frequently in large systems are the disk drives,

and not all disk drives have been engineered for enterprise class workloads and reliability. How often will disk drives fail? Were disk drive failures simulated on a fully loaded system during benchmark execution? In fact, what are the reliability statistics for each component of the system – the MTBF, uptime percentages, and other RAS statistics in a real situation? Can these be verified through references that have had the respective platform in production for at least 18 months?

15. Did you test and report concurrent, in-flight queries?

Was the number of concurrent, in-flight queries measured and reported for you and your competitor?

There are many ways to meet system performance requirements, but execution of a single query is not a sufficient measure. For example, suppose it can be demonstrated that three large queries can execute serially in ten seconds, 30 seconds, and three minutes, for a total of three minutes and forty seconds. Suppose further that a different system has a longer elapsed time for the three queries, but can perform them if submitted simultaneously in less time. Which is the more valid benchmark? Appliance vendors seem to focus their benchmark timings on single queries, executing one at a time. A more accurate and predictive benchmark would model concurrency, which produces very random I/O. This is where the cost differential of inexpensive disk drives in an appliance would show its downside. A concurrent query benchmark demonstrates not only query speed, but load balancing of a mixed workload, which is a reality in production.

16. Benchmark average age of your hardware and competitor's

When benchmarks are conducted with a prospective customer, what is the average age of your hardware? What is the average age of the hardware you are benchmarking against?

Because appliance vendors don't maintain their own benchmark labs of competing vendors, they rely on the goodwill of their prospective clients to set up and benchmark in the competitor's system. It stands to reason that this equipment is at least a few

years old (it's unlikely they would entertain another vendor shortly after a purchase or upgrade). With CPU speed doubling every two years, comparing a three to four-year-old platform would put the older unit at a distinct disadvantage. CPU speed isn't everything, but all of the other components and technology are also subject to dramatic performance improvements.

17. Are competitors' benchmarks conducted by experts?

Are your benchmarks of your competitors conducted by parties who are experts with the system?

In fact, were benchmarks actually conducted at all, or with the same precision as the appliance vendor's, or are the numbers an estimation or extrapolation based on past performance and/or anecdotal evidence? In any case, if the benchmark was actually conducted, were the qualifications for the person(s) conducting it of equal skill and experience to those conducting the appliance benchmark?

One thing to consider is that if an organization is considering a competing offer from an appliance vendor, head-to-head as a replacement, it's reasonable to assume that the existing system may not be performing optimally in the first place and those conducting the benchmark

may not be more effective at the benchmark than they are with the production system. The solution to this is to engage someone from the existing database organization to supervise the benchmarking activity or an independent third party with adequate background.

18. Describe tuning in your benchmark

Were you able to improve performance further or reduce resource utilization through tuning? If so, describe the tuning of your system for the benchmark?

Appliance vendors often claim that their systems require no tuning, but even the preparation of the data, the physical database schema or the partitioning methods are examples of tuning. However, when benchmarking only a narrow set of functions,

particularly those that scan large tables and do not use indexes, the value of tuning is minimal, but in production, the breadth of functionality needed will very likely be much greater.

19. Describe tuning in other systems

Describe the tuning of the other system(s) for the benchmark.

What expert advice was sought in determining what tuning, if any, was necessary?

In particular, an existing system may already be tuned for known workloads, which may be inappropriate for the set of tests in a benchmark.

20. Describe projected load on system when in production

What will the load be on the system when the intended application is finally running?

If there is already a data warehouse in place, and the appliance is being considered for re-platforming, this question is easy to answer. However, if it's a new installation, great care should be taken, including seeking the advice of experienced data warehouse professionals who can evaluate and extrapolate the requirements. However, each installation is unique, and it's not simple to forecast the

requirements with a high level of precision. Data warehousing is one area where it's always best to not paint yourself into a corner by meeting minimal requirements. An appliance may turn out to be too limited for the task at hand.

... the true test of a system is how well it executes the tasks it will ultimately be required to perform.

21. Who created the benchmark?

Who created the actual benchmark? Was the benchmark based on a single or set of generic queries, a set of queries specific to a given customer or prospective customer, a set of generic benchmark queries from a third party, or something else?

Generic benchmarks are useful (to a point) when applied evenly to all the participants, but the true test of a system is how well it executes the tasks it will ultimately be required to perform.

22. Define the term “other databases”

When you compare your performance to “other databases” without specifying which ones, should a customer construe that to mean in all cases or just in some?

