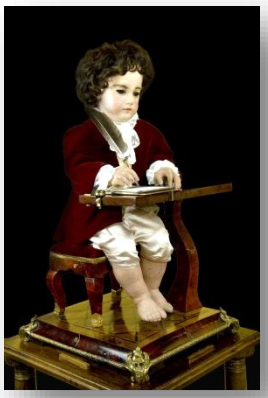
Introduction to Digital Computers, Servers and Storage

Craig Ball



In 1774, a Swiss watchmaker named Pierre Jaquet-Droz built an ingenious mechanical doll resembling a barefoot boy. Constructed of 6,000 handcrafted parts and dubbed "L'Ecrivain" ("The Writer"), Jaquet-Droz' automaton uses quill and ink to handwrite messages in cursive, up to 40 letters long, with the content controlled by interchangeable cams. The Writer is a charming example of an early programmable computer.

The monarchs that marveled at Jaquet-Droz' little penman didn't need to understand how it worked to enjoy it. Lawyers, too, once had little need to understand the operation of their clients' information systems in order to conduct discovery. But as the volume of electronically stored information (ESI) has exploded and the forms and sources of ESI continue to morph and multiply, lawyers conducting electronic discovery cannot ignore the watchwork. New standards of competence demand that lawyers and litigation support personnel master certain fundamentals of information technology and electronic evidence.

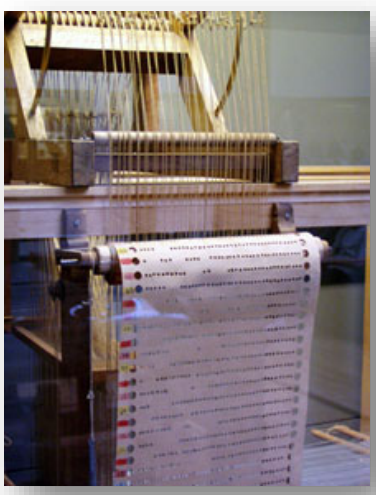
Data, Not Documents

Lawyers—particularly those who didn't grow up with computers—tend to equate data with documents when, in a digital world, documents are just one of the many forms in which electronic information exists. Documents akin to the letters, memos and reports of yore account for a dwindling share of electronically stored information relevant in discovery, and documents generated from electronic sources tend to convey just part of the information stored in the source. The decisive information in a case may exist as nothing more than a single bit of data that, in context, signals whether the fact you seek to establish is true or not. A Facebook page doesn't exist until a request sent to a database triggers the page's assembly and display. Word documents, PowerPoint presentations and Excel spreadsheets lose content and functionality when printed to screen images or paper.

With so much discoverable information bearing so little resemblance to documents, and with electronic documents carrying much more probative and useful information than a printout or screen image conveys, competence in electronic discovery demands an appreciation of data more than documents.

Introduction to Data Storage Media

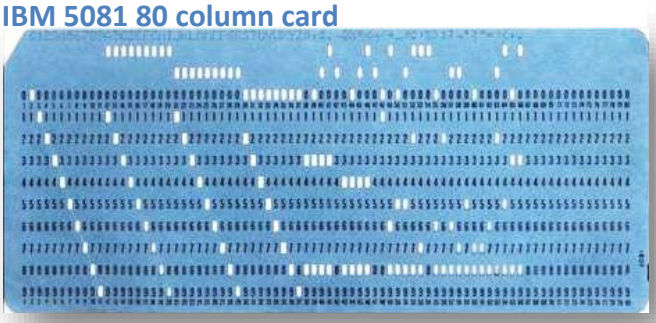
Mankind has been storing data for thousands of years, on stone, bone, clay, wood, metal, glass, skin, papyrus, paper, plastic and film. In fact, people were storing data in binary formats long before the emergence of modern digital computers. Records from 9th century Persia describe an organ playing interchangeable cylinders. Eighteenth century textile manufacturers employed perforated rolls of paper to control looms, and Swiss and German music box makers used metal drums or platters to store tunes. At the dawn of the Jazz Age, no self-respecting American family of means lacked a player piano capable (more-or-less) of reproducing the works of the world's greatest pianists.



Whether you store data as a perforation or a pin, you're storing binary data. That is, there are two data states: hole or no hole, pin or no pin. Zeroes or ones.

Punched Cards

IBM 5081 80 column card



In the 1930's, demand for *electronic* data storage led to the development of fast, practical and cost-effective binary storage media. The first of these were punched cards, initially made in a variety of sizes and formats, but ultimately standardized by IBM as the 80 column, 12 row (7.375" by 3.25") format (right) that dominated computing well into the 1970's. [From 1975-79, the author spent many a midnight in the basement of a computer center at Rice University typing program instructions on these unforgiving punch cards].

The 1950's saw the emergence of magnetic storage as the dominant medium for electronic data storage, and it remains so today. Although optical and solid state storage are expected to ultimately eclipse magnetic media for local storage, magnetic storage will continue to dominate network and cloud storage well into the 2020s, if not beyond.

