

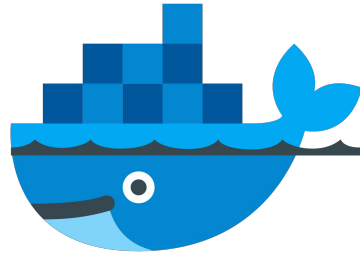
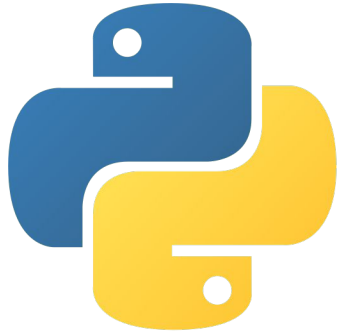
Готовим PostgreSQL в эпоху DevOps. Опыт 2ГИС

Павел Молявин, 2ГИС

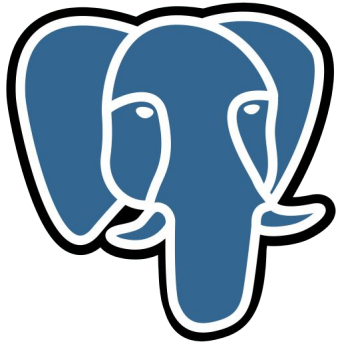


- Отдел веб-разработки

