

Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Fast Track to Solaris 10 Adoption: ZFS Technology A Sun Expert Exchange Discussion

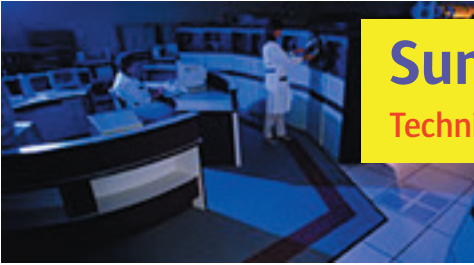
ZFS represents a dramatic advance in file system management and maintenance by automating common administrative tasks, protecting data from corruption, and providing virtually unlimited scalability. ZFS uses virtual storage pools, which make it easier to expand or contract file systems simply by adding more drives. This significantly reduces costs by simplifying storage administration and allowing resources to be shared among file systems.

This summary includes highlights of the hour-long Q&A,* organized into the following sections:

- General Information Pages 2-5
- Documentation & Training Page 6
- Performance Issues Pages 7-12
- Installation & Configuration Pages 13-15
- Compatibility Issues Pages 16-19
- Functionality & Usability Issues Pages 20-22

In addition to questions and answers, you'll also find references and links to additional resources provided by Sun.

*Note: The information contained in this transcript, taken directly from a live Sun Expert Exchange event, has been edited for clarity and adherence to trademark guidelines.



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



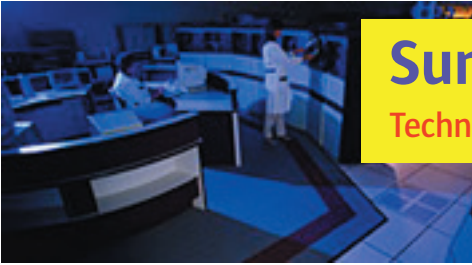
Fast Track to Solaris 10 Adoption: ZFS Technology General Information

1. Who came up with the name ZFS, and what does the “Z” stand for?
2. When is the Solaris 10 OS slated for full release?
3. Is ZFS based on the “Advanced File System,” which is the base for SAM-FS and QFS?
4. What is the package name for ZFS? SUNWZFS?
5. Will ZFS replace the Solaris Volume Manager and Veritas Volume Manager?
6. What are drawbacks in using ZFS, or situations that might not be suited for ZFS?
7. Is ZFS the strategic direction for Solaris filesystems?
8. When you say everything has a 64-bit checksum, what is the unit of data that is check summed?
Block, K, etc?
9. Is it accurate to say that, from a storage hardware perspective, JBOD is looking like the preferred ZFS platform?
10. Any thoughts on porting ZFS to Linux, AIX, or HP-UX?
11. In a container environment, what is ZFS’s role? Is there a ZFS for each container?
12. How about the access control and security features for ZFS?
13. Are there any plans to make ZFS a distributed file system in the future?
14. Is ZFS going to be read/writable when you have booted into single user mode from a cdrom or jumpstart server?
15. How do you limit the storage available to a file system?
16. What’s the thinking on using ZFS inside an N1 Grid Container? ZFS on ZFS — NFS?
17. Do I need to use command mount in Solaris OS with ZFS?
18. Does ZFS have multipathing load balancing i.e., powerpath/veritas multipathing/etc.?
19. Is there a reason SAMFS and QFS are not available at install time? Will they be?
20. How are ZFS and QFS related?
21. Will ZFS replace traditional volume management?
22. When will ZFS be released via Solaris Express?
23. Does ZFS + Samba put Sun in a position to better compete with NAS vendors like NetApp?
24. I recently went to a lecture on the Solaris 10 OS, and one of the engineers said I could download the ZFS package from Sun Software Express. Is this true?
25. Is ZFS available separately? Can I convert from HFS to ZFS on the Solaris 9 OS, for example?
26. Is ZFS in the base Solaris 10 OS, or will it be additional cost?
27. Why go to ZFS? What does it offer compared to UFS and VFS?
28. When will ZFS be included in the Solaris 10 OS? We were told first in late summer 2004, then early 2005, then May 2005.

Q: Who came up with the name ZFS, and what does the “Z” stand for?

A: ZFS was originally the project name where the letter “Z” stood for Zettabyte, which indicated the virtually unlimited scalability of the file system. Now the “Z” no longer stands specifically for Zettabyte. Given the several advanced capabilities offered by this technology, the letter “Z” could stand for the “Zen File system” or “the last alphabet on file systems” or anything else.

Q: When is the Solaris 10 OS slated for full release?



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



A: The current schedule is to ship the Solaris 10 OS at the end of calendar year 2004.

Q: Is ZFS based on the “Advanced File System,” which is the base for SAM-FS and QFS?

A: No, ZFS is not based on “Advanced File System.” ZFS is Sun’s next generation file storage solution, designed from the ground up to meet the modern needs of a general purpose, host-based file system.

Q: What is the package name for ZFS? SUNWZFS?

A: This is yet to be finalized, but internally we’re using SUNWZFS for now.

Q: Will ZFS replace the Solaris Volume Manager and Veritas Volume Manager?

A: In the sense that ZFS does not require a separate volume manager, yes. If you’re still using UFS or VxFS, ZFS will not replace their respective volume managers.

Q: What are drawbacks in using ZFS, or situations that might not be suited for ZFS?

A: ZFS works well in all situations.

Q: Is ZFS the strategic direction for Solaris filesystems?

A: Yes, for Solaris OS.

Q: When you say everything has a 64-bit checksum, what is the unit of data that is checked? Block, K, etc?

A: Each block. In ZFS, we support multiple block sizes where each block can be from 512 bytes to 128K (maybe larger in the future).

Q: Is it accurate to say that, from a storage hardware perspective, JBOD is looking like the preferred ZFS platform?

A: Yes. With access to all disks in a system (not hidden behind a RAID controller), we can provide functionality like self-healing data, priority/deadline scheduling, etc.

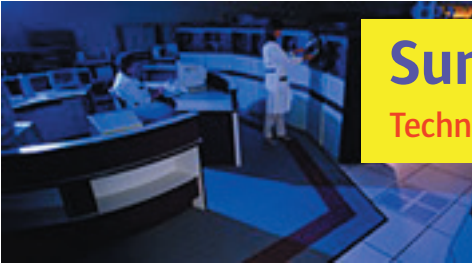
Q: Any thoughts on porting ZFS to Linux, AIX, or HPUX?

