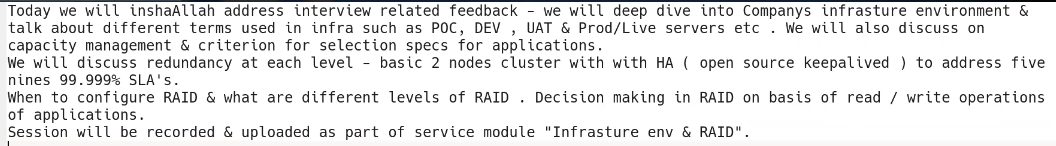
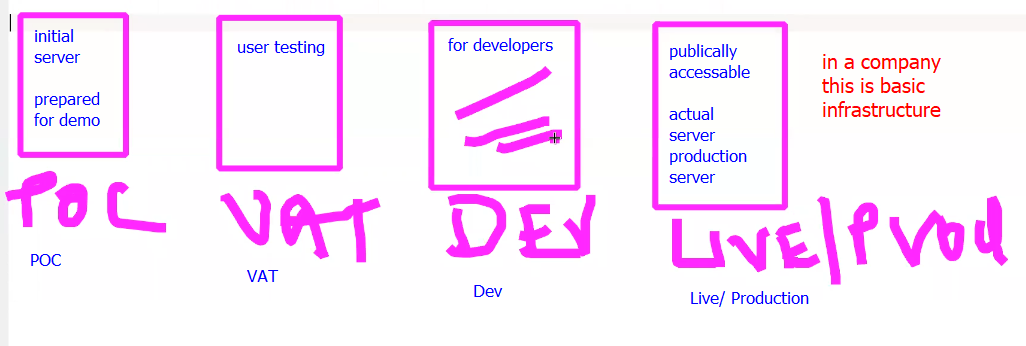
Lecture 20

**Infrastructure Setup/Capacity Management/HA/RAID**



* 
* Capacity management

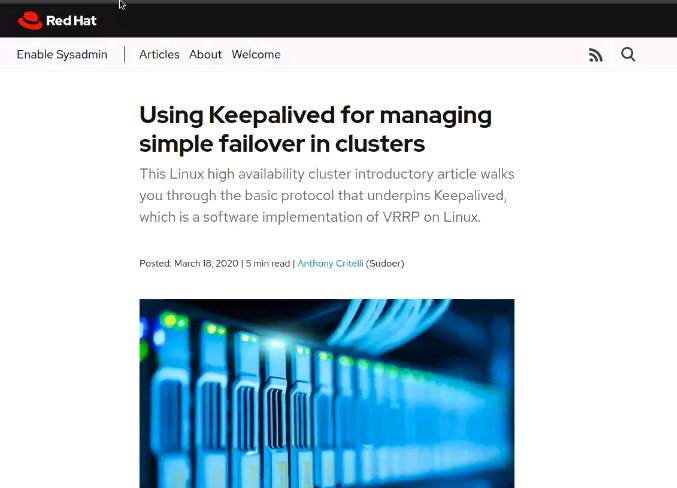
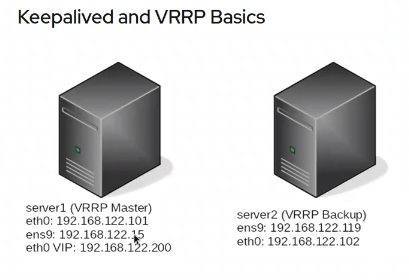
Capacity management is an important aspect of Linux administration in an organization. It involves ensuring that the organization's resources are efficiently utilized to meet the current and future demands of its users. The following are some steps involved in capacity management:

1. Performance monitoring: Monitoring the system's performance helps identify any bottlenecks in the system that may need optimization.
2. Capacity planning: Capacity planning involves estimating the resources required to meet current and future demands. It involves analyzing historical data to predict future resource requirements.
3. Resource allocation: Once the resource requirements are estimated, resources can be allocated to ensure that they are used efficiently. This involves allocating resources to different applications and services based on their priorities and usage patterns.
4. Optimization: Continuous optimization of the system helps ensure that it is running at optimal performance. This involves analyzing performance data and identifying areas that need improvement.
5. Disaster recovery planning: In the event of a system failure, having a disaster recovery plan in place can help minimize downtime and ensure that the system is restored quickly.

