

Sun Expert Exchange

“Fast Track to Solaris 10 Adoption: N1 Grid Containers”

Wednesday, July 21, 2004

Magnus (Q): Does Sun Cluster support zones? Is it possible to cluster between zones in different physical machines?

Dan Price (A): The Sun Cluster team and the Zones team are working together to make sure that Sun Cluster and Zones will work together seamlessly. We're still working out the details of how the actual clustering will work... :)

pj (Q): If I have a DB zone, an Application Zone and a Web Server zone, instead of running everything in the global zone. What are the performance penalties because all communication now has to go across zones via the network?

Andrew Tucker (A): Cross-zone network communication is handled in the IP layer of the kernel, rather than going out over the wire, so performance shouldn't be significantly different than when running the applications together in the global zone (and should be much better than running on separate systems, assuming sufficient CPU, memory, etc. is available).

Q: There is a new mount point in Solaris 10, which is /device why ??

Dan Price (A): I think you mean /devices, right?-- /devices is the front-end to a new kernel-provided filesystem called 'devfs'. This is part of a large-scale rearchitecture of the way we probe and attach hardware devices on the system. This has led to improved boot performance and has enabled a whole raft of further device-management improvement projects, most of which are still in the works. Generally speaking, for an admin or driver developer, /devices is just an implementation detail, and you don't need to worry about it.

jkofman (Q): How use it is to change from NIS+ to LDAP in Solaris 10

David Comay (A): There are tools in Solaris 9 and 10 to aid in transitioning from NIS or NIS+ to LDAP. Please see the documentation at <http://docs.sun.com/> for more information on these tools.

Peter%20Baer%20Galvin (Q): To clarify the LU question - can you use LU on a system that includes zones, but the zones will not be upgraded? Or can you not LU the global zone either?

David Comay (A): In the current Software Express releases, zones need to be uninstalled in order to use LU to upgrade the system. However, this is being addressed at the current time and LU will be the mechanism by which zones themselves are upgraded.

tim (Q): Just a note if anyone asks about running other OSes inside a zone. I know Sun says you can't but I have. Mostly just because Sun said it couldn't be done :) What I did was to compile up the bochs ("box") app inside a zone. This is a x86 system emulator and I successfully ran FreeDOS inside a zone. Now running x86 code on a Sparc isn't the most efficient use of resources but you can do it. I would be interested to see how it runs on a Solaris x86 system.

Dan Price (A): Sure, this certainly works! Generally we tell people "other OS's don't run inside zones" to emphasize that zones is not actually a virtual machine technology-- basically, you've started your own virtual machine...

Q: Has or will SMCC consider hardware support to assist virtualization in its SPARC platform for grids as PMMU's have assisted in virtualizing memory - or is this concept more directed to domains?

Andrew Tucker (A): Zones don't require any hardware support, and work equally well across all platforms supported by Solaris 10. Sun is also continuing to invest in improving its domain technology on future SPARC-based platforms.

Q: Would it be possible to run an MS Windows based program on a SPARC based system using either software or the Hardware PCI card from SUN, and running the session in its own container session. Second, could the MS Windows container interact with a software application such as Citrix Metaframe>

David Comay (A): We haven't tested the Sun PC card inside a zone yet but it should be possible. We will work on testing this soon.

ChuckReed (Q): When you say that all IPC is virtualized, does that mean I can allocate more shared memory to one zone vs. another and does that amount get pulled from the total allocated to the global zone, meaning it would have to be preallocated to the global zone first?

Andrew Tucker (A): IPC limits can now be specified on a per-project basis, so shared memory can be allocated for individual applications without affecting the global zone. We're also working on providing per-zone limits for locked memory (including ISM) and IPC limits in general.

Tim (Q): I am convinced that zones has growth potential. Is there any limitations; other than thinking of it as a VM clone, that I should be aware of?

Dan Price (A): Thanks! We are too. Since I don't know anything about your environment, it's hard to know what you might need which we haven't provided. Remember, it *isn't* a virtual machine, and so the benefits and limitations are different (don't forget that virtual machines have their own problems). I think the best thing you could do to answer your question is to give Zones a spin via the Solaris Express program, and let us know via the Zones Bigadmin Forum what is or isn't working for you.

Q: Where do you see customers using grid containers as opposed to simply using the resource management features of Solaris 9? Why?

David Comay (A): Grid containers is really the combination of both the resource management features available in Solaris 9 (and extended and improved in Solaris 10), as well as the isolation and security features provided by the zones facility in Solaris 10. You can use each independently, but we feel that the combination provides the fullest solution.

Prosser (Q): If a user logs into a non-global zone is there any way for that user to zlogin to the global zone?

Andrew Tucker (A): No, any access to the global zone must be through network services (e.g., ssh). If those services are disabled (or the non-global zone has no network interface) then there will be no way to go from the non-global zone to the global zone.

sundmc (Q): What is the process for applying patches when zones are in use?

David Comay (A): For OS or Solaris patches. the procedure will be to apply the patch from the global zone and the patch tools will automatically upgrade all the zones on the system. For unbundled software installed in a zone, the procedure will be to apply the patch within the zone itself.

