Designing Storage Architectures for Digital Collections 2019
Library of Congress
Common Terms for Attendees

In conjunction with the 2019 DSA meeting at the Library of Congress, here are several common terms provided by attendees. Please note, these items are not intended as exhaustive or prescriptive; they are merely intended to provide useful context related to the meeting topics.

BaFe or Barium Ferrite is a highly magnetic material, has a high packing density, and is a metal oxide. Studies of this material date at least as far back as 1931, and it has found applications in magnetic card strips, speakers, and magnetic tapes. One area in particular it has found success in is long-term data storage; the material is magnetic, resistant to temperature change, corrosion and oxidization.

(Wikipedia, accessed 9/05/2019)

Bit Error Rate: In digital transmission, the number of **bit errors** is the number of received bits of a data stream over a communication channel that have been altered due to noise, interference, distortion or bit synchronization errors.

The **bit error rate** (**BER**) is the number of bit errors per unit time. The **bit error ratio** (also **BER**) is the number of bit errors divided by the total number of transferred bits during a studied time interval. Bit error ratio is a unitless performance measure, often expressed as a percentage.

The **bit error probability** is the expectation value of the bit error ratio. The bit error ratio can be considered as an approximate estimate of the bit error probability. This estimate is accurate for a long time interval and a high number of bit errors.

(Wikipedia accessed 9/05/2019)

Block (data storage): In computing (specifically data transmission and data storage), a block, sometimes called a physical record, is a sequence of bytes or bits, usually containing some whole number of records, having a maximum length, a *block size*. Data thus structured are said to be *blocked*. The process of putting data into blocks is called *blocking*, while *deblocking* is the process of extracting data from blocks. Blocked data is normally stored in a data buffer and read or written a whole block at a time. Blocking reduces the overhead and speeds up the handling of the data-stream. For some devices, such as magnetic tape and CKD disk devices (a type of IBM mainframe storage device), blocking reduces the amount of external storage required for the data. Blocking is almost universally employed when storing data to 9-track magnetic tape, NAND flash memory, and rotating media such as floppy disks, hard disks, and optical discs.

Most file systems are based on a block device, which is a level of abstraction for the hardware responsible for storing and retrieving specified blocks of data, though the block size in file systems may be a multiple of the physical block size. This leads to space inefficiency due to internal fragmentation, since file lengths are often not integer multiples of block size, and thus the last block of a file may remain partially empty. This will create slack space. Some newer file systems, such as Btrfs and FreeBSD UFS2, attempt to solve this through techniques called block suballocation and tail merging. Other file systems such as ZFS support variable block sizes.

Block storage is normally abstracted by a file system or database management system (DBMS) for use by applications and end users. The physical or logical volumes accessed via *block I/O* may be devices internal to a server, directly attached via SCSI or Fibre Channel, or distant devices accessed via a storage area network (SAN) using a protocol such as iSCSI, or AoE. DBMSes often use their own block I/O for improved performance and recoverability as compared to layering the DBMS on top of a file system.

(Wikipedia, accessed 8/28/2018)

Block-level storage is a concept in cloud-hosted data persistence where cloud services emulate the behavior of a traditional block device, such as a physical hard drive. It is a form of Network Attached Storage (NAS).

Storage in such is organized as blocks. This emulates the type of behavior seen in traditional disk or tape storage. Blocks are identified by an arbitrary and assigned identifier by which they may be stored and retrieved, but this has no obvious meaning in terms of files or documents. A filesystem must be applied on top of the block-level storage to map 'files' onto a sequence of blocks.

Amazon EBS (Elastic Block Store) is an example of a cloud block store. Cloud block-level storage will usually offer facilities such as replication for reliability, or backup services.

Block-level storage is in contrast to an object store or 'bucket store', such as Amazon S3 (Simple Storage Service), or to a database. These operate at a higher level of abstraction and are able to work with entities such as files, documents, images, videos or database records.

Instance stores are another form of cloud-hosted block-level storage. These are provided as *part of* an 'instance', such as an Amazon EC2 (Elastic Compute Cloud) service. As EC2 instances are primarily provided as compute resources, rather than storage resources, their storage is less robust. Their contents will be lost if the cloud instance is stopped. As these stores are part of the instance's virtual server they offer higher performance and bandwidth to the instance. They are best used for temporary storage such as caching or temporary files, with persistent storage held on a different type of server.

