



transforming data into understanding[®] Series

DATA WAREHOUSE APPLIANCES

BALANCING PRICE AND PERFORMANCE

By James Newland

The best solution for your requirements is not necessarily the most expensive one. In fact, many businesses overspend on their data warehouse, purchasing a system that is overkill for their requirements. Worse still, even expensive DW appliances can fail to meet your performance requirements.

In Part 2 of the series Data Warehouse Appliances, we will help you identify the tradeoffs between price and performance. By understanding these issues you can make intelligent decisions about how best to satisfy your requirements while maximizing your data warehouse budget. Some of the areas of tradeoff are:

- Speed
- Implementation Architecture
- System Availability
- Storage Efficiency
- Usable Terabytes

Should you need more information on any of these items or any other data warehouse appliance considerations, please give us a call – our consultants are happy to help.

Speed

Speed is the reason you are purchasing a DW appliance solution, right? But how fast is fast enough? You may not need the absolute fastest solution on the market.

Speed comes at a price and you should consider carefully how much speed you actually need. If your reports are currently taking 3 hours to run and one DW appliance can get them to you in 30 seconds while another in 1 minute (twice as long as the former!) which should you choose?

Remember that the performance improvement DW appliance companies offer over traditional storage solutions is sometimes orders of magnitude. Given the improvement over your existing solution, two or three times difference between the DW appliances may not make a big difference at all.

Implementation Architecture

How does your DW appliance company propose to implement their solution? There are more than a few ways to do it.

The mother lode for DW appliance companies is a 'floor sweep' of your existing hardware and software. This means that they come in and replace everything you have with their solution. Many vendors will push for this option as it means more revenue for them as well as hurting a competitor.

This may be the right thing to do, especially when the current solution meets very few of your requirements and has a high maintenance cost. This is also the most expensive and perhaps painful solution. The time to implement this solution can be long and the cost high. An implementation that takes too long risks making the hardware purchased at the beginning of the project obsolete before the implementation is completed!

An alternative is to offload some of the processing to an appliance system. This setup allows you to keep your existing solution and still take advantage of the performance of a high-powered system. For example, data that is analyzed or reported on frequently can be moved to a DW appliance. In this scenario you get blazing reports at a fraction of the cost. On the other hand, you will have to duplicate data for many of these solutions to work. In addition, you may only be delaying the inevitable floor sweep if your existing solution is outdated.

System Availability

Some systems are 'highly available' while others are less so. The term 'highly available' means that the system has a very low rate of unplanned outages. Most high-availability systems have essentially 100% up-time. This is a desirable trait in a data warehouse. It can also be expensive.

Highly available systems generally have no single points of failure. 'Share-nothing' architecture, redundant storage and redundant systems are common methods for high availability. Individual component failures reduce performance, but system availability is maintained. Component replacement usually can be completed without a system outage or downtime. To get assurance of a highly available system you will have to look at established solutions with proven track records of implementing highly available systems.

Established solutions tend to be more expensive than the new solutions as the newer solutions compete mainly on price. Not everyone needs a guaranteed highly available system though. If you can stand a few unexpected outages it might be wise to spend a little less money here for gains in other areas.

Storage Efficiency

There are significant differences in the way DW appliances compress data for storage. Some of this difference is architectural. Even within a storage architecture type, there can be differences.

Columnar storage architecture tends to be more aggressive at compressing data than row-based storage architecture. All other things being equal, you will get more data stored per terabyte on column-based system than a row-based system. Despite these differences, this should not be one of your major deciding factors in which system to choose. Row and column-based systems are good at different tasks and that difference will be more important to you in the long-term than storage efficiency.

Within storage architecture types (row and column) there can be significant differences in data storage efficiency. Once you have decided on which architecture is right for your business, this could play an important part in your decision. The ratio of raw data to stored data can range from 1:2.5 up to 1:10. Many factors contribute to the wide range including model design, redundant data, indexes and view design. Efficiently using storage space helps you minimize costs in the long-run. Each additional gigabyte of drive required is another gigabyte of load and backup. With concerns about 'greener' systems, power consumption and cooling requirements should also be considered. Take a look at the different DW appliances within a storage architecture type and let efficiency be one of your acceptance criteria.

Usable Terabytes

Be careful when examining the number of terabytes storage the DW appliance is offering. The number of terabytes of storage is NOT the number of usable terabytes. As you fill disks towards their capacity, you will experience a significant performance hit with most DW appliances. The magnitude of the effect is dependent on several architectural factors, including size of the disk and I/O. In addition, disk overhead may take up a significant chunk of your storage space, up to 25% of raw disk space.

Most DW appliance companies will be up front about usable terabytes vs. raw storage. However, it is worth asking the question.

Conclusion

As stated in Part I of this series, DW Appliance Architecture, your requirements should drive your DW appliance evaluation. Once you have settled on the major requirements, however, it is worth taking a look at the various claims – speed, storage capacity, compression, etc. and deciding where you can sacrifice a little performance for major cost savings.

About Datric, Inc.

Datric, founded in 2000, is headquartered in Charlotte, NC, USA. Datric is a premier provider of data integration and collaboration solutions, specializing in SAP and data warehousing. Datric consultants have experience both architecting and implementing data integration systems for many Fortune 100 companies. We combine a simple, time-honored approach to business with leading-edge technology to help you achieve your goals. For more information about our products, services and clients, please visit us at <http://datric.com>