

DuckCorp Infrastructure - Review #518

Review branch backup

2017-04-03 13:58 - Pierre-Louis Bonicoli

| | |
|---|-------------------------------|
| Status: In Progress | Start date: 2017-04-03 |
| Priority: Normal | |
| Assignee: Pierre-Louis Bonicoli | |
| Category: | |
| Target version: | |
| Branch: backup | |
| Description This is a start: the committed configuration backups only PostgreSQL databases hosted on Touthirou. Other hosts/directories will be added latter. | |
| Related issues: Related to DuckCorp Infrastructure - Review #519: Review burp role In Progress 2017-04-03 Related to DuckCorp Infrastructure - Enhancement #497: Change Backup System In Progress 2017-05-09 | |

History

#1 - 2017-04-03 13:58 - Pierre-Louis Bonicoli

- Branch set to backup

#2 - 2017-04-03 14:00 - Pierre-Louis Bonicoli

- Related to Review #519: Review burp role added

#3 - 2017-04-03 17:02 - Marc Dequènes

- Status changed from New to In Progress

I did not look at the role yet, first remarks about what i can see from the outside.

The database backup rewrites what `/usr/local/sbin/srv_backup_databases` did. So this script is most probably to be improved, but it was complex because it involved using `fifo` to avoid storing local data. On Orfeo we lack space and this is not gonna work as the DSPAM database is **huge**. So we need to use the feature to read from `fifos`.

We don't have any playbook separation, we're using tags at the moment. Is there any plan to merge it to `dc.yml` or do you think we should arrange differently?

Where is defined variable 'backup_excludes'?

I see "ratelimit:" without any value; I don't recall this is a valid YAML syntax. Is it accepted?

Every 5 minutes is not a lot of stress? I guess it is simply retries if for eg the network is down?

More later.

#4 - 2017-04-03 18:20 - Pierre-Louis Bonicoli

Marc Dequènes wrote:

I did not look at the role yet, first remarks about what i can see from the outside.

The database backup rewrites what `/usr/local/sbin/srv_backup_databases` did. So this script is most probably to be improved, but it was complex because it involved using `fifo` to avoid storing local data. On Orfeo we lack space and this is not gonna work as the DSPAM database is **huge**. So we need to use the feature to read from `fifo`s.

Fifos are used again.

We don't have any playbook separation, we're using tags at the moment. Is there any plan to merge it to `dc.yml` or do you think we should arrange differently?

I plan to add `include backup.yml` in `dc.yml`.

Where is defined variable `'backup_excludes'`?

variable removed.

I see `"ratelimit:"` without any value; I don't recall this is a valid YAML syntax. Is it accepted?

yes (None), `ratelimit` removed anyway.

Every 5 minutes is not a lot of stress? I guess it is simply retries if for eg the network is down?

We don't want to miss the short windows. There is a bug, the cron task is executed on `Toushirou` which use Paris timezone.

More later.

#5 - 2017-04-05 06:24 - Marc Dequènes

- Category set to Service :: Backup

Pierre-Louis Bonicoli wrote:

Fifos are used again.

Could you make the global catalog a fifo too please?

I plan to add include backup.yml in dc.yml.

Ok, fine.

I see "ratelimit:" without any value; I don't recall this is a valid YAML syntax. Is it accepted?

yes (None), ratelimit removed anyway.

I know None, but without value at all?

Every 5 minutes is not a lot of stress? I guess it is simply retries if for eg the network is down?

We don't want to miss the short windows. There is a bug, the cron task is executed on Touthirou which use Paris timezone.

So TZ can cause problems when asking the server if it is the right time? Not sure to get it.

I think the BACKUP_DIR should be specific to the task, just in case we have concurrent backups, and it's clearer too.

To be clean, the list of excluded databases should be computed with a parameter. *postgres* and *template0* are always to be excluded, but *bacula* is an admin-specific choice.

Also it would be nice to be able to kill all remaining processes in the *post* action, but be sure to only touch the ones we started in this script. This can be done in a future PR though.

Ok, the rest is fine, I'm gonna read the role now.

#6 - 2017-04-05 12:15 - Pierre-Louis Bonicoli

Marc Dequènes wrote:

Pierre-Louis Bonicoli wrote:

Fifos are used again.

Could you make the global catalog a fifo too please?

Done

I see "ratelimit:" without any value; I don't recall this is a valid YAML syntax. Is it accepted?

yes (None), ratelimit removed anyway.

I know None, but without value at all?

None or without value at all, that's the same.

