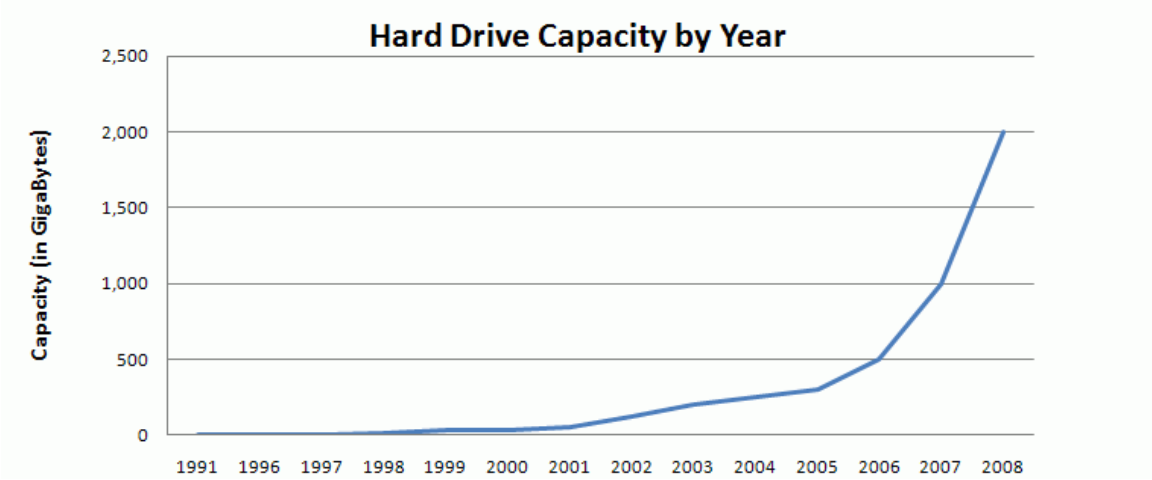
Why Are Server Backups So Painful?

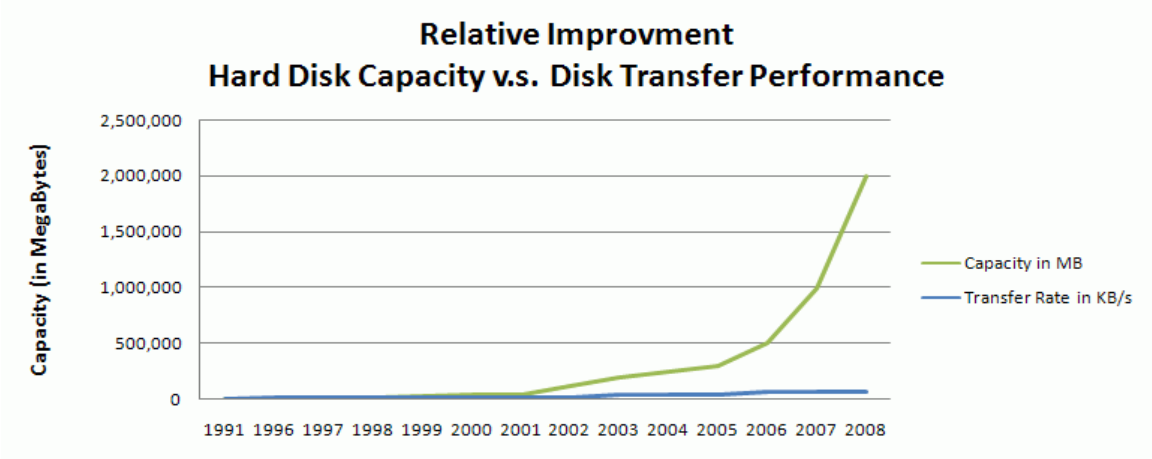
Why backups have such a dramatic impact on server performance

