PostgreSQL Clustering with Red Hat Cluster Suite

Devrim GÜNDÜZ

PostgreSQL Contributor

Twitter: @DevrimGunduz

Red Hat Certified Engineer

Community: devrim@PostgreSQL.org

Personal: devrim@gunduz.org

Before we start...

Thanks to who involved in this conference.

...who is this guy?

- I have been contributing to PostgreSQL over the last 8 years.
- I'm not a hacker, I work on RPMs and website.
- I rarely break RPMs, but break website more often.
- Working as a PostgreSQL DBA right now at markafoni.
- Live in Istanbul.
- Have a son.

Agenda

- **Goals**
- **Why Red Hat Cluster Suite (RHCS)?**
- **Before initializing setup...**
- **Choosing right hardware (including network)**
- **Setting up RHCS**
- **Setting up PostgreSQL**
- **Failover, switchover...**
- **Postgres-XC**
- **Questions?**

Dec 6, 2010

PostgreSQL European Conference, 2010 - Stuttgart

Agenda

- **Goals**
- **Why Red Hat Cluster Suite (RHCS)?**
- **Before initializing setup...**
- **Choosing right hardware (including network)**
- **Setting up RHCS**
- **Setting up PostgreSQL**
- **Failover, switchover...**
- **Postgres-XC**
- **Questions?**

Goals

- Active/passive clustering
- Having a redundant system
  - Data redundancy
  - Network redundancy
  - Server and power redundancy
- Maximum uptime
- Service failover (=PostgreSQL)
- Data integrity

Agenda

- Goals
- **Why Red Hat Cluster Suite (RHCS)?**
- Before initializing setup...
- Choosing right hardware (including network)
- Setting up RHCS
- Setting up PostgreSQL
- Failover...
- Postgres-XC
- Questions?

Why Red Hat Cluster Suite (RHCS)?

- Open Source Clustering Solution
- Developed by Red Hat, **with the community**
- Available through (Red Hat Network) RHN, but also available via the CentOS repositories (*unsupported by Red Hat, or supported by 3rd party support companies*)

Why RHCS?

- RHEL 5 Advanced Platform provides RHCS+GFS.
- It is the only open source clustering solution that has decent support.
- Use at least RHEL 5.4. All versions prior to that are broken in various ways.
- Minimizes downtime

Why RHCS?

- Support wide range of hardware
- Application/Service Failover - Create n-node server clusters for failover of key applications and services
- Load Balancing - Load balance incoming IP network requests across a farm of servers
- TGIOS! (Thanks God It is Open Source)

RHCS Overview

- Supports up to 16 nodes (RHCS 5). The upper limit is 8 for RHEL 3 and 4.
- All PostgreSQL nodes can access to the same storage, but they don't use it at the same time.
- Automatic failover
- http://www.redhat.com/cluster_suite/
- http://sources.redhat.com/cluster/wiki/ (Development site)

What else?

- RHCS avoids cancer.
- It helps peace in the world.
- RHCS cannot be used as a replica. If you want to hear about replicas, this is not the right talk.
- RHCS does not run on Windows.
- It does not do “multimaster” clustering.
- Postgres-XC? We will talk about it later.

Agenda

- Goals
- Why Red Hat Cluster Suite (RHCS)?
- **Before initializing setup...**
- Choosing right hardware (including network)
- Setting up RHCS
- Setting up PostgreSQL
- Failover,...

  Postgres-XC
- Questions?

Before initializing setup

- Make sure that you have at least a RHCE or similar around.

Agenda

- Goals
- Why Red Hat Cluster Suite (RHCS)?
- Before initializing setup..
- **Choosing right hardware (including network)**
- Setting up RHCS
- Setting up PostgreSQL
- Failover, switchover...
- Caveats
- Questions?

Hardware and software requirements

- Minimum hardware: An hardware that Red Hat Enterprise Linux can run.
- Typical hardware : Depends on your needs. *See related threads in pgsql-performance mailing list.*
- SAN : Storage is the most important part – Use RAID arrays.

Hardware and software requirements

- RHCS is built on GFS.
- GFS is built on LVM.
- PostgreSQL :-)
- Each node needs to have 1GB ram (not for PostgreSQL, it is for RHCS)
- Decent fibre channel switch to storage, decent ethernet switches for internal and external communications.

