



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Fast Track to Solaris 10 Adoption: Solaris Grid Containers A Sun Expert Exchange Discussion

Solaris Grid Containers isolate software applications and services using flexible, software-defined boundaries. Each application can be given a private environment, virtually eliminating error propagation, unauthorized access, and unintentional intrusions.

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

- General Information Pages 2-6
- Documentation & Training Page 7
- Performance Issues Pages 8-13
- Installation & Configuration Pages 14-16
- Compatibility Issues Pages 17-20
- Functionality & Usability Issues Pages 20-23

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: Solaris Grid Containers General Information

1. Can you explain the features of Solaris Grid Containers?
2. Does the Solaris Grid Container feature cost extra, or is it included with the Solaris 10 OS?
3. Just to be clear, are zones and Solaris Grid Containers the same thing?
4. 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?
5. I've heard that a 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, will it mean that other zones will be shutdown as well?
6. You mentioned something about how much money people can save with Solaris Grid Container technology. How does this save me money?
7. What other kind of metrics are you tracking for this feature? Performance?
8. When do you expect 64-bit technology to be in full swing?
9. Are you using the terms "zones" and "containers" interchangeably?
10. Is "Live Upgrade" integrated into the Solaris 10 OS and if it is, is it zone "aware"?
11. Does Solaris Grid have a scheduler to run jobs according to a schedule and any met conditions?
12. Are Solaris Grid Containers similar to logical partitions?
13. IBM has been pushing Micro-Partitioning 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?
14. How does this compare to VMware?
15. What kind of adoption are you seeing for Solaris Grid Containers?
16. What are the common hurdles in IT organizations in getting people to use Solaris Grid Containers?
17. Will Solaris get a GUI update, and are you considering integrating the new 3D interface?
18. Typically not all functionality and features are available for the GA release. Has Sun posted a functionality roadmap for the Solaris 10 OS with GA features and the update features? If so, where can we see it?
19. How soon is the Solaris 10 OS for AMD64 going to be available?
20. How do you compare zones with IBM'S LPAR technology?
21. Where do you see customers using Solaris Grid Containers as opposed to simply using the resource management features of the Solaris 9 OS? Why?

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

A: To get the whole story, check the main [Solaris 10 OS Web site](#) — there's a feature article on Solaris Grid Containers linked there. The short take: you can divide any system that can run Solaris 10 into multiple application environments — literally thousands per system.

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 does



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



not 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: Does the Solaris Grid Container feature cost extra, or is it included with the Solaris 10 OS?

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

Q: Just to be clear, are Zones and Solaris Grid Containers the same thing?

A: The concept of Zones was introduced in the Solaris 9 OS with the Solaris Resource Manager. Solaris 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: 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?

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 that vendors licensing 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.

Q: I've heard that a 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, will it mean that other zones will be shutdown as well?

A: 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.

Q: You mentioned something about how much money people can save with Solaris Grid Container technology. How does this save me money?

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

Q: What other kind of metrics are you tracking for this feature? Performance?

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 in which we think containers can be beneficial.

Q: When do you expect 64-bit technology to be in full swing?



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



A: The first big 64-bit technology wave was really in the mid-to-late 90s. That's 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 beyond 1GB 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.

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

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 Solaris Grid Container) is really a superset of zones, Solaris Resource Manager, and some other technologies, adding up to a highly partitioned system. So you can think of zones as one of the specific technologies that make up our overall container strategy.

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

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

Q: Does Solaris technology have a scheduler to run jobs according to a schedule and any met conditions?

A: I guess maybe you're thinking about Solaris Grid Engine? That's a separate product, and I believe it has the features you're talking about.

Q: Are Solaris Grid Containers similar to logical partitions?

A: On a broad level, yes. LPARs, Dynamic System Domains, and Solaris Grid Containers all provide virtual partitioning on a single system. We think containers, especially in combination with domains, provides many advantages.

Q: IBM has been pushing Micro-Partitioning 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?

A: The major advantages of containers are: runs on any system, no system performance hit, dramatically simplified management, thousands of containers on a system vs. 10 per CPU with Micro-Partitioning.

Q: How does this compare to VMware?

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 are running a single OS instance, and creating 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 be-



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



tween 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.).

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

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, and fault and resource isolation this can add to their existing environments.

Q: What are the common hurdles in IT organizations in getting people to use Solaris Grid Containers?

