Storage Industry Terms:

ATA over Ethernet (AoE): is a network protocol developed by the Brantley Coile, the founder of Coraid, designed for simple, high-performance access of SATA storage devices over Ethernet networks. It is used to build storage area networks (SANs) with low-cost, standard technologies.

Cloud storage: A model of networked online storage where data is stored on virtualized pools of storage which are generally hosted by third parties. Hosting companies operate large data centers; and people who require their data to be hosted buy or lease storage capacity from them and use it for their storage needs. The data center operators, in the background, virtualize the resources according to the requirements of the customer and expose them as storage pools, which the customers can themselves use to store files or data objects. Physically, the resource may span across multiple servers.

Cloud storage services may be accessed through a web service application programming interface (API), or through a Web-based user interface.

Directly Attached Storage DAS: In the old days all mass storage devices such as disk drives etc were attached directly to the nearest computer (or were located inside the same box) so there was no need for a term describing this. But since the technologies which enable storage networking have become more prevalent, the term DAS has been used to describe those parts of a wider storage network in which this local connection is still used.

Fixed storage: also known as read-only memory, read-only storage, ROM - memory whose contents can be accessed and read but cannot be changed

Flash memory: is a non-volatile computer storage chip that can be electrically erased and reprogrammed. It was developed from EEPROM (electrically erasable programmable read-only memory) and must be erased in fairly large blocks before these can be rewritten with new data. The high density NAND type must also be programmed and read in (smaller) blocks, or pages, while the NOR type allows a single machine word (byte) to be written and/or read independently.

Enterprise multi-level cell: (eMLC) eMLC is a “souped-up” version of MLC flash with a controller and software that remedies some of the shortcomings of MLC; it’s becoming more popular in enterprise solid-state products.

Ethernet: A family of computer networking technologies for local area networks (LANs) commercially introduced in 1980. Systems communicating over Ethernet divide a stream of data into individual packets called frames. Each frame contains source and destination addresses and error-checking data so that damaged data can be detected and re-transmitted.

Exabyte: a measure of the size of storage. It's equal to 1,024 x 1,024 x 1,024 x Gigabytes = just over 1 billion Gigabytes. This is not yet a practical unit of storage for most real people, but if information keeps expanding at current rates, we'll start to get more familiar with it towards the end of the first decade of the 21st century.

Fibre Channel: Fibre-Channel is an interface standard for connecting computers to mass storage devices such as disk drives and tape libraries. Developed more than a decade after SCSI, which it was intended to replace for high performance applications, the Fibre-Channel standard was specified around faster data throughput speeds, and longer distances because of its use of fiber-optic cable. The differences in spelling were done deliberately by the standard creators. In the meantime, new versions of SCSI have caught up in speed terms, so the performance differences are now blurred.

Gigabyte: (G)or (GB): A measure of the size of storage. It's equal to 1,024 Megabytes. This is about one and a half times the capacity of a standard music CD-ROM, or about a million times the capacity of a standard 3.5 inch floppy disk.

Garbage Collection: Garbage Collection is an important background process in flash SSD controllers. Some editors and software vendors (who don't understand flash technology) mistakenly attribute a long term slow down in some SSDs to fragmentation - when really the issue is the ratio of resources allocated to Garbage Collection. In products which have reserved enough CPU power, internal R/W bandwidth and over-provisioning this "performance degradation" does not occur - or is minimal. For example systems from Violin Memory.

The "Garbage Collection Process" eliminates the need to perform erasure of the whole block prior to every write. The "Garbage Collection process" accumulates data marked for erase as "Garbage" and perform whole block erase as space reclamation in order to reuse the block.

Input/Output Operations Per Second: (IOPS) pronounced eye-ops) is a common performance measurement used to benchmark computer storage devices like hard disk drives (HDD), solid state drives (SSD), and storage area networks (SAN). As with any benchmark, IOPS numbers published by storage device manufacturers do not guarantee real-world application performance.

IOPS can be measured with applications, and is primarily used with servers to find the best storage configuration.

Hardware virtualization: Hardware virtualization or *platform virtualization* refers to the creation of a virtual machine that acts like a real computer with an operating system. Software executed on these virtual machines is separated from the underlying hardware resources.

In hardware virtualization, the *host machine* is the actual machine on which the virtualization takes place, and the *guest machine* is the virtual machine. The words *host* and *guest* are used to distinguish the software that runs on the actual machine from the software that runs on the virtual machine.