A: No plans of porting to AIX and HPUX. Porting to Linux is currently being investigated.

Q: In a container environment, what is ZFS’s role? Is there a ZFS for each container?

A: Basically, the rules for ZFS with containers would be just as with any other file system; shared read-only is doable, multiple writers would be risky. You’d most likely have one storage pool on the system, with separate ZFS filesystems for each container.

Q: How about the access control and security features for ZFS?



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



A: ZFS will support NFSv4-style access control lists (ACLs), in addition to the standard POSIX user/owner/group modes.

Q: Are there any plans to make ZFS a distributed file system in the future?

A: This is currently being investigated.

Q: Is ZFS going to be read/writable when you have booted into single user mode from a cdrom or jumpstart server?

A: Yes.

Q: How do you limit the storage available to a file system?

A: ZFS offers virtually unlimited data scalability. ZFS is a 128-bit file system. It will provide about 16 billion-billion times more capacity than the currently available 32-bit and 64-bit file systems. In other words, ZFS is designed to support more storage, more file systems, more snapshots, more directory entries, and more files than can be created in the foreseeable future.

Q: What's the thinking on using ZFS inside an N1 Grid Container? ZFS on ZFS — NFS?

A: From the container's perspective, ZFS is just another file system; it would be made available and appears just as UFS, QFS, and so on.

Q: Do I need to use command mount in Solaris OS with ZFS?

A: Simply by creating a ZFS filesystem, it will be available in the namespace (i.e., mounted). Of course, you will also be able to turn off this feature and manually use the mount(1m) command.

Q: Does ZFS have multipathing load balancing i.e., powerpath/veritas multipathing/etc.?

A: Yes. ZFS sits on top of multipathing capability MPxIO which is part of the Solaris 10 OS.

Q: Is there a reason SAMFS and QFS are not available at install time? Will they be?

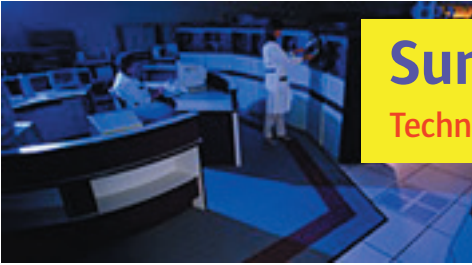
A: Those products are additional cost items on the Sun price list, which is why they aren't available in the core install.

Q: How are ZFS and QFS related?

A: ZFS and QFS are separate products; although there's overlapping functionality, there's no shared code, for example. We see there being different places where each will play; this is briefly touched upon in today's sun.com cover story. Expect more guidance and positioning of different filesystem technologies as ZFS gets closer to release.

Q: Will ZFS replace traditional volume management?

A: In a word, yes. ZFS incorporates the functions typically handled by separate volume manager software, which makes for simplified system administration.



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Q: When will ZFS be released via Solaris Express?

A: We are working hard to get ZFS into a build of Solaris Express prior to our general ship of the Solaris 10 OS, but that schedule is still not firmed up. Please join the Solaris Express program and stay tuned.

Q: Does ZFS + Samba put Sun in a position to better compete with NAS vendors like NetApp?

A: Yes, ZFS will drastically improve ease of management as well as performance. Among other things, this will improve Sun's position in the NAS space (both with NFS and Samba).

Q: I recently went to a lecture on the Solaris 10 OS, and one of the engineers said I could download the ZFS package from Sun Software Express. Is this true?

A: ZFS is not in Software Express at the current time.

Q: Is ZFS available separately? Can I convert from HFS to ZFS on the Solaris 9 OS, for example?

A: There are no current plans to offer ZFS on previous versions of the Solaris OS. You will need to install the Solaris 10 OS.

Q: Is ZFS in the base Solaris 10 OS, or will it be additional cost?

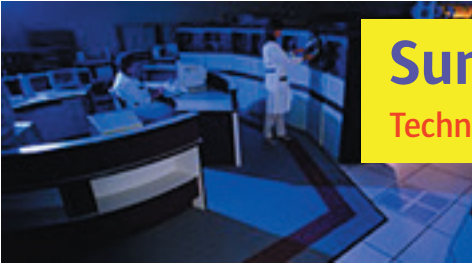
A: There are no current plans to charge for ZFS.

Q: Why go to ZFS? What does it offer compared to UFS and VFS?

A: ZFS offers a number of enhancements compared to traditional filesystems. You can find an overview on the sun.com feature story.

Q: When will ZFS be included in the Solaris 10 OS? We were told first in late summer 2004, then early 2005, then May 2005.

A: ZFS will be in the Solaris 10 OS when we ship the product. The current projection is the end of 2004.



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Fast Track to Solaris 10 Adoption: ZFS Technology Documentation & Training

1. Can you walk us through the command line to create a ZFS file system? I am curious what the command line is to implement a RAID1 layout?
2. This will be a big selling feature for moving to the Solaris 10 OS. When/where will there be detailed information I can present to my managers?
3. Will information about storage on ZFS be available through kstat?
4. Is there any literature that explains in detail the performance enhancements of the Solaris 10 OS?
5. The only mention of ZFS on the net so far seems to be marketing material. Are there real docs online somewhere now? If so, where? If not, when?

Q: Can you walk us through the command line to create a ZFS file system? I am curious what the command line is to implement a RAID1 layout?

A: We haven't finalized the exact command line syntax, but you can expect to do something like:
zpool create pool_name disk1 disk2 #create a ZFS storage pool with two striped disks
ZFS create pool_name/fs_name #create a ZFS filesystem
You would use a different "zpool create" command to make a mirrored pool. You would do something like: zpool create pool_name mirror disk1 disk2

Q: This will be a big selling feature for moving to the Solaris 10 OS. When/where will there be detailed information I can present to my managers?

A: On Sept. 15, 2004, we did a major content refresh on our Solaris 10 OS site (www.sun.com/solaris/10). We have content there for both business and technical audiences.

Q: Will information about storage on ZFS be available through kstat?

A: There will be extensive run-time performance data available from ZFS. We're investigating how best to provide this — kstat and DTrace are two (not mutually exclusive) possibilities.

Q: Is there any literature that explains in detail the performance enhancements of the Solaris 10 OS?