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" — when people realize how much money they can save with this technology, they'll be eager to try it out. We've seen significant adoption among finance and telecom customers.

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

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

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

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. 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 OS builds stands a very good chance of being at GA. A more formal roadmap of features would need to be discussed with your Sun support team under NDA.

Q: How soon is the Solaris 10 OS for AMD64 going to be available?

A: There will be one release date for both platforms; we will ship at the end of this calendar year.

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

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



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



is greater; however, it means more OS instances to manage and a lower degree of sharing.

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

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



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Fast Track to Solaris 10 Adoption: Solaris Grid Containers Documentation & Training

1. When can we expect the first training classes and instructional manuals to be available?
2. Do you intend to have (or do you already have) an internal or external reference site running significant instances of certain pervasive products such as Oracle, WebLogic, WebSphere App Server, etc. in highly active containers? This would build confidence that this technology is acceptable for mission-critical applications.
3. Which hardware course is suitable for me? My finances are limited, I can only afford one course at this present time.
4. Where can I see the Solaris Grid Container demo?

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

A: Training classes are currently under development, but you can get complete access to the documentation today. If you visit [our page on the BigAdmin site](#), you can download the manual today. Keep watching BigAdmin to find out when classes are available.

Q: Do you intend to have (or do you already have) an internal or external reference site running significant instances of certain pervasive products such as Oracle, WebLogic, WebSphere App Server, etc. in highly active containers? This would build confidence that this technology is acceptable for mission-critical applications.

A: Yes. We have thousands of customers using the Solaris 10 OS today (more than 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: Which hardware course is suitable for me? My finances are limited, I can only afford one course at this present time.

A: Please check our course schedules at the [Sun Education site](#).

Q: Where can I see the Solaris Grid Container demo?

A: See <http://www.sun.com/solaris/10> — we'll also send participants a follow-up email with details. Of course, the "real" demo is the Solaris 10 OS itself, which you can download today via our [Software Express program](#).



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Fast Track to Solaris 10 Adoption: Solaris Grid Containers Performance Issues

1. 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 incurred from all communication going across zones via the network?
2. Just a note: if anyone asks about running other operating systems inside a zone — I know Sun says you can't but I have. What I did was compile up the Bochs ("box") app inside a zone. This is an x86 system emulator, and I successfully ran FreeDOS inside a zone. Now running x86 code on a SPARC system isn't the most efficient use of resources, but you can do it. I would be interested to see how it runs on a x86 system.
3. 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 pre-allocated to the global zone first?
4. I am convinced that zones has growth potential. Are there any limitations; other than thinking of it as a VM clone, that I should be aware of?
5. Where do you see customers using Solaris Grid Containers as opposed to simply using the resource management features of the Solaris 9 OS? Why?
6. Do Solaris Grid Containers have any implications/benefits for massive multiplayer game networks?
7. I am seeing that Solaris technology will or could potentially be set up to be fault tolerant between containers. Any feedback on this?
8. Is a zone panic isolated to the zone or does it affect the entire server?
9. Let's 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.
10. How does this compare to User-Mode Linux?
11. Can I move a zone from one physical machine to another? If so, does it require a reboot?
12. Will zones help if an application "triggers" a CPU crash (panic)?
13. Does each container have separate cron daemons running so that users of each zone can have crons kick off?
14. I'm surprised that the Solaris Grid Console software isn't mentioned within the answers. It has a GUI and can move Solaris Grids easily. When will it be available for the Solaris 10 OS?
15. Do IP Filters (or packet filter or Sun's equivalent) work in zones to isolate them from each other?
16. Does the grid container get tied to a CPU? If so, how would failover work?
17. The discussion so far seems based on a single system. Can multiple systems make up a grid definition and can they interact?
18. Is one able to capture and save Container configuration information so that, for example, it could be easily migrated to another physical server?
19. 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?
20. 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?
21. Are the memory regions isolated per zone? Is this isolation available on both the Solaris OS and x86/AMD?
22. 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's being used for compiles and the newest compiler needs it), but you don't want to disturb zone B?
23. Where can I get the list of files being copied from the global zone when a zone is created?



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



24. What criteria would I look at to determine what apps to place in containers as compared to separate domains?
25. With a grid container, are all kernel resources virtualized? For example, are the kernel IPC data structures or other fixed-size kernel data structures like the proc table?
26. 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?
27. 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?
28. Is there a per-server or per-process limit on disk I/O requests that an application can make?

