CSG2132 Module 1 Lecture Notes

Enterprise Data Centres (EDCs)

An EDC is usually a private facility operated by a single organisation that centralises its IT infrastructure, as well as the storage, management, and dissemination of corporate data. They may be located onsite, but often offsite at a location with connectivity, power, and security advantages. For large organisation's EDCs are essential for driving business operations and maintaining business continuity.

Internet-facing data centres support a small number of browser-based applications and unlimited users, examples being those operated by big players such as Google, Amazon Web Services, Microsoft, Apple, and Facebook. Enterprise data centres on the other hand, service fewer users (organisational staff and authorised 3rd parties), with a wide range of desktop and enterprise applications; both off-the-shelf and custom.

EDCs are custom-built to meet the organisation's needs. They are generally expensive to design, build and maintain, and are mainly utilised by large organisation able to exploit significant economies of scale and to provide highly structured control over how IT services and business data are deployed to and interacted with by users.

Organisations that most favour enterprise data centres are those to which strict data security and compliance requirements are legislatively applied, and hence, third-party enterprise service providers are not preferred. Such organisations may include those related to banking, medical, government services, mining, oil and gas and others.

An EDC is made up of four (4) main components:

1. **Facility** – the usable space available for IT equipment. Facility needs to support the operational parameters and tolerances of the equipment being used, such as manufacturer-specified temperature and humidity ranges.
2. **Support infrastructure** – the equipment that keeps resources running securely with maximum uptime possible. These include uninterruptible power sources (UPS), environmental controls, and physical security systems.
3. **IT equipment** – the actual servers, storage hardware, cables and racks, and firewalls.
4. **Operations staff** – to monitor maintain EDCs 24/7.

This unit focuses on the Data Storage components and infrastructure of EDCs.

Physical Data Storage

EDC data storage options fall into three (3) main categories:

- Hard Disk Drives (HDD)
- Solid State Drives (SSD)
- Tape Drives

HDD

As an EDC data storage option, HDD has been the standard for decades. Constantly improving technologies over this time such as perpendicular recording technology, bit patterned media recording (BPMR) and Shingled magnetic recording (SMR) have facilitated huge storage capacities, at a very low cost per gigabyte. This is what makes HDD options desirable to this day.

However, incessant demand for data access speeds and operational reliability have seen HDD begin to reach the outer frontier of their capabilities. Their access speeds have a definite ceiling which emerging demand has begun to exceed, they tend to be quite power- hungry, and generate significant noise and heat when clustered. And being mechanical devices with moving parts, they are always subject to damage, wear and tear over time.

It's these limitations and disadvantages in the face of ever-increasing performance demands that have seen the emerge of SSD as a EDC storage option.

SSD

Solid State Drives (SSD) are in effect an integrated circuit in which there are no moving parts. Rather than using disks, motors and read/write heads, SSDs employ NAND (non-volatile) flash memory, which are chips that retain data even when the power is turned off.

SSDs work similarly smartphone or tablet storage, but are considerably faster to start up, faster, shut down, and transfer data. SSDs are manufactured in much smaller form factors that HDD and consume considerably less power as well (hence their use in mobile devices) . They make less noise and tend to be more reliable over time due to not having mechanical components.

SSDs still have a finite operational life however, although not for the same reasons as HDD. SSD memory banks have a finite in life expectancy due to the existence of a limit of read/write operations before they cease to work. However, read/write operations are carefully managed by built-in logic designed to dynamically manage these operations in ways that extend SSD operational life.

Although more expensive than HDD, and with much lower storage capacities at the present time, SSD has proven a highly desirable storage option for high-speed, low-latency, heavy-usage, random data access applications such as real-time database processing, AI, big data, and e-commerce where read/write speeds and data throughput are critical.

Tape

Tape storage was the go-to solution for data backup in the /70s, and is still in use now four decades later, evolving considerably in this time. Tape storage has persisted all this time for several key reasons. Firstly, it's generally cheaper than disk storage by a significant degree. Secondly, it usually does not necessitate expensive IT infrastructure upgrades by an organisation to use as is the case for high speed HDD arrays and SSD. Tape also makes for an excellent mass data backup, redundancy and disaster recovery option, when backups can be transferred offsite and stored in specialised facilities for a very long time indeed.

However, tape, being an old technology, has certain shortcomings as you would expect. Firstly, read/write times tend to be very slow when compared to disk options. Secondly, a higher risk of data read issues have also been frequently reported over the years, although this is often attributed to poor management procedures than the actual tape systems themselves. Thirdly, tape-based backup/restore processes must be completed sequentially, which can complicate backup and restoration procedures.

Enterprise Data Storage Solution Types

In an enterprise datacentre, a key portion of the infrastructure is devoted making data available to the network. Multiple servers, users or services may need access to the same pools of data shared across the network. It is not always feasible for a server to be directly attached to the storage that it requires. Therefore, Enterprise data storage systems are usually employ some combination of three (3) different types of storage:

Direct Attached Storage (DAS)

Direct Attached Storage (DAS) is storage directly attached to a server, e.g. HDDs installed on full-featured computers. At the enterprise level, DAS reliability is enhanced and extended using data replication technology such RAID (Redundant Array Of Independent Disks). The main attraction of DAS is normally the cheapest and easiest solution to set up and maintain. However, the DAS approach is limited to the available disk slots on the server assets. This makes increasing storage capacity beyond this outer limit difficult or expensive.

Network Attached Storage (NAS)

Network Attached Storage (NAS) is a storage array that acts as a discreet, independent node on the network rather than embedded with a network node such as a server. All servers on the network will see the NAS as its own local DAS, even though it is not. NAS usually reside within their own devices that contain the hardware and software (logic) to house the storage devices and act as a 'file server' of sorts. The key advantage of NAS is that its capacity is easily scalable by simply adding more NAS nodes to the network. The disadvantage of NAS is that RW performance is highly dependent on network speed, which may well deteriorate as additional devices impact on network performance in the form of congestion. This, any plan to add NAS nodes may necessitate commensurate upgrades of the underlying network transmission hardware as well. At the enterprise level, NAS will generally use SAS (Serial Attached SCSI) enabled drives that are faster and more reliable than SATA drives, essential in high speed and high availability data transfer I/O requirements such as banking transactions and Ecommerce.

Storage Area Network (SAN)

Storage Area Network (SAN) allow attached client nodes to stored data at the block level. Block level storage places HDD/SSD assets in a remote chassis accessible via Fibre Channel or iSCSI. In effect therefore, SAN storage is seen as a locally attached disk to network servers, which they then interact with accordingly, even though they are not local to the servers at all. Servers can even be configured to boot from SAN rather than local disk. SAN offers flexibility in data storage management because the location and configuration of physical storage devices is irrelevant to users. Since servers only have direct access only to the SAN device (remote chassis), and not to the storage that device manages, SAN expansions and reconfigurations can be done without impacting attached servers. This is perfect for 'live' expansions. SANS usually employ one of three transmission protocols, these being Fibre Channel (FC), FC over Ethernet (FCoE) and iSCSI. Fibre Channel (FC) is a high-speed data transfer protocol that runs from 1 to 128 gigabit per second rates, providing in-order, lossless delivery of raw block data. It usually employs optical fibre cables within and between network nodes, but can also run on copper. iSCSI (Internet Small Computer Systems Interface) is an IP-based storage networking standard that allows block-level access to SAN devices by carrying SCSI commands over a TCP/IP network and allow location-independent data storage and retrieval.