A: Check our Solaris 10 OS Web site (www.sun.com/solaris/10) for content including details on DTrace and other areas where we get performance boosts. Benchmarks and specific details will be available when we ship the Solaris 10 OS at the end of 2004.

Q: The only mention of ZFS on the net so far seems to be marketing material. Are there real docs online somewhere now? If so, where? If not, when?

A: We're working on publishing more in-depth documentation. For now, you can find some details on my Weblog at <http://blogs.sun.com/ahrens>.



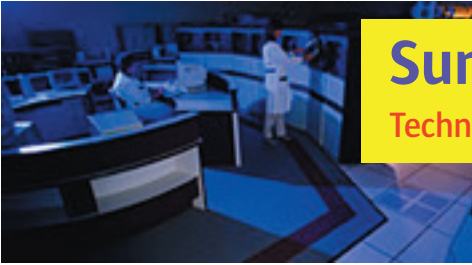
Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Fast Track to Solaris 10 Adoption: ZFS Technology Performance Issues

1. What is HA Storage+?
2. How does MPXIO stack up against DMP and powerpath?
3. The sun.com front-page story mentions "Explicit I/O priority with deadline scheduling." Does this mean you can control disk I/O priorities at the process-level, e.g., to tame disk hogging or runaway processes?
4. Re: finding disks based on their content rather than their path name, does this mean that devices become path agnostic? For example, if c1t3d0 moves to c2t3d0, will ZFS find it and do the right thing?
5. Under ZFS (pooled storage), could one identify which physical disk(s) a file resides on?
6. Will ZFS have ZFSdump / flash / snapshot capabilities?
7. ZFS may greatly simplify administration, but what kind of performance does it give?
8. VxFS grows file systems quickly, but takes a long time to shrink them. Is the same true with ZFS?
9. How does ZFS handle storage hotspots?
10. There has been much file system work going on in the Linux world with the release of ReiserFS v4. Looking at ZFS and reiserfs, does ZFS support atomic commits? What about extensibility/ plugins for enhancing ZFS? Also, what about security? Does ZFS help me in these aspects? Finally can it deal with millions of small files efficiently while still handling large gigabytes files?
11. How does ZFS manage blocks?
12. Given that all storage is in a single pool, the pool is raw storage and file systems are sharing space and bandwidth; does this mean that different filesystems can be "interleaved" on the same disk?
13. Some storage vendors have the concept of alternate or dynamic pathing; is this feature included in ZFS?
14. Will there be an SNMP trap daemon for problems similar to mdlogd?
15. How does the QOS in ZFS work? Can you specify a default QOS for the creation of new volumes (as opposed to current volume managers having to specify RAID5/Striping)?
16. With current volume managers, unless you micromanage them, after multiple growths of volumes/ filesystems, you may wind up with a fragmented disk/LUN where there are little chunks of your filesystem spread all across your disk/LUN. Does ZFS handle this situation automatically?
17. What do you say about striping in ZFS?
18. Does each host have its own storage pool, or are multiple hosts sharing the same storage pool?
19. What administration task should someone expect to be running most often once ZFS is fully implemented on a system?
20. Can I evacuate a disk LUN from a ZFS file system if there is still room available with the LUN missing, e.g., migrate from one set of disks to another set while running?
21. How often are checks of ondisk data executed? The documentation indicates self-healing capability, which seems to imply active consistency checking.
22. In the event of disaster recovery, how does ZFS help me sort things out as I bring a system back online?
23. Will ZFS commands be integrated into Solaris Volume manager or will it be a completely separate tool?
24. How is alternate pathing handled?
25. Will you be able to import/export/split storage pools, in order to move the entire pool or merely specific filesystems to other systems?



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



26. What kind of performance hit will there be with ZFS vs. UFS vs. SDS vs. StorEdge systems?
27. Is ZFS redundancy handled via mirroring at the block level? For example, is it like RAID1, where usable space equals 50 percent of raw space? Or is something like RAID5 possible, where loss is less?
28. How does ZFS handle a disk failure in a pool ?
29. The currently available docs state that ZFS will detect corrupt blocks in mirrored or RAID configurations. Without a volume manager, how does ZFS know about mirrors and raid storage?
30. There are advanced features in ZFS that are outside the standard UFS attributes (e.g., mirroring and snapshots). Will there be a “ZFSdump” (or other mechanism) that will preserve such information across backups?
31. How is ZFS going to reduce administration task time? Is it based on a specific technology that will, for example, have fsck run faster? Even with the storage pool, data corruption can happen.
32. Are ZFS file systems shrinkable? How about fragmentation? Any need to defrag them?
33. I have a workgroup of two Ultra60s, an UltraSPARC laptop, and four PCs. All machines are networked and drives mounted. Will ZFS allow me to treat all storage as a single pool?
34. Does ZFS still use inodes?
35. Is ZFS bootable, and if not, when will it be?

Q: What is HA Storage+?

A: This is a level of cluster compatibility. In order to satisfy this, the storage has to be available from both servers in the cluster, and only one server at a time can be reading/writing the data. In failover, it has to happen such that no data is lost. This is what we currently provide in ZFS. Higher levels provide things like simultaneous access (active-active) from both servers in the cluster. We don't provide this, but we're investigating it.

Q: How does MPXIO stack up against DMP and powerpath?

A: Quite well, but this is not a ZFS-specific question. Please look for MPxIO and its documentation/marketing materials on the sun.com site.

Q: The sun.com front-page story mentions “Explicit I/O priority with deadline scheduling.” Does this mean you can control disk I/O priorities at the process-level, e.g., to tame disk hogging or runaway processes?

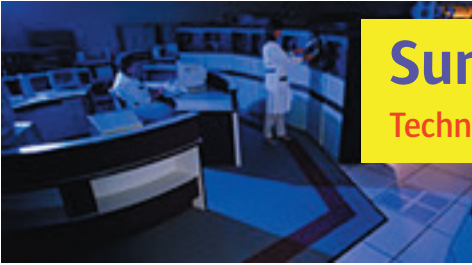
A: Yes, the technology is there, and we have been able to demonstrate excellent behavior in the face of situations like the one you describe. We are currently trying to figure out how to let the user/administrator specify adjustments to the priority/deadline on a process/user/zone basis.

Q: Re: finding disks based on their content rather than their path name, does this mean that devices become path agnostic? For example, if c1t3d0 moves to c2t3d0, will ZFS find it and do the right thing?