TELUSARCH (Q): Do you intend to have (or now have) an internal or external reference site running significant instances of certain pervasive products such as Oracle, Weblogic, Websphere App Srvr etc. in highly active containers. This would build confidence that this technology is acceptable for mission critical applications.

Larry Wake (A): Yes. We have thousands of customers using Solaris 10 today (over 420,000 licensed systems), including many running the sorts of mission-critical apps you mention in containers. We're also looking at making systems available outside our firewall so that more people can test their applications on systems they might not normally have access to.

Q: In the Grid containers will this integrate with Oracle or do you have your own software solution

Andrew Tucker (A): Assuming you're talking about the Oracle database software, existing software should run unmodified in a container (a number of Solaris Express customers have verified this).

gameBoy (Q): Do N1 Grid Containers have any implications/benefits for massively multiplayer game networks?

Dan Price (A): I went to a cool talk about this at JavaOne-- Sun is even developing some novel new designs for multiplayer game networks. We think that zones fits well with this, as it may allow game hosting providers to create more dynamic provisioning systems, and also to host multiple customers in a highly isolated fashion on the same nodes in the network.

Tim (Q): In upgrading from Sol 8 to 10 will I still be able to mirror my volumes cleanly across vm that I configure?

David Comay (A): Yes, this is supported on upgrade.

Q: Will a list of what system attributes, features, sub-structures, daemons, etc. are "global" and which can/will be provided as a "local" copy to a given container? I saw the answer for the "cron" question and cron appears to be zoneable for example?

Dan Price (A): It sounds like you are asking: "What does zones virtualize, and what doesn't it?" This is pretty well covered in the documentation in BigAdmin-- see <http://www.sun.com/bigadmin/content/zones/>, but we'd encourage you to try it out for yourself! As for cron-- it hasn't been modified to "be zoneable". It "just works" as do most apps.

JasonY (Q): What's the advantage of using containers compared to using domains? Can we run multiple containers under each domain?

Andrew Tucker (A): This is similar to the answer regarding LPARs; domains partition the physical hardware to run separate OS instances, while containers allow multiple applications to share a single OS instance while still remaining isolated. Yes, you can run multiple containers in each domain.

Tim (Q): going back to the "kick" container question...I am seeing that N1 will or could be setup potentially to be a fault tolerant between containers. Any feedback on this?

David Comay (A): Containers are already fault tolerant between themselves since application failures in one zone will not affect another zone. Even better, Zone and the new Predictive Self Healing technology (available in the current Software Express) will be used together so certain hardware faults may cause only one particular zone to be affected rather the whole system.

TELUSARCH (Q): If I want to run a software product such as Veritas FS on an E25K but I want to run it only on one CPU in a container that is in a 10 CPU domain. Is it becoming more clear how software vendors will view licensing costs for my one CPU need?

Larry Wake (A): This is almost a two-part question -- I think VxFS would be an example of an app that would run in the global zone, providing services to all containers. But generally speaking, we expect vendors that license on a per-CPU basis will see containers the same way as they do domains, since a container can be tied to specific CPUs and this assignment cannot be changed from within the container. This is how Oracle, for example, defines a "hard partition" today, so it would make sense for them to extend the same licensing policy to containers.

philly (Q): Would N1 Grid and Grid Provisioning be a good grid solution for popping fortran jobs on a set of racked 1U slaves to run? Would N1 Grid not be needed in this case, just the Grid Provisioning part?

Andrew Tucker (A): Containers are useful when there's a need to isolate multiple applications running on the same system, either in terms of resource requirements or configuration, security, namespace, etc.. Generally, I'd expect HPC

or compute intensive fortran apps to have significant resource requirements, but not necessarily need the namespace isolation. I'd suggest looking at your app's requirements and figuring out what works best for you.

Sandie (Q): In a previous answer, it's mentioned that lost password can only be reset with a CD. If that happens to the root password of one zone only, by resetting it using the CD, it will mean that other "zones" will be shutdown as well. Is that correct?

Dan Price (A): Ahh, I was talking about recovering the root password for the global zone. Non-global zones are very different. If you lose the root password to your zone, the global zone administrator can easily reset it, since the global zone administrator can always login to the zone using `zlogin(1M)`, usually even when the zone is damaged. This makes zones ideal for hosted and managed environments where users might be prone to losing their passwords, or making other similar mistakes.

chasm (Q): will S10 support Sun PCI (personal computer cards)

A: Yes

tim (Q): Thanks. This was very good!

A: Thank you. We'd encourage you to continue to come to these live chat events; we have our August event posted and will soon communicate dates for more events in the fall on other Solaris 10 topics.

Q: is the "core dump" will be "zone aware"?

David Comay (A): Each zone can have their own `coreadm(1M)` settings. Also, the global zone can be configured to have copies of all the zones core files.

apollof (Q): Which 3rd party software would run in a "install in global, run multiple instances in non-global zones" mode?

Larry Wake (A): We don't have specific names or numbers on this yet, but we think it would be a significant number of existing apps.