- Отдел веб-разработки
- Infrastructure & Operations



Stack

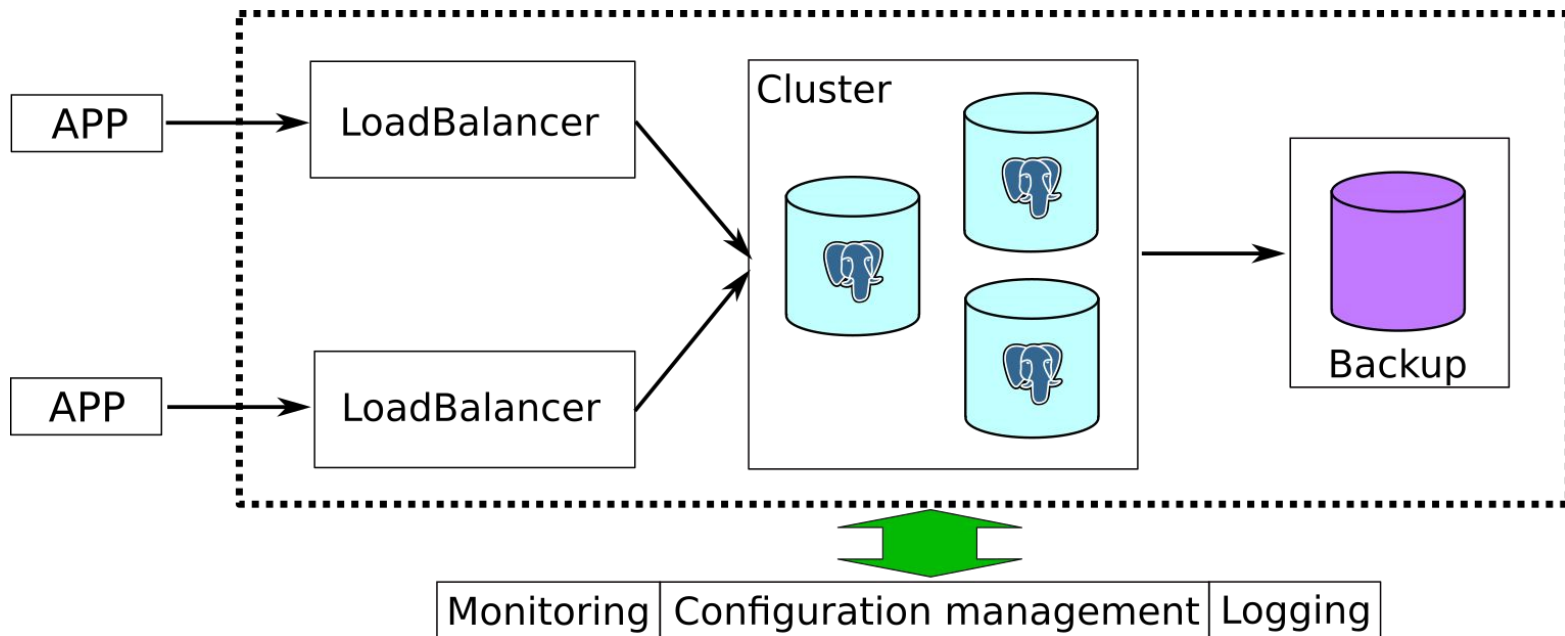


Командное селфи

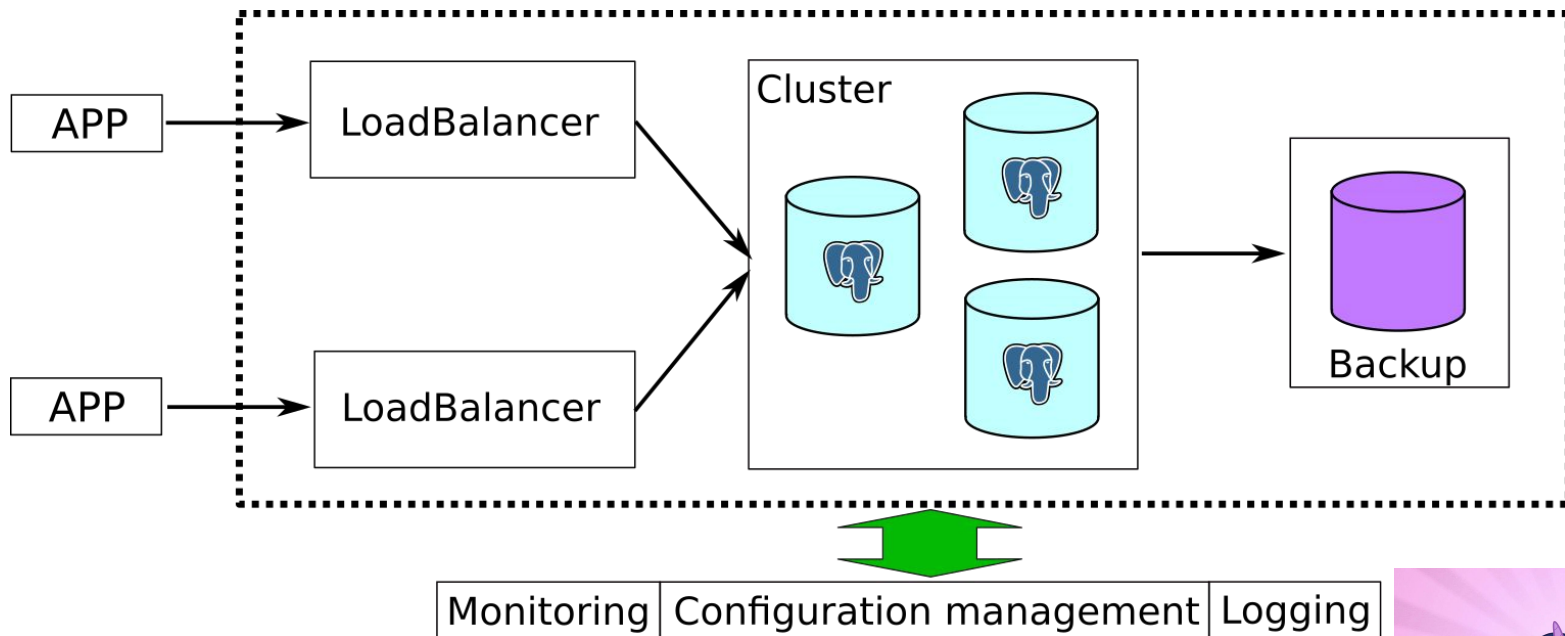
Командное селфи