A: Absolutely!

Q: Under ZFS (pooled storage), could one identify which physical disk(s) a file resides on?

A: We have no command that will tell you this. However, you could figure it out using DTrace and the I/O provider and seeing which disks get I/O traffic from a given file.



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Q: Will ZFS have ZFSdump / flash / snapshot capabilities?

A: ZFS supports an unlimited number of snapshots, which can be created in constant time. It also supports writable snapshots.

Q: ZFS may greatly simplify administration, but what kind of performance does it give?

A: Performance is very important for ZFS. You may know that with the Solaris 9 OS, UFS with logging was significantly faster than VxFS 3.5. A benchmark paper is published on this. Now with ZFS, Sun is ready to launch its next-generation file system and, once again, performance will be a key attribute. Currently, we do not have the numbers. Please stay tuned for information on ZFS performance numbers.

Q: VxFS grows file systems quickly, but takes a long time to shrink them. Is the same true with ZFS?

A: Short answer: yes. Long answer: it depends. If the disks that you remove have lots of data on them, and the data is not replicated on other disks in the pool, then we have to copy that data onto disks that are not being removed. This takes time (there is no magic). If there is little data on the disks, it won't take long.

Q: How does ZFS handle storage hotspots?

A: Being a copy-on-write filesystem, ZFS is much less susceptible to hot spots than traditional, statically-laid-out filesystems.

Q: There has been much file system work going on in the Linux world with the release of ReiserFS v4. Looking at ZFS and reiserfs, does ZFS support atomic commits? What about extensibility/plugins for enhancing ZFS? Also, what about security? Does ZFS help me in these aspects? Finally can it deal with millions of small files efficiently while still handling large gigabytes files?

A: ZFS will perform well with many small files as well as extremely large files. We'll be looking in to exposing some of the internal modularity of ZFS in future releases.

Q: How does ZFS manage blocks?

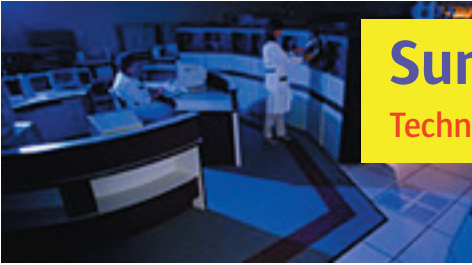
A: We have our own algorithms that minimize fragmentation and maximize performance.

Q: Given that all storage is in a single pool, the pool is raw storage and file systems are sharing space and bandwidth; does this mean that different filesystems can be "interleaved" on the same disk?

A: Yes. You can also have multiple pools on the same system.

Q: Some storage vendors have the concept of alternate or dynamic pathing; is this feature included in ZFS?

A: This is handled by MPxIO, which is part of the Solaris 10 OS. So, ZFS will transparently handle multipathing (active-active).



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Q: Will there be an SNMP trap daemon for problems similar to mdlogd?

A: Yes, eventually. We're not sure if it will be in the first release.

Q: How does the QOS in ZFS work? Can you specify a default QOS for the creation of new volumes (as opposed to current volume managers having to specify RAID5/Striping)?

A: Currently, each pool has its own redundancy, as specified by the administrator. We are in the process of designing functionality that will allow you to have multiple types of redundancy in a single storage pool (for example, both mirroring and RAID-5 in a single pool). This may or may not make it into the initial release.

Q: With current volume managers, unless you micromanage them, after multiple growths of volumes/filesystems, you may wind up with a fragmented disk/LUN where there are little chunks of your filesystem spread all across your disk/LUN. Does ZFS handle this situation automatically?

A: Yes, with ZFS you can create as many filesystems as you like, drawing their storage from a single pool. Filesystems always use exactly as much space as they need, so there's no need to manually grow or shrink them.

Q: What do you say about striping in ZFS?

A: ZFS implements dynamic striping, which provides better performance and more flexibility than the static striping (RAID0) that volume managers use.

Q: Does each host have its own storage pool, or are multiple hosts sharing the same storage pool?

A: ZFS is a local filesystem; thus each host would have its own storage pool. You can use NFS to access the storage pool from remote hosts.

Q: What administration task should someone expect to be running most often once ZFS is fully implemented on a system?

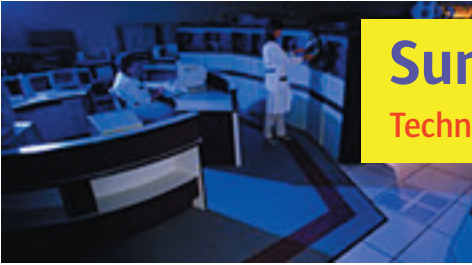
A: That depends a lot on what applications you're using with ZFS. If you're using it on a home directory server, I imagine you'd most frequently be running the "ZFS create" command to create new ZFS filesystems for new users. You would occasionally run the "zpool add" command to add more storage (e.g., disks) to the pool.

Q: Can I evacuate a disk LUN from a ZFS file system if there is still room available with the LUN missing, e.g., migrate from one set of disks to another set while running?

A: Yes, that's fully supported.

Q: How often are checks of ondisk data executed? The documentation indicates self-healing capability, which seems to imply active consistency checking.

A: With checksums, we check data integrity on every I/O. For redundant configurations, we only read



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



data from one location, but we will provide a background “disk scrubber” that you can run to validate data continuously.

Q: In the event of disaster recovery, how does ZFS help me sort things out as I bring a system back online?

A: In the event of a disaster, ZFS will be able to find any disks that are part of a storage pool, based on their content rather than their path name.

Q: Will ZFS commands be integrated into Solaris Volume manager or will it be a completely separate tool?

A: It will be a completely separate tool since ZFS includes its own functionality that is a superset of traditional volume managers.

Q: How is alternate pathing handled?

A: Alternate pathing is handled through MPXIO, which is a standard feature in the Solaris 10 OS.

Q: Will you be able to import/export/split storage pools, in order to move the entire pool or merely specific filesystems to other systems?

A: We can import/export entire pools between machines (and even between SPARC and x86 platforms) with incredible simplicity (one command). All filesystems that are part of that storage pool are migrated with it automatically.

Q: What kind of performance hit will there be with ZFS vs. UFS vs. SDS vs. StorEdge systems?

