Business Continuity and Disaster Recovery

Ed Crowley

Ch 12

Topics

- Disaster Recovery
- Business Impact Analysis
- MTBF and MTTR
- RTO and RPO
- Redundancy
- Failover
- Backup Sites
- Load Balancing
- Mirror Sites

Disaster Recovery Methods

- When implementing disaster recovery, it is important to form a disaster recovery plan (DRP) or business continuity plan (BCP) that describes organizational plans.
- Requires understanding how critical the application or server is and then determining the proper disaster recovery method for it.
- When creating a DRP, it is first necessary to focus on mission critical applications and servers.

Malden Mills Fire

- Fire destroys plant
- Employee retention critical
- Salaries paid for 3 months
- Disaster, Recovery, and Damage Control costs
- Plan for potential disaster impacts

- BIA Video (12minutes)
- https://www.youtube.com/watch?v=bMkyV4bMCM4

DR Planning Considerations

- If the failure of a system results in the organization's failure to operate and generate income, that system would be considered mission critical.
  - These systems need to be identified and backed by a proper disaster recovery method.
- Another DRP consideration is where to place the disaster recovery center.

Disasters

- Geographic diversity should be taken into account when planning for a disaster that may impact a particular geographic region.
- Disasters come in many forms, including natural disasters, so placing the disaster recovery center in a location that is 1000 miles away might prevent the same natural disaster from destroying both the primary data center and the disaster recovery center.

MTBF & MTTR

- Mean time between failures (MTBF)
- Mean time to repair (MTTR).
- MTBF is the average time a device will function before it fails.
- MTBF can be used to determine approximately how long a hard drive will last in a server.
  - It can also be used to plan how long it might take for a particular hardware component to fail and thereby help with the creation of a DRP.
- MTTR, on the other hand, is the average time that it takes to repair a failed hardware component.

RTO and RPO

- Recovery time objective (RTO) amount of time between an outage and the restoration of the service.
- Recovery point objective (RPO) maximum amount of time in which data can be lost for a service due to a major incident.
- For example, if you back up your system overnight, then the recovery point objective becomes the end of the previous day.
- VMWare Use Case Disaster Recovery
- https://www.youtube.com/watch?v=5MSyB2TNO_A

Redundancy

- One of the factors that should be considered and that can help meet expected RTO and RPO is redundancy.
- A redundant system can be used to provide a backup to a primary system in the case of failure.
- Redundant components protect the system from failure and can include power supplies, switches, network interface cards, and hard disks.
- RAID (redundant array of independent disks) an example of a redundant system.

Redundancy

- A redundant component means you actually have more of that component than you need.
- For example, a virtualization host computer might have two power supplies to make it redundant, but it can actually function with a single power supply.
- Redundant does not mean that there is not an impact to performance if a component fails; it means that service can be restored to working condition (although the condition may be at a degraded state), without the need for external components.
- Redundancy differs from fault tolerance in that fault tolerance allows the system to tolerate a fault and continue running in spite of it.

Heartbeat and Failover

- Failover uses a constant communication mechanism between two systems called a heartbeat.
- As long as this heartbeat continues uninterrupted, failover to the redundant system will not initiate.
- If the heartbeat between the servers fails, the redundant system will take over processing for the primary system.
- If the primary system becomes operational again, the organization can initiate a failback.
- A failback is the process of restoring the processing back to the original state before the failure of the primary system.

Multisite Configuration

- To help reduce downtime, an organization can set up and configure a multisite environment.
- Using a multisite configuration is more expensive but helps provide a more advanced business continuity plan.
- In order to utilize a multisite configuration, the organization needs to establish a backup site where they can easily relocate their computer equipment if a disaster occurs at their primary location and data center.
- The backup site needs to be either another location that the company owns and has available to them to implement additional equipment or a space they purchase or rent from another provider for an annual or monthly fee.
- In either case the organization needs to have a secondary location that it can use to host the computer system in case of a disaster.

Three Types Of Backup Sites

- Cold site
- Warm site
- Hot site

- Difference between each site is determined by the administrative effort to implement and maintain them and the costs involved with each type.