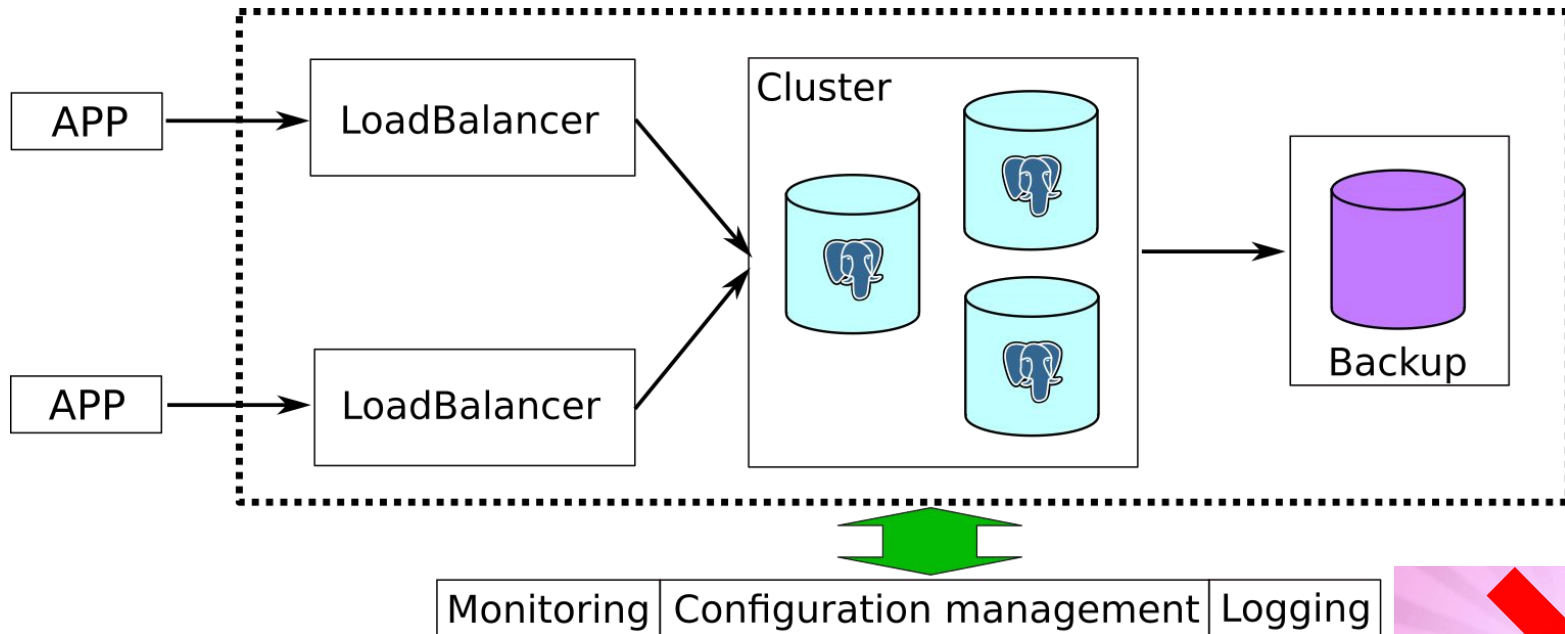
О чем это всё



О чем это всё



О чем это всё



Постановка задачи

Постановка задачи



Постановка задачи



App

App

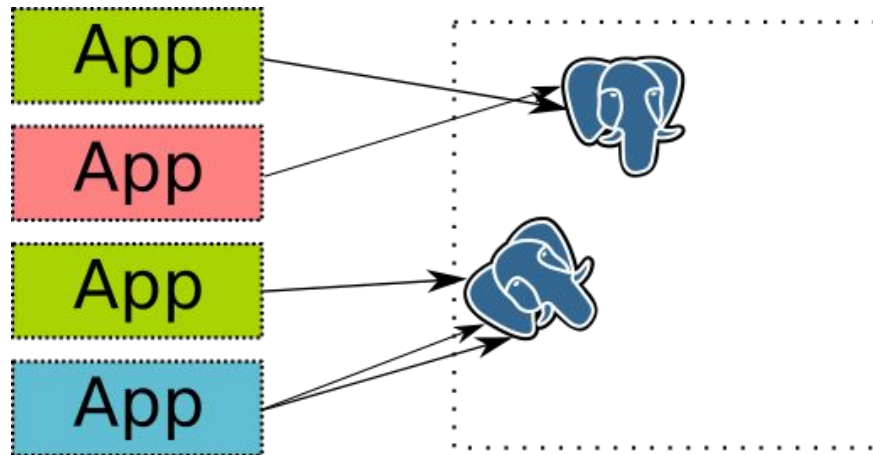
App

App

