Overview 2

Hi and welcome back. This lecture will be on the following. We'll be talking about multiple ha pairs, flexor rays on tap select. Uh, we'll see what a metro cluster is and we'll start introducing storage by means of physical storage. Now if you want to have more than one ha pair, that means that you'll be scaling out your cluster. Adding a second ha pair to the same plusser means that uh, you have the same configuration environment so you still have a data network management network obviously, but now you've got more than one [inaudible]. This also means that in a single age a you can have a switch lists interconnect and what you see here is that if obviously if you have more than two nodes, you can't, you can't have a point to point cluster interconnect. So you definitely need switches and these switches should be dedicated for that network.

There will be no client traffic transferring across this network. This is purely for the cluster like we discussed in the previous slide. Also, you can see that both HTA payers have their own physical storage. Note three for example, we'll never be able to excess physically access the disks of note too because simply because it's not cable that way, so no two or node one share the same disks. Node three and node four also share the same disks. It doesn't mean that node one is enabled to excess volumes that are hosted by the second ha pair, but it will never be able to access the discs physically and obviously each age APR has its own Ha interconnect network. Now you can have a maximum of up to 12 ha pairs per cluster depending on the protocol you use in a San Environment. You can have fewer ha pairs in your cluster.

Also, depending on the controller type you use, if you use the all flash arrays, you can have up to six ha pairs, but the smaller raise only allow for four haps percent environment. Then the flexor aid, as we discussed it, will receive loans that are presented from, for example, IBM storage arrays or HP or other stories you raise. These lungs are presented via the private channel protocol and the Flexray will present the lens via an aggregate. So your clients will be able to write two volumes in the aggregate, which are constructed of array lands that are offered from other vendors, storage environments. Then we have on depth select, which is the software defined storage environment. Uh, you can use it in a private cloud or in the Ed or in the Amazon web service cloud. Uh, so your virtual machines, which, uh, which is what, uh, ontop select is running, can be running as an actual cluster, uh, but will be running on top of hypervisors.

And the Nice thing is that the hypervisor can have a raid controller or other storage that, that the hypervisor uses and then the hypervisor will run the virtual machine and um, allow that virtual machine to use VM dks that are actually configured on the hypervisor as the disks for the simulator. So an ha pair, we'll have discs that are, um, presented by the hypervisor. A big difference with physical nodes is that the haps Perez in the on tap select environment do not share their storage. So node one in this case has private storage note to his private storage even though they form an ha pair, which is absolutely contradictory to the original uh ha pair with physical machines.

And finally we've got the metro cluster which allows you to run in two different data centers with a maximum of 300 kilometers distance. And the um, the important thing here is that you have two

networks. The first network is for the configuration of the storage virtual machines on this one to the other cluster, um, which is in the different, in the other data center. And another thing is that we have the fiber channel switches which are responsible for transferring energy ram mirroring and transferring the mirror data to the other side. So whatever we put in here is also put in there. So your data is absolutely mirrored 100% as zero data loss. Also the envy rem is mirrored from node one and node two to the other nodes in the other cluster. So in a metro cluster environment, you can lose an entire data center without losing access to your data.

Now let's have a look at storage. When we talk about storage, we can think of physical storage and logical storage, as far as Netapp is concerned. First of course we start with physical storage because you can't have anything logical without you having a physical environment first. Now physical storage can be subdivided in two types of disk drives. We've got hard disk drives and we've got solid state drives and we can have arrays that only have or ha pairs or single node clusters that only use high disk drives. We can have an all flash arrays which only uses solid state drives and we can have a combination of the two. So we can have an intermediate array that that does have hard disk drives and it also runs as is ds in the same this group or aggregates. When you put two types of disks, um, hard disk drives and SSDs in the same aggregate, you would talk about a hybrid pool or a flash pool. Um, obviously this is good for performance, whereas SSDs in all flash arrays are even better for performance.

Now, what you see in this picture is two nodes with an ha interconnect. So obviously this is an ha pair. Uh, you also see that there's an aggregate which is a hosted or owned by node one and physically no two is able to access the discs, but no to is not the owner of the disks a node one. However, if node one fails, then the aggregate would be taken over by new too if he wanted that. Another thing that, um, um, sticks out here is that there is something which is called [inaudible]. This is just an example, right? So all the disks that we have here, originally, these discs are spare disks. So they're available for, uh, this particular note. So they have been owned by this node and there are no spare disks. The minute you put them in an aggregate, they become data disks or parity disks, right?

And this is called a raid group, or in other words, a stripe. So you can have a stripe of 16 disks or 18 discs and every stripe will have its own parity discs as a parenting level is either rate for ray DP or ray tech. Um, raid for obviously stands for one parity disk raid. DP stands for dual parity or double parity discs. And rate tech stands for a triple erasure coding, which means three parity discs. This is always a trade off. Either you go for performance and space or you go for resilience. Uh, so the more parity this you have, the less space you have in the aggregate, but the more resilient you you are. So there's always a trade of you should go for the best of both worlds. And NETAPP has defaults for that. So for example, if you've got says devices and you use double parity, then your rate group size will be 16 discs, but you can change that. The minimum number of raid groups in an aggregate is one. So you always have raid group zero and the more disk you put it in the aggregate, the more rate groups you will have. Another important thing you should know for the exam. Resizing the aggregate means that you can add disks but you cannot shrink the aggregate so you can't remove disks logically from the aggregates.

Then, um, by means of summary, the discs that you add to a controller or to a controller pop pair is uh, an unassigned set of disks. They're not owned by any of the two notes. The minute you run the

disk assigned command, then the disc will be owned by one note or by the other node in the 80 [inaudible]. If the disks are assigned, then the automated automatically becomes spare disks and spare disks are there for two purposes. One is you can create an aggregate using spare disks and two, you can use the sped is to replace a broken disk.

Another thing which is important to know is that Netapp runs the aggregates but not just as a bunch of disks, but it contains an aggregated, contains a file system. Now file system as you may or may not know as a collection of I nodes and data blocks in this little cloud in here because this is a very oversimplified picture of a file system, eh, it means in this little cloud you have a lot of eye nose and pointers in direct pointers to ultimately to the data blocks. So there's a structure of administration inside this cloud, but for now you have to know that there is a starting point, which is the route I node of the file system. And there's the end point, which is the data book. So you can have thousands of files, all these files we'll have I nodes that have pointers to data books and you can only access them if the root I knowed is there and offers you a path to the [inaudible] inside this little cloud, how the file system is per aggregate. And another thing you have to know for the exam is that the size of logical blocks that has written to the disks inside an aggregate is four k. So if you have a stripe, for example, of a 16 discs and you write a data to these discs, then all the disks will be used because per disc only four k is written. So let's take a small break and then continue with the overview.