Overall, capacity management is critical to ensure that the organization's Linux systems are running efficiently and meeting user demands.

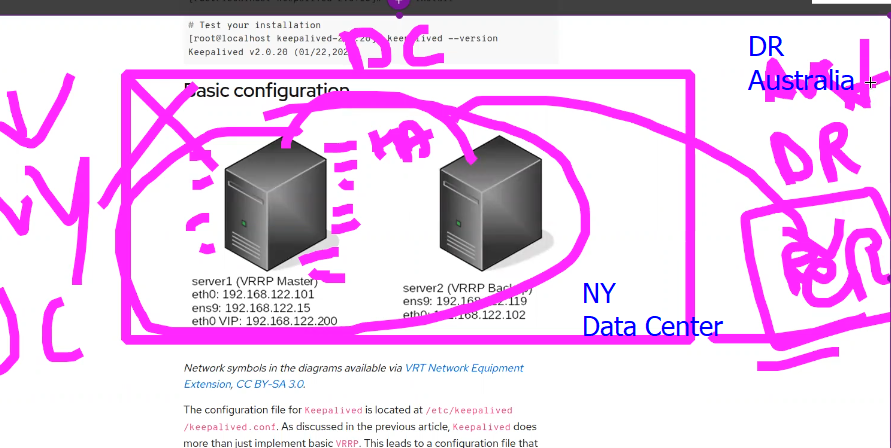
* **Dev Server** or any SERVER (company infrastructure)
* Need to engage with app team 🡪 if server is migrated.
* Read white paper for requirements.
* For **redundancy** 
  + Redundancy refers to the practice of adding extra components or systems to a design or system to improve its reliability and availability. The goal of redundancy is to minimize the risk of system failure and ensure that critical functions can continue in the event of a failure.
* Sir Kazim told, minimum 3 HDDs are present in a company, on a server (bear metal) 🡪 to rum operating systems
* At least two Network Cards 🡪 with two ports on each.
* It is to be connected to SAN (for storage) 🡪 2 HBA (Host But Adopter) card , fibre channel 🡪 2 ports each.
* 
* Fibre cables
* 
* SAN (Storage Area Network).
* 
* .
* 1 Management port (only in servers) for Ethernet Cable 🡪 if server is not accessible 🡪 this port can be connected through ILO to know what’s going on in server. With IDC
  + **IDC** Connect refers to a type of high-speed data communication interface that is used to connect computing devices and storage systems in data centers. IDC stands for Insulation Displacement Connector, which is ***a type of connector*** that uses a special mechanism to make a connection without needing to strip the wire insulation.
  + Integrated Lights-Out, or **iLO**, is a proprietary embedded server management technology by Hewlett-Packard Enterprise which provides out-of-band management facilities. The physical connection is an Ethernet port that can be found on most ProLiant servers and microservers of the 300 and above series..
* Ports 80, 443
* 2 SNPS (Power Supply)
* 2 PDC (Power Distribution Unit)
* 2 UPS
* *Manager decides which* ***brand*** *is to be purchased.*
* 2 mother boards (large servers) (it is called hot swappable) 🡪 means can be switched in running machine.

Project scenario

* xyz.abc,com web site is to be launched.
* 2 servers are purchased.
* Aik **cluster** bny ga (in dono servers ko mila kr) 🡪 *two node cluster.*
* This is minimum requirement
* Company never deploys 1 server 🡪 so that *downtime is avoided*.
* Bothe servers are placed and connected to IDC.
* Connected to these servers with “management port” **remotely.**
* Suppose, Server 1 IP 4.4.4.4 and Server 2 IP 8.8.8.8 and application virtual IP is 2.2.2.2 (Server IPS are in background)
* ***2.2.2.2 (Live IP) is for connection from internet***. 🡪 it floats on server 1 or server 2 in case of failover.
* Virtual IP will also work as Load Blancer, Failover etc.
* While accessing xyz.abc.com is accessed from outside, the request will go to server 1 and if it is down it will go to server 2.
* This is called cluster.
* Some application has its own cluster.
* If not, Linux cluster is to be implemented ( an open source cluster is **keepalived**)
* 
* .
* 
* .
* Graphical user interface, text, application, email