See the following python snippet:

```
import yaml; print(yaml.load("{'1':, '2':null, '3':None, '4':~}"))
```

outputs:

```
{'1': None, '3': 'None', '2': None, '4': None}
```

Every 5 minutes is not a lot of stress? I guess it is simply retries if for eg the network is down?

We don't want to miss the short windows. There is a bug, the cron task is executed on Touthirou which use Paris timezone.

So TZ can cause problems when asking the server if it is the right time? Not sure to get it.

No: cron task is executed on Touthirou which is using PARIS timezone, Korutopi is using Japan timezone.

I think the BACKUP_DIR should be specific to the task, just in case we have concurrent backups, and it's clearer too.

There is a lock on the client and a lock on the server, concurrent backups aren't possible.

To be clean, the list of excluded databases should be computed with a parameter. *postgres* and *template0* are always to be excluded, but *bacula* is an admin-specific choice.

Done

Also it would be nice to be able to kill all remaining processes in the *post* action, but be sure to only touch the ones we started in this script. This can be done in a future PR though.

Implemented. I guess we should use systemd units instead of bash script/cron task :-/ Besides systemd time can use UTC timezone.

#7 - 2017-04-11 12:10 - Marc Dequènes

- *Category deleted (Service :: Backup)*

Pierre-Louis Bonicoli wrote:

None or without value at all, that's the same.

Thanks for clarifying.

I guess it was not the same in the past in Ruby, but all languages seem to use libyaml now so same behavior.

No: cron task is executed on Toshiro which is using PARIS timezone, Korutopi is using Japan timezone.

Not very easy to setup then.

Implemented. I guess we should use systemd units instead of bash script/cron task :-/ Besides systemd time can use UTC timezone.

In the "jobs directory will be ignored" part, why a mkdir and not a touch? It worked with a file with Bacula so it would be nice current config in user data dirs (anywhere on DC) would work too.

I guess using UTC would be better for our heads and avoid calculation errors.

As for using systemd I'll be in favor but I'm still looking for a solution for non-linux systems (like the Hurd). There is a systemd-shim package, but only builds on linux-any systems and is orphaned, so...

#8 - 2017-05-09 16:03 - Pierre-Louis Bonicoli

Marc Dequènes wrote:

In the "jobs directory will be ignored" part, why a mkdir and not a touch? It worked with a file with Bacula so it would be nice current config in user data dirs (anywhere on DC) would work too.

Both file and directory are allowed/recognized.

#9 - 2017-05-09 16:06 - Pierre-Louis Bonicoli

- Related to Enhancement #497: Change Backup System added

#10 - 2017-07-16 08:42 - Marc Dequènes

New review on current code.

Why are cron_*, timer_* and keep_* variables not backup_ prefixed like other such variables. Looks like PG config stuff, and even if now I know there is no such thing in our Ansible config, it might happen one day.

Except from this the config in backup.yml files is clear.

The play on Orfeo:Toushirou is I guess supposed to setup the generic file backup, but I see no roles or tasks defined.

I will discuss the rest in [#519](#) because it is related to the role's API.

#11 - 2017-07-16 10:10 - Marc Dequènes

If i understand well, include_vars in client_specific_conf_from_file.yml loads the files in playbooks/files/burp/Nicecity_*.

So what I see is a way to prepare the parameters for burp using *_vars parameters and a higher level template. This is not bad but does not allow to share config between similar hosts easily (for example all DB hosts). Why not generate the entry in backups for all hosts of a group instead? (using moving the content of playbooks/files/burp/Nicecity_Orfeo_postgres_db.yml into group_vars/postgresql_servers/backup.yml in the backups tree for example)

#12 - 2017-08-26 09:51 - Marc Dequènes

So I've got this error:

```
TASK [burp : Render BURP clientconfdir configuration] *****
*****
fatal: [Toushirou -> nicecity.duckcorp.org]: FAILED! => {"changed": false, "failed": true, "msg": "AnsibleUndefinedVariable: 'dict object' has no attribute u'keep_db_mysql'"}
*****
```

Unfortunately the file `playbooks/files/burp/Nicecity_Toushirou_root_db.yml` is both caught is the file backup and database backup, and `playbooks/templates/burp/duckcorp_ccd.conf` used for file backup does not cope well with the defined parameters because this file is made with the database backup in mind.

As the client name is used for the user doing the backup, I'm not sure how to solve this.

#13 - 2017-08-26 10:53 - Marc Dequènes

I pushed the `backup_duck` branch with my test changes. I just moved Toushirou's db config file away and back again to work around the previous bug and was able to update all configs.