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 incurred from all communication going across zones via the network?

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 from 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: Just a note: if anyone asks about running other operating systems inside a zone — I know Sun says you can't but I have. What I did was compile up the Bochs ("box") app inside a zone. This is an x86 system emulator, and I successfully ran FreeDOS inside a zone. Now running x86 code on a SPARC system isn't the most efficient use of resources, but you can do it. I would be interested to see how it runs on a x86 system.

A: Sure, this certainly works. Generally we tell people "other operating systems 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: 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 pre-allocated to the global zone first?

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.

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

A: Since I don't know anything about your environment, it's hard to know what you might need that we haven't provided. Remember, it isn't a virtual machine, 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. Please let us know via the Zones BigAdmin Forum what is or isn't working for you.

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



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



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

Q: Do Solaris Grid Containers have any implications/benefits for massive multiplayer game networks?

A: Sun is 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.

Q: I am seeing that Solaris technology will or could potentially be set up to be fault tolerant between containers. Any feedback on this?

A: Containers are already fault tolerant between themselves, since application failures in one zone will not affect another zone. Even better, Zones 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.

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

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 (system-wide) kernel, or a catastrophic hardware fault, the kernel will panic, and all zones will be affected.

Q: Let's 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.

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.`

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

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.

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

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 we plan to improve this in the future.

Q: Will zones help if an application "triggers" a CPU crash (panic)?

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).



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Q: Does each container have separate cron daemons running so that users of each zone can have crons kick off?

A: Yes, each container has its own cron daemon and can be configured with its own crontabs.

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

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

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

A: Not at the present time. However, the same effect can be achieved either by setting up "reject" routes from the global zone (see the route(1M) main 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?

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.

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

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

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

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: 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 the Solaris OS 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 it only takes seconds, not minutes.

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



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



is a weakness?

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 the Solaris 10 OS, 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: Are the memory regions isolated per zone? Is this isolation available on both the Solaris OS and x86/AMD?

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.

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's being used for compiles and the newest compiler needs it), but you don't want to disturb zone B?

A: For the present, all zones need the same patch level for Solaris OS 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: Where can I get the list of files being copied from the global zone when a zone is created?

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: What criteria would I look at to determine what apps to place in containers as compared to separate domains?

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 manage different OS instances on each domain.

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

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 con-



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



tainer, 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: 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?

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: 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 were able to gain control (superuser access) over the global zone, he could 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. However, the overall zone design prevents 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.

Q: Is there a per-server or per-process limit on disk I/O requests that an application can make?

A: Not at present. However, we plan on continuing to add additional resource controls, and this may include disk I/O requests. Please review the Resource Management and Zones AnswerBook resources for more info.



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Fast Track to Solaris 10 Adoption: Solaris Grid Containers Installation & Configuration

1. How useful is it to change from NIS+ to LDAP in the Solaris 10 OS?
2. Has or will SMCC consider hardware support to assist virtualization in its SPARC platform for grids as PMMUs have assisted in virtualizing memory, or is this concept more directed to domains?
3. Will the “core dump” be “zone aware”?
4. Can the zones be configured to make use of the dynamic reconfiguration capabilities of an F15, i.e., if a CPU is added from a less active domain, can the zone automatically use it, or does it have to be defined to the zone?
5. It's my understanding that the container inherits patch levels of the Solaris OS. Is this true of products as well? MQ products, for example, are famous for installing in system directories.
6. How difficult is it to upgrade from the Solaris 8 OS to the Solaris 10 OS? Do I need to go to the Solaris 9 OS as an intermediate step?
7. 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, memory, etc.) that one zone uses?
8. 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?
9. Can JASS and FixModes be installed in a global zone to secure all zones at the same time?
10. What are the minimum system requirements (OS, memory, CPU) to run on a master and slaves?
11. Let's say I want to run a Web site with a three-tier configuration in a single machine. The three tiers would be on three different subnets. Configuration as follows: two Web servers as two zones, with a physical NIC; two application servers as two zones with a physical NIC assigned; and two db servers as two 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?
12. When will the Solaris 10 OS be released for x86, and will ZFS be available in zone configurations?
13. Can a single network interface work with multiple containers on the same server?

Q: How useful is it to change from NIS+ to LDAP in the Solaris 10 OS?