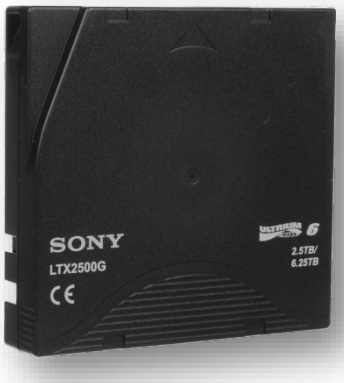
Tape

The earliest popular form of magnetic data storage was magnetic tape. Spinning reels of tape were a clichéd visual metaphor for computing in films and television shows from the 1950s through 1970's. Though the miles of tape on those reels now resides in cartridges and cassettes, tapes remain an enduring medium for backup and archival of electronically stored information. The LTO-6 format introduced in 2012 natively holds 2.5 terabytes of uncompressed data (6.2 TB compressed) and delivers a transfer rate of 160 megabytes per second. Since most data stored on backup tape is compressed, the actual volume of ESI on tape may be 2-3 times greater than the native capacity of the tape.

Magnetic tape was the earliest data storage medium for personal computers including the pioneering Radio Shack TRS-80 and the very first IBM personal computer, the model XT.

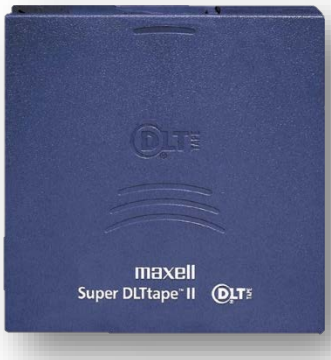
While tape isn't as fast or capacious as hard drives, it's proven to be more durable and less costly for long term storage; that is, so long as the data is being *stored,* not *restored*.

LTO-6 Ultrium Tape



Sony AIT-3 Tape

SDLT-II Tape



Chronology of Magnetic Tape Formats for Data Storage (Wikipedia)

1951 – UNISERVO
1952 - IBM 7 track
1958 - TX-2 Tape System
1962 – LINCtape
1963 – DECtape
1964 - 9 Track
1964 – MagCard Selectric typewriter
1966 - 8-Track Tape
1972 - QIC
1975 - KC Standard, Compact Cassette
1976 - DC100
1977 - Commodore Datasette
1979 – DECtapeII
1979 - Exatron Stringy Floppy
1983 - ZX Microdrive
1984 - Rotronics Wafadrive
1984 - IBM 3480
1984 - DLT
1986 - SLR
1987 - Data8
1989 - DDS/DAT
1992 - Ampex DST
1994 - Mammoth
1995 - IBM 3590
1995 - Redwood SD-3
1995 - Travan
1996 - AIT
1997 - IBM 3570 MP
1998 - T9840
1999 – VXA
2000 - T9940
2000 - LTO Ultrium
2003 - SAIT
2006 - T10000
2007 - IBM 3592
2008 - IBM TS1130
2011 - IBM TS1140

For further information, *see* Ball, Technology Primer: Backups in Civil Discovery at http://www.craigball.com/Ball_Technology%20Primer-Backups%20in%20E-Discovery.pdf

Floppy Disks

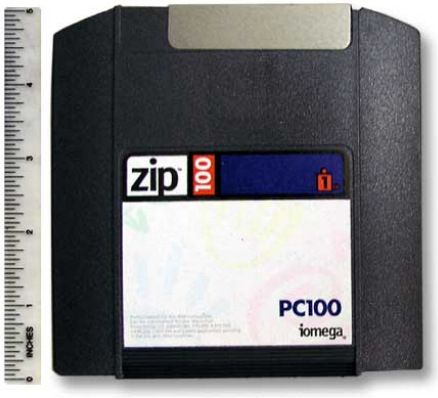
It's rare to encounter a floppy disk today, but floppy disks played a central role in software distribution and data storage for personal computing for almost thirty years. Today, the only place a computer user is likely to see a floppy disk is as the menu icon for storage on the menu bar of Microsft Office applications. All floppy disks have a spinning, flexible plastic disk coated with a magnetic oxide (e.g., rust). The disk is essentially the same composition as magnetic tape in disk form. Disks are **formatted** (either by the user or pre-formatted by the manufacturer) so as to divide the disk into various concentric rings of data called **tracks**, with tracks further subdivided into tiny arcs called **sectors**. Formatting enables systems to locate data on physical storage media much as roads and lots enable us to locate homes in a neighborhood.

Though many competing floppy disk sizes and formats have been introduced since 1971, only five formats are likely to be encountered in e-discovery. These are the 8", 5.25", 3.5 standard, 3.5 high density and Zip formats and, of these, the 3.5HD format 1.44 megabyte capacity floppy is by far the most prevalent legacy floppy disk format.