Crash (Q): Can containers be spread across multiple nodes in a single cluster, but still have a single global container?

Andrew Tucker (A): Sort of. With the Sun Cluster software, you'll be able to associate a clustered application with a container, and that application will run within that container regardless of which node it is running on. You'll still need to configure the containers on each node (though we may be providing software to make that easy).

mdv (Q): Is a zone panic isolated to the zone or does it affect the entire server?

Dan Price (A): Because zones don't have independent kernels, there really is no such thing as a "zone panic." If there is a bug in the (systemwide) kernel, or a catastrophic hardware fault, the kernel will panic, and all zones will be affected. That said, we've made great strides in Solaris 10 in reducing the set of faults which are fatal. The PCI drivers have been hardened, and the new predictive self healing features can offline CPUs and DIMMs which seem it predicts are failing. All of this should add up to greater overall system reliability.

whammond (Q): One thing I would like to do is develop an application on a downlevel version of the OS and then test on all widely used versions of the OS on one piece of hardware. However I also need to access all the hardware resources of the system. Can I do this?

David Comay (A): If you're using zones, all containers have the same OS level so you cannot use it for this purpose. However, Sun Domains which is available on some of our servers, does support the ability to run multiple OS releases on the same machine.

DebK (Q): You mentioned financial savings. What other kind of metrics are you tracking for this feature? Performance?

Larry Wake (A): Performance is certainly one metric (and of course that really tracks right back to savings) -- by not imposing a significant system overhead, we allow customers to do more with the systems they buy. Ease of administration and security would be two other areas we think containers can be of benefit.

philly (Q): Does N1 run Fortran apps in containers and does it perform well?

Andrew Tucker (A): Yes, there's no difference in the application environment, so Fortran apps will run as well as C, C++, Java, Perl, etc.. There's no performance difference between running within a zone and running "natively" (aside from any overhead due to sharing resources with other apps running on the same system).

Magnus (Q): Lets say I share /usr with all zones. Is it possible to "exclude" certain directories under the /usr filesystem? I don't want to share /usr/openwin with the zones, but I need it globally?

David Comay (A): One way of doing this as part of the zone configuration is to create an empty directory in the global zone and configure a loopback mount for the zone on top of the directory in question: `add fs set dir=/usr/openwin set special=/empty set type=lofs add options ro`

Magnus (Q): What is the best way to backup a zone? Should a backup client be installed per zone, or should it be done globally at /zone/1, /zone/2 etc?

Dan Price (A): It's really up to you. We think either method is OK, and will just depend on your needs.

Crash (Q): Dan mentioned something about how much money people can save with grid container technology. How does this save me money?

A: The primary use case for N1 Grid Containers is for server consolidation; customer are using it to reduce the number of servers in their datacenter and the costs associated for managing them.

trent (Q): How does this compare to User-Mode Linux?

Andrew Tucker (A): UML creates multiple OS instances, each of which can run an application. With containers, we have a single OS instance, with isolated application environments. This results in a more powerful administrative model (we think), as well as performance advantages. See also my previous answer about VMware.

raffi (Q): when do you expect 64 bit technology in full swing

Larry Wake (A): The first big 64-bit technology wave was really in the mid to late 90s. This is when you saw "big iron" move to 64 bits in a big way, for large databases, image analysis, design, and so on. The second wave is now. You're seeing even high-volume client systems going over 1 GB of memory on a regular basis; we're all hitting the wall on 32-bit addressability, especially in the x86 world. This is why we're excited about AMD64, which lets customers take advantage of existing 32-bit x86 apps with extremely high, economical performance, but now can add in 64-bit applications and/or work within a system address space the readily breaks the 32-bit barrier.

tim (Q): Is it possible to *kick* a container from one physical server/domain to another, eventually while it is running ?

David Comay (A): Not at the present time. We are looking to be able to migrate zones from one server to another.

Q: Can the zones be configured to make use of the dynamic reconfiguration capabilities of an F15, i.e. if cpu is added from less active domain, can the zone automatically use it or does it have to be defined to the zone?

Dan Price (A): Actually, we have some neat auto-sizing technology in Solaris 10. We introduced a new system daemon which can resize resource pools based on policy; it's also smart enough to cope with DR. It is called 'pooled'. Since Zones can be bound to resource pools, you can take advantage of this. So for example, you could have a resource pool of size "1 to 5" cpus, and as CPUs get added or removed, pooled will adjust your resource pools.

Q: is zone will help if an application "triggers" a cpu crash (panic,...)?

Andrew Tucker (A): If the kernel panics, either due to kernel software problems or unrecoverable hardware failure, then the entire system (including all zones) will go down. On the other hand, if a hardware failure or software fault can be restricted to a single zone, only that zone will be rebooted (which happens very quickly).

Prosser (Q): does each container have separate cron daemons running so that users of each zone can have crons kick off?

David Comay (A): Yes, each container has its own cron daemon and can be configured with their own crontabs.

Q: How useful would N1 be for a dedicated data server such as Sybase ASE?

Larry Wake (A): The main benefits of containers would apply: resource isolation, security isolation, and hardware fault isolation. Depending on your site's requirements, containers may also simplify administration, make resource accounting easier, and drive higher levels of hardware utilization by making consolidation simpler.