  Description automatically generated
* .
* Graphical user interface

  Description automatically generated with medium confidence
* .
* 
* .
* Graphical user interface, text, application, email

  Description automatically generated
* If keepalived service is stopped 🡪 the floating will go to server 2
* A **floating IP** address is an IP address that can be quickly and easily moved from one network interface to another within a network. It is typically used in high-availability configurations, where multiple servers or nodes are used to provide redundancy and failover protection for critical applications or services.
* This is an example of ***High Availability.***
* Disaster Recovery
* 
* .

**RAID**

* RAID stands for Redundant Array of Inexpensive Disks or Redundant Array of Independent Disks, and it is a method of combining multiple hard disk drives (HDDs) or solid-state drives (SSDs) into a single logical unit. The purpose of RAID is to provide increased storage reliability, performance, and capacity.
* For ***Hard Disk Redundancy*** 🡪 RAID is used 🡪 minimum 2 Hard Disks are used.
* For Network Redundancy 🡪 Bonding
* HBA Redundancy 🡪 multi pathing
* .
* Controller Card for RAID
* 
* .
* 2 cards for Redundancy are required.
* iLO can be used to configure server remotely.
* 1st step is RAID 🡪 it must do step before installation.
* RAID Levels
* 0, 1, 5, 1+0, 6
* Mostly these levels are used 🡪 **interview question**, what is the difference b/w RAID level 1+0 (nested level) and RAID level 5 is.
* There are two types of RAID
  + Hardware RAID
    - RAID Card
  + Software RAID
    - Embedded on motherboard.

Definition by ChatGPT

1. Hardware RAID: In a hardware RAID configuration, a dedicated RAID controller is used to manage the RAID array. The RAID controller is a separate hardware component that sits between the computer's disk interface and the physical disks. Hardware RAID can provide better performance, reliability, and scalability than software RAID, but it is also more expensive.
2. Software RAID: In a software RAID configuration, the RAID functionality is handled by the operating system or software. Software RAID can be implemented using the computer's existing hardware and is therefore less expensive than hardware RAID. However, it can also be less reliable and provide lower performance.
3. Hybrid RAID: Hybrid RAID is a combination of hardware and software RAID. It uses a dedicated RAID controller to manage the RAID array but also relies on software to perform certain functions. Hybrid RAID can provide better performance and reliability than software RAID, but it is also more expensive than pure software RAID.
4. Network RAID: Network RAID allows multiple computers to share a single RAID array over a network. This can be useful in situations where data needs to be shared between multiple computers and a single RAID array is more efficient than multiple independent arrays.

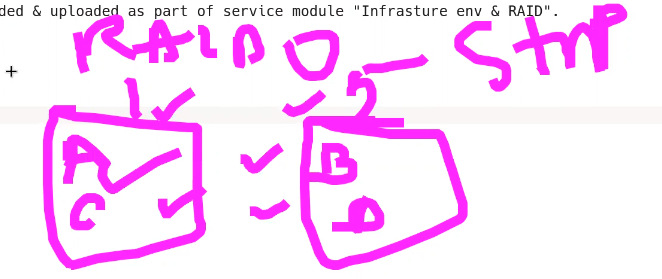
Each RAID type has its own advantages and disadvantages, and the choice of RAID type depends on the specific needs of the system or application.

* ***RAIDS are configured from RAID BIOS.***

***Critical thing is which level of the RAID is to be used.***

**RAID Levels**

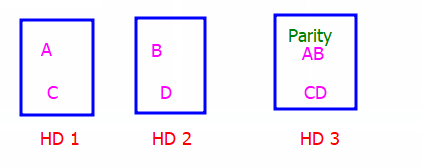
**RAID 0:** Data is **striped** (Striping is a technique used in RAID (Redundant Array of Independent Disks) where data is split into blocks and stored across multiple disks in the array. Each block is written to a different disk, with the goal of improving the read and write performance of the array.) across multiple disks for improved performance, but there is **no redundancy or fault tolerance.** RAID 0 requires at least **two disks** and provides increased speed and capacity.