The Zip Disk was one of several proprietary “super floppy” products that enjoyed brief success before the high capacity and low cost of recordable optical media (CD-R and DVD-R) and flash drives rendered them obsolete.

8", 5.25" and 3.5" Floppy Disks

Zip Disk



8" Floppy Disk in Use

Optical Media

The most common forms of optical media for data storage are the CD, DVD and Blu-ray disks in read only, recordable or rewritable formats. Each typically exists as a 4.75” plastic disk with a metalized reflective coating and/or dye layer that can be distorted by a focused laser beam to induce pits and lands in the media. These pits and lands, in turn, interrupt a laser reflected off the surface of the disk to generate the ones and zeroes of digital data storage. The practical difference between the three prevailing forms of optical media are their native data storage capacities and the availability of drives to read them.

A **CD** (for **C**ompact **D**isk) or **CD-ROM** (for **CD R**ead **O**nly **M**edia) is read only and not recordable by the end user. It's typically fabricated in factory to carry music or software. A **CD-R** is recordable by the end user, but once a recording session is closed, it cannot be altered in normal use. A **CD-RW** is a re-recordable format that can be erased and written to multple times. The native data storage capacity of a standard-size CD is about 700 megabytes.

A **DVD** (for **D**igital **V**ersitile **D**isk) also comes in read only, recordable (**DVD±R**) and rewritable (**DVD±RW**) iterations and the most common form of the disk has a native data storage capacity of approximately 4.7 gigabytes. So, one DVD holds the same amount of data as six and one-half CDs.

By employing the narrower wavelength of a blue laser to read and write disks, a dual layer **Blu-ray** disk can hold up to about 50 gigabytes of data, equalling the capacity of about ten and one-half DVDs. Like their predecessors, Blu-ray disks are available in recordable (BD-R) and rewritable (CD-RE) formats

Though ESI resides on a dizzying array of media and devices, by far the largest complement of same occurs within three closely-related species of computing hardware: *computers, hard drives* and *servers.* A server is essentially a computer dedicated to a specialized task or tasks, and both servers and computers routinely employ hard drives for program and data storage.

Conventional Electromagnetic Hard Drives

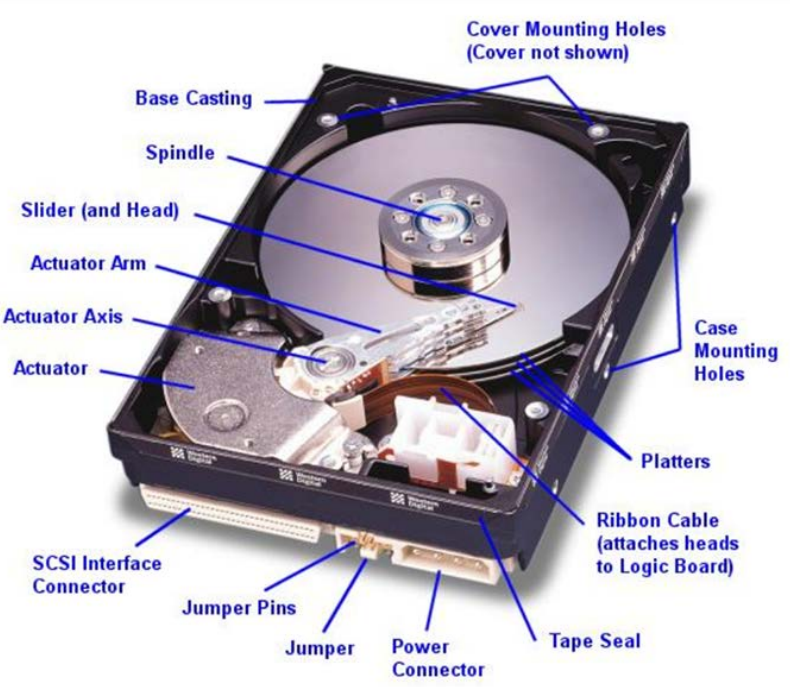
A hard drive is an immensely complex data storage device that's been engineered to appear deceptively simple. When you connect a hard drive to your machine, and the operating system detects the drive, assigns it a drive letter and—presto!—you've got trillions of bytes of new storage! Microprocessor chips garner the glory, but the humdrum hard drive is every bit a paragon of ingenuity and technical prowess.

A conventional personal computer hard drive is a sealed aluminum box measuring (for a desktop system) roughly 4" x 6" x 1" in height. A hard drive can be located almost anywhere within the case and is customarily secured by several screws attached to any of ten pre-threaded mounting holes along the edges and base of the case. One face of the case will be labeled to reflect the drive specifications, while a printed circuit board containing logic and controller circuits will cover the opposite face.

A conventional hard disk contains round, flat discs called **platters**, coated on both sides with a special material able to store data as magnetic patterns. Much like a record player, the platters

have a hole in the center allowing multiple platters to be stacked on a spindle for greater storage capacity.