A: There should be no performance hit, only performance gains. The architecture in ZFS removes just about all constraints on I/O order, which allows us to run the disks much closer to full I/O capacity than current filesystems/volume managers.

Q: Is ZFS redundancy handled via mirroring at the block level? For example, is it like RAID1, where usable space equals 50 percent of raw space? Or is something like RAID5 possible, where loss is less?

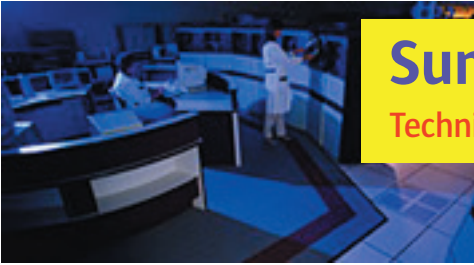
A: ZFS will provide mirroring as well as RAID5-like redundancy.

Q: How does ZFS handle a disk failure in a pool ?

A: If a disk fails, we recover the data if possible and return EIO if not.

Q: The currently available docs state that ZFS will detect corrupt blocks in mirrored or RAID configurations. Without a volume manager, how does ZFS know about mirrors and raid storage?

A: ZFS incorporates the functionality of a volume manager, thus it implements mirrors, striping, and raid itself.



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Q: There are advanced features in ZFS that are outside the standard UFS attributes (e.g., mirroring and snapshots). Will there be a “ZFSdump” (or other mechanism) that will preserve such information across backups?

A: Complicated storage administration concepts automated and consolidated into straightforward language, reducing administrative overhead by up to 80 percent; All data is protected by 64-bit checksums, resulting in 99.9999999999999999 percent error detection and correction; 128-bit file system; 16 billion billion times the capacity of 32- or 64-bit file systems; Copy-on-write eliminates the need for fsck or other recovery mechanisms ZFS is POSIX compliant. Hence, applications run without modification including backups.

Q: How is ZFS going to reduce administration task time? Is it based on a specific technology that will, for example, have fsck run faster? Even with the storage pool, data corruption can happen.

A: It should never be necessary to run fsck, but the more significant administrative speedups actually come from the simplification of functions such as filesystem allocation, resizing, and so on. Also, ZFS has features that make recovery of accidentally-deleted files much easier than we see in other comparable technologies.

Q: Are ZFS file systems shrinkable? How about fragmentation? Any need to defrag them?

A: ZFS file systems can be dynamically resized; they can grow or shrink as needed. The allocation algorithms are such that defragmentation is not an issue.

Q: I have a workgroup of two Ultra60s, an UltraSPARC laptop, and four PCs. All machines are networked and drives mounted. Will ZFS allow me to treat all storage as a single pool?

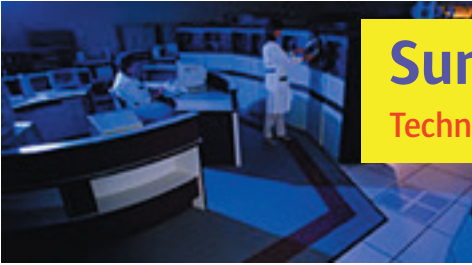
A: If all disks are available locally to one machine (e.g., cXtXdX), then, yes. ZFS is not a distributed filesystem that can access raw storage over the network.

Q: Does ZFS still use inodes?

A: Like all filesystems, ZFS does have an on-disk structure which represents and keeps track of each file. However, unlike UFS, this structure is dynamically allocated.

Q: Is ZFS bootable, and if not, when will it be?

A: It won't be in the first ZFS release; look for it in a Solaris 10 OS update.



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Fast Track to Solaris 10 Adoption: ZFS Technology Installation & Configuration

1. What types of QOS features does ZFS have in terms of data availability? Can it be configured with a minimum service time for requests?
2. If an existing Solaris 8 OS server is upgraded to the Solaris 10 OS using the install upgrade process, will it be possible to include a migration to ZFS at that time?
3. Will I be able to upgrade a server running the Solaris 9 OS to the Solaris 10 OS with ZFS? If so, what are some of the caveats?
4. How do I configure swap for ZFS?
5. How does ZFS set up? How does it choose RAID type for instance?
6. Will it be possible for me to easily migrate my VxVM volumes to ZFS? If so, how?
7. If you configure ZFS to a single pool, would you be able to use part of that pool as additional swap space when the need arises?
8. Can ZFS be set up to prefer certain devices over others in its storage pool? If so, is this set at the pool or filesystem level?
9. In reading the available information on ZFS and your visions of it, I get a little confused about how I should be planning my storage systems if I plan to use ZFS. If you were planning about 6 TB of storage, what would be most in line with the ZFS thinking and why?
10. What are the tuning parameters you're offering initially? Anything like direct I/O, concurrent I/O, read/write release behind, etc. mount options?
11. Is storage grouped into a small number of storage pools with intuitive names? Is each storage pool represents QOS type?
12. What kind of code is managing the "storage pool"? Would it become a single point of failure? In traditional LVM, one volume failure doesn't affect others. What happens when "storage pool manager" has a problem?
13. What conversion tools if any will be needed?

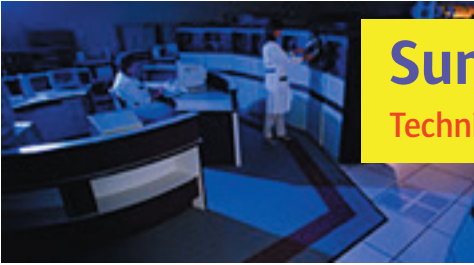
Q: What types of QOS features does ZFS have in terms of data availability? Can it be configured with a minimum service time for requests?

A: We have the ability to inherit a priority for I/Os, which are then taken into account when doing our disks I/O scheduling. This scheduling takes into account both deadlines and priority with the goal of minimizing overall system latency and maximizing throughput to the disks. We are investigating what kind of interface to provide to the administrator to further control the priority given to I/O requests from an application/user/zone.

Q: If an existing Solaris 8 OS server is upgraded to the Solaris 10 OS using the install upgrade process, will it be possible to include a migration to ZFS at that time?

A: ZFS can be added to the system as part of the upgrade, but existing UFS filesystems will remain UFS. Data migration tools are planned, but will not be available when ZFS first ships.

Q: Will I be able to upgrade a server running the Solaris 9 OS to the Solaris 10 OS with ZFS? If so, what are some of the caveats?



