

Sun Expert Exchange

“Fast Track to Solaris 10 Adoption: ZFS Technology”

Wednesday, September 15, 2004

Bob (Q): Several mentions were made to HA Storage+. What is this?

Bill Moore (A): This basically 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 currently provide this, but are investigating it.

Ryan (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?

Matt Ahrens (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

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

Bill Moore (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.

mark (Q): Can you have separate pool's per system?

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

Fazal Majid (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)?

Bill Moore (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.

TrogL (Q): I'm still confused. Suppose I have three E450's. Would ZFS allow me to integrate storage across all three boxes into one big "poor man's SAN"?

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

drsolaris8 (Q): What is the package name for zfs? SUNWzfs?

Matt Ahrens (A): This is yet to be finalized, but internally we're using SUNWzfs for now.

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

Larry Wake (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.

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

Smita Thakur (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.

Tao (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?

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

mlf (Q): Re: finding disks based on their content rather than their path name. Does this mean that devices become path agnostic? E.g. If c1t3d0 moves to c2t3d0, will ZFS find it and do the right thing?

Matt Ahrens (A): Absolutely!

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

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

walterg (Q): Will ZFS have zfsdump / flash / snapshot capabilities?

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

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

Smita Thakur (A): Performance is very important for ZFS. You may know that with Solaris 9, 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.

drsolaris8 (Q): Where can I find the matrix for Hardware support?

A: See the Solaris Express website: <http://sun.com/solaris-express/> -- click on "get the software" link and look at the Hardware Requirements section at the bottom of that page.

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

Matt Ahrens (A): ZFS will work just fine in a SAN.

Ben (Q): VxFS grows file systems quickly but take a long time to shrink file systems. Is the same true with ZFS?

Bill Moore (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.

lance (Q): How does ZFS handle storage hotspots?

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

DebK (Q): But there are no plans to back-port ZFS to Solaris 8 or 9 through Kernel Updates, right?

Brian Ellefritz (A): That is correct.

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

Matt Ahrens (A): ZFS works great in all situations :-)

Dario (Q): In Solaris Volume Manager for Solaris 9 you can build replicas for pools, should you use replicas state database for Solaris 10(zfs)?

Bill Moore (A): With ZFS, there is no separate state database. All state in ZFS is maintained internally with an appropriate level of replication.

JimSmall (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 gigabyte files?

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

lance (Q): How will ZFS play with Samba?

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

jdefer (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?

Bill Moore (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.

Ben (Q): If an existing Solaris 8 server is upgraded to Solaris10 using the install upgrade process will it be possible to include a migration to zfs at that time?

Larry Wake (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.

Dario (Q): How does zfs management blocks?

Matt Ahrens (A): We have our own algorithms which minimize fragmentation and maximize performance.

walterg (Q): Is ZFS the strategic direction for Solaris filesystems?

Brian Ellefritz (A): Yes, for Solaris.

walterg (Q): Who came up with the name ZFS and what does the "Z" stand for?

Smita Thakur (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 the letter "Z" does not stand 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. The list can go on & on :-)

gheagerty (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?

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

Fazal Majid (Q): My production systems run Oracle9 on NetApp filers. How would you compare ZFS and WAFL? Some concepts seem similar, like copy-on-write.

Matt Ahrens (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.

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

Bill Moore (A): Yes. Also, you can have multiple pools on the same system as well.

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

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

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

Matt Ahrens (A): with ZFS, you'll be able to designate "hot space", which is reserved for use in the case of a disk failure. Rather than leaving an entire disk unused, thus reducing performance, ZFS reserves a little bit of each disk in the pool.

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

Bill Moore (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).

mark (Q): Are quotas supported on ZFS ?

Larry Wake (A): Yes.

apaton (Q): REF Brian Ellefritz (A): It works best on raw storage because we can apply our Predictive Self Healing functions to it. Do you mean JBOD devices, If so most storage arrays have/sell have RAID 0/1/ 5, controllers with large caches, how can ZFS self tune when there an additional hardware layer?