The platters rotate at high speed—typically 5,400, 7,200 or 10,000 rotations per minute—driven by an electric motor. Data is written to and read from the platters by tiny devices called **read/write heads** mounted on the end of a pivoting extension called an **actuator arm** that functions similarly to the tone arm that carried the phonograph cartridge and needle across the face of a record. Each platter has two read/write heads, one on the top of the platter and one on the bottom. So, a conventional hard disk with three platters typically sports six surfaces and six read/write heads.



Unlike a record player, the read/write head never touches the spinning platter. Instead, when the platters spin up to operating speed, their rapid rotation causes air to flow under the read/write heads and lift them off the surface of the disk—the same principle of lift that operates on aircraft wings and enables them to fly. The head then reads the magnetic patterns on the disc while flying just .5 millionths of an inch above the surface. At this speed, if the head bounces against the surface, there is a good chance that the head will burrow into the surface of the platter, obliterating data, destroying both read/write heads and rendering the hard drive inoperable—a so-called "head crash."

The hard disk drive has been around for more than 50 years, but it was not until the 1980's that the physical size and cost of hard drives fell sufficiently for their use to be commonplace.

Introduced in 1956, the IBM 350 Disk Storage Unit pictured was the first commercial hard drive. It was 60 inches long, 68 inches high and 29 inches deep (so it could fit through a door). It held 50 magnetic disks of 50,000 sectors, each storing 100 alphanumeric characters. Thus, it held 4.4 megabytes, or enough for about two cellphone snapshots today. It weighed a ton (literally), and users paid $130.00 per month to *rent* each megabyte of storage.

Today, that same $130.00 *buys* a 4 terabyte hard drive that stores *3 million times* more information, weighs less than three pounds and hides behind a paperback book.

Over time, hard drives took various shapes and sizes (or "form factors" as the standard dimensions of key system components are called in geek speak). Three form factors are still in use: 3.5" (desktop drive), 2.5" (laptop drive) and 1.8" (iPod and microsystem drive, now supplanted by solid state storage).

Hard drives connect to computers by various mechanisms called "interfaces" that describe both how devices "talk" to one-another as well as the physical plugs and cabling required. The five most common hard drive interfaces in use today are:

PATA for **P**arallel **A**dvanced **T**echnology **A**ttachment (sometimes called EIDE for **E**xtended **I**ntegrated **D**rive **E**lectronics):

SATA for **S**erial **A**dvanced **T**echnology **A**ttachment

SCSI for **S**mall **C**omputer **S**ystem Interface

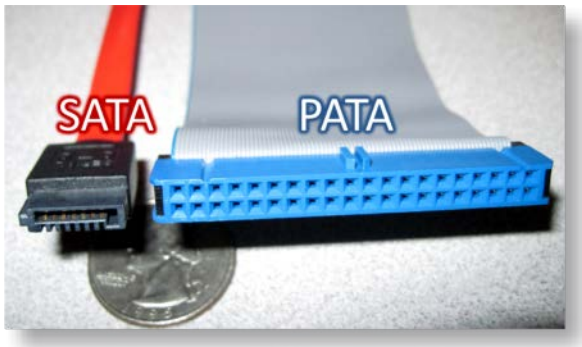
SAS for **S**erial **A**ttached **S**CSI

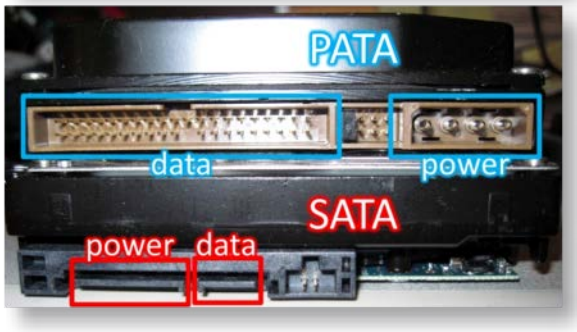
FC for **F**ibre **C**hannel

Though once dominant in personal computers, PATA drives are rarely found in machines manufactured after 2006. Today, virtually all laptop and desktop computers employ SATA drives for local storage. SCSI, SAS and FC drives tend to be seen exclusively in servers and other applications demanding high performance and reliability.

From the user's perspective, PATA, SATA, SCSI, SAS and FC drives are indistinguishable; however, from the point of view of the technician tasked to connect to and image the contents of the drive, the difference implicates different tools and connectors.

The five drive interfaces divide into two employing parallel data paths (PATA and SCSI) and three employing serial data paths (SATA, SAS and FC). Parallel ATA interfaces route data over multiple simultaneous channels necessitating 40 wires where serial ATA interfaces route data through a single, high- speed data channel requiring only 7 wires. Accordingly, SATA cabling and connectors are smaller than their PATA counterparts (see photos, right).





Fibre Channel employs optical fiber (the spelling difference is intentional) and light waves to carry data at impressive speeds. The premium hardware required by FC dictates that it will be found in enterprise computing environments, typically in conjunction with a high capacity/high demand storage device called a **SAN** (for Storage Attached Network) or a **NAS** (for Network Attached Storage).