Each database is different. When an appliance vendor claims its performance is 10X – 100X better than the “other” databases, this could mean anything from Access to Teradata Database. Logically, the statement, “We outperform the class of databases composed of SlowDB, SlowerDB, and Teradata Database, the statement may actually be true, but it doesn’t mean that it outperforms *each* of the databases. Evaluate these statements very closely.

23. Do you use any third-party benchmarks?

Do you provide any third-party benchmark data, such as TPC-H, Winter, OSDB (Open Source Database, PolePosition, etc.)?

Reputable third parties construct and certify robust benchmarks for database vendors. While the TPC-C benchmark is the most well known, and is controlled by the TPC Council, its transaction processing nature makes it an unreliable test for data warehousing. The TPC Council has worked to improve a benchmark for “decision support” processing, the TPC-H (though it is still far from useful for data warehousing). The TPC is also finalizing a new DSS benchmark called TPC-DS, which will probably become available in 2007. The Winter Corporation provides in-depth research and analysis in the Very Large Database (VLDB) market.

24. Do you recommend offloading from the EDW with your appliance?

Divide and conquer: Do you recommend that customers use the data warehouse appliance to offload or summarize parts of a larger data warehouse?

This is an approach that was used by the merchant databases to work around their inability to provide a suitable platform for enterprise data warehousing. In the long run, it just created more silos of information that rapidly got out of sync with the rest of the data, adding multiples of maintenance and most importantly, severely limiting the range of analysis possible by arbitrarily partitioning very fine-grained data from summarized and aggre-

gated data. Despite this well-documented drawback, the appliance vendors pitch this approach when they are unable to secure the enterprise data warehouse account. For example:

Netezza: Aggregation is one of many areas where minimal latency allows the NPS system to be used in new ways to improve business performance. For example, some Netezza customers use the NPS system to build aggregate tables for other systems that are used as “marts” in the data center. This approach tends to be ten to fifty times faster than using incumbent systems (e.g., Teradata, IBM, Oracle, etc.) and helps boost data center productivity.

Note the use of the phrase “tend to be.” Also the range, “ten to fifty times faster” – would it be reasonable to say that I am six to thirty feet tall or that my child’s GPA is 3.5 to 17.5? And last, the phrase, “Teradata, IBM, Oracle” leads one to believe that this approach is ten to fifty times faster than each of them. Teradata’s ability to create virtual marts has to be infinitely faster, because it takes no time at all. These are the kinds of claims that require constant scrutiny.

Datallegro: Learn how you can “divide and conquer” your data warehouse by segmenting a specific business function or requirement and offloading it from your enterprise data warehouse to a low cost, high-performance appliance. By offloading these tasks from the enterprise data warehouse to the appliance, companies are greatly reducing the need for an expensive upgrade of the enterprise data warehouse. In addition, the specialized nature and advanced technology of the appliance enables the above processes to run significantly faster, often by two orders of magnitude.

The passage refers to “companies,” but at the time of writing, it was unclear whether this vendor had any clients actually in production or moving beyond the proof of concept phase. Again, notice the misleading language – “often” by two orders of magnitude. That would be 100 times faster. The questions to ask are, “Where is the benchmark,” and, “Faster than whom?”

Conclusion

Appliance vendors don't have a monopoly on tortured benchmark statistics or marketing hyperbole. All vendors in competitive markets engage in it to some extent, though newer entrants tend to be the worst offenders. But database performance for data warehousing is a very complicated issue because data warehouses have to operate in so many different modes, often simultaneously. If there were a single, simple measure for capturing all of the variations, like annual rainfall or miles per gallon, there would be no room for exaggeration. The point of this paper is to help you, the customer, critically evaluate these messages and be aware of common tactics to tilt the scales in favor of the messenger. It is important to be able to see through the claims and not end up with a pig in a poke.

Data warehouse appliances are not a new idea. Teradata pioneered the idea more than two decades ago. The success of Teradata validated the idea and now others are offering variations on it, further validating its utility. But Teradata has also demonstrated that it requires a great deal more than a good appliance to succeed in the largest and most important data warehouses. It takes staff, services, partners, and most of all, experience. These are all factors that should be part of any benchmark.

About the Author

Neil Raden is the founder of Hired Brains, Inc., <http://www.hiredbrains.com>. Hired Brains provides consulting, systems integration and implementation services in Business Intelligence, Data Warehousing, and Performance Management for clients worldwide. Hired Brains Research provides consulting, market research, product marketing and advisory services to the Business Intelligence, Data Warehousing and Semantic Technology industries. Based in Santa Barbara, CA, Raden is an active consultant and widely published author and speaker. He welcomes your comments at nraden@hiredbrains.com.