At one time, block-level storage was provided by Storage Area Networks (SAN) and NAS provided file-level storage. With the shift from on-premises hosting to cloud services, this distinction has shifted. Even block-storage is now seen as distinct servers (thus NAS), rather than the previous array of bare discs.

Wikipedia, accessed 9/05/2019).

Blockchain is a growing list of records, called *blocks*, that are linked using cryptography. Each block contains a cryptographic hash of the previous block,[6] a timestamp, and transaction data (generally represented as a Merkle tree).

By design, a blockchain is resistant to modification of the data. It is "an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way".[7] For use as a distributed ledger, a blockchain is typically managed by a peer-to-peer network collectively adhering to a protocol for inter-node communication and validating new blocks. Once recorded, the data in any given block cannot be altered retroactively without alteration of all subsequent blocks, which requires consensus of the network majority. Although blockchain records are not unalterable, blockchains may be considered secure by design and exemplify a distributed computing system with high Byzantine fault tolerance. Decentralized consensus has therefore been claimed with a blockchain.

(Wikipedia, 9/05/2019)

Ceph (software) is a free-software storage platform, implements object storage on a single distributed computer cluster, and provides interfaces for object-, block- and file-level storage. Ceph aims primarily for completely distributed operation without a single point of failure, scalable to the exabyte level, and freely available.

Ceph replicates data and makes it fault-tolerant, using commodity hardware and requiring no specific hardware support. As a result of its design, the system is both self-healing and self-managing, aiming to minimize administration time and other costs.

(Wikipedia, accessed 9/05/2019)

Error Correction Code (ECC) is used for controlling errors in data over unreliable or noisy communication channels. The central idea is the sender encodes the message with a redundant in the form of an ECC.

(Wikipedia, accessed 9/05/2019)

Failure rate is the frequency with which an engineered system or component fails, expressed in failures per unit of time. It is often denoted by the Greek letter λ (lambda) and is highly used in reliability engineering.

The failure rate of a system usually depends on time, with the rate varying over the life cycle of the system. For example, an automobile's failure rate in its fifth year of service may be many times greater than its failure rate during its first year of service. One does not expect to replace an exhaust pipe, overhaul the brakes, or have major transmission problems in a new vehicle.

In practice, the mean time between failures (MTBF, 1/λ) is often reported instead of the failure rate. This is valid and useful if the failure rate may be assumed constant – often used for complex units / systems, electronics – and is a general agreement in some reliability standards (Military and Aerospace). It does in this case *only* relate to the flat region of the bathtub curve, which is also called the "useful life period". Because of this, it is incorrect to extrapolate MTBF to give an estimate of the service lifetime of a component, which will typically be much less than suggested by the MTBF due to the much higher failure rates in the "end-of-life wearout" part of the "bathtub curve".

The reason for the preferred use for MTBF numbers is that the use of large positive numbers (such as 2000 hours) is more intuitive and easier to remember than very small numbers (such as 0.0005 per hour).

The MTBF is an important system parameter in systems where failure rate needs to be managed, in particular for safety systems. The MTBF appears frequently in the engineering design requirements, and governs frequency of required system maintenance and inspections. In special processes called renewal processes, where the time to recover from failure can be neglected and the likelihood of failure remains constant with respect to time, the failure rate is simply the multiplicative inverse of the MTBF (1/λ).

(Wikipedia, accessed 8/27/2018).

FAIR: A set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable. See force11.org.

Fixity checking: The practice of algorithmically reviewing digital content to insure that it has not changed over time. See: Fixity Survey Report – An NDSA Report - https://ndsa.org/documents/Report_2017NDSAFixitySurvery.pdf

Flash memory is an electronic (solid-state) non-volatile computer storage medium that can be electrically erased and reprogrammed.

Toshiba developed flash memory from EEPROM (electrically erasable programmable read-only memory) in the early 1980s and introduced it to the market in 1984. The two main types of flash memory are named after the NAND and NOR logic gates. The individual flash memory cells exhibit internal characteristics similar to those of the corresponding gates.

While EPROMs had to be completely erased before being rewritten, NAND-type flash memory may be written and read in blocks (or pages) which are generally much smaller than the entire device. NOR-type flash allows a single machine word (byte) to be written – to an erased location – or read independently.