A: There are tools in both Solaris 9 and 10 operating systems 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.

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

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



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Q: Will the “core dump” be “zone aware”?

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

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

A: Actually, we have some neat auto-sizing technology in the Solaris 10 OS. We introduced a new system daemon that can resize resource pools based on policy; it’s also smart enough to cope with DR. It’s called “poold.” 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 are added or removed, poold will adjust your resource pools.

Q: It’s my understanding that the container inherits patch levels of the Solaris OS. Is this true of products as well? MQ products, for example, are famous for installing in system directories.

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 the Solaris 10 OS, each zone can have its own version and patch level.

Q: How difficult is it to upgrade from the Solaris 8 OS to the Solaris 10 OS? Do I need to go to the Solaris 9 OS as an intermediate step?

A: You can upgrade directly from the Solaris 8 OS to the Solaris 10 OS; there’s no need for an intermediate step. Difficulty is relative, of course; it depends on the software and drivers installed on your system. We guarantee that applications running on the Solaris 8 OS will run on forward releases; we plan to make this guarantee even simpler for developers to take advantage of with the Solaris 10 OS.

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, memory, etc.) that one zone uses?

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.

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?

A: Global TCP/IP settings as set via /etc/system and ndd(1M) are global and not currently settable on a per-zone basis. Please note that, with the new TCP/IP stack in the Solaris 10 OS, 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 to 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.



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



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

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 that JASS-ified all of your zones. I'm not 100 percent sure, but I believe that you don't need to use FixModes, starting with the Solaris 8 OS, as we've fixed all of the modes in the base product.

Q: What are the minimum system requirements (OS, memory, CPU) to run on a master and slaves?

A: By master, I assume you mean the "global" zone. The requirements will be the same as for running the Solaris 10 OS 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.

Q: Let's say I want to run a Web site with a three-tier configuration in a single machine. The three tiers would be on three different subnets. Configuration as follows: two Web servers as two zones, with a physical NIC; two application servers as two zones with a physical NIC assigned; and two db servers as two 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?

A: The zones on a system can be set up 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.

Q: When will the Solaris 10 OS be released for x86, and will ZFS be available in zone configurations?

A: The Solaris 10 OS is being developed concurrently for both SPARC and x86; release is planned for the end of this year. You can [download a preview of the Solaris 10 OS](#) today. 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 in the future to directly administer ZFS storage pools from a zone.

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

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 zone's software takes care of this for you).



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Fast Track to Solaris 10 Adoption: Solaris Grid Containers Compatibility Issues

1. Does Sun Cluster support zones? Is it possible to cluster between zones in different physical machines?
2. 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?
3. 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, running the session in its own container session? Second, could the MS Windows container interact with a software application such as Citrix Metaframe?
4. In upgrading from the Solaris 8 OS to the Solaris 10 OS, will I still be able to mirror my volumes cleanly across VM that I configure?
5. Will the Solaris 10 OS support Sun PCI (personal computer cards)?
6. 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?
7. Does Solaris technology run Fortran apps in containers, and does it perform well?
8. How useful would Solaris technology be for a dedicated data server such as Sybase ASE?
9. How does Solaris Grid Containers interact with the DFS?
10. Can you use "Flash Archive" with zones?
11. Does Veritas Clustering work in the Solaris 10 OS? If not, when will it work?
12. Do products such as Veritas VxVM/VxFS play well with containers?
13. We are currently on the Solaris 8 OS. Are there any special migration considerations that we must be aware of in order to implement Solaris Grid Containers on the Solaris 10 OS?
14. Can Zones be upgraded using LU?
15. Do you anticipate any issues with third-party software vendors with respect to Solaris Grid Containers?
16. I have an E10K and F15K today with multiple domains running critical applications. How do I use Solaris Grid Containers in this environment?
17. Is there full Solaris support for AMD Opteron processor-based systems?
18. Does Sun Cluster 3 work with Solaris Grid Containers?
19. How do you handle the multiple net addresses with one network adapter? Or do you need multiple adapters?
20. Do you have any idea about the adoption rate for EDA vendors like Synopsys and Cadence to deploy their tools on the Solaris 10 OS? I believe that the Solaris 8 OS is the supported version currently.
21. Have grid, SMC and resource management been consolidated, i.e., are they managed separately or together?
22. Is `no_exec_user_stack` available for x86, and can it also restrict exec in zones?