Data Sets Are Growing Rapidly



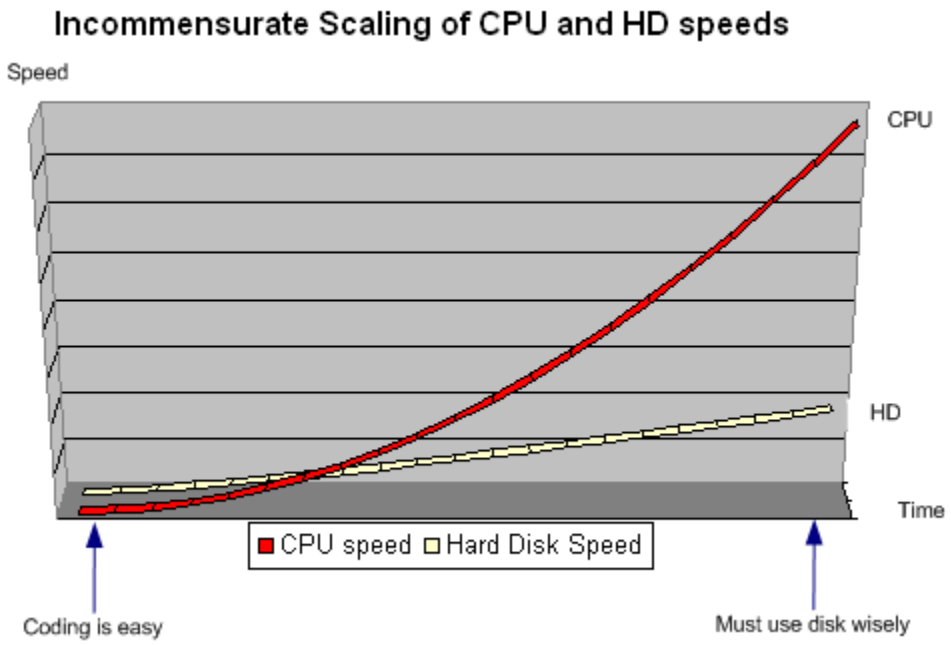
Data sets are growing rapidly, and storage is always the most precious commodity a server has. Above is a graph of hard disk capacity availability as it has grown over the years. These numbers are compiled using data from http://www.mattscomputertrends.com/harddiskdata.html. If you haven't ever seen Matt's Computer Trends, then you are missing out. According to his site, Matt spends his spare time searching old computer magazines to find historical specs on computer components, and the awesome data he has compiled over the last few years reflects this.

Disk Performance Improvements Are Not Keeping Up with Capacity Growth



Above is a graph depicting the relative capacity of hard disks vs. performance. Comparing performance of hard disks is not completely straightforward, so hard disk transfer rate is

measured in KB/s is used. These performance measurements are hard to come by. The transfer rates used here are from an excellent Tom's Hardware article that covers 15 years of hard disk performance and actually benchmarks the disks to get real performance numbers. While the article is dated from the fall of 2006 and the fastest drive it covers is a 10,000 RPM Western Digital SATA drive, the numbers don't matter much when put on a scale compared to disk capacity. Even if we tossed in a 15,000-RPM SAS drive manufactured in 2008, the transfer speed would not be noticeable when compared with the growth in capacity, and the Tom's article does a good job of driving this point home.



Above is a graph from the UCLA computer science department used in UCLA's Computer Science and Engineering (CSE 111) Introduction to Operating Systems lectures. The graph depicts the growth of CPU performance vs. hard disk performance. Clearly, disk I/O is the most precious resource a computer has.

Backup Technology vs. Hard Disk Capacity

Date	Capacity	Cost	Typical File Counts on Servers
1986 GNU Tar was Born	70 MB	$3,592	Hundreds or Thousands
1990 Legato Networker 1.0	270 MB	$3,500	Thousands
1999 rsync was born	32,000 MB	$299	100,000 - 500,000
2005 CDP for Linux Invented by R1Soft	300,000 MB	$300	1,000,000+

2008 Hard Disk	2,000,000 MB	$319	1,000,000+

This table above shows the evolution of backup software technology vs. hard disk technology. It also shows the typical number of files found on a server. A real challenge for typical backup software is that they have to examine every file on the server to determine if it has changed. While this examination is done using variety of techniques discussed later, the sheer volume of files that need to be indexed by the backup application is a massive bottleneck. For example, consider that unless you use R1Soft's Continuous Data Protection product, your backup application probably has to crawl the entire tree of files on your server. For some servers, this is not that bad. For others, it is very time consuming. Factor in the time to compute deltas or actually read file data, and it is no wonder that we only back up our servers up on a daily basis without Continuous Data Protection technology.