



Craig S. Mullins

Database Performance Management

[Return to Home Page](#)

April 1998

Computing : News & Review

The Age of the VHDB

By Craig S. Mullins

Databases are growing in size. There is no denying that simple fact of life. I have talked to hundreds of DBAs and visited many sites and not a single one of them report that their databases are getting smaller. This has been true for many years, but the pace at which database sizes are growing is at an all time high.

For several years now database experts have used the acronym VLDB, or very large database, to refer to the largest databases in production environments. However, some production databases are now approaching a petabyte in size. Refer to Figure 1 for an idea of how large a petabyte is and where we will inevitably go from there. It demeans a database of this size to call it "large"; this is a "huge" database. Hence the updated terminology VHDB, or very huge

database.

Figure 1. Storage Abbreviations

Abbrev.	Term	Amount
KB	Kilobyte	1,024 bytes
MB	Megabyte	1,024 KB
GB	Gigabyte	1,024 MB
TB	Terabyte	1,024 GB
PB	Petabyte	1,024 TB
EB	Exabyte	1,024 PB
ZB	Zettabyte	1,024 EB
YB	Yottabyte	1,024 ZB

There are many factors influencing organizations to support and foster this growth. People are creating data warehouses to enable analytical processing on vast amounts of historical data. And they are creating a lot of indexes on this data to enable rapid data access. Indexes require even more storage space, further increasing the overall size of the database.

Data mining is fast becoming a requirement whereby heuristic algorithms are applied to historical data to automatically discover patterns in the data that can be exploited for competitive advantage. The more data there is, the better the quality of the data is, and the quality of the pattern discovery algorithms determines the value of the data mining applications. So people are inclined to store more data for a longer period of

time.

Hardware improvements also spur this growth along. The hard drive in my laptop computer is bigger than the first mainframe hard drives I worked with years ago. The ability to cheaply store multiple gigabytes of information enables the creation, storage, and access of these VHDBs. Since the cost is so minimal, why not store more data?

But, unfortunately the speed of access has not kept up with the volume of storage available. The amount of storage space on a disk drive has grown nearly three orders of magnitude in the past 25 years. But the data exchange rate has changed only one order of magnitude in that same time. The increase in storage space vastly outpaced the increases in disk access speed. This causes hardware and DBMS vendors to keep pace by requiring additional main storage, caching data in memory, enabling parallel data access, and other techniques. This complicates database administration.

The net result of increasing database size is that the largest production databases are unmanageable even with the best tools that money can buy. Manageability includes, but is not limited to:

- database schema management (database change, migration, editing, etc.)
- backup and recovery
- contingency planning (disaster recovery)

- effective database access (SQL coding)
- performance management
 - SQL analysis & tuning
 - system analysis & tuning
 - database parameter tuning
- utility processing: reorganization, load, unload, etc.
- data purge and archival
- capacity planning
- data integrity verification
- security and authorization
- data movement (replication, transformation, propagation, etc.)

As database administrators struggle with ever increasing database sizes, the task of day-to-day administration becomes more and more difficult because of the added complexities of dealing with VHDBs. Some administration tasks may not make sense at all. If you can rebuild the entire database quicker than it would take to backup and recover it, then why bother with a backup and recovery strategy? Instead, deal with a recreation strategy. The DBA tasks that can be accomplished are dependent upon many factors including:

- your individual environment (hardware, software, DBMS, applications, etc.)
- available staffing
- availability requirements (e.g. 24x7)
- concurrent workload requirements
- DBA tools you have in house

- your overall budget (even if tools are available if you can't get budget, you can't use them to help)

The bottom line is that you need to be aware of the pace of growth for your databases and plan for what resources are required to manage your ever-increasing database portfolio. Failure to do can result in system downtime, lost data, and system degradation.

From [Computing News and Review](#), April 1998.

© 1999 Mullins Consulting, Inc. All rights reserved.
[Home](#). Phone: 281-494-6153 Fax: 281-491-0637