* + Separate controller is not required.
  + Can work on OS level.
  + RAID 0 is striping.
  + 
  + So, if one HD is failed, data is lost.
  + If more then two HDs are used, the performance will improve proportionally. 🡪 but *it will increase the cost also.*
  + Its advantage is that its performance is fast.
  + In case of using two HDs of 500 Tb 🡪 usable capacity would be 1000Tb
  + ***If we require fast write operation, 🡪 we will use RAID 0.***

**RAID 1**: Data is **mirrored** across two or more disks for redundancy. If one disk fails, data can still be accessed from the other disk(s). RAID 1 requires at least **two disks** and provides increased ***fault tolerance***.

* + Its advantage is redundancy.
  + In case of two 500 Tb HDs 🡪 useable capacity will remain 500 Tb because it mirrors the data.
  + If 1 HD fails 🡪 no data loss.

**RAID 5:** Data is **striped** across multiple disks, and **parity** (In RAID 5, parity information is distributed across all of the disks in the array, so there is no single dedicated parity disk.) data is distributed across all disks. If one disk fails, data can be reconstructed from the remaining disks using the parity information. RAID 5 requires at least three disks and provides both increased performance and redundancy.

* + Minimum 3 HDs 🡪 stripe and parity
  + It is mostly used in professional environment.
  + 3 HDs of 500 Gb = 1.5 Tb 🡪 but it will be considered as 1 Tb capacity. (Total Storage – 1 HD)
  + 
  + No data loss if 1 HD is failed.
  + If 2 HDs are failed, data is loss
  + Cost factor (3 HDs)

**RAID 1+0** (also known as RAID 10): Data is mirrored across two or more sets of disks, and then striped across those sets for increased performance. RAID 1+0 requires at least four disks and provides both increased performance and redundancy.

* + Minimum 4 HDs 🡪 cost factor
  + Fault tolerance
  + Great striping performance
  + Great writ performance
  + Hardware RAID controller is must.
  + Data stripe and mirror

Interview Question Which level is best?

* + - If great read performance is required (DNS Server) 🡪 RAID 5
    - For great write and read performance RAID 1+0

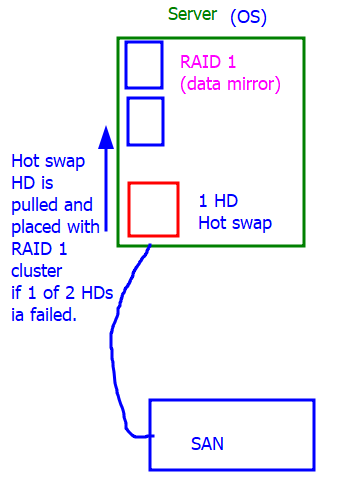
**RAID 6:** Data is striped across multiple disks, and two sets of parity data are distributed across all disks. If two disks fail, data can still be reconstructed from the remaining disks using the parity information. RAID 6 requires at least four disks and provides greater fault tolerance than RAID 5.

* + 2 fault tolerance
  + Dual parity
  + 4 HDs

Each RAID level has its own advantages and disadvantages, and the choice of RAID level depends on the specific needs of the system or application. RAID 0 is good for improving performance but has no redundancy, while RAID 1 provides redundancy but reduces capacity. RAID 5 provides both performance and redundancy but requires more disks than RAID 1 or 0. RAID 1+0 provides both performance and redundancy but requires more disks than RAID 5. RAID 6 provides the highest level of fault tolerance but requires at least four disks and has lower write performance compared to RAID 5.

* .

Personal experience of Sir Kazim

* For OS operation (not much data is to be write)
* Data is saved on SAN.
* Total 3 HDs
  + 2 HDs RAID 1
    - *Mirror* because OS needs to be up always.
  + 1 HD Hot Swap
    - Push & pull the HD
* If 1 HD is failed, the Hot Swap HD is pulled and placed with 2 HDs cluster (RAID 1). 🡪 no downtime as a result of this architecture.
* The Hot swap HD will be repaired and replaced as hot swapper.
* 
* For OS always use RAID 1