VM satus meaning and troubleshuting

It is to me a little of a mystery how SLURM decides how to mark each node when I execute *sinfo*, this problem has bugged me enough to make me write this document in which I describe some of the states you may get into, what they mean and a way to move from that state to another. This is mainly because I haven't find a good documentation about power saving mode besides the link that you will find at the bottom of this document.

down* fg5

SLURM has never (?) seen this VMs or you manually issued the shutdown command, just start the VMs manually, when it finishes booting it will connect to the server at will be put in idle state:

```
/opt/thiao/bin/Resume fg5
```

You may need to execute that command the very first time you install and enable Thiao's integration for power saving mode.

idle fg[0-1]

The VMs are there doing nothing, ready to start executing a job at any moment. If it remains in this state for more than SuspendTime seconds, the VM will be shut down (see idle~).

idle~ fg1

The node is in **power saving mode** (it was shut down because it was idle for too much time), SLURM will resume (create) the VM when a job arrives to the queue.
If there are jobs in the queue but the **VMs are not started** automatically you can force the VM creation isuing the power_on command:

```
scontrol update nodename=fg[0-1] state=power_up
```

It seems that that happens because the VMs are still being pooled after being put in power saving (effectively shutdown by OpenNebula through Thiao!), and at some point (after ???? seconds) they are marked as not responding (though it doesn't seems to reflect in the information show in sinfo or scontrol), if that happens the only thing you get is something like this:

```
# grep responding /var/log/slurm-llnl/slurmctld.log | tail
[2011-11-07T11:34:35] error: Nodes fg[0-3,7] not responding
[2011-11-07T11:38:38] Node fg0 now responding
[2011-11-07T11:38:47] Node fg1 now responding
[2011-11-07T11:54:35] error: Nodes fg[2-3,7] not responding
```

There must be some parameter to alter this behaviour, if you know which is it send me an email to update this document.

idle* fg2

fg2 was in *down** state, got to this state after executing:

```
scontrol update nodename=fg2 state=resume
```

SLURM doesn't issue the resume command, only the VM was marked as not responding. I'ts needed to start the VM manually to take it to the idle state:

```
/opt/thiao/bin/Resume fg2
```

idle# fg[0-3]

The VMs where started manually asking slurm to power up the nodes:

```
scontrol update nodename=fg[0-1] state=power_up
```

After booting the nodes should be in idle or alloc status, if they are not, check the VM templates, specially the contextualization section.

alloc# fg[0-3]

The VM was assigned a job but is not responding, it will be put in this state if it was resumed to execute a job... until it finishes booting and connects to the server.
If it remains in this state for too long, check that the VM template is correct, specially the phase of contextualization (it happened to me that the IP and hostname of some VMs overlapped making them stay in this state indefinitely because the wrong VM was being resumed or the same one was being resumed twice).

Additional resources

Remember that there are some differences between OpenNebula 2 and 3 that can break your VM template files, also there may be some differences in the behaviour of SLURM 2.2 and 2.3. For this OpenNebula 2.2 and SLURM 2.2 where used since those are the versions available in Debian/testing at the moment of writing this software.

OpenNebula's contextualization guide

http://opennebula.org/documentation:archives:rel2.2:cong

Managing virtual machines

http://opennebula.org/documentation:archives:rel2.0:vm_guide

VM template

http://opennebula.org/documentation:archives:rel2.0:template
http://opennebula.org/documentation:rel3.0:vm_guide

SLURM's power saving guide

https://computing.llnl.gov/linux/slurm/power_save.html