Choosing the right hardware: Network

- Multicast traffic must be supported / enabled in network switches.
  - Testing: ping -t 1 -c 2 224.0.0.1
- Cluster services may not work if they don't respond to ICMP echo requests.

Choosing the right hardware: Fencing

- Fencing: Disconnection of a node from the cluster's shared storage (RHCS docs)
- It cuts off I/O from share storage to ensure data integrity.
- System **must** have a supported fencing device.

.

Choosing the right hardware: Fencing

- Power fencing : Uses a power controller to power off an inoperable node.
- Fibre Channel switch fencing : Disables the Fibre Channel port that connects storage to an inoperable node.
- GNBD fencing :Disables an inoperable node's access to a GNBD server.
- Other fencing :Several other fencing methods that disable I/O or power of an inoperable node, including IBM Bladecenters, PAP, DRAC/MC, HP ILO, IPMI, IBM RSA II, and others.

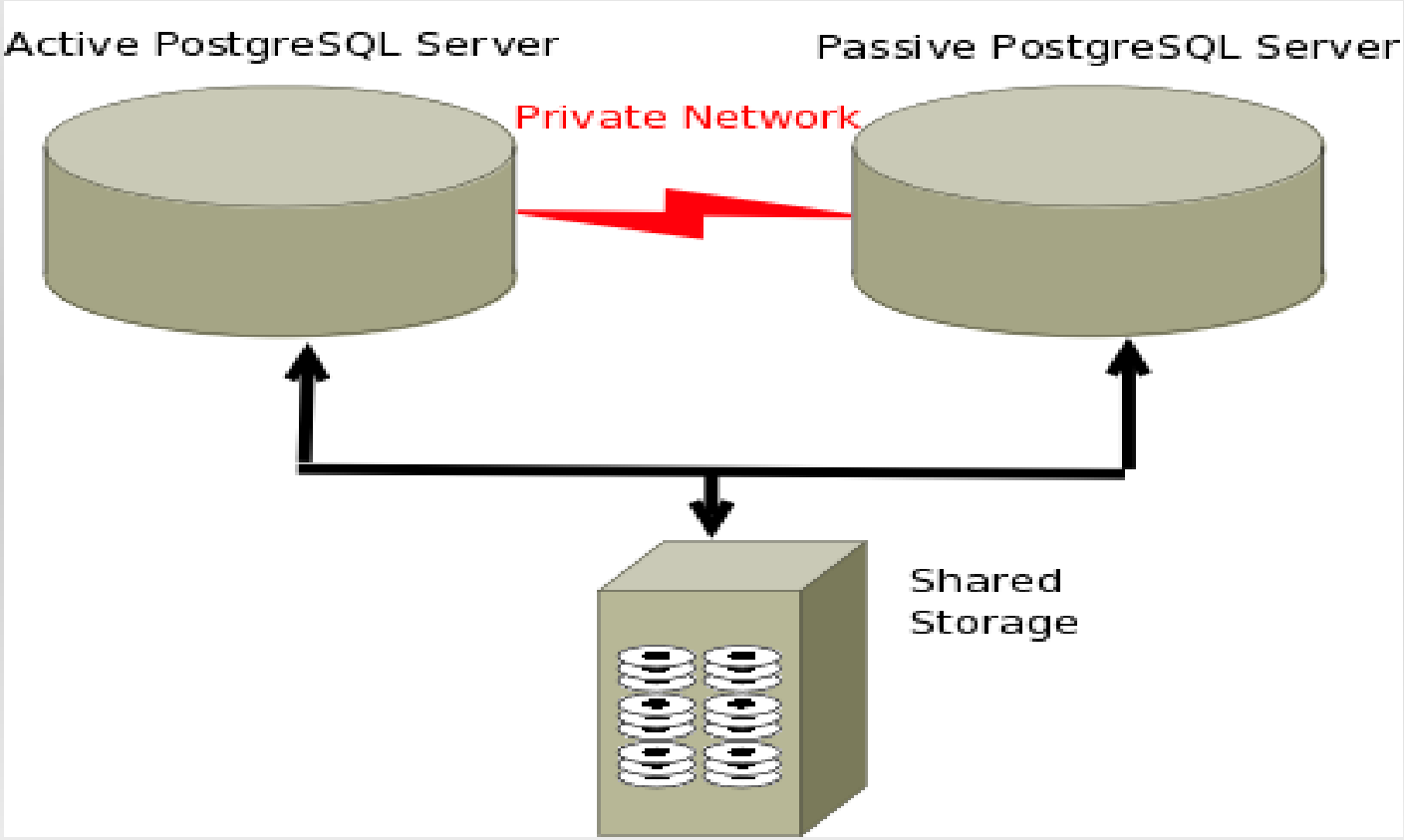
Design and howtos

- We need two servers that has been setup identically.
  - Only OS and PostgreSQL will run
  - Same PostgreSQL versions.
- Using GFS, all data will be mounted from the storage. GFS is not a requirement, but we would better be safe.
- When node1 goes down, node2 will act as “active” server by announcing specified cluster ip.
- When node1 comes back, the process is reverted.