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



A: To migrate your storage from UFS to ZFS, you will need to do a backup (using a POSIX-compatible backup utility) from the UFS filesystem and then restore onto ZFS.

Q: How do I configure swap for ZFS?

A: We have not published those processes yet; this will be described in the ZFS documentation when we make that feature available in a Solaris 10 OS build.

Q: How does ZFS set up? How does it choose RAID type for instance?

A: The administrator specifies what type of RAID they want. We will provide control such that you can either have detailed control over which disks are replicated with which, or you can just have a “bag of disks,” and we will do our best to match disks up based on the administrator’s request.

Q: Will it be possible for me to easily migrate my VxVM volumes to ZFS? If so, how?

A: Right now, it’s dump and restore; migration tools are on the project list, but won’t be there until after ZFS first ships.

Q: If you configure ZFS to a single pool, would you be able to use part of that pool as additional swap space when the need arises?

A: Yes

Q: Can ZFS be set up to prefer certain devices over others in its storage pool? If so, is this set at the pool or filesystem level?

A: The filesystem/pool interface follows a malloc/free paradigm. Any such control would be at the pool level. The device preference is currently controlled by our adaptive algorithms. We are exploring ways to let the administrator give us “hints” to tune these algorithms, but we have not finalized what will be in the final version.

Q: In reading the available information on ZFS and your visions of it, I get a little confused about how I should be planning my storage systems if I plan to use ZFS. If you were planning about 6 TB of storage, what would be most in line with the ZFS thinking and why?

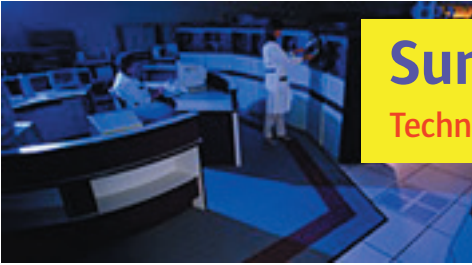
A: In most situations, we configure all storage into a single pool. This makes it simplest for the administrator and allows all filesystems on the system to share both space and bandwidth.

Q: What are the tuning parameters you’re offering initially? Anything like direct I/O, concurrent I/O, read/write release behind, etc. mount options?

A: Our goal is to make such tuning parameters unnecessary — ZFS should be able to obtain optimal performance with a minimum of “wacky knobs.”

Q: Is storage grouped into a small number of storage pools with intuitive names? Is each storage pool represents QOS type?

A: ZFS is quite flexible, and you can configure your system that way if you wish. We can support any number of storage pools, and if you want to configure each pool according to QOS, we support such



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



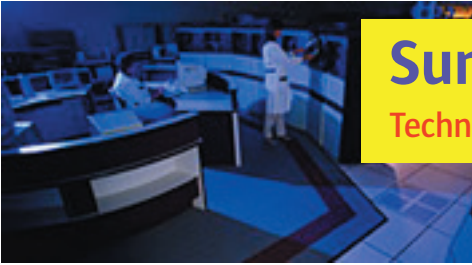
a configuration. And each pool (as well as each filesystem) has a name that you pick, so they are as intuitive as you wish to make them.

Q: What kind of code is managing the “storage pool”? Would it become a single point of failure? In traditional LVM, one volume failure doesn’t affect others. What happens when “storage pool manager” has a problem?

A: It all depends on how you configure your storage pool. If you configure it such that you have redundant copies of your data, we can recover in the case of a failure. As long as at least one copy of your data is online, we can service any requests to it. If you’re referring to a software failure (bug, panic, etc.), then that may cause your system to reboot, just like a failure in traditional LVM kernel code.

Q: What conversion tools if any will be needed?

A: If you want to convert your data from an existing filesystem to ZFS, you will have to copy the data from your existing filesystem over to ZFS. At some future date, we plan on providing a mechanism to migrate your data to ZFS online.



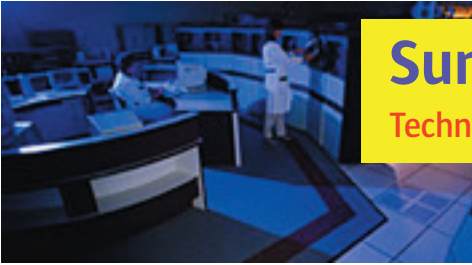
Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Fast Track to Solaris 10 Adoption: ZFS Technology Compatibility Issues

1. Will ZFS be “open” enough that my Windows or Linux OS on the same box will be able to at least mount ZFS in read-only mode?
2. Where can I find the matrix for Hardware support?
3. How will ZFS work with a SAN? Any potential problems among different vendor solutions?
4. But there are no plans to back-port ZFS to the Solaris 8 or 9 operating systems through kernel Updates, right?
5. In Solaris Volume Manager for the Solaris 9 OS you can build replicas for pools; should you use replicas state database for the Solaris 10 OS (ZFS)?
6. How will ZFS play with Samba?
7. Trying to connect some numbers: 128-bit file system, 64-bit checksum, and 64-bit processor. Does 64-bit processor limit the checksum size? Will a 128-bit processor be required to take full advantage of ZFS?
8. My production systems run Oracle 9 on NetApp filers. How would you compare ZFS and WAFL? Some concepts seem similar, like copy-on-write.
9. Are quotas supported on ZFS ?
10. Can I convert UFS and VXFS to ZFS ?
11. Will ZFS support quotas?
12. Does ZFS support iSCSI?
13. Are you working at all with Veritas to add this to the NetBackup application — or other ISVs — for use in backup-to-disk? Or with Solstice backup, for that matter?
14. Will Live Upgrade be integrated so I can convert my file systems from UFS to ZFS as part of the upgrade process?
15. Will ZFS support Oracle data files? Is there “quick I/O” or ODM support?
16. Will ZFS work on EOL hardware such as the Ultra 5?
17. Will ZFS be cluster aware? How will storage pools be migrated? Will Sun cluster and VCS support this at a future time?
18. When will Oracle certify ZFS?
19. Are you working with major database and other application vendors, such as Oracle, to make sure they will work out of the box? Is there going to be any benchmark like TPC-C/D with Oracle when the Solaris 10 OS is released?
20. Can UFS work with ZFS?
21. Will ZFS work with Sun Cluster on a shared disk ?
22. I would like to know whether there is a RAC Testing-Scenario based on FireWire-shared-disk on the roadmap for x86 Solaris OS and Oracle. This would provide the opportunity to test database clusters without expensive costs for Fiber Channel Hardware.
23. Can we expect any new I/O tools (such as iostat) to be introduced with ZFS or just updates to the current tools?
24. Does ZFS provided mirroring/RAID5 capabilities in the storage pool, or is it designed to run on top of SEVM/VXVM?
25. Is ZFS hardware specific? Does it only run on Sun storage or can it be run on EMC, etc.?
26. What volume manager tool works with ZFS?
27. Will or does ZFS support multi-reader, multi-writer type file systems through a SAN?