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

A: The Sun Cluster team and the zones team are working together to make sure that Sun Cluster and zones will work together seamlessly.



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



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?

A: In the current Software Express releases, zones need to be uninstalled in order to use LU to upgrade the system. This is being addressed currently, and LU will be the mechanism by which zones themselves are upgraded.

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, running the session in its own container session? Second, could the MS Windows container interact with a software application such as Citrix Metaframe?

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.

Q: In upgrading from the Solaris 8 OS to the Solaris 10 OS, will I still be able to mirror my volumes cleanly across VM that I configure?

A: Yes, this is supported on upgrade.

Q: Will the Solaris 10 OS support Sun PCI (personal computer cards)?

A: Yes.

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?

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.

Q: Does Solaris technology run Fortran apps in containers, and does it perform well?

A: Yes, there's no difference in the application environment, so Fortran apps will run as well as C, C++, Java code, 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).

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

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 Solaris Grid Containers interact with the DFS?

A: We're still working out the details of how these will best be integrated. At minimum, this will work like any other file system; the global zone administrator will configure storage and determine which file



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



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.

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

A: Not at the current time, but this is something we may support in the future.

Q: Does Veritas Clustering work in the Solaris 10 OS? If not, when will it work?

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

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

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.

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

A: Looking at it first from a broad view, if you have an application that runs on the Solaris 8 OS, it should work on the Solaris 10 OS. 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 OS 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.

Q: Can Zones be upgraded using LU?

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

Q: Do you anticipate any issues with third-party software vendors with respect to Solaris Grid Containers?

A: Most third-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: I have an E10K and F15K today with multiple domains running critical applications. How do I use Solaris Grid Containers in this environment?



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



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 you the ability to create more than 100,000 separate application environments.

Q: Is there full Solaris support for AMD Opteron processor-based systems?

A: The roadmap of supporting Solaris Grid Containers is the same for the UltraSPARC and AMD Opteron-based platforms.

Q: Does Sun Cluster 3 work with Solaris Grid Containers?

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

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

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 they can be used.

Q: Do you have any idea about the adoption rate for EDA vendors like Synopsys and Cadence to deploy their tools on the Solaris 10 OS? I believe that the Solaris 8 OS is the supported version currently.

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 the Solaris 10 OS back in February.

Q: Have grid, SMC and resource management been consolidated, i.e., are they managed separately or together?

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

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

A: no_exec_user_stack is a global tunable. It affects all zones on a system and currently is not settable per-zone. Intel systems do not support this feature, although AMD64-based systems should support it when AMD64 support is available.



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Fast Track to Solaris 10 Adoption: Solaris Grid Containers Functionality & Usability Issues

1. 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?
2. If a user logs into a non-global zone, is there any way for that user to zlogin to the global zone?
3. What is the process for applying patches when zones are in use?
4. In the grid containers, will this integrate with Oracle, or do you have your own software solution?
5. What’s the advantage of using containers compared to using domains? Can we run multiple containers under each domain?
6. Would Solaris Grid and Grid Provisioning be a good grid solution for popping Fortran jobs on a set of racked 1U slaves to run? Would Solaris Grid not be needed in this case, just the Grid Provisioning part?
7. Can containers be spread across multiple nodes in a single cluster, but still have a single global container?
8. 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?
9. What would be the best practice for patching a server that’s running zones? Would you have to stop all non-global zones first?
10. Is there a way to quickly re-provision/move a zone?
11. If I forget the root password, how can I recover it?
12. Is Solaris administration done via the command line or GUI?

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?

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 simply using resource management due to concerns over hardware and administrative costs. The idea behind containers is to make this easier.

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

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.

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

A: For OS or Solaris OS 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.



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



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

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).

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

A: 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.

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

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 to need the namespace isolation. I'd suggest looking at your app's requirements and figuring out what works best for you.

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

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).

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?

A: It's really up to you. We think either method is OK; it just depends on your needs.

Q: What would be the best practice for patching a server that's running zones? Would you have to stop all non-global zones first?

A: Certain patches (such as the Kernel Update) will require the zones to be shutdown, but most will not. 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 there a way to quickly re-provision/move a zone?

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 the situation going forward.



Sun Expert Exchange

Technical Knowledge Base for Sun Inner Circle Members



Q: If I forget the root password, how can I recover it?

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`), then go to `/a/etc` and blank out the password part of root's password in the shadow file.

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

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