

Storage Challenge: Where Will All That Big Data Go?

Neal Leavitt



Big data creates numerous exciting possibilities for organizations, but first they must figure out where they're going to store all that information.

A key topic in technology today is big data.

Market research firm Aberdeen Group found that data growth has averaged about 35 percent annually in recent years, as Figure 1 shows.

According to Aberdeen senior research analyst Dick Csaplár, this means the amount of storage necessary to hold all this information doubles about every two years, barring the development of new technologies.

Big data has brought with it some big problems. One in particular is how organizations are going to store and keep up with this tsunami of information.

New storage technologies won't be available in the near future to help stem the tide, as they are still in the research stage, noted Mike Matchett, senior analyst with market research firm the Taneja Group.

For example, holographic storage—a radically different technology that promises vastly increased capacity—is still at least 15 years away from mainstream usage, added Pankaj Kumar, chief storage architect for Intel's Storage Division.

This means that existing technologies—led by the venerable hard drive—will have to step up. In fact, there are new advances that could significantly increase hard drives' storage capacity.

Cloud storage will provide ways for many organizations to handle increasing amounts of information.

Object storage will make it easier for storage to scale as needed.

And solid-state drives (SSDs) will offer faster performance, which is important for the many organizations that need to quickly access and make sense of their data.

However, attempts to cope with the growing volume of data will face challenges.

THE DATA FLOOD

The chief reasons why data volumes are increasing are more information shared between business partners; more channels for collecting data from customers, such as mobile applications; more Internet-connected devices; and more online sources of information, such as social media.

According to Intel's Kumar, many companies are gathering increasing

amounts of data for analysis to help them improve their products and otherwise gain a competitive marketplace advantage.

New analysis techniques are encouraging this trend, noted Matchett.

And the volume of machine-generated data—created by sources such as sensors—is growing and is now bigger than the amount of human-generated information, said Ken Wood, director of strategy in Hitachi Data Systems' Office of Technology and Planning.

Also, the amount of unstructured data is increasing rapidly. This includes high-definition video and still images taken by the rising number of mobile devices with cameras, noted Simon Robinson, senior analyst for market research firm 451 Research.

COPING WITH THE DELUGE

Several approaches have emerged for handling and managing large volumes of data.

Tools like Apache Hadoop help organizations work with the growing amount of information they are storing, said Currie Munce,

vice president for SSD development at hard-drive maker HGST, formerly Hitachi Global Storage Systems and now a division of Western Digital. Hadoop is an open source software framework allowing for the distributed processing of large datasets across clusters of computers using simple programming models.

Technologies that increase efficiency such as deduplication—which eliminates redundant data—and compression are also helping, said Intel's Kumar.

In addition, Munce said, shifts to alternative storage architectures are rapidly emerging. For example, scale-out architectures use multiple low-cost servers and other nodes to create a storage pool whose size and performance can be increased as necessary.

Hard drives

Hard drives, Munce noted, will continue to be the primary storage technology until at least 2020.

"They're ubiquitous, practical, and cheap," he explained.

Hard drives have an almost unassailable advantage in terms of density and cost per bit of data stored, noted University of California, San Diego (UCSD) professor Steven Swanson.

"There's no other technology on the horizon that has any chance of displacing disks for bulk data storage," he said.

As they have in the past, vendors are gradually increasing hard drives' capacity.

Kumar said that 5-terabyte hard drives are expected to debut at the end of 2013 or the beginning of 2014. Currently, the largest commercial drives hold 4 terabytes.

Three new hard-drive approaches promise to provide more capacity.

Nanolithography. This technology—which creates storage media much in the same way that lithography imprints circuit designs

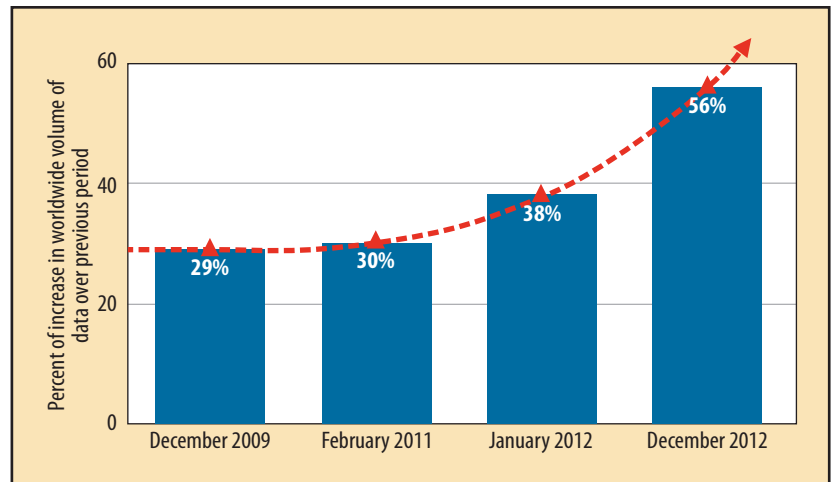


Figure 1. In the past few years, the amount of data that businesses store annually has increased significantly. (Source: Aberdeen Group.)

on chips—doubles hard-disk-drive capacity via two innovative nanotechnologies.