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Q: Will ZFS be “open” enough that my Windows or Linux OS on the same box will be able to at least mount ZFS in read-only mode?

A: In our initial release, we will not have support for other operating systems. We have thought about providing this, though.

Q: Where can I find the matrix for Hardware support?

A: Go to the Software Express Web site, click the “get the software” link, and look at the “hardware requirements” section at the bottom of that page.

Q: How will ZFS work with a SAN? Any potential problems among different vendor solutions?

A: ZFS will work just fine in a SAN.

Q: But there are no plans to back-port ZFS to the Solaris 8 or 9 operating systems through kernel Updates, right?

A: That is correct.

Q: In Solaris Volume Manager for the Solaris 9 OS you can build replicas for pools; should you use replicas state database for the Solaris 10 OS (ZFS)?

A: With ZFS, there is no separate state database. All states in ZFS are maintained internally with an appropriate level of replication.

Q: How will ZFS play with Samba?

A: Samba can be used to share ZFS data just as with any local filesystem.

Q: Trying to connect some numbers: 128-bit file system, 64-bit checksum, and 64-bit processor. Does 64-bit processor limit the checksum size? Will a 128-bit processor be required to take full advantage of ZFS?

A: All these values are independent of the processor size. ZFS will work on any system supported by Solaris 10 system, which includes Pentium, AMD64 (both 32- and 64-bit mode), UltraSPARC and SPARC64.

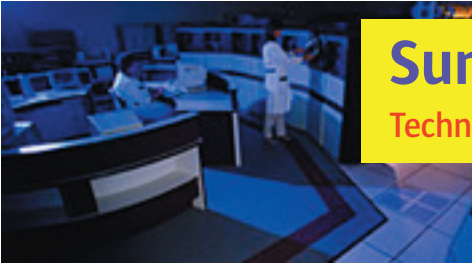
Q: My production systems run Oracle 9 on NetApp filers. How would you compare ZFS and WAFL? Some concepts seem similar, like copy-on-write.

A: One difference is that ZFS runs on the host, so Oracle doesn't need to go over the network to get to a NetApp box.

Q: Are quotas supported on ZFS ?

A: Yes.

Q: Can I convert UFS and VXFS to ZFS ?



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



A: Right now, it'd be via dump and restore. Migration tools are on the roadmap for sometime after the initial ZFS release.

Q: Will ZFS support quotas?

A: Yes, ZFS will support setting both a quota and reservation on each filesystem.

Q: Does ZFS support iSCSI?

A: Yes.

Q: Are you working at all with Veritas to add this to the NetBackup application — or other ISVs — for use in backup-to-disk? Or with Solstice backup, for that matter?

A: We are working with all the vendors who support Solaris OS platform.

Q: Will Live Upgrade be integrated so I can convert my file systems from UFS to ZFS as part of the upgrade process?

A: It's under investigation, but it won't be part of the initial release.

Q: Will ZFS support Oracle data files? Is there “quick I/O” or ODM support?

A: We will support Oracle data files. Also, we are trying to make every I/O a “quick” I/O. We don't believe that an application should have to play special tricks or give us special flags to get full performance out of their system, and we are currently benchmarking with Oracle to make sure this will be true.

Q: Will ZFS work on EOL hardware such as the Ultra 5?

A: ZFS will work on any system supported by the Solaris 10 OS, which does include the Ultra 5.

Q: Will ZFS be cluster aware? How will storage pools be migrated? Will Sun cluster and VCS support this at a future time?

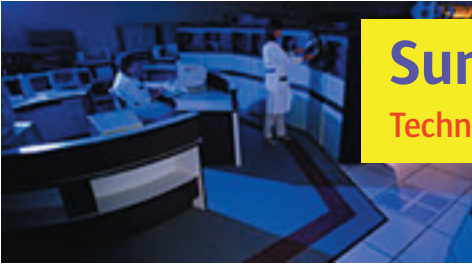
A: Currently, we support HA storage+, which means that ZFS can use shared disks, as long as only a single machine is using the pool at a time. We are investigating making it more aware of a cluster environment to relax this constraint in the future.

Q: When will Oracle certify ZFS?

A: Oracle has already publicly stated their support for the Solaris 10 OS; details as to when this will include specific features such as ZFS are TBD. I'd recommend checking with Oracle on this — among other things, it gives them reinforced feedback that customers are looking for this from them.

Q: Are you working with major database and other application vendors, such as Oracle, to make sure they will work out of the box? Is there going to be any benchmark like TPC-C/D with Oracle when the Solaris 10 OS is released?

A: Yes, we are actively working with the database vendors on certification and benchmarking. Oracle



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



has already stated their intent to support the Solaris 10 OS.

Q: Can UFS work with ZFS?

A: UFS and ZFS can co-exist on the same system.

Q: Will ZFS work with Sun Cluster on a shared disk ?

A: ZFS will support HA Storage+. This allows a storage pool that is connected to multiple hosts to “fail over” to a good host. Only one host can access the storage pool at a time.

Q: I would like to know whether there is a RAC Testing-Scenario based on FireWire-shared-disk on the roadmap for x86 Solaris OS and Oracle. This would provide the opportunity to test database clusters without expensive costs for Fiber Channel Hardware.

A: The Solaris 10 OS supports FireWire disks, and ZFS works fine on top of them. If you wish to configure Oracle on top of such a configuration, that would be supported.

Q: Can we expect any new I/O tools (such as iostat) to be introduced with ZFS or just updates to the current tools?

A: All the current tools report I/O information from ZFS. In addition, we have “zpool iostat,” which reports more detailed and ZFS-specific information.

Q: Does ZFS provided mirroring/RAID5 capabilities in the storage pool, or is it designed to run on top of SEVM/VXVM?