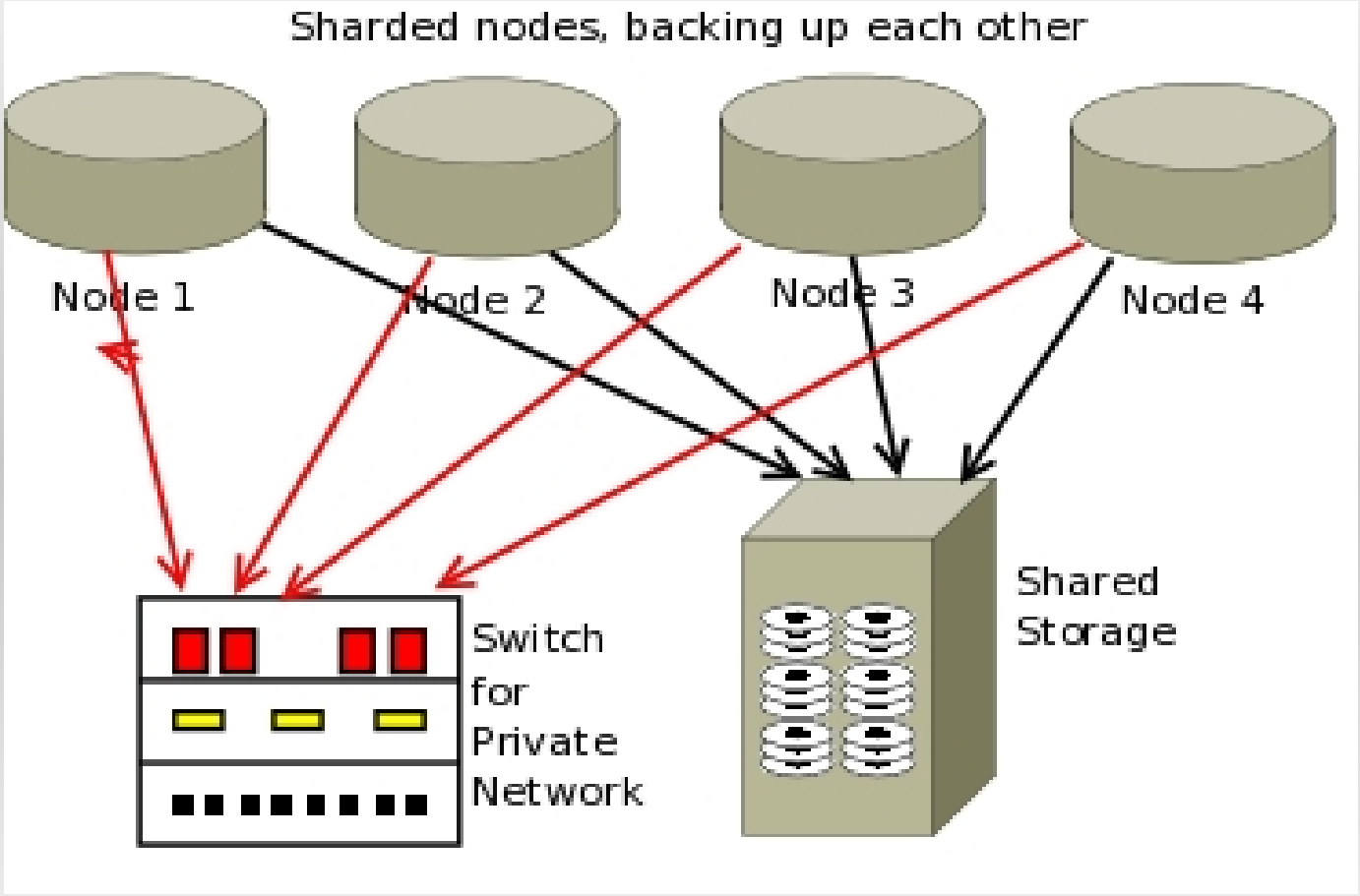
Hardware and software requirements

- http://www.redhat.com/cluster_suite/hardware/
- Check this list **before** you purchase the hardware.
- HP Blade Servers have been proved to be working well with RHCS. Recommended.
- Make sure that you have updated firmware.

