



## How Big Data and the Cloud Will Affect You

**ANYONE WHO READS PUBLICATIONS THAT DEAL WITH DATA PROCESSING AT NEARLY ANY LEVEL HAS PROBABLY SEEN THE TERM “BIG DATA” A LOT IN THE PAST YEAR OR SO. WHAT IS IT AND HOW WILL IT AFFECT YOUR BUSINESS?**

“Big Data” (BD from now on) is a collection of data sets so large and complex that existing database management tools often cannot be used to process it. Data processing applications are said to be “scalable” if a process that was developed for a small data set can be used for larger data sets. But there are limits to scalability.

The BD issue is a larger and more severe problem for large corporations and government agencies but it also affects small and medium-size businesses that increasingly find themselves awash in data that they hope to use to identify business trends, determine quality of research, and link related documents.

The increasing use of mobile devices, apps, and cloud-based computing exacerbate the problem. Microsoft recently surveyed more than 280 decision makers in midsize and large organizations on the topic of big data. Among their findings: Transforming BD into insights has become a top priority. This isn't exactly a new concept. The understanding that piles of data do not equate to knowledge is antique. In other words, the immense number of trees can obscure the view of the forest.

Current limits on the size of data sets that can be processed in a reasonable amount of time are measured in exabytes of data. Yes, that's a lot of data. Far more than any medium or small business will have. But even smaller organizations are seeing increases in data that, while they could be handled easily by a mainframe computer, are taxing the limits of desktop systems commonly used in these settings.

The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s and we are currently creating 2.5 quintillion bytes of data every day.

### Taming the Flow

TO SOME EXTENT, HARDWARE IS PROVIDING, IF NOT A SOLUTION TO THE PROBLEM, AT LEAST A WAY TO MITIGATE ITS EFFECTS.

Sixty-four-bit computers have been available since the 1960s (although in those days the computers were mammoths, both in terms of size and cost). Intel introduced a 64-bit chip nearly a decade ago and even Windows XP had a 64-bit version, but few people used it.

Today, all Macs are 64-bit systems. Many notebook computers have 64-bit processors. Even some tablets come with 64-bit CPUs. Even so, some businesses try to save a few dollars by continuing to use 32-bit hardware. If speed is important to you, selecting a computer with a 64-bit processor is the first step you should take.

The 64-bit hardware opens the door to the next consideration that can improve performance: Memory. The 32-bit systems are limited to 4GB of RAM, of which only about 3GB will be available to applications. By contrast, 64-bit systems can address a virtually unlimited amount of memory. If you're a stickler for facts, there is a limit. It's about

4 petabytes, but good luck finding any mainboard with enough memory slots!

The reason that more memory makes a system faster is easy to understand: When a computer's memory is full, any new operating instructions



cannot be placed in memory until what's currently there is written out to the disk. This is called “swapping” and the more times instructions have to be swapped out to disk and then swapped back in the next time they're needed, the slower the computer will run.

Some software manufacturers (Adobe, for example) have already transitioned some of their high-end applications to run only on 64-bit architecture. These applications are so computationally intensive that on a 32-bit system they would be unacceptably slow. Rather than try to maintain 2 versions, Adobe simply decided that the users of these applications will understand the need

for faster hardware. And, for the most part, this seems to be true.

## A Third Speed Trick

HARD DISK DRIVES HAVE BECOME FASTER BUT THEY ARE STILL THE SLOWEST COMPONENTS IN THE COMPUTER IN TERMS OF DATA TRANSFER. THE BOOT PROCESS AND LOADING PROGRAMS ARE DISK-INTENSIVE OPERATIONS.

That's why the fastest and most expensive disk drives come with large amounts of cache memory. Increasingly, though, savvy computer buyers are specifying solid-state disk drives that will hold the computer's operating system and applications.

Unlike standard drives, solid-state drives (SSD) have a limited number of write cycles and eventually the solid-state memory becomes exhausted. SSDs typically arrive with more than the stated amount of memory. This allows for a certain percentage of the drive to be taken out of service over time and still allow the SSD to maintain its rated size. So although SSDs are a smart way to speed a computer's startup operation and decrease the amount of time required to load programs, it's not a particularly good choice for storing data that changes a lot.