Nanoimprinting and molecular self-assembly help create hard drives with tinier storage islands than were possible in the past, which increases areal density.

Helium drives. HGST has announced the first helium-filled hard-drive platform, which it plans to release by the end of this year.

Rather than air, the drives are filled with helium, which is less dense and thus reduces drag on the disk platters.

According to HGST's Munce, this lets the drives support seven platters, rather than the maximum of five found in air-filled drives.

And with the reduced mechanical friction, the write arm can move more precisely to place bits closer together.

Analyst Fang Zhang of market research firm IHS said these capabilities could increase a drive's overall storage capacity by 25 to 50 percent.

Munce said the drives would also reduce energy consumption by 23 percent.

IHS predicts that more than 100 million helium-filled drives will ship by 2016, especially if today's high production costs are reduced.

HAMR. Heat-assisted magnetic recording (HAMR) technology, currently under development, shows promise as a way to increase hard-drive capacity.

HAMR uses lasers to heat high-stability storage media.

The technique uses disks made with iron-platinum and other alloys that enable greater storage density than materials currently used. However, these materials must be heated so that they are sufficiently magnetically malleable to store data.

In laboratory tests, hard-drive maker Seagate Technology used HAMR to increase a drive's areal storage density to 1 terabyte per square inch. Today's drives offer maximum densities of 620 gigabytes per square inch.

Seagate said HAMR could yield hard drives that store 6 terabytes of data in the near future and eventually a maximum of perhaps 60 terabytes.

Seagate said its plans to release its first HAMR products in 2016.

Solid-state memory

A major challenge for organizations is quickly accessing and moving huge quantities of data.

SSDs help with this because they have no moving parts and thus can access data faster than

hard drives, which must rotate to a given position before being able to read information.

“SSDs will serve as the working memory for big-data analysis,” said UCSD’s Swanson.

Their data-retrieval efficiency will be particularly useful for applications that access information irregularly, such as those that analyze large social networks, he noted.

SSDs are 10 times more expensive per gigabyte of capacity than hard drives. However, consumer SSD prices have been dropping while their capacity has been increasing.

SSD maker Micron Technology recently released the 960-gigabyte

avoid having to provide their own storage.

Cloud storage serves as a provisioning and storage model, providing on-demand, pay-as-you-go access to resources, noted Hu Yoshida, Hitachi Data Systems’ vice president and chief technology officer.

This is particularly useful for smaller organizations that collect large quantities of data but that don’t have the resources to store it on site.

Object storage

As businesses collect more information—particularly unstructured data such as multimedia files—administrators are having

high-speed SSDs provide, systems require faster bus technology than has been used in the past.

This has led to the increased popularity of the PCI Express (PCIe) bus, said David Reinsel, a group vice president with market research firm IDC.

Previously, SSDs and hard drives most commonly used the Serial ATA interface. SATA’s latest revision offers data rates that are much faster than a hard drive can read or write but not much faster than the latest SSDs’ speeds.

PCIe is becoming popular for SSDs because it provides better performance, scalability, and flexibility than SATA, said UCSD’s Swanson.

New hard-drive approaches promise to provide more capacity.

M500, which sells for about \$600, said company marketing director Kevin Kilbuck.

Researchers are working on new SSD approaches. For example, 3D techniques could stack multiple layers of NAND circuits on each chip, greatly increasing storage density. The first 3D NAND devices are expected in 2015.

Several vendors are working on resistive RAM technology. For example, SanDisk and Toshiba have jointly manufactured a 64-gigabyte test chip that uses ReRAM.

ReRAM applies electric current to a material, thereby changing its resistance. The resistance state can then be measured as either a binary data one or zero, enabling fast information writing and reading.

Vendors hope to have ReRAM ready for general adoption by 2017 or 2018.

Cloud storage

Many organizations are placing their data in the cloud—frequently in large datacenters—as a way to

trouble managing, indexing, accessing, and securing material stored in traditional, hierarchical file systems. The challenge is maintaining hierarchical organization and central data indices.