Q: How does N1 grid containers interact with the DFS?

Andrew Tucker (A): We're still working out the details of how these should best be integrated. At minimum, this will work like any other file system; the global zone administrator will configure storage and determine which file systems are available in a given zone. We'd also like to allow a global zone administrator to assign a pool of storage to a zone, letting the zone administrator decide how to carve that into file systems. Since DFS is still under development, we're not sure how much of this will be available initially.

TELUSARCH (Q): It's my understanding that the Container inherits patch levels of the Solaris O/S. Is this true of products as well, e.g. MQ products are famous for installing in system directories?

David Comay (A): If those products are in a "shared" area such as /usr, then yes the Zones will inherit the same patch level. But for most unbundled products including the version of Java Enterprise System that will ship in conjunction with Solaris 10 each zone can have their own version and patch level.

rku (Q): How difficult to upgrade Solaris 8.0 to Solaris 10. Do we need to upgrade to 9 before 10?

Larry Wake (A): You can upgrade directly from Solaris 8 to Solaris 10; there's no need to go to Solaris 9 as an intermediate step. Difficulty is relative, of course; it depends on the software and drivers installed on your system. We guarantee (<http://sun.com/solaris/programs/guarantee.html>) that applications that run on Solaris 8 will run on forward releases; we plan on making this guarantee even simpler for developers to take advantage of with Solaris 10.

CarpeDM (Q): I'm surprised that the N1 Grid Console software isn't mentioned within the answers. It has a GUI, can move N1 Grids easily, ... When will it be available for Solaris 10?

A: Most of our docs and discussions of Solaris 10 are synced to the currently available features being shipped via the builds in the Solaris Express program. N1 Grid Console is a future capability.

Q: Can you use "Flash Archive" with zones?

David Comay (A): Not at the current time but this is something we may support in the future.

tim (Q): Do IP Filters (or packet filter or Sun's equivalent) work in zones to isolate them from each other ?

David Comay (A): Not at the current time. However, the same effect can be achieved by either setting up "reject" routes from the global zone (see the route(1M) man page for details) or by setting up IPsec in the global zone and configuring it to deny traffic the IP addresses configured in the zones you are trying to separate.

Q: Does the grid container get tied to a cpu ?If so how would failover work?

Andrew Tucker (A): Yes, a container can be bound to an individual CPU (or set of CPUs). If that CPU needs to be taken offline due to hardware issues, the container will be unbound and the (global zone) administrator will be notified.

Reed (Q): Is each zone tied to physical resources (specific CPUs) or do all zones share all the resources within the box dynamically? Is it possible to limit the amount of resources (CPU, Mem, etc.) one zone uses?

Andrew Tucker (A): Zones can either be configured to share resources, or the system resources can be partitioned and each zone can be assigned a specific set of resources (e.g., CPUs). In the case where the resources are shared, the proportion each zone receives can be configured. For example, the fair-share CPU scheduler can be used to assign each zone a share of the overall CPU in the system. This allows the resources to be divided to almost arbitrary granularity.

DebK (Q): Are you using the terms "zones" and "containers" interchangeably?

Dan Price (A): Great question! The terminology is consistent, but can be tough to sort out. Here goes: A *zone* is a way of partitioning the system to provide namespace and security isolation. The end effect is something similar to a virtual machine. A *container* (or N1 Grid Containers) is really a superset of Zones, Solaris Resource Manager, and some other technologies which add up to a highly partitioned system. So you can think of Zones as one of the specific technologies which make up our overall container strategy.

ChuckReed (Q): What would be the best practice for patching a server running zones (would you have to stop all non-global zones first)?

David Comay (A): Certain patches (like the Kernel Update) will require the zones to be shutdown. Most patches will not require zones to be shutdown. For Solaris patches (as opposed to unbundled and layered products), the procedure will be to apply the patch in the global zone and all zones that have been installed will have the patch applied to them automatically.

Q: Is "Live Upgrade" integrated into Solaris 10 and if it is, is it zone "aware"?

Larry Wake (A): Live Upgrade came in in Solaris 8, and remains in through Solaris 10. Currently (i.e., what you see via Software Express), it is not zone-aware, but plans are to make it so by Solaris 10 release at the end of this year.

philly (Q): Does N1 have a scheduler to run jobs per a schedule and any met conditions?

Dan Price (A): I guess maybe you're thinking about N1 Grid Engine? That's a separate product, and I do believe it has the features you are talking about. See <http://www.sun.com/software/gridware/>.

Q: The discussion so far seems based on a single system. Can multiple systems make up a Grid definition and can they interact?

Andrew Tucker (A): Multiple systems can be tied together using other software in the N1 Grid product suite (particularly the clustering software and N1 Grid Engine). The N1 Grid Containers feature only applies within a single system (allowing that system to be subdivided to run multiple applications in isolation).

Q: Are N1 Grid Containers similar to logical partitions?

Larry Wake (A): On a broad level, yes. LPARs, Dynamic System Domains and N1 Grid Containers all provide virtual partitioning on a single system. We think containers, especially in combination with domains, gives many advantages as discussed above.