Cold Site Least Expensive

- Cold site does not include any backup copies of data from the organization's original data center.
- When an organization implements a cold site, they do not have readily available hardware at the site; they only have the physical space and network connectivity for recovery operations.
- Since there is no hardware set up and ready to use at the backup site, it takes longer…

Hot Site

- Can contain a real-time synchronization between the original site and the backup site and can be used to completely mirror the original data center.
- Operationally duplicates original site of the organization and has readily available hardware and a near-complete backup of the organization's data.
- If the original site is impacted by a disaster, the hot site is available for the organization to quickly relocate to, with minimal impact on the normal operations of the organization.

Warm Site

- Between a cold site and a hot site.
- It has readily available hardware but on a much smaller scale than the original site or a hot site.
- Warm sites will also have backups at the location, but they may not be complete backups or they might be a few days old.

Acceptable RTO

- Drives the choice.
- A hot site might have an RTO of a few hours, whereas a cold site might have an RTO of a day or more.
- A hot site provides faster recovery time but is also at a much higher cost than a warm site.
- While a cold site is the least expensive to set up, it also takes the longest to implement in the event of a disaster.

Backups and Recovery

- Selecting appropriate backup solution is critical.
- Backup is simply the process of copying and archiving data so that the data is available to be restored to either the original location or an alternate location should the original data be lost, modified, or corrupted.

Backups serve two primary purposes.

- Restore data that is lost because either it was deleted or it became corrupt.
- Enable recovery of data from an earlier time frame.

Data Retention

- An organization should have a data retention policy that specifies how long data needs to be kept.
  - For example, if an organization has a data retention policy that specifies all data must be kept for two weeks, an end user who needs to have a document restored from ten days ago could do so.

Backups

- When selecting a backup policy, several factors need to be taken into consideration.
- First, the organization must determine how the backups will be stored, whether on tape or DVD-R media, to a dedicated hard disk, or to a cloud-based storage system.
- If they are storing data on tapes or DVD-R media, they need to determine if the backups should be stored at an off-site location.
- Storing backups at an off-site location or in the cloud allows for recovery of the data in the event of a disaster.
- After choosing a media type, the next step is to choose the style of backup.

Three Backup Styles

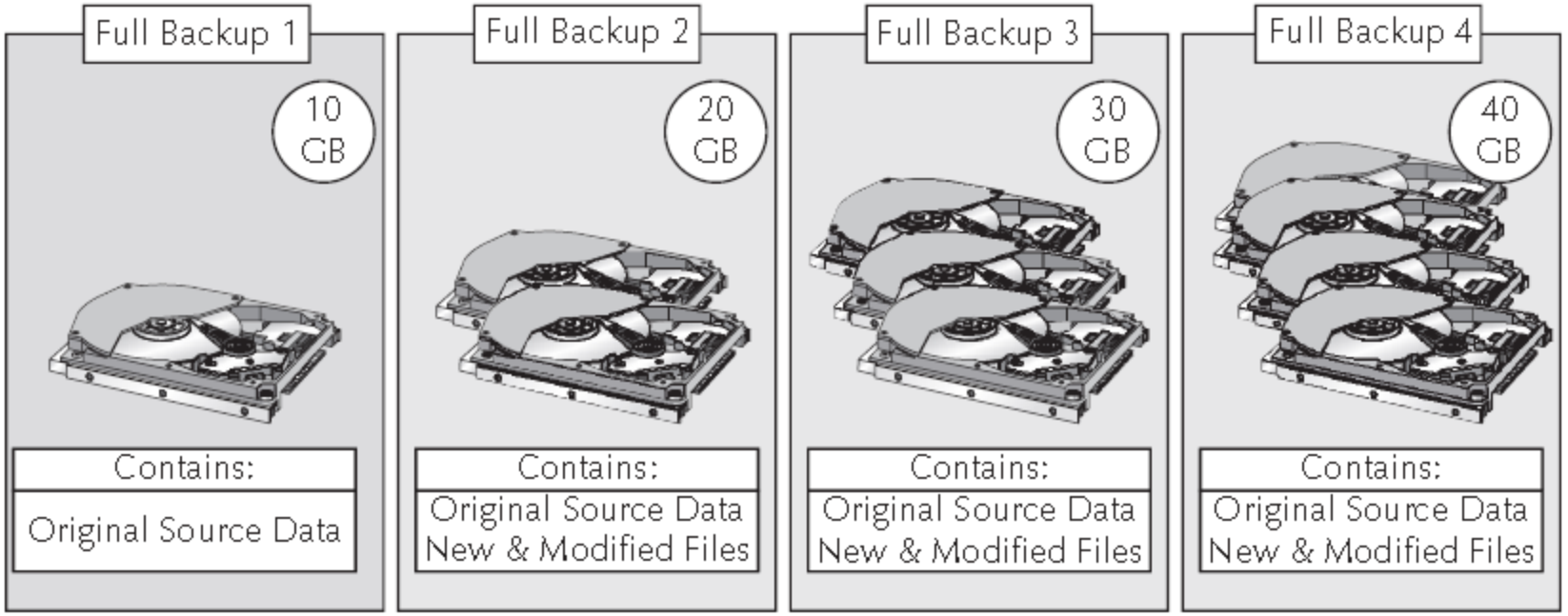
- Full
- Incremental
- Differential.