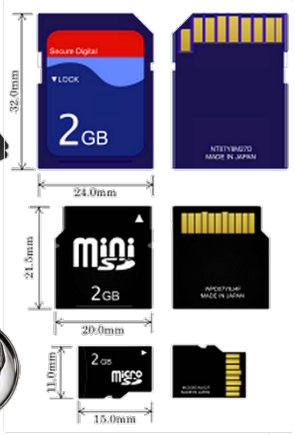
It's easy to become confused between hard drive interfaces and external data transfer interfaces like USB or FireWire seen on external hard drives. The drive within the external hard drive housing will employ one of the interfaces described above (except FC); however, to facilitate external connection to a computer, a device called a **bridge** will convert data written to and from the hard drive to a form that can traverse a USB or FireWire connection. In some compact, low-cost external drives, manufacturers dispense with the external bridge board altogether and build the USB interface right on the hard drive's circuit board.

Flash Drives, Memory Cards, SIMs and Solid State Drives

Computer memory storage devices have no moving parts and the data resides entirely within the solid materials which compose the memory chips, hence the term, "solid state." Historically, rewritable memory was volatile (in the sense that contents disappeared when power was withdrawn) and expensive. But, beginning around 1995, a type of non-volatile memory called NAND flash became sufficiently affordable to be used for removable storage in emerging

SmartMedia Card

Compact Flash Media

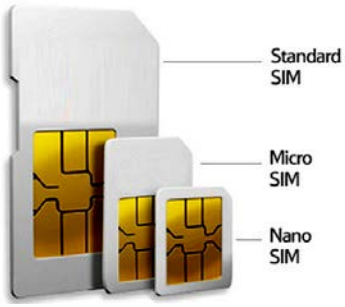


Secure Digital (SD) Media

applications like digital photography. Further leaps in the capacity and dips in the cost of NAND flash led to the near-eradication of film for photography and the extinction of the floppy disk, replaced by simple, inexpensive and reusable USB storage devices called, variously, SmartMedia, Compact Flash media, SD cards, flash drives, thumb drives, pen drives and memory sticks or keys.

USB Flash Drives

A specialized form of solid state memory seen in cell phones is the **Subscriber Identification Module** or **SIM card**. SIM cards serve both to authenticate and identify a communications device on a cellular network and to store SMS messages and phone book contacts.



SIM Cards

As the storage capacity of NAND flash has gone up and its cost has come down, the conventional electromagnetic hard drive is rapidly being replaced by **solid state drives** in standard hard drive form factors. Solid state drives are significantly faster, lighter and more energy efficient than conventional drives, but they currently cost anywhere from 10-20 times more per gigabyte than their mechanical counterparts. All signs point to the ultimate obsolescence of mechanical drives by solid state drives, and some products (notably tablets like the iPad and Microsoft Surface or ultra-lightweight laptops like the MacBook Air) have eliminated hard drives altogether in favor of solid state storage.

Currently, solid state drives assume the size and shape of mechanical drives to facilitate compatibility with existing devices. However, the size and shape of mechanical hard drives was driven by the size and operation of the platter they contain. Because solid state storage devices have no moving parts, they can assume virtually any shape. It's likely, then, that slavish adherence to 2.5" and 3.5" rectangular form factors will diminish in favor of shapes and sizes uniquely suited to the devices that employ them.

With respect to e-discovery, the shift from electromagnetic to solid state drives is

inconsequential. However, the move to solid state drives will significantly impact matters necessitating computer forensic analysis. Because the NAND memory cells that comprise solid state drives wear out rapidly with use, solid state drive controllers must constantly reposition data to insure usage is distributed across all cells. Such “wear leveling” hampers techniques that forensic examiners have long employed to recover deleted data from conventional hard drives.

RAID Arrays

Whether local to a user or in the Cloud, hard drives account for nearly all the electronically stored information attendant to e-discovery. In network server and Cloud applications, hard drives rarely work alone. That is, hard drives are ganged together to achieve greater capacity, speed and reliability in so-called **R**edundant **A**rrays of **I**ndependent **D**isks or **RAID**s. In the SAN pictured at left, the 16 hard drives housed in trays may be accessed as **J**ust a **B**unch **o**f **D**isks or **JBOD**, but it’s far more likely they are working together as a RAID

RAIDs serve two ends: redundancy and performance. The redundancy aspect is obvious—two drives holding identical data safeguard against data loss due to mechanical failure of either drive—but how do multiple drives improve **performance**? The answer lies in splitting the data across more than one drive using a technique called **striping**.

A RAID improves performance by dividing data across more than one physical drive. The swath of data deposited on one drive in an array before moving to the next drive is called the "stripe." If you imagine the drives lined up alongside one-another, you can see why moving back-and-forth the drives to store data might seem like painting a stripe across the drives. By striping data, each drive can deliver their share of the data simultaneously, increasing the amount of information handed off to the computer’s microprocessor.