Q: How do Grid Containers work with physical resources connected externally via SCSI or Fibre Channel? Is each HBA limited to one grid (or zone)?

David Comay (A): Zones can share the HBA of the system. In general, Zones use higher level abstractions than specific hardware resources. For example, Zones are assigned their own "root" file system which can off of a specific HBA. Other zones have their own root file system which also might be on the same HBA and even the same disk. However, the partitioning technology keeps the zones root file systems distinct, and processes in one zone cannot access files in other zones' root file systems.

Q: Does Veritas Clustering work in Solaris 10 ? If not, when ?

A: Veritas is a key Sun ISV partner and we are working with them on Solaris 10. We cannot commit them to any roadmap for their products but we would expect in general that their products be available around our general availability date.

TELUSARCH (Q): IBM has been really pushing Micro Partitioning to us lately and the technology and relative power of the Power5 line looks impressive. So for a go forward strategy we are looking at the potential of Containers and IBM's Micro Partitioning. Can you point out the advantages of Containers over Micro Partitioning?

Larry Wake (A): The major advantages we see of containers: runs on any system; no system performance hit; dramatically simplified management; thousands of containers on a system vs. 10 per CPU with micropartioning. (By the way, if it's only ten, why is it called "micro"? Does that mean we should call ours "pico-partioning"? :-)

Iwithers (Q): Is one able to capture and save Container configuration information so , for example, it could be easily migrated to another physical server ?

Andrew Tucker (A): Yes, the tools for managing the container (zone) configuration allow the configuration to be written out and transferred between machines. We're also working on ways to make this easier, particularly if the contents of the zone (files, etc.) can be placed on shared storage.

Q: how does this compare to vmware?

Andrew Tucker (A): VMware virtualizes the physical machine (creating a "virtual machine" that can run stock operating systems), and runs a separate OS instance on each VM. With containers, we running a single OS instance, and create isolated application environments on top of that - basically virtualizing the OS environment rather than the physical machine. Since we're not introducing new layers of software between the process and the physical hardware (container boundaries are implemented using the same OS checks already existing for cross-process security), containers have none of the performance issues present with a VM. In addition, a system with containers has only a single OS instance to administer (patch, upgrade, etc.).

Reed (Q): Can I move a zone from one physical machine to another? If so, does it require a reboot?

Dan Price (A): You can't at this time (although you can roll your own solution by having two identical zones on separate machines). We realize this is inconvenient, and plan to improve this in the future.

Q: Do products such as Veritas VxVM/VxFS play well with containers?

David Comay (A): VxVM/VxFS can be configured in the global zone and then zones can use the resulting file systems. This is the recommended strategy in general with volume managers like Solaris Volume Manager and VxVM.

ChuckReed (Q): Is there a way to quickly reprovision/move a Zone?

Dan Price (A): Initially zones will need to be uninstalled and then reinstalled in the other location, but we realize this is inconvenient, and are working to improve this going forward.

Sandie (Q): How is N1 Grid Container different from domain concepts? Does it provide hardware redundancy all the way through?

Larry Wake (A): See discussion earlier in the transcript.

Q: What kind of adoption are you seeing for N1 Grid Containers?

Larry Wake (A): We're seeing massive interest. One of the unexpected responses we're seeing is that customers running even just a single application on a system are considering running it in a container because of the high degree of security, fault and resource isolation this can add to their existing environment.

Magnus (Q): Can each zone have different TCP/IP settings and performance settings? Can these be controlled with Resource Manager? Does Resource Manager also control bandwidth for zones network?

David Comay (A): Global TCP/IP settings as set via /etc/system and ndd(1M are global and not currently settable on a per zone basis. Note that with the new TCP/IP stack in Solaris 10, many of these settings no longer need to be changed. If there are other settings that you feel you need to set per-zone, please let us know on the Zones BigAdmin forum. Yes, it is possible control the bandwidth that a zone uses. This can be done by using the bundled IPQoS functionality and configuring bandwidth parameters for each of the IP addresses that are configured for a particular zone.

gameBoy (Q): Can you say a bit more about zones?

Dan Price (A): Zones are great! But seriously, what would you like to know about?

sjafri (Q): If I forgot the root password. How can I recover it ?

Dan Price (A): Recovering isn't usually possible. But you can reset it. Boot from CDRom, or over the net. Instead of allowing the installer to proceed, pull up a terminal window. Mount the root disk (usually mount /dev/dsk/c0t0d0s0 /a) and then go to /a/etc and blank out the password part of root's password in the shadow file.

Q: When you say a "zone boots" does that imply that each zone is running a separate instance of the OS like domains on your larger servers.

A: Each zone runs on a single version of Solaris for that server; the zone contains the application processes and a small amount of OS resources required for that zone; booting a zone essentially reboots the application(s) and takes seconds, not minutes.

Peter%20Baer%20Galvin (Q): The biggest weakness I've spotted on Zones is that there is still a global /etc/system (for some variables). Can you comment on the future of /etc/systems if you agree that it is a weakness?