Results:

- PG backup on Orfeo worked fine
- MySQL backup on Toushirou hang: Burp hang when running the pre script while manually run it works well, no idea what's going on. I disabled the crontab for the time being
- file backup works well

So I need to check things more in depth and also see if automatic cron backups are fine. I will also add the other hosts soon.

#14 - 2017-08-28 18:33 - Marc Dequènes

The exclude paths were not taken into account. I pushed a change to fix this and other related problems. It does not have to be the final solution but I hope to get the backup done properly without taking ages (and more TB than Pilou is OK to provide).

#15 - 2017-08-29 01:38 - Pierre-Louis Bonicoli

Marc Dequènes wrote:

- MySQL backup on Toushirou hang: Burp hang when running the pre script while manually run it works well, no idea what's going on. I disabled the crontab for the time being

Fixed by `8d12debba60888f655a3d246a4af2d082a8c2b1d`.

#16 - 2017-08-29 01:49 - Pierre-Louis Bonicoli

Marc Dequènes wrote:

So I've got this error:

```
TASK [burp : Render BURP clientconfdir configuration]
fatal: [Toushirou -> nicecity.duckcorp.org]: FAILED! => {"changed": false, "failed": true, "msg": "AnsibleUndefinedVariable: 'dict object' has no attribute u'keep_db_mysql'"}
```

submodule wasn't up to date, this error was fixed by `406a38a47748886eb9ade4e540dc8e05d1fa325b`.

#17 - 2017-08-29 04:49 - Marc Dequènes

Could you have a look at [#10](#) and [#11](#) remarks when you have time too (no urgency).

I rebased my branch to get your fixes. There's no urgency to review it, but keep in mind these are production settings (instead of the original branch), so please deploy using it. I did this while you're busy but of course everything is up to discussion.

#18 - 2017-08-29 05:26 - Marc Dequènes

MySQL backup is working fine, thanks a lot :-).

#19 - 2017-08-29 11:49 - Pierre-Louis Bonicoli

Marc Dequènes wrote:

New review on current code.

Why are `cron_*`, `timer_*` and `keep_*` variables not `backup_*` prefixed like other such variables. Looks like PG config stuff, and even if now I know there is no such thing in our Ansible config, it might happen one day.

Updated (f2a79cc24f954fda184a22cf31c40255a272186f/053921e500b6948692245441ba7748b4da5accfd).

The play on Orfeo:Toushirou is I guess supposed to setup the generic file backup, but I see no roles or tasks defined.

Fixed (df297d537c2449928359ec67906604c1f46be4d4)

I will discuss the rest in [#519](#) because it is related to the role's API.

#20 - 2017-08-29 12:24 - Pierre-Louis Bonicoli

Marc Dequènes wrote:

If i understand well, `include_vars` in `client_specific_conf_from_file.yml` loads the files in `playbooks/files/burp/Nicecity_*`.

So what I see is a way to prepare the parameters for burp using `*_vars` parameters and a higher level template. This is not bad but does not allow to share config between similar hosts easily (for example all DB hosts). Why not generate the entry in backups for all hosts of a group instead? (using moving the content of `playbooks/files/burp/Nicecity_Orfeo_postgres_db.yml` into `group_vars/postgresql_servers/backup.yml` in the backups tree for example)

In order to share configuration:

```
# One client is able to have many client configurations:

# ... defined by variables
- include: client_specific_conf.yml
  with_dict: '{{ client.value.backups }}'
  loop_control: { loop_var: 'backup' }
  when: '"backups" in client.value'

# ... and/or defined in YAML files
- include: client_specific_conf_from_file.yml
  # FIXME: one client is able to define only one configuration using YAML file
  # TODO: username should not be key OR use add one level in backups dict
  with_fileglob: 'burp/{{ burp_clients._server }}_{{ inventory_hostname }}_{{ client.key }}_*.yml'
  loop_control: { loop_var: 'backup_conf_path' }
  when: '"backups" not in client.value'
```

- variables can be used
- symlinks can be used (when configuration is defined in YAML file).

#21 - 2017-09-29 18:11 - Marc Dequènes

Beware the branch is still in the original repository, so it should be ported to the new Open Infra one.

As for sharing, I think it will be easier for me to grasp it when we have a real life example. Thanks for the explanation though.

It would be nice to finish polishing the remaining few things and merge. The other improvements can be done later. This would avoid too much divergence and more future work,

#22 - 2017-09-29 18:12 - Marc Dequènes

- Assignee changed from Marc Dequènes to Pierre-Louis Bonicoli