But, when you stripe data across drives, Information is lost if any drive in the stripe fails. You gain performance, but surrender security.

This type of RAID configuration is called a **RAID 0**. It wrings maximum performance from a storage system; but it's risky.

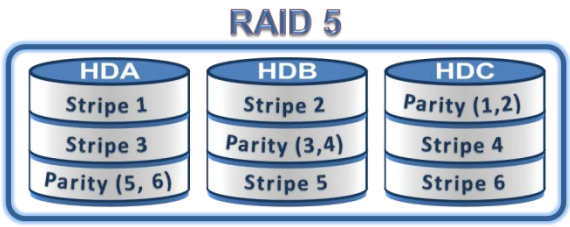
If RAID 0 is for gamblers, **RAID 1** is for the risk averse. A RAID 1 configuration duplicates everything from one drive to an identical twin, so that a failure of one drive won't lead to data loss. RAID 1 doesn't improve performance, and it requires twice the hardware to store the same information.

Other RAID configurations strive to integrate the *performance* of RAID 0 and the *protection* of RAID 1.

Thus, a "RAID 0+1" mirrors two striped drives, but demands four hard drives delivering only half their total storage capacity, Safe and fast, but not cost-efficient. The solution lies in a concept called ***parity,*** key to a range of other sequentially numbered RAID configurations. Of those other configurations, the ones most often seen are called **RAID 5 and RAID 7.**

To understand parity, consider the simple equation 5 + 2 = 7. If you didn't know one of the three values in this equation, you could easily solve for the missing value, *i.e.,* presented with "5 + __ = 7," you can reliably calculate the missing value is 2. In this example, "7" is the ***parity value*** or checksum for "5" and "2."

The same process is used in RAID configurations to gain increased performance by striping data across multiple drives while using parity values to permit the calculation of any missing values lost to drive failure. In a three drive array, any one of the drives can fail, and we can use the remaining two to recreate the third (just as we solved for 2 in the equation above).



In this illustration, data is striped across three hard drives, HDA, HDB and HDC. HDC holds the parity values for data stripe 1 on HDA and stripe 2 on HDB. It's shown as "Parity (1, 2)." The parity values for the other stripes are distributed on the other drives. Again, any one of the three drives can fail and all of the data is recoverable. This configuration is RAID 5 and, though it requires a minimum of three drives, it can be expanded to dozens or hundreds of disks.

Computers

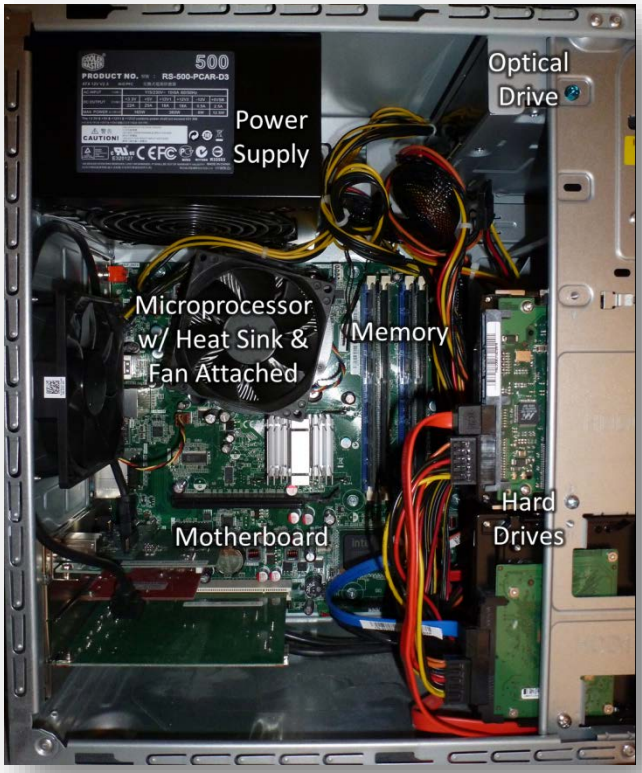
Historically, all sorts of devices—and even people—were “computers.” During World War II, human computers—women for the most part—were instrumental in calculating artillery trajectories and assisting with the challenging number-crunching needed by the Manhattan Project. Today, laptop and desktop personal computers spring to mind when we hear the term “computer;” yet smart phones, tablet devices, global positioning systems, video gaming platforms, televisions and a host of other intelligent tools and toys are also computers. More precisely, the **central processing unit** (CPU) or **microprocessor** of the system is the “computer,” and the various input and output devices that permit humans to interact with

the processor are termed **peripherals.** The key distinction between a mere calculator and a computer is the latter's ability to be programmed and its use of **memory** and **storage**. The physical electronic and mechanical components of a computer are its **hardware,** and the instruction sets used to program a computer are its **software.** Unlike the interchangeable cams of Pierre Jaquet-Droz' mechanical doll, modern electronic computers receive their instructions in the form of digital data typically retrieved from the same electronic storage medium as the digital information upon which the computer performs its computational wizardry.