Andrew Tucker (A): We're working on eliminating the need for tunables in /etc/system; our view is that these should either be automatically determined (without any need for administrative control), or turned into dynamically controllable parameters such as those provided by the resource controls facility. In Solaris 10, we've removed the need for the most prominent usage of /etc/system: the System V IPC (semaphore, shared memory, message passing) tunables are now either removed (if they could just be made dynamic) or are per-project resource controls. In either case, there's no need to configure these on a system-wide basis. Other /etc/system parameters still exist, but we're working on this for the future.

Q: We are currently on Solaris 8. Are there any special migration considerations that we must be aware of in order to implement N1 Grid Containers on Solaris 10?

Larry Wake (A): Looking at it first from a broad view: if you have an application that runs on Solaris 8, it should work on Solaris 10. (Caveats in this area would be if you're doing something seriously out of bounds, such as directly trolling through kernel memory or using other types of undocumented interfaces, but even then the odds are high that what you're doing will continue to work.) Looking specifically at containers, the rule of thumb is that unprivileged programs should work in a container with no problem. Apps that need root privileges may need to consider whether certain functions need to run from the global zone, or whether they're candidates for consolidation at all. Also, the new Solaris privilege model may make it possible for apps that used to need root access to run at a lower privilege level and thus run in a container successfully.

Jay (Q): Are the memory regions isolated per zone? Is this isolation available on both Solaris and x86/AMD?

Dan Price (A): We do plan to do this in a (near) future release; in fact the memory isolation will be available even without using zones. Zones can already be bound to resource pools, and in a future release you will be able to associate a memory set with a resource pool. We're also working on some other memory-related limits and controls. And yes, it will all work on all platforms!

pj (Q): Can zones be upgraded using LU?

David Comay (A): In the current Solaris Express release, upgrading a zone via Live Upgrade is not yet supported. However, this is coming and in fact, Live Upgrade will be the primary upgrade mechanism for zones.

DebK (Q): Will white papers, best practices documents, and other "extra" helpful materials be available to customers soon? If so, where?

A: We've already posted an Administration Guide and other documents on our Big Admin site at <http://www.sun.com/bigadmin/content/zones/> We'll continue to populate that site with new resources through the release date, so bookmark that page!

CTUniAdmin (Q): Do you anticipate any issues with 3rd party software vendors, in regards to N1 Grid containers?

Andrew Tucker (A): Most 3rd party software (excluding kernel software, such as file systems and volume managers) will just work in a zone; the standard application environment is unchanged. We're working with vendors who represent exceptions to this rule (e.g., Veritas) to make sure they have versions of their software that works with containers.

Q: What are the common hurdles in IT organization to get people to use N1 Grid Containers?

Dan Price (A): It's a new approach to an old problem, so the biggest hurdle will probably be mindset. Our experience has been that "money talks" and when people realize how much money they can save with this technology, they are eager to try it out. We've seen significant adoption among Finance and Telco customers.

Magnus (Q): What happens when you create two non-global zones and they both share /usr, but then you need to apply some patch because zone A needs it (maybe because it is being used for compiles and the newest compiler needs it), however, you don't want to disturb zone B?

David Comay (A): For the present time, all zones need the same patch level for Solaris packages and the patch commands will automatically keep zones in sync. Patch levels for unbundled and layered products like the compilers and Java Enterprise System can be at "different" levels in different zones.

Q: Will Solaris get a GUI update and are you considering integrating the new technology 3D interface

Larry Wake (A): Solaris 9 introduced GNOME support; in Solaris 10 this will be significantly enhanced with the delivery of the GNOME-based Java Desktop System on Solaris as well as Linux. Integrating in 3D functionality such as what's been demonstrated with Project Looking Glass is still under discussion.

Q: Is N1 administration done via the command line or GUI?

A: Currently N1 Grid Containers administration is via command line; a GUI option is planned for a future release.

Q: Security and usability are said to be inversely proportional. Containers add security and isolation, but also complexity. What do you see as the "killer app" or ideal use for containers where the benefits outweigh the complexity?

Andrew Tucker (A): We've identified a number of possible uses: traditional data center server consolidation (databases, etc.), web hosting, developer use (dividing development from production, or allowing developers to share machines), etc.. There are already lots of folks doing server consolidation using simply resource management, due to concerns over hardware and administrative costs. The idea behind containers is to make this easier.

ShuChih (Q): Where can I get the list of files being copied from the global zone when a zone is created?

David Comay (A): The list of files copied from the global zone come from the packaging database which can be seen in /var/sadm/install/contents. Some files are not copied - those in packages marked with a SUNW_PKG_HOLLOW pkginfo(4) attribute and others (editable and volatile) are copied from an archive so they're installed in a factory default condition.

Q: I have an E10K and F15K today with multiple Domains running critical applications. How do I use N1 Grid Containers in this environment?

Larry Wake (A): Containers and domains are very complementary; you can use containers within a domain to further isolate applications and obtain an even finer level of granularity. The combination of containers and domains on a system like the F15K would give the ability to create over 100,000 separate application environments!

philly (Q): Is there full N1 support for Opterons?

