Glossary of Terms - LC Storage Meeting 2015
(Not intended as exhaustive or prescriptive, but intended to provide context for discussions in the meeting)
(Definitions are excerpted from Wikipedia, unless otherwise indicated)

ARM: A family of reduced instruction set computing (RISC) instruction set architectures for computer processors. This design approach means ARM processors require significantly fewer transistors than typical complex instruction set computing (CISC) x86 processors in most personal computers. This approach reduces costs, heat and power use.

ASIC: An integrated circuit developed for a particular use, as opposed to a general-purpose device.

BER (Bit Error Rate): Number of bit errors per unit time. The BER is the number of bit errors divided by the total number of transferred bits during a studied time interval. A unitless performance measure, often expressed as a percentage.

CRC: Cyclic redundancy check is an error-detecting code commonly used in digital networks and storage devices to detect accidental changes to raw data. Blocks of data entering these systems get a short *check value* attached, based on the remainder of a polynomial division of their contents. On retrieval the calculation is repeated, and corrective action can be taken against presumed data corruption if the check values do not match. CRCs are so called because the *check* (data verification) value is a *redundancy* (it expands the message without adding information) and the algorithm is based on *cyclic* codes. CRCs are popular because they are simple to implement in binary hardware, easy to analyze mathematically, and particularly good at detecting common errors caused by noise in transmission channels. Because the check value has a fixed length, the function that generates it is occasionally used as a hash function.

Cryptographic hash is designed to take a string of any length as input and produce a fixed-length hash value, and must be designed to withstand all known types of cryptanalytic attack.

Erasure coding (EC) is a method of data protection in which data is broken into fragments, expanded and encoded with redundant data pieces and stored across a set of different locations or storage media. http://whatis.com.

Ethernet: A family of computer networking technologies. Systems communicating over Ethernet divide a stream of data into shorter pieces called frames. Each frame contains source and destination addresses and error-checking data so that damaged data can be detected and re-transmitted.

Fibre Channel (FC): A high-speed network technology primarily used to connect computer data storage. Fibre Channel is standardized in the T11 Technical Committee of the International Committee for Information Technology , a ANSI-accredited standards committee. Fibre Channel was primarily used in supercomputers, but has become a common connection type for storage area networks (SAN) in enterprise storage.

Field-programmable gate array (**FPGA**) : An integrated circuit designed to be configured by a customer or a designer after manufacturing – hence "field-programmable."

Flash memory: Electronic non-volatile computer storage medium that can be electrically erased and reprogrammed.

Gate: In electronics, a **logic gate** is an idealized or physical device implementing a Boolean function; that is, it performs a logical operation on one or more logical inputs, and produces a single logical output.

Hadoop: **Apache Hadoop** is an open-source software framework written in Java for distributed storage and distributed processing of very large data sets on computer clusters built from commodity hardware. All the modules in Hadoop are designed with a fundamental assumption that hardware failures (of individual machines, or racks of machines) are commonplace and thus should be automatically handled in software by the framework. The core of Apache Hadoop consists of a storage part (Hadoop Distributed File System (HDFS)) and a processing part (MapReduce). Hadoop splits files into large blocks and distributes them amongst the nodes in the cluster. To process the data, Hadoop MapReduce transfers packaged code for nodes to process in parallel, based on the data each node needs to process. This approach takes advantage of data locality—nodes manipulating the data that they have on hand—to allow the data to be processed faster and more efficiently than it would be in a more conventional supercomputer architecture that relies on a parallel file system where computation and data are connected via high-speed networking.

Hard Disk Drive (HDD): A data storage device used for storing and retrieving digital information using one or more rigid ("hard") rapidly rotating disks (platters) coated with magnetic material.

Hard Error: (1) A permanent, unrecoverable error such as a disk read error. Contrast with soft error. **(2)** A group of errors that requires user intervention and includes disk read errors, disk not ready (no disk in drive) and printer not ready (out of paper). http://www.yourdictionary.com/computer/hard-error

Infiniband: A computer-networking communications standard used in high-performance computing, which features very high throughput and very low latency. It is used for data interconnect both among and within computers.

Low Cost Optical Networking (LCON): Optical networking is a means of communication that uses signals encoded onto light to transmit information among various nodes of a telecommunications network. Because it is capable of achieving extremely high bandwidth, it is an enabling technology for today's Internet and the communication networks that transmit the vast majority of all human and machine-to-machine information.

Mean Time To Data Loss (MTTDL): The average time before a loss of data happens in a given storage array. Mean time to data loss of a given RAID may be higher or lower than that of its hard disks. This depends on the type of RAID used.

NAND: A type of flash memory, most commonly used in USB drives and solid state drives (SSD).

OpenStack: A free and open-source cloud-computing software platform. Users primarily deploy it as an infrastructure-as-a-service (IaaS). The technology consists of a group of interrelated projects that control pools of processing, storage, and networking resources throughout a data center—which users manage through a web-based dashboard, through command-line tools, or through a RESTful API. OpenStack.org released it under the terms of the Apache License.

RAID: (originally **redundant array of inexpensive disks**, now commonly **redundant array of independent disks**) is a data storage virtualization technology that combines multiple physical disk drive components into a single logical unit for the purposes of data redundancy, performance improvement, or both. Data is distributed across the drives in one of several ways depending on the required level of redundancy and performance. The different schemas, or data distribution layouts, are named by the word RAID followed by a number, for example RAID 0 or RAID 1. Each schema, 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.

SAS: Serial Attached SCSI is a point-to-point serial protocol that moves data to and from computer storage devices such as hard drives and tape drives. SAS replaces the older Parallel SCSI (**S**mall **C**omputer **S**ystem **I**nterface, pronounced "scuzzy") bus technology and uses the standard SCSI command set. SAS is standardized in the T10 Technical Committee of the International Committee for Information Technology, an ANSI-accredited standards committee. SAS offers backward compatibility with SATA, versions 2 and later, allowing SATA drives to be connected to SAS backplanes. The reverse, connecting SAS drives to SATA backplanes, is not possible.

SATA: A computer bus interface that connects host bus adapters to mass storage devices such as hard disk drives and optical drives. SATA is marketed as a general-purpose successor to Parallel ATA and is now common in the consumer market, while the more expensive SAS is marketed for critical server applications**.**

SSD (Solid State Drive or Disk): A type of storage device that uses integrated circuit assemblies as memory to store data persistently. SSDs have no moving parts and uses electronic interfaces compatible with traditional hard disk drives. Most SSDs use NAND-based flash memory, which retains data without power. SSDs can also be built from RAM for faster access, but these require an additional power source to retain data after power loss. Hybrid drives contain a large hard disk drive and an SSD cache to improve performance of frequently accessed data.

Uncorretable bit error rate (UBER): A metric for data corruption rate equal to the number of data errors per bit read after applying any specified error-correction method. http://www.jedec.org/standards-documents/dictionary/terms/uncorrectable-bit-error-rate-uber