## The Future of BD

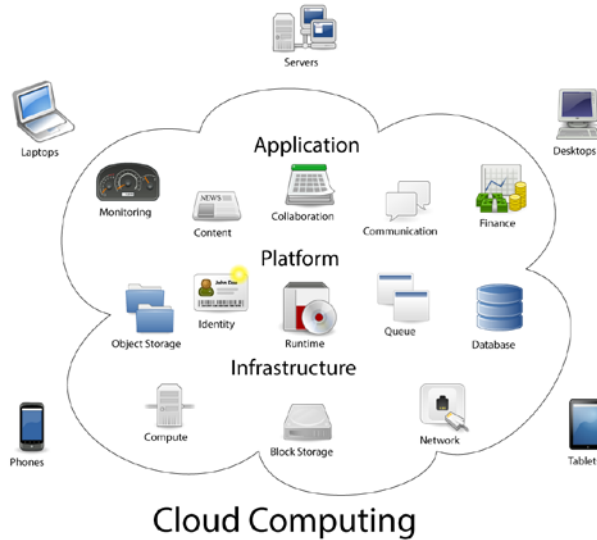
SCIENTIFIC USES OF DATA PROVIDE INSIGHTS INTO WHERE WE'RE GOING BECAUSE SCIENCE RUNS ON DATA BUT THE REST OF THE WORLD FOLLOWS SCIENCE'S LEAD. SO CONSIDER...

The Large Hadron Collider experiments used 150 million sensors that each delivered data 40 million times per second. Scientists were observing nearly 600 million collisions per second and a filtering mechanism eliminated all but about 0.001% of the data streams. Even then, the scientists found about 100 collisions per second that looked interesting. The data flow from the experiments represented 25 petabytes annually but then the data streams are replicated and expand to 200 petabytes.

If the process didn't discard all but 0.001% of the data, scientists would have to deal with about 500 exabytes per day, and that's before replication. Your business needs won't approach data flows of this magnitude for a long, long time but still you've undoubtedly seen increases in the amount of "stuff" you have on hand.

Even for home users, it's not uncommon to have several terabytes of data. The future is arriving quickly and bringing lots of data with it. Yet another driver for big data comes from the (rhetorical) sky.

## Cloud-Based Computing



IT'S BEEN A COUPLE OF YEARS SINCE IBM ANNOUNCED (1 MAR 2011) THE SMARTER COMPUTING FRAMEWORK TO SUPPORT A SMARTER PLANET. ALTHOUGH THE FRAMEWORK HAS SEVERAL COMPONENTS, ONE OF THE MOST CRITICAL IS CLOUD COMPUTING.

Most of us have never seen a cloud that computes. Nor have any of us ever published a desktop. Even so, the term "cloud computing" seems to have caught on just as "desktop publishing" did in the 1980s and apparently we will have to live with these terms.

At its most basic, cloud computing entails the use of software that is delivered "as a service" over a network (typically the Internet). Software as a service has earned the entertaining acronym "SaaS" and "cloud computing" is actually based on the use of a cloud symbol in network diagrams where the cloud represents the network.

Cloud computing also often involves storing a user's data remotely for access at any time from any location that has an Internet connection.

When dealing with cloud-based applications, network speed becomes an important consideration in addition to the speed of the remote computers where the applications are housed. Users often access applications via a Web browser or a light-weight desktop or mobile app. The promoters of cloud-based computing say that

this is a way to reduce maintenance, improve manageability, and move new applications into production faster.

Ironically, many of these same claims were made in the 1980s when applications were moved away from central mainframe computers to small computers on users' desks.

In fact, the concept of cloud computing dates back to the 1950s. As large mainframe computers became available, video terminals ("dumb terminals") were attached to them to allow individual users to run applications in a time-sharing system.

A 1960s-era computer scientist suggested that computation may someday be organized as a "public utility". That may bring to mind thoughts of the Internet. In 1966, Douglas

Parkhill's *The Challenge of the Computer Utility* examined the future of computing by discussing "elastic provisioning" of computing services that would be provided online essentially as a utility.

As far back as the 1950s, scientists were suggesting a world filled with computers (albeit dumb terminals). The future as seen from then included a dozen or so giant computing centers. In those days, centralization was the only option imaginable because of the expected cost.

More recently, Amazon has become a key player in cloud computing with its centers that many companies use. Even Microsoft is in the cloud with Windows 8, Office 2013, and SkyDrive.

Performance will be a key consideration if cloud computing is to catch on, so Amazon's centers routinely use only 10% of their capacity, allowing acceptable performance when activity spikes. In the 30 years that computers have been available on desktops, we have learned at least this fact: *Nobody likes waiting for a computer.* Ω

**n-Lighten.us**  
Division of William Blinn Communications  
179 Caren Ave., Worthington, Ohio 43085  
614/859.9359 • www.n-lighten.us