The NAND type operates primarily in memory cards, USB flash drives, solid-state drives (those produced in 2009 or later), and similar products, for general storage and transfer of data. NAND or NOR flash memory is also often used to store configuration data in numerous digital products, a task previously made possible by EEPROM or battery-powered static RAM. One key disadvantage of flash memory is that it can only endure a relatively small number of write cycles in a specific block.

Example applications of both types of flash memory include personal computers, PDAs, digital audio players, digital cameras, mobile phones, synthesizers, video games, scientific instrumentation, industrial robotics, and medical electronics. In addition to being non-volatile, flash memory offers fast read access times, although not as fast as static RAM or ROM. Its mechanical shock resistance helps explain its popularity over hard disks in portable devices, as does its high durability, ability to withstand high pressure, temperature and immersion in water, etc.[

(Wikipedia, accessed 9/5/2019)

HA: High Availability

HAMR: Heat-assisted magnetic recording is a magnetic storage technology for greatly increasing the amount of data that can be stored on a magnetic device such as a hard disk drive by temporarily heating the disk material during writing, which makes it much more receptive to magnetic effects and allows writing to much smaller regions (and much higher levels of data on a disk).

In February 2019, Seagate Technology announced that HAMR would be launched commercially in 2019, having been extensively tested at partners during 2017 and 2018. The first drives will be 16 TB, with 20 TB expected in 2020, 24 TB drives in advanced development, and 40 TB drives by around 2023. Its planned successor, known as heated-dot magnetic recording (HDMR), or bit-pattern recording, is also under development, although not expected to be available until at least 2025 or later. HAMR drives have the same form factor (size and layout) as existing traditional hard drives, and do not require any change to the computer or other device in which they are installed; they can be used identically to existing hard drives.

(Wikipedia, accessed 9/05/2019)

HDD: A **hard disk drive** (**HDD**), **hard disk**, **hard drive**, or **fixed disk**[b] is an electro-mechanical data storage device that uses magnetic storage to store and retrieve digital information using one or more rigid rapidly rotating disks (platters) coated with magnetic material. The platters are paired with magnetic heads, usually arranged on a moving actuator arm, which read and write data to the platter surfaces.[2] Data is accessed in a random-access manner, meaning that individual blocks of data can be stored or retrieved in any order and not only sequentially. HDDs are a type of non-volatile storage, retaining stored data even when powered off.

(Wikipedia, accessed 9/05/2019)

INSIC: Information Storage Industry Consortium

LTO: Linear Tape Open is a magnetic tape data storage technology originally developed in the late 1990s as an open standards alternative to the proprietary magnetic tape formats that were available at the time. Hewlett Packard Enterprise, IBM, and Quantum control the **LTO Consortium**, which directs development and manages licensing and certification of media and mechanism manufacturers.

(Wikipedia, accessed 9/05/2019)

LTFS Linear Tape File System is a file system that allows files stored on magnetic tape to be accessed in a similar fashion to those on disk or removable flash drives.

MAMR: Microwave Assisted Magnetic Recording is a type of hard disk drive technology.

Meantime between failures (MTBF): See Failure Rate

Merkle tree: In cryptography and computer science, a hash tree or Merkle tree is a tree in which every leaf node is labelled with the hash of a data block, and every non-leaf node is labelled with the cryptographic hash of the labels of its child nodes. Hash trees allow efficient and secure verification of the contents of large data structures.

(Wikipedia, accessed 10/28/2019)

NAND: See Flash Memory.

Network-attached Storage (NAS):

File-level computer data storage server connected to a computer network providing data access to a heterogeneous group of clients. NAS is specialized for serving files either by its hardware, software, or configuration. It is often manufactured as a computer appliance – a purpose-built specialized computer. NAS systems are networked appliances which contain one or more storage drives, often arranged into logical, redundant storage containers or RAID. Network-attached storage removes the responsibility of file serving from other servers on the network. They typically provide access to files using network file sharing protocols such as NFS, SMB/CIFS, or AFP. From the mid-1990s, NAS devices began gaining popularity as a convenient method of sharing files among multiple computers. Potential benefits of dedicated network-attached storage, compared to general-purpose servers also serving files, include faster data access, easier administration, and simple configuration.

NVMe (Non-Volatile Memory Express) is a logical device interface, which has been designed to capitalize on the low latency and internal parallelism of solid-state storage devices.