When you push the power button on your computer, you trigger an extraordinary, expedited education that takes the machine from insensible illiterate to worldly savant in a matter of seconds. The process starts with a snippet of data on a chip called the ROM BIOS storing just enough information in its Read Only Memory to grope around for the Basic Input and Output System peripherals (like the keyboard, screen and, most importantly, the hard drive). The ROM BIOS also holds the instructions needed to permit the processor to access more and more data from the hard drive in a widening gyre, "teaching" itself to be a modern, capable computer.

This rapid, self-sustaining self-education is as magical as if you lifted yourself into the air by pulling on the straps of your boots, which is truly why it's called "bootstrapping" or just "booting" a computer.

Computer hardware circa 2014 shares certain common characteristics. Within the CPU, a microprocessor chip is the computational "brains" of system and resides in a socket on the **motherboard**, a rigid surface etched with metallic patterns serving as the wiring between the components on the board. The microprocessor generates considerable heat necessitating the attachment of a heat dissipation device called a **heat sink**, often abetted by a small fan. The motherboard also serves as the attachment point for memory boards (grouped as modules or "sticks") called **RAM** for Random Access Memory. RAM serves as the working memory of the processor while it performs calculations; accordingly, the more memory present, the more information can be processed at once, enhancing overall system performance.



Other chips comprise a Graphics Processor Unit (GPU) residing on the motherboard or on a separate expansion board called a **video card** or **graphics adapter**. The GPU supports the display of information from the processor onto a monitor or projector and has its own complement of memory dedicated to superior graphics performance. Likewise, specialized chips on the motherboard or an expansion board called a **sound card** support the reproduction of audio to speakers or a headphone. Video and sound processing capabilities may even be fully integrated into the microprocessor chip.

The processor communicates with networks through an interface device called a **network adapter** which connects to the network physically, through a **LAN Port,** or wirelessly using a Wi-Fi connection.

Users convey information and instructions to computers using tactile devices like a keyboard, mouse or track pad, but may also employ voice or gestural recognition mechanisms.

Persistent storage of data is a task delegated to other peripherals: **optical drives** (CD-ROM and DVD-ROM devices), **floppy disk drives**, **solid-state media** (*i.e.,* thumb drives) and, most commonly, **hard drives**. .

All of the components just described require electricity, supplied by batteries in portable devices or by a **power supply** converting AC current to the lower DC voltages required by electronics.

From the standpoint of electronic discovery, it's less important to define these devices than it is to fathom the information they hold, the places it resides and the forms it takes. Parties and lawyers have been sanctioned for what was essentially their failure to inquire into and understand the roles computers, hard drives and servers play as repositories of electronic evidence. Moreover, much money spent on electronic discovery today is wasted as a consequence of parties' efforts to convert ESI to paper-like forms instead of learning to work with ESI in the forms in which it customarily resides on computers, hard drives and servers.

Servers

Servers were earlier defined as computers dedicated to a specialized task or tasks. But that definition doesn't begin to encompass the profound impact upon society of the so-called **client-server** computing model. The ability to connect local "client" applications to servers via a network, particularly to **database servers**, is central to the operation of most businesses and to all telecommunications and social networking. Google and Facebook are just enormous groupings of servers, and the Internet merely a vast, global array of shared servers.

Local, Cloud and Peer-to-Peer Servers

For e-discovery, let's divide the world of servers into three realms: Local, Cloud and Peer-to-Peer server environments.

"Local" servers employ hardware that's physically available to the party that owns or leases the servers. Local servers reside in a computer room on a business' premises or in leased equipment "lockers" accessed at a co-located data center where a lessor furnishes, *e.g.,* premises security, power and cooling. Local servers are easiest to deal with in e-discovery because physical access to the hardware supports more and faster options when it comes to preservation and collection of potentially responsive ESI.

"Cloud" servers typically reside in facilities not physically accessible to persons using the servers, and discrete computing hardware is typically not dedicated to a particular user. Instead, the Cloud computing consumer is buying services via the Internet that emulate the operation of a single machine or a room full of machines, all according to the changing needs of the Cloud consumer. Web mail is the most familiar form of Cloud computing, in a variant called SaaS (for Software as a Service). Webmail providers like Google, Yahoo and Microsoft make e-mail accounts available on their servers in massive data centers, and the data on those servers is available solely via the Internet, no user having the right to gain physical access to the machines storing their messaging.

"Peer-to-Peer" (P2P) networks exploit the fact that any computer connected to a network has the potential to serve data across the network. Accordingly, P2P networks are decentralized; that is, each computer or "node" on a P2P network acts as client and server, sharing storage space, communication bandwidth and/or processor time with other nodes. P2P networking may be employed to share a printer in the home, where the computer physically connected to the printer acts as a print server for other machines on the network. On a global scale, P2P networking is the technology behind file sharing applications like BitTorrent and Gnutella that have garnered headlines for their facilitation of illegal sharing of copyrighted content. When users install P2P applications to gain access to shared files, they simultaneously (and often unwittingly) dedicate their machine to serving up such content to a multitude of other nodes.