Bill Moore (A): We will utilize a disk (either raw or RAID device) based on its observed performance characteristics. If we cannot see the raw disks, we cannot provide self-healing data since we have no way to tell the RAID controller to read from the other side of a mirror, the parity disk, etc.

drsolaris8 (Q): Can I convert UFS and VXFS to ZFS ?

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

dlbewley (Q): I hope this is relevant. Using SVM on Sol9 all disk objects are named with digits. Will we be able to use text labels in 10?

Smita Thakur (A): It is on the roadmap. Please stay tuned for information in the future.

TrogL (Q): will there be an SNMP trap daemon for problems similar to mdlogd?

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

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

Matt Ahrens (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 on to ZFS.

JonM (Q): Can users see into the snapshots, and would be able to copy files out to perform their own "restores"?

Larry Wake (A): Yes.

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

Bill Moore (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.

walterg (Q): Will ZFS support quotas?

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

TrogL (Q): does ZFS support iSCSI?

Smita Thakur (A): Yes it will.

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

Bill Moore (A): I take it to mean that you're talking about data redundancy, not QOS w.r.t. performance. 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 in the initial release.

jdefer (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?

Matt Ahrens (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.

Bob (Q): Any thoughts on porting ZFS to Linux, AIX or HP-UX?

Smita Thakur (A): No plans of porting to AIX and HP-UX. Porting to Linux is currently being investigated.

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

Smita Thakur (A): We are working with all the vendors who support Solaris platform.

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

Larry Wake (A): It's under investigation, but won't be part of the initial release.

Dario (Q): what do you say about striping in zfs?

Matt Ahrens (A): ZFS implements dynamic striping, which provides better performance and more flexibility than the static striping (RAID 0) that volume managers use.

wenestvedt (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. (It reminds me of why disks are chopped into slices now instead of installing on one vast filesystem.)

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

Bob (Q): Will ZFS support Oracle data files? Is there 'quick i/o' or ODM support?

Bill Moore (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 that this will be true.

mnmezi (Q): As you probably already know, previous Solaris x86 releases have a hard time recognizing Network Interface Cards. (The majority of my systems are x86s with some Sparcs.) How will Solaris 10 handle this problem. Will it be as easy as it is with Linux systems?

Brian Ellefritz (A): Today's topic is about ZFS so we're not covering networking issues today (although we can assure you that Solaris 10 has dramatically improved hardware support on x86 platforms)

mjsmith (Q): Does each host have it's own storage pool or are multiple hosts sharing the same storage pool?

Matt Ahrens (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.

Dario (Q): Can I use the command vmstat for analyze performance of disk in Solaris 10?

Brian Ellefritz (A): Yes; in fact with DTrace in Solaris 10 you can do more advanced analysis than vmstat.

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

Matt Ahrens (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 (eg. disks) to the pool.

JW (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?)

Bill Moore (A): Yes, that's fully supported.

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

Bill Moore (A): With checksums, we check data integrity on every I/O. For redundant configurations, we only read data from one location, but we will provide a background "disk scrubber" that you can run to validate data continuously.

honeysuckle (Q): In a container environment what is ZFS's role? Is there a ZFS for each container?

Larry Wake (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.

Dario (Q): How do I configure swap for ZFS?

Brian Ellefritz (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 build.

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

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

timR (Q): Are there any plans to make zfs a distributed file system in the future?

Smita Thakur (A): This is currently being investigated.

ShawnFerry (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

Matt Ahrens (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 (eg. in a SAN).

lance (Q): How does zfs setup? How does it choose RAID type for instance/

Bill Moore (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 job to match disks up based on the administrator's request.

hkoby (Q): Does Solaris 10 support ORACLE Cluster File System?

Smita Thakur (A): Please check with Oracle if this product will be supported on Solaris 10.

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

Matt Ahrens (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.

canalla (Q): Will ZFS have a backup command like ufsdump?

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

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

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

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

Brian Ellefritz (A): Yes.

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

Smita Thakur (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.

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

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

rem (Q): How is alternate pathing handled?

Matt Ahrens (A): Alternate pathing is handled through MPXIO, which is a standard feature in Solaris 10.

Ben (Q): Is there a gui to manage the storage pool?