Active/Passive Cluster Overview



Sharding, multiple nodes



Agenda

- Goals
- Why Red Hat Cluster Suite (RHCS)?
- Before initializing setup...
- Choosing right hardware (including network)
- **Setting up RHCS**
- Setting up PostgreSQL
- Failover
- Postgres-XC
- Questions?

Setting up RHCS

- Before we start:
  - Do **NOT** edit contents of cluster.conf manually.
  - If you choose to edit cluster.conf manually, make sure that xml version numbers are identical on each node.
  - If you think that you will screw up things, ask someone else.
  - Be patient. This is not a plug-and-play solution.

Services that needs to start on boot

- clvmd
- cman
- gfs
- rgmanager
- PostgreSQL

- chkconfig is your friend.

RPM packages

- Cluster packages must be installed prior to setup.
- Perl-Crypt-SSLeay package is essential for fencing mechanism to function properly.

Setting up Cluster

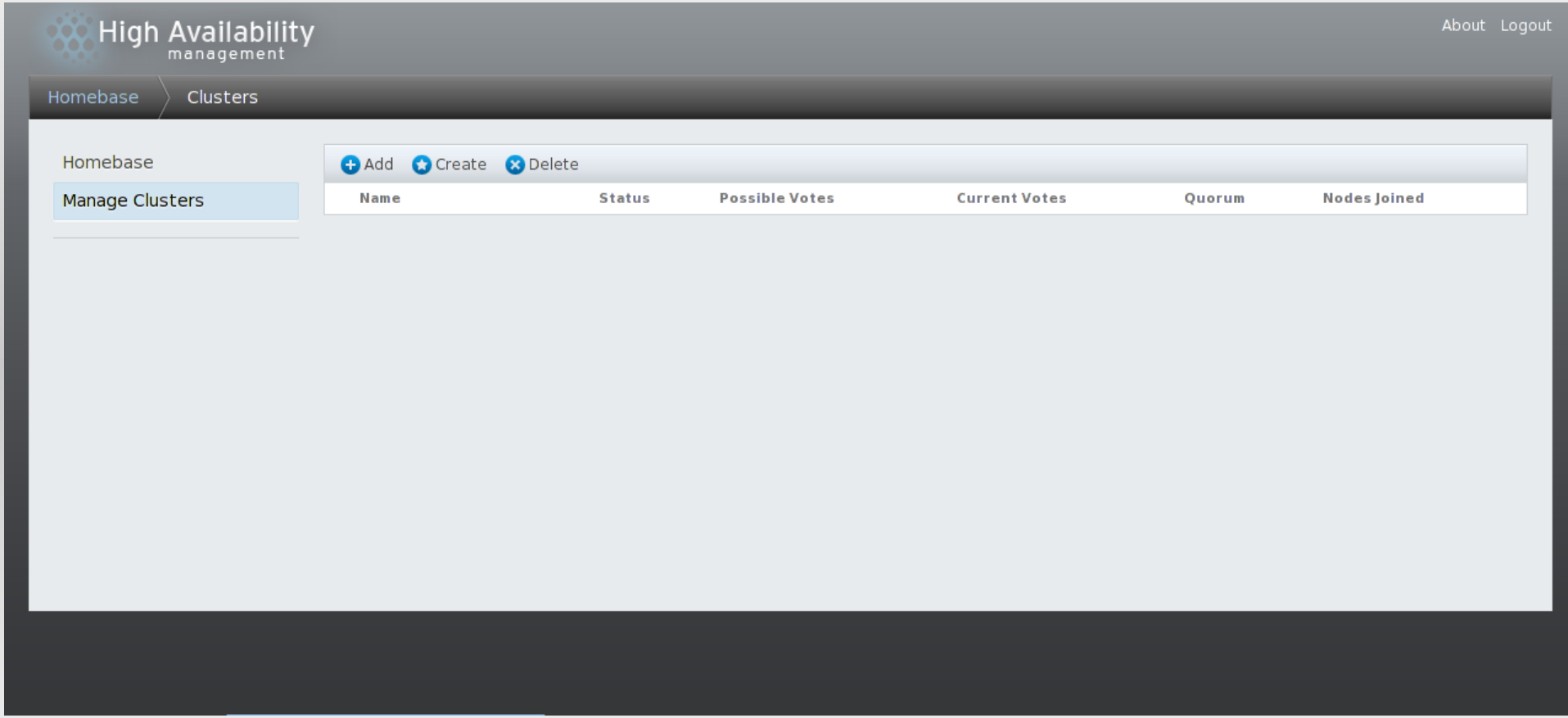
- RHEL 5 provides system-config-cluster (scc), which is not supported in RHEL 6.
- Use only very recent versions of scc, otherwise you may screw up your cluster.
- scc helps you versioning your cluster configuration. Make sure that it is the same in all nodes.
- clusterssh will be your best friend during setup.