Virtual Servers

Though we've so far spoken of server hardware, *i.e.,* physical devices, servers may also be implemented virtually, through software that emulates the functions of a physical device. Such "hardware virtualization" allows for more efficient deployment of computing resources by enabling a single physical server to host multiple virtual servers.

Virtualization is the key enabling technology behind many Cloud services. If a company needs powerful servers to launch a new social networking site, it can raise capital and invest in the hardware, software, physical plant and personnel needed to support a data center, with the

attendant risk that it will be over-provisioned or under-provisioned as demand fluctuates. Alternatively, the startup can secure the computing resources it needs by using virtual servers hosted by a Cloud service provider like Amazon, Microsoft or Rackspace. Virtualization permits computing resources to be added or retired commensurate with demand, and being pay-as-you-go, it requires little capital investment.

It's helpful for attorneys to understand the role of virtual machines (VMs) because the ease and speed with which VMs are deployed and retired, as well as their isolation within the operating system, can pose unique risks and challenges in e-discovery, especially with respect to implementing a proper legal hold and when identifying and collecting potentially responsive ESI.

Server Applications

Computers dedicated to server roles typically run operating systems optimized for server tasks and applications specially designed to run in a server environment. In turn, servers are often dedicated to supporting specific functions such as serving web pages (Web Server), retaining and delivering files from shared storage allocations (File Server), organizing voluminous data (Database Server), facilitating the use of shared printers (Print Server), running programs (Application Server) or handling messages (Mail Server). These various server applications may run physically, virtually or as a mix of the two.

Network Shares

Sooner or later, all electronic storage devices fail. Even the RAID storage arrays previously discussed do not forestall failure, but instead afford a measure of redundancy to allow for replacement of failed drives before data loss. Redundancy is the sole means by which data can be reliably protected against loss; consequently, companies routinely back up data stored on server NAS and SAN storage devices to backup media like magnetic tape or online (*i.e.,* Cloud) storage services. However, individual users often fail to back up data stored on local drives. Accordingly, enterprises allocate a "share" of network-accessible storage to individual users and "map" the allocation to the user's machine, allowing use of the share as if it were a local hard drive. When the user stores data to the mapped drive, that data is backed up along with the contents of the file server. Although **network shares** are not local to the user's computer, they are typically addressed using drive letters (*e.g.,* M: or T:) as if they were local hard drives.

Practice Tips for Computers, Hard Drives and Servers

Your first hurdle when dealing with computers, hard drives and servers in e-discovery is to identify potentially responsive sources of ESI and take appropriate steps to inventory their relevant contents and preserve them against spoliation. As the volume of ESI to be collected and

processed bears on the expense and time required, it's useful to get a handle on data volumes and distribution as early in the litigation process as possible.

Start your ESI inventory by taking stock of physical computing and storage devices. For each machine or device holding potentially responsive ESI, you may wish to collect some or all of the following information:

- Manufacturer and model
- Serial number and/or service or asset tag
- Operating system
- Custodian
- Location
- Type of storage (don't miss removable media, like SD cards)
- Aggregate storage capacity (in MB, GB or TB)
- Encryption status
- Credentials (user IDs and passwords), if encrypted
- Prospects for upgrade or disposal
- If you'll preserve ESI by drive imaging, it's helpful to identify device interfaces.

For servers, further information might include:

- Purpose(s) of the server (*e.g.,* web server, file server, print server, etc.)
- Names and contact information of server administrator(s)
- Time in service and data migration hostory
- Whether hardware virtualization is used
- RAID implementation(s)
- Users and privileges
- Logging and log retention practices
- Backup procedures and backup media rotation and retention
- Whether the server is "mission critical" and cannot be taken offline or can be downed.

When preserving the contents of a desktop or laptop computer, it's typically unnecessary to sequester any component of the machine other than its hard drive(s) since the ROM BIOS holds little information beyond the rare forensic artifact. Before returning a chassis to service with a new hard drive, be sure to document the custodian, manufacturer, model and serial number/service tag of the redeployed chassis, retaining this information with the sequestered hard drive.

The ability to fully explore the contents of servers for potentially responsive information hinges upon the privileges extended to the user. Be sure that the person tasked to identify data for preservation or collection holds administrator-level privileges.

Above all, remember that computers, hard drives and servers are constantly changing while in service. Simply rebooting a machine alters system metadata values for large numbers of files. Accordingly, you should consider the need for evidentiary integrity before exploring the contents of a device, at least until appropriate steps are taken to guard against unwitting alteration. Note also that connecting an evidence drive to a new machine effects changes to the evidence unless suitable write blocking tools or techniques are employed.