Larry Wake (A): Not at this time -- it's simple enough that there doesn't look to be a need for it.

Ryan (Q): What are those "predictive self healing" functions?

Smita Thakur (A): Predictive Self Healing has two main components: Solaris Fault Manager — A new service that receives data relating to hardware and software errors and automatically and silently diagnoses the underlying problem. Once a problem is diagnosed, an extensible set of agents automatically responds by off-lining the faulty component. Solaris Service Manager — Makes services, rather than processes, into first-class citizens in the Solaris environment, permitting self-healing to be implemented for services. Service descriptions, which include full dependency information for start, stop, and restart, are provided for base Solaris services, and applications can be converted with a few lines of XML. Predictive Self Healing offers the following benefits: - Improved system and service availability through predictive diagnosis and isolation of faulty components. - Automated diagnosis and restart of components happens automatically and in milliseconds rather than hours or days. - Simplified, administration model for managing services, reducing cost of ownership - Fast and easy repair of problems with links to knowledge articles - Scalable architecture can be rapidly upgraded and adapted to new problems without requiring downtime

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

Bill Moore (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.

jrs52 (Q): What's the thinking on using ZFS inside an N1 Grid Container? ZFS on ZFS - NFS ... ?

Larry Wake (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 would.

bernie (Q): Are there any limitations or special consideration 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?

Matt Ahrens (A): ZFS will recognize any style partitions that Solaris 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.

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

Brian Ellefritz (A): Next week along with our Network Computing online event (Tuesday), we'll be doing a major content refresh on our Solaris 10 site (www.sun.com/solaris/10). We'll have content there for both business and technical audiences. That site will continue to be built out in the coming months and weeks. If you don't see what you need there next week, please email me at brian.ellefritz@sun.com.

Dmitry (Q): when will Oracle certify ZFS?

Larry Wake (A): Oracle has already publicly stated their support for Solaris 10; 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.

jdefer (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?

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

Dario (Q): do I need to use command mount in Solaris with ZFS?

Matt Ahrens (A): ZFS will be so easy to administer, you won't even need to use mount(1m) :-). Simply by creating a ZFS filesystem, it will be available in the namespace (ie. mounted). Of course, you will also be able to turn off this feature and manually use the mount(1m) command.

TrogL (Q): what kind of performance hit will there be with ZFS vs ufs vs. SDS vs. Storange?

Bill Moore (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.

Ryan (Q): By self healing, I meant the ability to recover data from some type of checksum mechanism.

Smita Thakur (A): With ZFS all data is protected by 64-bit checksums resulting in 99.999999999999999999 percent error detection and correction. When any data is read, the checksum is verified to ensure that the data which the application wrote is what it gets back. If a checksum error is detected in a mirrored pool, the correct data will be read from the other side of the mirror, and the corrupt data will be repaired.

Tao (Q): Are you working with major DB and other application vendor, like 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 S10 is released ?

Brian Ellefritz (A): Yes, we are actively working with the db vendors on certification and benchmarking. Oracle has already stated their intent to support Solaris 10.

TrogL (Q): will information about storage on ZFS be available through kstat?

Matt Ahrens (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.

mark (Q): Does ZFS have multipathing load balancing ie. powerpath/veritas multipathing/etc ?

Smita Thakur (A): Yes. ZFS sits on top of multipathing capability MPxIO which is part of Solaris 10.

JonP (Q): Will it be possible for me to easily migrate my VxVM volumes to ZFS, and, if so, how?

Larry Wake (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.

lance (Q): Is ZFS redundancy handled via mirroring at the block level? I.e., is it like RAID 1 where usable space = 50% of raw space? Or is something like RAID 5 possible where "lossage" is less?

Matt Ahrens (A): ZFS will provide mirroring as well as RAID5-like redundancy.

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

Bill Moore (A): Quite well. :) If a disk fails, we recover the data if possible and return EIO if not.

coop (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?

Brian Ellefritz (A): Yes

MichaelMessina (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?

Matt Ahrens (A): With 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.

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

Larry Wake (A): It's extremely simple: issue a single command which changes the size of the filesystem.