A: The roadmap of supporting N1 Grid Containers is the same for the UltraSPARC and Opteron platforms.

Q: Is there additional pricing for the N1 Grid Container feature or does it included for free with Solaris 10?

A: N1 Grid Containers is a part of Solaris 10 and included in the pricing of that product.

seg (Q): Typically not all functionality and features are available for the GA release. Has Sun posted a functionality roadmap for Solaris 10 with GA features and the update features and if so, where can we see it?

Larry Wake (A): Probably the best way to gauge what will be available at GA will be to track what's already in the Software Express releases (<http://sun.com/solarisexpress/>). Although any feature list being discussed is subject to change until it's truly released, anything that's already integrated into the early Solaris 10 builds stands a very good chance of being there at GA. A more formal roadmap of futures would need to be discussed with your Sun support team under NDA.

Roy (Q): Does sun cluster 3 work with grid containers?

Andrew Tucker (A): We're actively working on this; support for containers should be ready in an update to SC 3.1.

TerryKozziel (Q): What criteria would I look at to determine what apps to place in containers as compared to separate domains

Dan Price (A): The criteria I would look at are: application size, the fault boundary and level of isolation required. For application size: what are the resource requirements? If your application requires only a small amount of resources (let's say, 1 CPU), zones will definitely save you money, since domains are typically at a rougher granularity (the system board). For the fault boundary, and isolation: what are the consequences of a fault? A domain is going to give you hardware-level isolation. The downside is that you'll have to managed different OS instances on each domain.

Q: My question is : Which Hard ware Course is suitable for me? My financial is limited, I can only afford one course at this present time !!!1 Thank you

A: Please check our course schedules at the Sun Education site at www.sun.com/education

Q: With a Grid container, are all kernel resources virtualized? For example, the kernel IPC data structures or other fixed sized kernel data structures like the proc table?

Andrew Tucker (A): System V IPC is virtualized; this means that two applications running in different containers can use the same IPC key without conflicts. In other cases, we've kept the same basic data structures, but filtered access; for example, searching /proc in a container will only see processes from that container, but the actual kernel data structure contains all processes. In general, we've made decisions based on the virtualization requirements and performance concerns, and reworked subsystems that seemed to require it.

Q: How do you handle the multiple net addresses with one network adapter ? or do you need multiple adapters

David Comay (A): Each zone can have one or more IP addresses assigned to it. When a zone boots, the system will automatically "plumb" logical interfaces that use a given physical adapter. So no, multiple network adapters are not required but can be used.

Q: Hi, Do you have any idea about the adoption rate for EDA vendors like Synopsys and Cadence to deploy their tools on Solaris 10? Currently I believe that Solaris 8 is the supported version. Thanks, Jose' Ramos Unix Admin

Larry Wake (A): We're getting a very enthusiastic response from software vendors, especially because of technologies such as zones and DTrace and some of the very significant performance gains we're seeing. Cadence announced their endorsement of Solaris 10 back in February.

Jay (Q): How soon is Solaris 10 for AMD64 going to be available?

A: We are offering one Solaris and one release date on both platforms; we will ship at the end of this calendar year.

Magnus (Q): Can each zone be on a different subnet with its own NIC? If so, how do I add a default route for each zone within the zone?

David Comay (A): Each zone can be on its own subnet but at the present time, the global zone itself must have an IP address on each of those subnets as well. Note that the network interface in the global zone can be in a "down" state and so only the local zone with the *other* IP address on the physical interface will be using it. The global zone can add a default route tied to each such interface so non-global zones can have their own default route.

Q: Have Grid, SMC and Resource management been consolidated? ie. are they managed separately or together.

Andrew Tucker (A): We're consolidating and integrating these features over time. Resource management (formerly provided by an unbundled product, SRM) is now part of Solaris as of Solaris 9. The Grid Containers functionality is similarly an integral part of Solaris 10. Other parts of the N1 Grid product suite (e.g., the N1 Grid Service Provisioning software) is separate, but we're working on tightly integrating these so that, e.g., the Service Provisioning software knows how to configure containers.

Jenn (Q): What are the new features in Solaris 10 that make it different from the previous version?

A: There are a ton of new features; the key ones are discussed in the table at www.sun.com/solaris/10.

Magnus (Q): Can JASS and FixModes be installed in a global zone to secure all zones at the same time?

Dan Price (A): JASS can be installed on a per-zone basis (and JASS has been enhanced to cope with zones properly), so you could easily write a script which JASS-ified all of your zones. I'm not 100% sure but I believe that you don't need to use FixModes starting with Solaris 8, as we've fixed all of the modes in the base product.

philly (Q): What's the min sys requirements (os, mem, cpu) to run on a master and slaves?

David Comay (A): By master, I assume you mean the "global" zone. The requirements will be the same as for running Solaris 10 in general. For "slave" or "non-global" zones, there are no hard requirements other than about 70MB of disk space and some amount of memory.

ChuckReed (Q): Just to be clear are Zones and N1 Grid Containers the same thing?