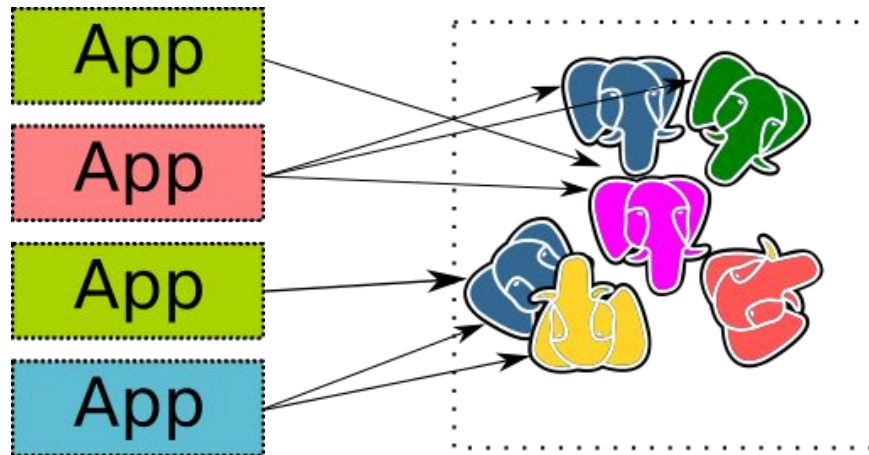
Постановка задачи



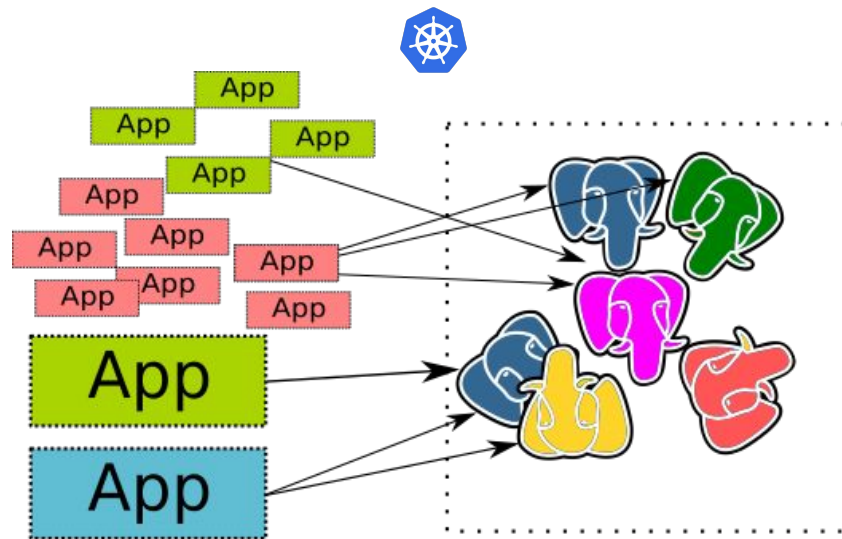
Постановка задачи



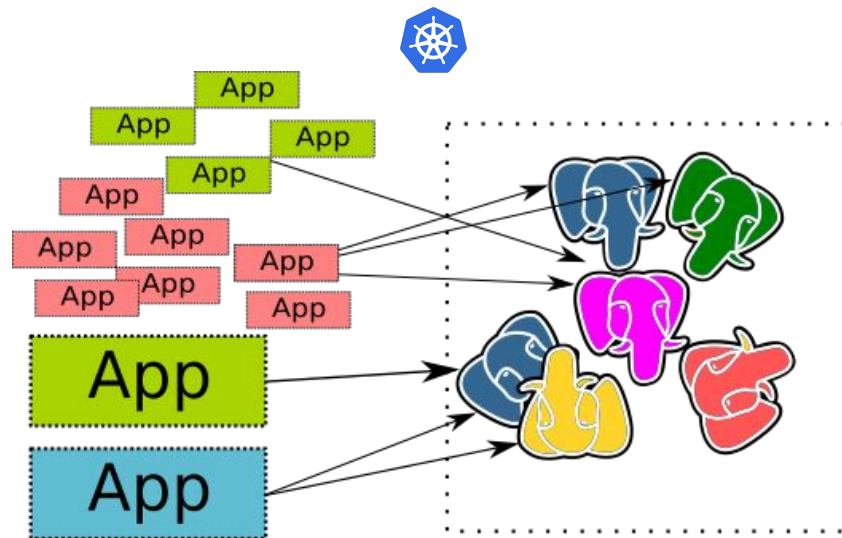
Постановка задачи



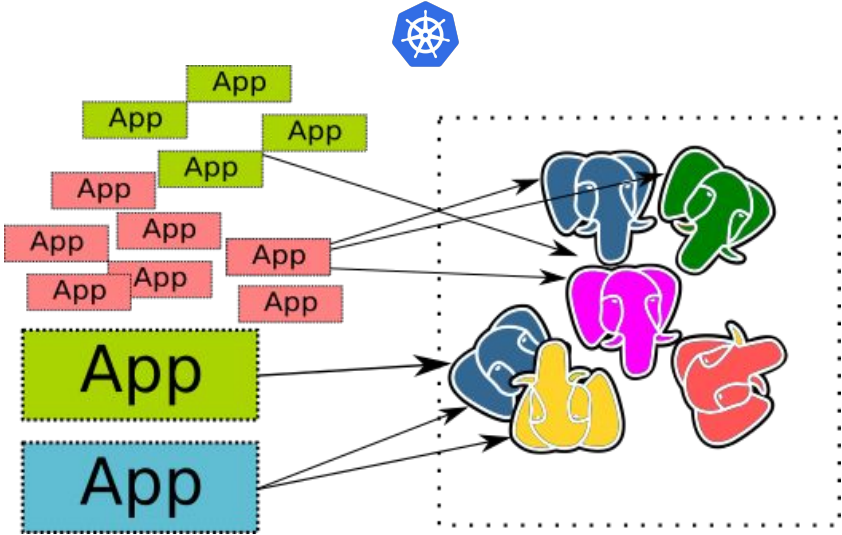