A: ZFS incorporates the capabilities of volume managers — e.g., mirroring, striping, raid. Thus a volume manager (e.g., SVM, VxVM) is not necessary.

Q: Is ZFS hardware specific? Does it only run on Sun storage or can it be run on EMC, etc.?

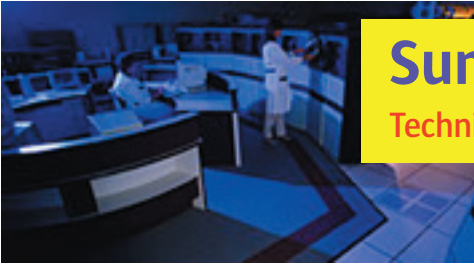
A: ZFS works on any underlying storage that the Solaris 10 OS is accessing. It works best on raw storage because we can apply our Predictive Self-Healing functions to it.

Q: What volume manager tool works with ZFS?

A: ZFS has its own built-in functionality that replaces traditional volume managers.

Q: Will or does ZFS support multi-reader, multi-writer type file systems through a SAN?

A: Not in the current release. We are looking at ways in which we could do this in the future, but it will not be in the Solaris 10 OS.



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Fast Track to Solaris 10 Adoption: ZFS Technology Functionality & Usability Issues

1. Can you have separate pools per system?
2. I'm still confused. Suppose I have three E450s. Would ZFS allow me to integrate storage across all three boxes into one big "poor man's SAN"?
3. Will a GUI be available for the administrative functions?
4. Can disks be designated as spares to be used only in case of disk failure, or is it like RAID5 where a disk failure results in no data loss, but lessened performance?
5. I hope this is relevant. Using SVM on the Solaris 9 OS, all disk objects are named with digits. Will we be able to use text labels in the Solaris 10 OS?
6. Can users see into the snapshots, and can they copy files out to perform their own "restores"?
7. Are there caps on how large a filesystem can grow? I would hate to see one filesystem run away and not only fill itself, but also eat all the other disk space connected to the system.
8. Can I use the command `vmstat` to analyze performance of the disk in the Solaris 10 OS?
9. In a shared storage type environment how would I move a ZFS volume from one host to another? Like exporting a Veritas disk group?
10. Will ZFS have a backup command like `ufsdump`?
11. Is there a GUI to manage the storage pool?
12. Are there any limitations or special considerations concerning ZFS and x86 fdisk partitions? Does ZFS exist in an x86 fdisk partition so it is possible to have ZFS, NTFS, FAT, and ext3 filesystems all on the same disk?
13. What degree of control does an administrator have over how the pool is allocated? For example, can I direct certain filesystems to stay on certain hardware devices?
14. What happens when I need to grow/shrink a ZFS?
15. Where is the storage pool managed? At the host? On the arrays? Somewhere in-between?

Q: Can you have separate pools per system?

A: Yes, you can have multiple pools attached to one host. Typically though, you'd only need one.

Q: I'm still confused. Suppose I have three E450s. Would ZFS allow me to integrate storage across all three boxes into one big "poor man's SAN"?

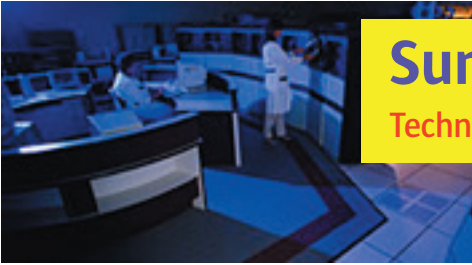
A: No, ZFS is a local filesystem (for the time being). To access storage attached to a different host, use NFS.

Q: Will a GUI be available for the administrative functions?

A: GUI will not be available in the initial release. It will be available later. But administration with ZFS is so simple that all one needs to do is state the intent and ZFS will take care of the rest.

Q: Can disks be designated as spares to be used only in case of disk failure, or is it like RAID5 where a disk failure results in no data loss, but lessened performance?

A: with ZFS, you'll be able to designate "hot space," which is reserved for use in the case of a disk



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



failure. Rather than leaving an entire disk unused, thus reducing performance, ZFS reserves a little bit of each disk in the pool.

Q: I hope this is relevant. Using SVM on the Solaris 9 OS, all disk objects are named with digits. Will we be able to use text labels in the Solaris 10 OS?

A: It is on the roadmap.

Q: Can users see into the snapshots, and can they copy files out to perform their own “restores”?

A: Yes.

Q: Are there caps on how large a filesystem can grow? I would hate to see one filesystem run away and not only fill itself, but also eat all the other disk space connected to the system.

A: Short answer: yes. Each filesystem has a hard cap, which can be dynamically changed by the administrator.

Q: Can I use the command vmstat to analyze performance of the disk in the Solaris 10 OS?

A: Yes; in fact with DTrace in the Solaris 10 OS, you can do more advanced analysis than with vmstat.

Q: In a shared storage type environment how would I move a ZFS volume from one host to another? Like exporting a Veritas disk group?

A: You can use HA Storage+ to enable automatic fail-over. Additionally, you can easily export the pool from one host and import it to another without re-cabling if it is attached to multiple hosts (e.g., in a SAN).

Q: Will ZFS have a backup command like ufsdump?

A: You'd back up ZFS at a higher level, with any POSIX-compliant backup utility.

Q: Is there a GUI to manage the storage pool?

A: Not at this time, but it's simple enough that there doesn't appear to be a need for it.

Q: Are there any limitations or special considerations concerning ZFS and x86 fdisk partitions? Does ZFS exist in an x86 fdisk partition so it is possible to have ZFS, NTFS, FAT, and ext3 filesystems all on the same disk?

A: ZFS will recognize any style partitions that the Solaris OS does. That said, we greatly prefer EFI labels because they support large disks and partitions and are much better designed than most other disk labeling schemes.

Q: What degree of control does an administrator have over how the pool is allocated? For example, can I direct certain filesystems to stay on certain hardware devices?



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



A: With the ZFS pooled storage model, all filesystems in a pool draw storage from all disks in the pool. This way, ZFS can make optimal use of all available disk bandwidth.

Q: What happens when I need to grow/shrink a ZFS?

A: It's extremely simple: issue a single command that changes the size of the filesystem.

Q: Where is the storage pool managed? At the host? On the arrays? Somewhere in-between?

A: The storage pool is managed in the Solaris OS, using simple, straightforward command-line tools.