(Wikipedia, accessed 9/05/2019)

Object storage (also known as **object-based storage**) is a computer data storage architecture that manages data as objects, as opposed to other storage architectures like file systems which manage data as a file hierarchy, and block storage which manages data as blocks within sectors and tracks. Each object typically includes the data itself, a variable amount of metadata, and a globally unique identifier. Object storage can be implemented at multiple levels, including the device level (object-storage device), the system level, and the interface level. In each case, object storage seeks to enable capabilities not addressed by other storage architectures, like interfaces that can be directly programmable by the application, a namespace that can span multiple instances of physical hardware, and data-management functions like data replication and data distribution at object-level granularity.

Object-storage systems allow retention of massive amounts of unstructured data. Object storage is used for purposes such as storing photos on Facebook, songs on Spotify, or files in online collaboration services, such as Dropbox.[3]

(Wikipedia, accessed 8/27/2018)

POSIX (Portable Operating System Interface) is a family of standards specified by the IEEE Computer Society for maintaining compatibility between operating systems. POSIX defines the application programming interface (API), along with command line shells and utility interfaces, for software compatibility with variants of Unix and other operating systems.

(Wikipedia, accessed 9/05/2019)

Preservation Storage:

The use of storage technology for digital preservation has changed dramatically over the last twenty years. During this time, there has been a change in practice. Previously, the norm was for storing digital materials using discrete media items, e.g. individual CDs, tapes, etc., which are then migrated periodically to address degradation and obsolescence. Today, it has become more common practice to use resilient IT storage systems for the increasingly large volumes of digital material that needs to be preserved, and perhaps more importantly, that needs to be easily and quickly retrievable in a culture of online access. In this way, digital material has become decoupled from the underlying mechanism of its storage. With this come consequent benefits of allowing different preservation activities to be handled independently.

(Digital Preservation Coalition Handbook
https://www.dpconline.org/handbook/organisational-activities/storage
accesed 8/27/2018)

RAID (**Redundant Array of Independent Disks**, originally **Redundant Array of Inexpensive Disks**) is a data storage virtualization technology that combines multiple physical disk drive components into one or more logical units for the purposes of data redundancy, performance improvement, or both.[1]

Data is distributed across the drives in one of several ways, referred to as RAID levels, depending on the required level of redundancy and performance. The different schemes, or data distribution layouts, are named by the word "RAID" followed by a number, for example RAID 0 or RAID 1. Each scheme, or RAID level, provides a different balance among the key goals: reliability, availability, performance, and capacity. RAID levels greater than RAID 0 provide protection against unrecoverable sector read errors, as well as against failures of whole physical drives.

(Wikipedia, accessed 8/27/2018)

REST is a software architectural style that defines a set of constraints to be used for creating Web services. Web services that conform to the REST architectural style, called *RESTful* Web services (RWS), provide interoperability between computer systems on the Internet.

(Wikipedia, accessed 9/05/2019)

SMR (Shingled Magnetic Recording) – a hard disk drive technology.

Solid State Memory: See Flash Memory.

SSD: See Flash Memory.

Storage Area Network (SAN):

A **storage area network** (**SAN**) or **storage network** is a Computer network which provides access to consolidated, block level data storage. SANs are primarily used to enhance storage devices, such as disk arrays and tape libraries, accessible to servers so that the devices appear to the operating system as locally attached devices. A SAN typically is a separate network of storage devices not accessible through the local area network (LAN) by other devices.

The cost and complexity of SANs dropped in the early 2000s to levels allowing wider adoption across both enterprise and small to medium-sized business environments.

A SAN does not provide file abstraction, only block-level operations. However, file systems built on top of SANs do provide file-level access, and are known as shared-disk file systems.

T10000 – one of several tape formats created by StorageTek, These are commonly used with large computer systems, typically in conjunction with a robotic tape library. The most recent format is the T10000. StorageTek primarily competed with IBM in this market, and continued to do so after its acquisition by Sun Microsystems in 2005 and as part of the Sun Microsystems acquisition by Oracle in 2009.

(Wikipedia, accessed 9/05/2019)

Zero Trust: Zero Trust, Zero Trust Network, or Zero Trust Architecture refer to security concepts and threat model that no longer assumes that actors, systems or services operating from within the security perimeter should be automatically trusted, and instead must verify anything and everything trying to connect to its systems before granting access. The term was coined by a security analyst at Forrester Research.

(The Secret Security Wiki https://doubleoctopus.com/security-wiki/network-architecture/zero-trust/, accessed 9/16/2019)