A: The concept of Zones was introduced in Solaris 9 with the Solaris Resource Manager. N1 Grid Containers elevates the capabilities of these techniques. Some of the nomenclature is a bit confusing; hopefully the technical documents are helpful in sorting this out.

Q: How do you compare zones with IBM'S LPAR technology ?

Andrew Tucker (A): Although both allow running applications in isolation on the same hardware, these are very different technologies. Zones support multiple applications within the same operating system instance; there's one kernel, one set of patches, etc., and a user in the "global zone" has visibility into the entire system (across all zones). In addition, multiple zones can share the same physical hardware (CPUs, memory, I/O, etc.). LPARs (and Sun's Domains) partition the physical machine, allowing a separate operating system instance to run on each partition. The degree of isolation is greater, however it means more OS instances to manage and a lower degree of sharing.

Magnus (Q): Lets say I want to run a website with a 3 tier configuration in a single machine. The 3 tiers would be on 3 different subnets. Config as follows: 2 websevers as 2 zones, with a physical NIC. 2 applicationservers as 2 zones with a physical NIC assigned, and 2 db servers as 2 zones also with a physical NIC. What would be the best practice to set up? Can IP-filter in the global zone filter traffic between the web zone interfaces ce0:1, ce0:2, the app zone interfaces ce1:1, ce1:2, and db zone interfaces ce2:1, ce2:1 ?

David Comay (A): The Zones on a system can be setup with different subnets. However at the current time, IP Filter cannot be used to filter between zones. An alternate mechanism is to set up a "reject" route in the global zone for the relevant subnets. Another alternative is to configure IPsec from the global zone to deny traffic between certain zones on the system.

Magnus (Q): When will Solaris 10 be released for x86, and will ZFS be available in Zone configurations?

Larry Wake (A): Solaris 10 is being developed concurrently for both SPARC and x86; release is planned for end of this year. You can download a preview of Solaris 10 today; see <http://sun.com/solarisexpress> . At the initial release, zones will be able to access ZFS volumes created in the global zone; we're exploring how to make it possible to directly administer ZFS storage pools from a zone in the future.

Q: Can a single network interface works with multiple containers on the same server ?

Dan Price (A): Yes, and this is the default mode. We create a "logical" network interface (which is a long-standing Solaris feature) atop the existing NIC , and then assign that to the zone. For example, zone "blue" might be assigned hme0:3 and zone "red" might be assigned hme0:5 (the zones software takes care of this for you).

Magnus (Q): What if a hacker somehow gains control over the global zone. He could wipe out all zones and get access to disk and CPU. How can we be sure zoneadmd can't be broken from a zone?

A: This is really two questions. Obviously, if a hacker is able to gain control (superuser access) over the global zone, they can control the whole system - this is part of the architecture. Security sensitive installations should be careful to protect the global zone by limiting services, using firewalls, etc.. The overall zone design, though, is to prevent intruders from being able to go from a non-global zone to the global zone. This isn't enforced by zoneadmd (which just manages the zone lifecycle); it's enforced by the same kernel subsystems that control cross-process, uid, etc., access.

Magnus (Q): Is no_exec_user_stack be available for x86, and can it also restrict exec in zones?

David Comay (A): no_exec_user_stack is a global tunable and affects all of the zones on a system and currently is not settable per-zone. Intel systems do not support this feature although AMD64 systems should support it when AMD64 support is available.

ebp022c (Q): How do I get to see the N1 Grid Container demo?

Larry Wake (A): See <http://sun.com/solaris/10> -- we'll also send participants a followup email with details. Of course, the "real" demo is Solaris 10 itself, which you can download today via our Software Express program. See <http://sun.com/solarisexpress/> .

Q: Is there a per-server or per-process limit on disk i/o request that an application can make?

David Comay (A): Not at the present time. However, we plan on continuing to add additional resource controls and this may including disk i/o requests. Please review the Resource Management and Zones Answerbook for more info.

Q: Where do you see customers using grid containers as opposed to simply using the resource management features of Solaris 9? Why?

A: Grid containers is really the combination of both the resource management features available in Solaris 9 (and extended and improved in Solaris 10), as well as the isolation and security features provided by the zones facility in Solaris 10. You can use each independently, but we feel that the combination provides the fullest solution.

Q: Can you explain the features of N1 Grid Containers ?

Larry Wake (A): To get the whole story, check the main Solaris 10 web site: <http://sun.com/solaris/10> ; there's a feature article on N1 Grid Containers linked to there. The short take: you can divide any system that can run Solaris 10 into multiple application environments -- literally thousands per systems. Each container looks to apps and users like its own OS instance, with its own nodename, net address, process table, memory management and so on. One other major feature is what it *doesn't* have, which is system overhead; there is no significant performance impact associated with containers. Another key benefit is the dramatic reduction of management complexity associated with consolidating multiple applications on one OS instance.

Q: When can we expect the first training classes and instructional manuals to be available?

Dan Price (A): Training classes are currently under development, but you can get complete access to the documentation today. If you visit our site on BigAdmin, at <http://www.sun.com/bigadmin/content/zones/>, you can download the manual today. Keep watching BigAdmin to know when classes are available.