- Each style has its own set of advantages and disadvantages.
- Backup plan should include how the data is to be stored and, if the data is going to be stored off-site, how long it is kept off-site and how many copies are kept at the off-site facility.

Full Backup

- A full system backup backs up the entire system.
  - Makes a copy of all the data and files on the drive in a single process.
- Takes up the most space on storage media because it does a full drive copy every time the backup is executed.
  - So performing a full backup every day requires the same amount of space on the backup media as the drive being backed up.
- The benefit to a full backup is that an organization can take any of the backups from any day they were executed and restore data from a single backup media.

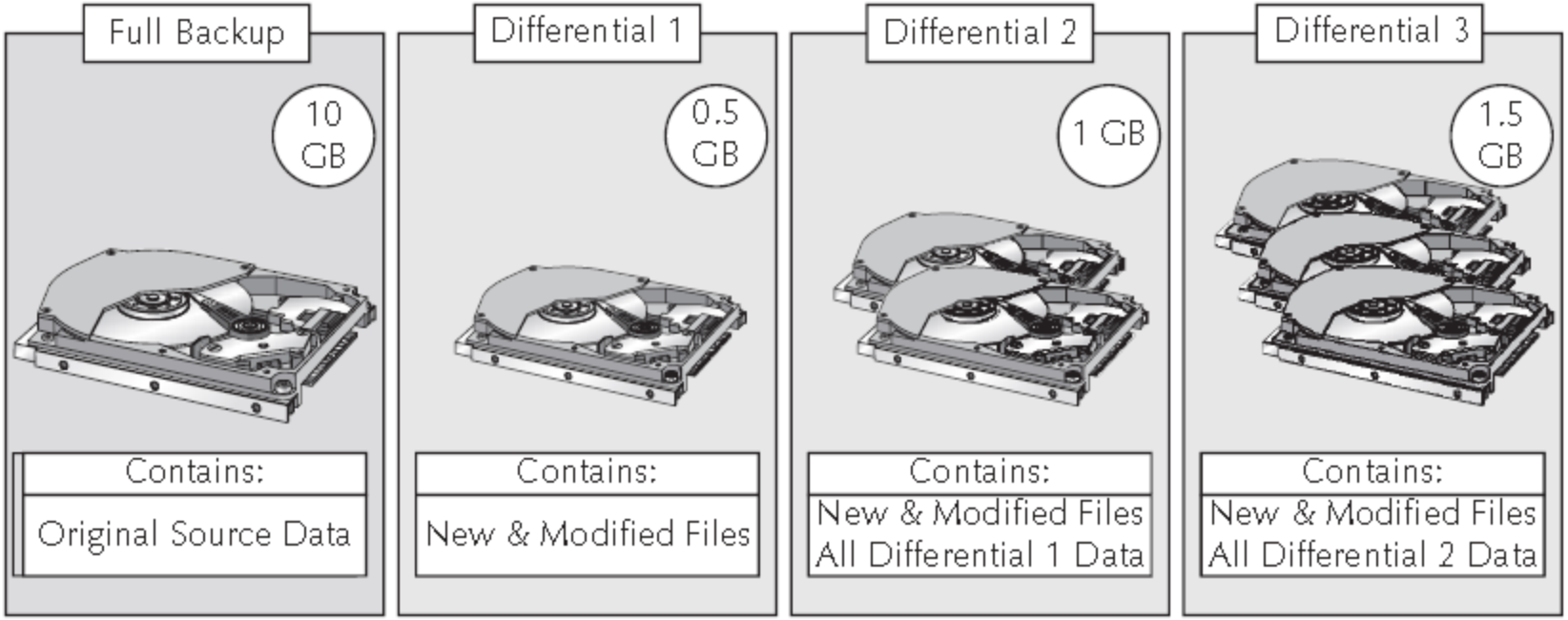
Full Backup



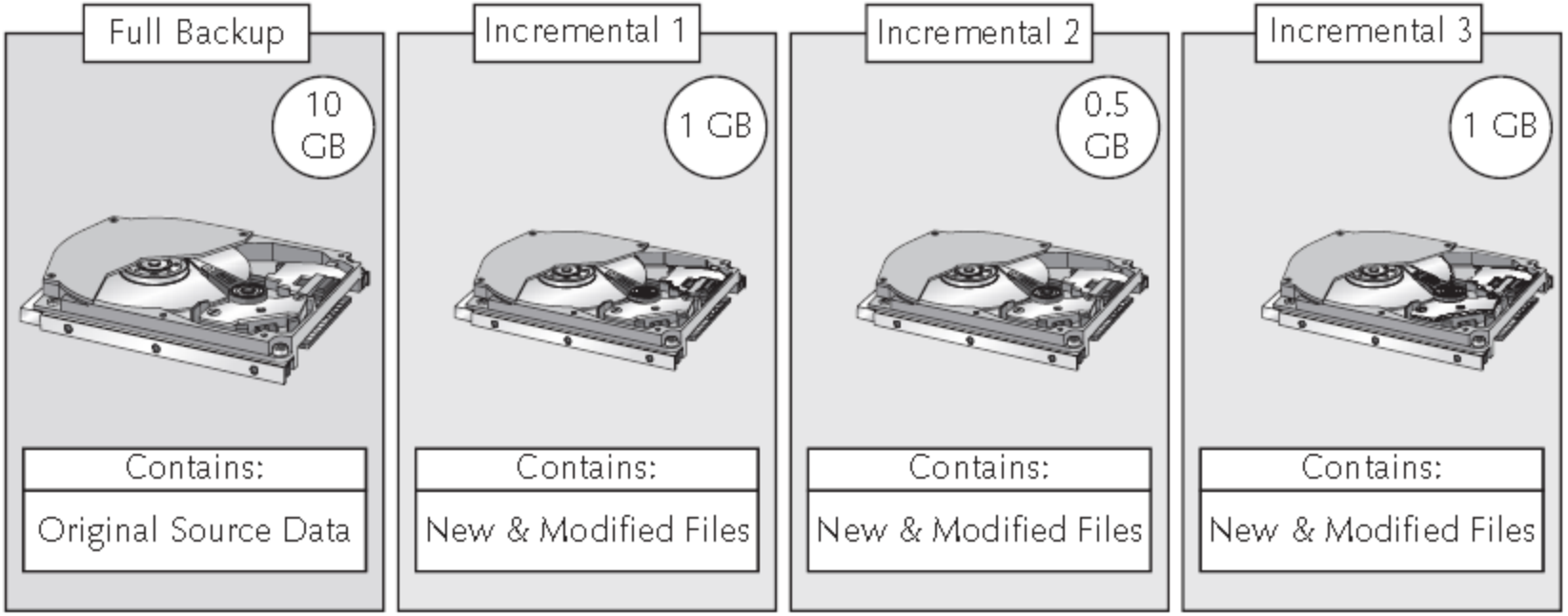
- How a full system backup would look after four backups.