Hypervisor: Also called virtual machine manager (VMM), is one of many hardware virtualization techniques that allow multiple operating systems, termed *guests*, to run concurrently on a host computer. It is so named because it is conceptually one level higher than a supervisory program. The hypervisor presents to the guest operating systems a virtual operating platform and manages the execution of the guest operating systems. Multiple instances of a variety of operating systems may share the virtualized hardware resources. Hypervisors are installed on server hardware whose only task is to run guest operating systems.

iSCSI: iSCSI is a software package which emulates SCSI protocols, but the connection method is via an IP network instead of a direct SCSI compatible cable.

Megabyte: A multiple of the unit byte for digital information storage or transmission with two different values depending on context: 1048576 bytes (2^{20}) generally for computer memory;[1][2] and one million bytes (10^6, see prefix mega-) generally for computer storage.[1][3] The IEEE Standards Board has decided that "Mega will mean 1 000 000", with exceptions allowed for the base-two meaning.[3] In rare cases, it is used to mean 1000×1024 (1024000) bytes.[3] It is commonly abbreviated as Mbyte or MB (compare Mb, for the megabit).

Non-volatile random access memory: NVRAM This is high-speed memory that's extremely fast like DRAM but can retain data when the power is turned off; it's used as a cache in some flash solid-state storage systems.

Real storage: The main memory in a virtual memory system

Multi-level cell (MLC): This is a NAND flash chip that stores two bits per cell; it's slower and doesn't last as long as SLC, but it's much cheaper.

NOR Flash Memory: A type of non-volatile computer memory, This arrangement is called "NOR flash" because it acts like a NOR gate: when one of the word lines is brought high, the corresponding storage transistor acts to pull the output bit line low. NOR Flash continues to be the technology of choice for embedded applications requiring a discrete non-volatile memory device. The low read latencies characteristic of NOR devices allow for both direct code execution and data storage in a single memory product

Network Attached Storage (NAS): A term used for RAID, tape and other mass storage systems which have an integral network connection such as ethernet or fibre-channel, These devices can be connected as a network resource rather than just attached to a particular server.

Local area network: (LAN) is a computer network that interconnects computers in a limited area such as a home, school, computer laboratory, or office building. The defining characteristics of LANs, in contrast to wide area networks (WANs), include their usually higher data-transfer rates, smaller geographic area, and lack of a need for leased telecommunication lines.

PCI Express: (PCIe) is a high-speed server bus technology that's used by a number of server-based solid-state storage products. PCIe is a high-speed server bus technology that's used by a number of server-based solid-state storage products.

Petabyte: A unit of information equal to one quadrillion (short scale) bytes, or 1000 terabytes. The unit symbol for the petabyte is PB. The prefix *peta* (P) indicates the fifth power to 1000:

- 1 PB = 1000000000000000B = 1000^5 B = 10^{15} B = 1 million gigabytes = 1 thousand terabytes

In traditional binary usage, there are 1 125 899 906 842 624 bytes in 1 petabyte.

RAID: Redundant Array of Inexpensive Disks. In the 1980's when this term was originated, by IBM, you could buy 2 types of disk drives, either low cost drives such as used in the average PC, or high speed high performance drives as used in digital video effects systems or mainframes. Because of the larger market for PC disks, they eventually became the technology drivers and effectively killed off the higher cost alternatives. However the original concept remains valid. You can create a disk array which looks

electronically just like a bigger ordinary disk, by attaching a bunch of disks working in parallel and connected to a RAID controller interface

Representational state transfer: (**REST**) is a style of software architecture for distributed hypermedia systems such as the World Wide Web. Conforming to the REST constraints is referred to as being "RESTful."

Storage Area Network (SAN): This term was first coined to describe networks based around fibre-channel, but nowadays SAN's can also be include other connection technologies such as ethernet and even the internet. SAN is really a networking concept in which the software has a knowledge of the quantity and value of data stored in mass storage devices and the characteristics of those storage devices. In a way SAN is a superset of traditional networking thinking.

In SAN systems it's not just enough to know that data is moving from A to B. The software also has to know about the backup strategy, data recovery and application software specific attributes to preserve and reconstruct the environment in case of a failure or system reconfiguration. This aims to automatically replicate many functions which were previously managed by people called systems administrators.

SCSI: A intelligent protocol which enables data blocks to be read from or sent at high speed to a storage device such as a disk or tape drive. Early implementations of SCSI used ribbon cable and industry standard logic levels.

Single-level cell: (SLC) This is a type of flash that stores a single bit in each chip cell; it's the fastest, most reliable, longest lasting and most expensive type of NAND flash.

Solid-state drive: (or disk) (SSD) is typically used to refer to solid-state storage that's packaged in a hard disk form factor.

Storage: The process of storing information in a computer memory or on a magnetic tape or disk

Virtualization: Virtualization, in computing, is the creation of a virtual (rather than actual) version of something, such as a hardware platform, operating system, a storage device or network resources.