

9.4 the Hard Way

Postgres cluster upgrade using hard links

Scope:

7 Clusters,

37 servers

667G of data (max)

2 datacenters

2 Operating Systems

Restrictions:

4 hour windows for upgrade

One Operating System

Restrictions:

Translates to 2 windows

4 Primary Clusters First

2 Secondary Clusters Second

1 Tertiary Cluster Whenever

Options:

Dump + Restore

- Time prohibitive
- 3.5 hours for initial `pg_dump`
- 4.5+ hours for master
- 3+ hours for slaves

Options:

Slony

- **Complex**
- **No inhouse expert**
- **High Stress Factor**
- **Large Set-Up and Tear-Down Phase**
- **Extra Hardware requirement**

Options:

pg_upgrade + Slave Rebuild

- Better!
- Well documented
- Well practiced
- 3+ hours for slaves
- Hard or Sym Links?

Options:

pg_upgrade + Hardlinks + Slave Rebuild

- **Almost Perfect**
- **No Easy Rollback**
- **Still - 3+ hours for slaves**

Options:

Isn't pg_upgrade + rsync super fast?

Options:

pg_upgrade + Hardlinks + rsync

- **Word of Mouth**
- **Devel Documentation only**
- **Both Success and Failure Stories**
- **Minutes, not Hours**

Planning, Testing, and Refinement

**Document every step, and every
command.**

BE VERBOSE.

Planning, Testing, and Refinement

1.5 months

Complete configuration overhaul

Countless Tests

Error, Fix, Reset, Repeat.

Planning, Testing, and Refinement

Starting Point:

<http://www.postgresql.org/docs/devel/static/pgupgrade.html>

We care about step 10.

Planning, Testing, and Refinement

In the Doc:

From a directory that is above the old and new database cluster directories, run this for each slave:

```
rsync --archive --delete --hard-links --size-only  
old_pgdata new_pgdata remote_dir
```

Planning, Testing, and Refinement

First Hurdle: rsync

**Symlinks, Permissions, Slow
and Tedious to Reset**

Planning, Testing, and Refinement

Solution:

Learn, Test, Move

Planning, Testing, and Refinement

Solution:

/srv/pg_data/data

/srv/pg_data/9.4/data

mv data 9.2data

Planning, Testing, and Refinement

Second Hurdle:

Configuration + Reset

Planning, Testing, and Refinement

Solution:

Fix Yer' Tech Debt.

DO NOT INIT THE SLAVE.

Planning, Testing, and Refinement

Third Hurdle:
Resuming Replication

Planning, Testing, and Refinement

Solution:

pg_start_backup before the rsync

per the normal slave build process

Planning, Testing, and Refinement

Fourth Hurdle:

Master OS Version Mismatch

Planning, Testing, and Refinement

Solution:

Swap On Upgrade

As if nothing was ever wrong.

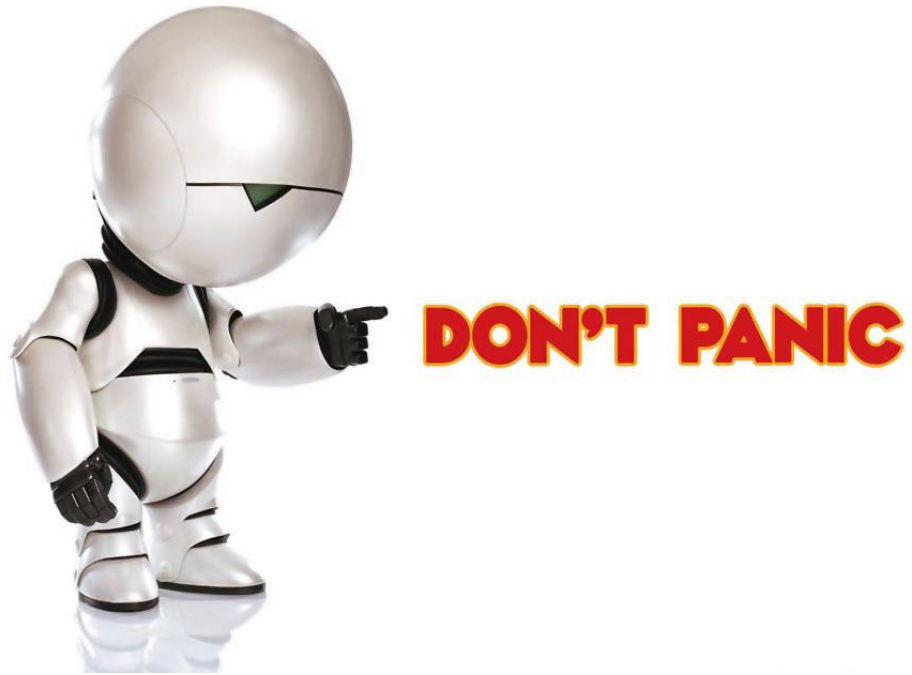
Day Of : First Window

Cluster 1: 3 hours, 667 G

Cluster 2: 28 minutes 100G

Cluster 3: 53 minutes 400G

Cluster 4: Let's not talk about that one.



Day Of : Second Window

Cluster 1: 3 Minutes, 80 G

Cluster 2: 35 minutes 320G

Day Of : Second Window

We Don't know why one cluster went slowly. (But faster than a rebuild)

Thanks to:
Bruce Momjian
Josh Berkus
Stephen Frost

Questions:

djmzsamantha@gmail.com