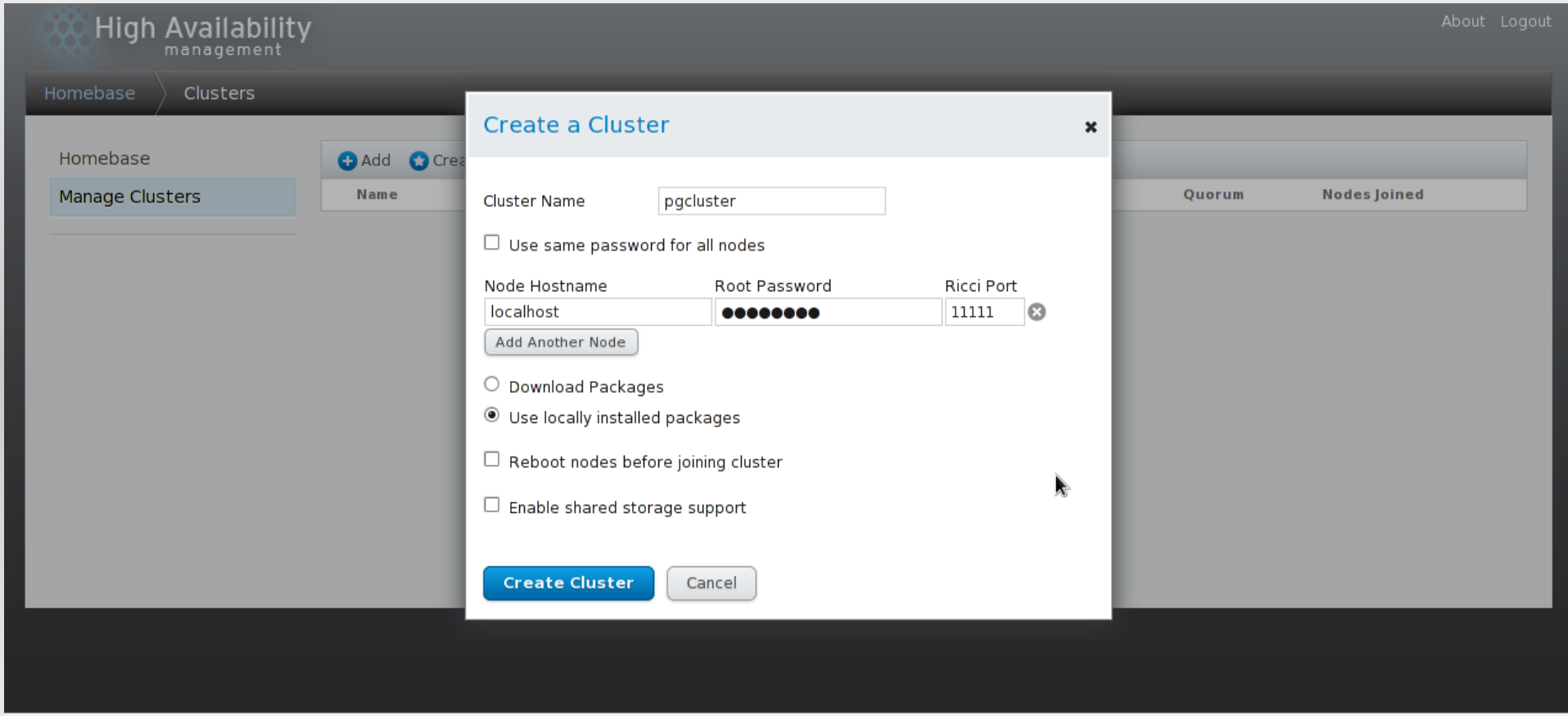
Features of Conga (copied from RHEL 6 docs)

