Bill Davidsen wrote:

> Dave Hansen wrote:
>
>> On Sat, 2006-03-25 at 04:33 +1100, Nick Piggin wrote:
>>
>>> Oh, after you come to an agreement and start posting patches, can you
>>> also outline why we want this in the kernel (what it does that low
>>> level virtualization doesn't, etc, etc)
>>
>>
>> Can you wait for an OLS paper? ;)
>>
>> I'll summarize it this way: low-level virtualization uses resource
>> inefficiently.
>>
>> With this higher-level stuff, you get to share all of the Linux caching,
>> and can do things like sharing libraries pretty naturally.
>>
>> They are also much lighter-weight to create and destroy than full
>> virtual machines. We were planning on doing some performance
>> comparisons versus some hypervisors like Xen and the ppc64 one to show
>> scaling with the number of virtualized instances. Creating 100 of these
>> Linux containers is as easy as a couple of shell scripts, but we still
>> can't find anybody crazy enough to go create 100 Xen VMs.
>
>
> But these require a modified O/S, do they not? Or do I read that
> incorrectly? Is this going to be real virtualization able to run any O/S?

This type is called OS-level virtualization, or kernel-level
virtualization, or partitioning. Basically it allows to create a
compartments (in OpenVZ we call them VEs -- Virtual Environments) in
which you can run full *unmodified* Linux system (but the kernel itself
-- it is one single kernel common for all compartments). That means that
with this approach you can not run OSs other than Linux, but different
Linux distributions are working just fine.

> Frankly I don't see running 100 VMs as a realistic goal

It is actually not a future goal, but rather a reality. Since os-level
virtualization overhead is very low (1-2 per cent or so), one can run
hundreds of VEs.

Say, on a box with 1GB of RAM OpenVZ [http://openvz.org/] is able to run about 150 VEs each one having init, apache (serving static content), sendmail, sshd, cron etc. running. Actually you can run more, but with the aggressive swapping so performance drops considerably. So it all mostly depends on RAM, and I'd say that 500+ VEs on a 4GB box should run just fine. Of course it all depends on what you run inside those VEs.

> , being able to run Linux, Windows, Solaris and BEOS unmodified in 4-5
> VMs would be far more useful.

This is a different story. If you want to run different OSs on the same box -- use emulation or paravirtualization.

If you are happy to stick to Linux on this box -- use OS-level virtualization. Aside from the best possible scalability and performance, the other benefit of this approach is dynamic resource management -- since there is a single kernel managing all the resources such as RAM, you can easily tune all those resources runtime. More to say, you can make one VE use more RAM while nobody else it using it, leading to much better resource usage. And since there is one single kernel that manages everything, you could do nice tricks like VE checkpointing, live migration, etc. etc.

Some more info on topic are available from
http://openvz.org/documentation/tech/

Kir.

>>
>> Anyway, those are the things that came to my mind first. I'm sure the
>> others involved have their own motivations.
>>
>> -- Dave
>>
>