Companies are thus turning to object storage, which stores data as variable-size objects rather than fixed-size blocks.

Users don’t find information in object storage systems based on its physical location on a disk drive, as is the case with traditional storage.

Instead, object storage uses unique identifier addresses to locate and identify data objects. This provides nonhierarchical, near-infinite address spaces.

Thus, object-storage systems scale easily, without making it more complicated to find information, said Sean Derrington, storage vendor Exablox’s senior director of products.

Data buses

To take advantage of the improved performance that today’s

Optical storage

Optical storage, the least expensive removable storage media, will not significantly help cope with big data in the near future, said Yoichiro Tanaka, senior manager for Toshiba’s Storage Products Application Engineering Department.

“To achieve higher densities enabling more storage, you’d have to utilize a holographic or multilayer solution, and these have difficult technical challenges that won’t be resolved anytime soon,” explained Wolfgang Schlichting, CEO of the Wolf Research consultancy.

Both magnetic and optical storage work by storing bits as distinct magnetic or optical changes on the recording medium’s surface.

Holographic approaches use light traveling at different angles to record data throughout the storage medium. It can capture multiple images in a single area, thereby boosting capacity.

Also, both magnetic and optical storage record data a bit at a time, while holographic storage records and reads in parallel.

However, holographic storage is complex, and creating economically

viable approaches suitable for widespread use has been challenging.

In addition, said Schlichting, stacking multiple optical disks in a single unit to increase capacity has not yet proven to be either affordable or productive.

STORAGE BARRIERS

Building more storage centers or better storage technology is expensive, noted IDC's Reinsel.

"People need cheap storage," he explained, "but the cost of storage isn't declining fast enough. This is a dynamic the industry must deal with. Providing storage at very low prices would likely mean compromising on performance, and no one [wants that]."

In addition, scaling storage quickly with no effect on performance has been challenging.

And for many organizations, storing information in the cloud raises significant security and privacy concerns, said Reinsel.

Moreover, moving to new, better storage platforms can be costly and painful, particularly for companies working with large datasets, noted the Taneja Group's Matchett.

To cope with growing information volumes, organizations will increasingly use techniques such as data compression, deduplication, object storage, and storage virtualization.

Nonetheless, this process will continue to be a challenge.

Data managers will thus have to become more discriminating about what to save and for how long, said the Taneja Group's Matchett.

It's also clear that current technologies won't be able to provide the necessary capacity or performance to handle the growing amount of data, said Micron's Kilbuck.

This will require new storage approaches, and economics will affect the development of these technologies. Given the cur-

rent volume of data, the cost per byte of various types of storage will be a primary driver in companies' decision making, noted Don Brown, senior architect and lead field engineer for big data applications vendor WibiData.

"Until we see things like holographic storage get far more price competitive and commoditized in the market, they're unlikely to penetrate the big data technology stacks other than as auxiliary systems," said Brown. "The value derived from the data has to be greater than the cost of storing and managing it."

Nonetheless, said the Aberdeen Group's Csaplar, "We're just beginning a revolution in storage, as all the drivers are there: rapidly rising demand, customers unhappy with their current solutions, and new technologies just arriving or on the horizon. The mix of storage options

for most companies will look very different five years from now." ■

Neal Leavitt is president of Leavitt Communications (www.leavcom.com), a Fallbrook, California-based international marketing communications company with affiliate offices in Brazil, France, Germany, Hong Kong, India, and the UK. He writes frequently on technology topics and can be reached at neal@leavcom.com.

Editor: Lee Garber, Computer; l.garber@computer.org

 Selected CS articles and columns are available for free at <http://ComputingNow.computer.org>.

SMART

SCIENCE, MATHEMATICS & RESEARCH FOR TRANSFORMATION

PART OF THE NATIONAL DEFENSE EDUCATION PROGRAM

SCIENCE, MATHEMATICS & RESEARCH FOR TRANSFORMATION

OPEN TO: Undergraduate, graduate, and doctoral students pursuing degrees in Science, Technology, Engineering, & Mathematics (STEM) fields

SMART Scholars receive:

- + Full tuition and educational fees
- + Generous cash stipend
- + Employment with Department of Defense facilities after graduation
- + Summer internships, health insurance, & book allowance

For more information and to apply, visit

[HTTP://SMART.ASEE.ORG](http://SMART.ASEE.ORG)

In accordance with Federal statutes and regulations, no person on the grounds of race, color, age, sex, national origin or disability shall be excluded from participating in, denied the benefits of, or be subject to discrimination under any program activity receiving financial assistance from the Department of Defense.