- One Web interface for managing cluster and storage
- Automated Deployment of Cluster Data and Supporting Packages
- Easy Integration with Existing Clusters
- Integration of Cluster Status and Logs

Conga (RHEL 6)

- 2 components: luci and ricci
- Luci: server side tool, communicates with ricci.
- Ricci: agent tool that runs on cluster members, and coammunicates with luci.
- TGIP (Thanks God It's Python!)


High Availability
management
About
Logout
Homebase
Clusters
Homebase
Manage Clusters
Add
Create
Delete
Name
Status
Possible Votes
Current Votes
Quorum
Nodes Joined


High Availability
management
About
Logout
Homebase
Clusters
Homebase
Manage Clusters
Add
Name
Quorum
Nodes Joined
Create a Cluster
Cluster Name
pgcluster
Use same password for all nodes
Node Hostname
Root Password
Ricci Port
localhost
11111
Add Another Node
Download Packages
Use locally installed packages
Reboot nodes before joining cluster
Enable shared storage support
Create Cluster
Cancel

cluster.conf example

- Let me run an editor first :)

Sample resource script

```
#!/bin/sh

start()
{
/etc/cluster/networkstart0
service postgresql-1 start
}

stop()
{
service postgresql-1 stop
/etc/cluster/networkstop0
}
...
```


networkstop0

```
#!/bin/bash
cp /etc/cluster/ifcfg-bond0.561:0 \
/etc/sysconfig/network-scripts/
        ifdown bond0.561:0
        rm -f /etc/sysconfig/network-
scripts/ifcfg-bond0.561:0
```

- ifcfg-bond0* file is a regular ifcfg file, which includes cluster ips.
- Trick: ONBOOT=no

networkstart0

```
#!/bin/bash
cp /etc/cluster/ifcfg-bond0.561:0 \
/etc/sysconfig/network-scripts/
        ifup bond0.561:0
        rm -f /etc/sysconfig/network-
scripts/ifcfg-bond0.561:0
```


Agenda

- Goals
- Why Red Hat Cluster Suite (RHCS)?
- Before initializing setup...
- Choosing right hardware (including network)
- Setting up RHCS
- **Setting up PostgreSQL**
- Failover, switchover...
- Postgres-XC
- Caveats
- Questions?

Setting up PostgreSQL

- No specific **tuning** needed.
- However, if you are using more than one node, you will want to be careful while sharing hardware resources.
- You will want to use **unix_socket_directory** parameter, so that each instance will use the same TCP/IP port, so that you'll avoid using workarounds in your application side.

Agenda

- Goals
- Why Red Hat Cluster Suite (RHCS)?
- Before initializing setup...
- Choosing right hardware (including network)
- Setting up RHCS
- Setting up PostgreSQL
- **Failover**
- Postgres-XC
- Questions?

Failover

- RHCS handles failover properly.
- It detects dead node, and moves service to the next machine, as configured in cluster.conf
- Once the dead machine is up, service is transferred back to it.
- ~30 seconds of downtime during this operation.

Agenda

- Goals
- Why Red Hat Cluster Suite (RHCS)?
- Before initializing setup...
- Choosing right hardware (including network)
- Setting up RHCS
- Setting up PostgreSQL
- Failover
- **Postgres-XC**
- Questions?

Postgres-XC

- A new synchronous and transparent clustering solution for PostgreSQL, providing both read and write scalability
- 0.9.3
- http://postgres-xc.sourceforge.net
- Can be used with or without RHCS, and it will work more or less like Oracle RAC.
- Under heavy development
- They have a talk tomorrow.

Agenda

- Goals
- Why Red Hat Cluster Suite (RHCS)?
- Before initializing setup...
- Choosing right hardware (including network)
- Setting up RHCS
- Setting up PostgreSQL
- Failover
- Postgres-XC
- **Questions?**

PostgreSQL Clustering with Red Hat Cluster Suite

Devrim GÜNDÜZ

PostgreSQL Contributor

Twitter: @DevrimGunduz

Red Hat Certified Engineer

Community: devrim@PostgreSQL.org

Personal: devrim@gunduz.org