Differential

- Differential backup backs up only those changes that were made since the last full backup was executed.
- In order to perform a differential backup, a full backup must first be performed.
- After the full backup is executed, every differential backup executed thereafter will contain only the changes made since the last full backup.
- One disadvantage to differential backups is that the time it takes to complete the backup will increase as files change between the last full backup.
- Another disadvantage is if the organization wants to restore an entire system to a particular point in time, they must first locate the last full backup taken prior to the point of failure and the last differential backup since the last full backup.



- How a differential backup looks after three days.
- An incremental backup also backs up only those files that have changed since the last backup was executed, but the last backup can be either a full backup or an incremental backup.
- Makes incremental backups faster and requires less space.



- Time it takes to perform a restoration is longer because both the last full backup and all the incremental backups must be restored.
- How an incremental backup would look after three backups.

Server Image

- In addition to backups, an organization has the option of capturing a server image.
- When capturing an image, the entire hard drive is captured block by block.
- Because the entire hard drive was captured, the image can be used to restore an entire server in the event of a disaster, allowing the image to be restored on new hardware.
- Creating an image of a server differs from the file-based backups discussed earlier in that the file-based backups only allow you to restore what was configured to be backed up, whereas an image allows for the entire restoration of the server, including files, folders, and operating system.

Replication

- Backups are sometimes confused with replication.
- The two differ in that backups are created to store unchanged data for a predetermined amount of time, whereas replicas are used to create a mirrored copy of the data between two redundant hardware devices.
- Replicas help to improve reliability and fault tolerance.
- When replicating data, the data is stored on multiple storage devices preferably at different locations so that if one location suffers a disaster the other location is available with the exact same data.

Two types of replication:

- synchronous
- asynchronous

Synchronous vs. Asynchronous

- Synchronous replication copies the data over the network to another device.
  - Allows for multiple copies of up-to-date data.
- Synchronous replication writes data to both the primary and secondary sites at the same time
  - Both locations have same data.
- Synchronous replication is more expensive than asynchronous replication and can impact the performance of the application that is being replicated.
- With asynchronous replication there is a delay before the data is written to the secondary site.
- New data can be accepted at the primary site without having to wait for the data to be written to the secondary site.
- If the primary site fails before the data can be replicated to the secondary site, then the data that had not yet been written to the secondary site may be lost.

Snapshots

- A snapshot simply captures the state of a virtual machine at the specific time when the snapshot was taken.
- While similar to a backup, a snapshot should not be considered a replacement for traditional backups.
- A virtual machine snapshot can be used to preserve the state and data of a virtual machine at a specific point in time.

Snapshot

- A snapshot can be taken before a major software installation, and if the installation fails or causes issues, the virtual machine can be restored to the state it was in when the snapshot was taken.
- A snapshot includes the state the virtual machine is in when the snapshot is created.
- Snapshot includes all the data and files that make up the virtual machine, including hard disks, memory, and virtual network interface cards.
- Snapshots and snapshot chains can be created and managed in a variety of different ways.
- It is possible to create snapshots, revert to any snapshot in the chain, and even delete snapshots.

Snapshots

- Although snapshots should not replace normal backup software, they are a good way to repeatedly revert a virtual machine to the same state without having to create multiple virtual machines.
- A snapshot keeps a delta file of all the changes after the snapshot was taken.
  - The delta file records the differences between the current state of the virtual disk and the state the virtual machine was in when the snapshot was taken.
- So if the snapshot is kept for long periods of time the file can grow and might become too large to remove.
  - Can cause performance issues for the virtual machine.
- If there is a need to keep a snapshot longer than a few days, it is recommended to create a full system backup.

High Availability

- High availability is a system design approach that ensures a system or component is continuously available for a predefined length of time.
- If the end user cannot access the service or application, it then becomes unavailable, commonly referred to as downtime.

Downtime comes in two different forms:

- scheduled downtime
- unscheduled downtime.

Downtime

- Scheduled downtime is downtime that has been predefined in a service contract that allows an administrator to perform routine maintenance on a system, like installing critical updates, firmware, or service packs.
- Unscheduled downtime usually involves interruption to a service or application due to a physical event, such as a power outage, hardware failure, or security breach.
- Most organizations exclude scheduled downtime from their availability calculation for an application or service as long as the scheduled maintenance does not impact the end users.
- Making sure that an IT department can meet availability requirements and that an application or service is always available to the end user is a critical component of the organization. In order to guarantee a certain level of availability for an application or service, fault tolerance can be employed.