Dario (Q): Can I forget to me of the file /etc/vfstab in Solaris 10?

Bill Moore (A): ZFS doesn't have to use /etc/vfstab. Each filesystem has a default mountpoint property which indicates where that filesystem should be automatically mounted. If you insist on putting things in /etc/vfstab, that is also supported, but not the recommended configuration.

mnmezi (Q): When is Solaris 10 slated for full release?

Brian Ellefritz (A): The current schedule is to ship Solaris 10 at the end of calendar year 2004.

timR (Q): Is ZFS based on the "Advanced File System", which is the base for SAM-FS and QFS ?

Smita Thakur (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.

Dario (Q): Can ufs work with zfs?

Larry Wake (A): UFS and ZFS can co-exist on the same system.

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

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

JW (Q): How are ZFS and QFS related?

Larry Wake (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 on in today's sun.com cover story. Expect more guidance and positioning of different filesystem technologies as ZFS gets closer to release.

BlazingSun27545 (Q): Is ZFS hardware specific? Does it only run on Sun storage or can it be run on EMC, etc

Brian Ellefritz (A): ZFS works on any underlying storage that Solaris is accessing. It works best on raw storage because we can apply our Predictive Self Healing functions to it.

canalla (Q): What volume manager tool works with ZFS?

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

Tao (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?

Matt Ahrens (A): Our goal is to make such tuning parameters unnecessary -- ZFS should be able to obtain optimal performance with a minimum of "wacky knobs".

Dario (Q): good afternoon, is storage grouped into small number of storage pools with intuitive names?, is each storage pool represents qos type?

Bill Moore (A): ZFS is quite flexible and you can configure your system in 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 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.

Ryan (Q): How does self healing work?

Brian Ellefritz (A): Our last Expert Exchange session was on Predictive Self Healing. The chat is archived at: [https://see.sun.com/Apps/DCS/mcp?r=703Z1342m5J0006304e4m32f403k7h042m4\\$0](https://see.sun.com/Apps/DCS/mcp?r=703Z1342m5J0006304e4m32f403k7h042m4$0)

gheagerty (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?

Bill Moore (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.

TrogL (Q): Does ZFS still use inodes?

Matt Ahrens (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.

court (Q): will ZFS replace traditional volume management?

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

JimSmall (Q): When will zfs be released via Solaris/Software Express?

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

Tao (Q): What kind of code (do you have a name for it, even if you don't call it "volume manager") 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 ?

Bill Moore (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.

JonM (Q): Does ZFS + Samba put SUN in a position to better compete with NAS vendors like NetApp ?

Matt Ahrens (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).

rrcarter (Q): I recently went to a lecture on solaris 10 and from talking to one of the engineers he said I could download the zfs package from sun software express. Is this true?

Larry Wake (A): ZFS is not in Software Express at the current time.

Yan (Q): Is it available separately? Convert from HFS to ZFS on Solaris 9, for example?

Brian Ellefritz (A): There are no current plans to offer ZFS on previous versions of Solaris. You will need to install Solaris 10.

JW (Q): Will or does ZFS support multi-reader, multi-writer type file system through a SAN?

Bill Moore (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 S10.

mlf (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?

Matt Ahrens (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>.

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

Brian Ellefritz (A): There are no current plans to charge for ZFS.

ShawnFerry (Q): Can we expect to see ZFS in an Express release before Solaris 10 ships?

Larry Wake (A): That's the current plan; stay tuned.

Yan (Q): can you get a brief on ZFS?

Smita Thakur (A): Please see today's sun.com cover story; if you need details beyond this, please contact your Sun sales/support team.

rrcarter (Q): why go to zfs and what does it offer compared to ufs and vfs

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

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

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

rem (Q): What conversion tools if any will be needed?

Bill Moore (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 you data to ZFS online.

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

Brian Ellefritz (A): ZFS will be in Solaris 10 when we ship the product. The current projection is the end of calendar 2004.

TrogL (Q): Sun has announced recently that Solaris 10 will be open source. How open is "open"?

Larry Wake (A): Today's session is on ZFS; we'll be giving more details on our open source plans in the near future.