Fault Tolerance

- Fault tolerance allows a computer system to function as normal in the event of a failure in one or more of the system's components.
- Fault-tolerant systems are designed for high availability and reliability by installing multiple critical components.
- For example, a virtualization host computer would have multiple CPUs, power supplies, and hard disks in the same physical computer.
- If one of the components were to fail, the spare component would take over without bringing the system down. However, having a system that is truly fault tolerant does result in greater expense because the system requires additional components to achieve fault-tolerant status.
- In addition to adding components to achieve fault tolerance, two or more computers can be connected together to act as a single computer.

Clustering

- Connecting multiple computers to provide parallel processing and redundancy is known as clustering.
- The computers are connected over a fast local area network (LAN), and each node (i.e., each computer used as a server) constituting the cluster runs its own operating system.
- Clusters can thereby improve performance and availability as compared to using a single computer.
- In addition to local clustering, there is also the ability to use geoclustering.

Geoclustering

- Geoclustering allows for the connection of multiple redundant computers while those computers are located in different geographical locations.
- So instead of having the nodes connected over a LAN, the nodes are connected over a wide area network (WAN) but still appear as a single highly available system.
- Geoclustering allows an organization to support enterprise-level continuity by providing a system that is location independent.

Redundancy

- Having an infrastructure that is redundant and highly available helps an organization provide a consistent environment and a more productive workforce.
- Determining which systems require the investment to be highly available is up to each organization.
  - There will be some systems or applications that do not need to be highly available and do not warrant the cost involved to make them so.
- One of the benefits of a public cloud model is that the cost of making the systems highly available falls on the cloud provider and allows the cloud consumer to take advantage of that highly available system.

Redundancy

- Determining which systems and which applications require redundancy can help reduce costs and administrative overhead.
- A standard needs to be established to help determine the availability required for each application.
- An organization might use a scale of 0 to 4 to rate the availability requirements of an application.
- In that scenario an application that has a rating of 0 would need to be available 99.99% of the time, whereas an application with a rating of 4 might only have to be available 98% of the time.
- Creating a scale allows an organization to prioritize their applications and appropriately distribute costs so that they can maximize their compute resources.

Multipathing

- Having a fault-tolerant system is a great start to achieving high availability, but it is not the only requirement.
- When planning for high availability, all aspects of the network must be considered.
- If the connection between the fault-tolerant systems is a single point of failure, then it is limiting the high availability of the system.
- Implementing multipathing allows for the configuration of multiple paths for connectivity to a storage device, providing redundancy for the system to connect to the storage device.

Load Balancing

- Another form of high availability is load balancing.
- Load balancing allows you to distribute a workload across multiple computers, networks, and disk drives.
- Load balancing helps to optimize workloads and resources, allowing for maximum throughput, and helps minimize response times for the end user.
- Load balancing can also help to create reliability with the use of multiple computers instead of a single computer and is delivered either with dedicated software or hardware.

Load Balancing

- Load balancing uses the resources of multiple systems to provide a single, specific Internet service; it can be used with a website or a file transfer protocol (FTP) site, for example.
- Load balancing can distribute incoming HTTP requests across multiple web servers in a server farm, which can help distribute the load across multiple servers to prevent overloading any single server.
- If one of the servers in the server farm starts to become overwhelmed, load balancing begins to distribute HTTP requests to another node in the server farm so that no one node becomes overloaded.

- In addition to using load balancing, some organizations may want to implement a mirror site.

Mirror Site

- A mirror site is either a hosted website or set of files that reside as exact copies of one another on multiple computers.
- This mirror copy ensures that the website or files are accessible from multiple locations to increase availability and reduce network traffic on the original site and is updated on a regular basis to reflect any changes in content from the original site.
  - A mirror site can be set up to reflect geographic discrepancies, making it faster to download from various places throughout the world.
- Sites that offer a large array of software downloads and have a large amount of network traffic can use mirror sites to meet the demand of the downloads and improve response time for the end user.

Questions?