Постановка задачи



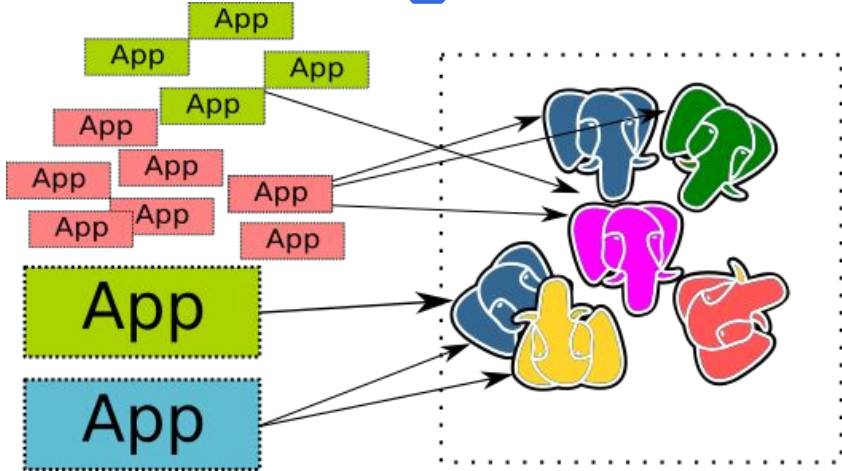
Постановка задачи



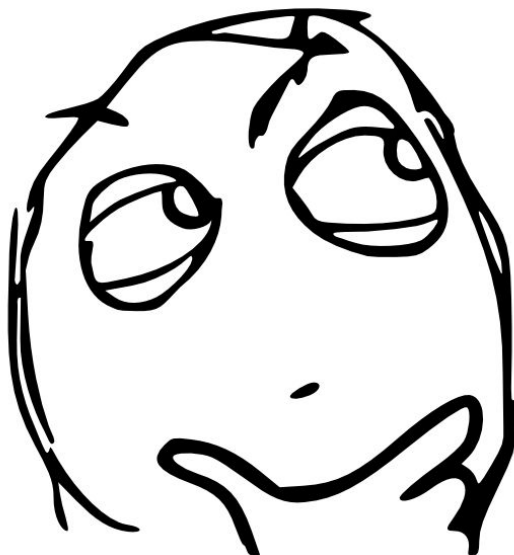
Постановка задачи



Постановка задачи



Что мы решили сделать



Кластер



Кластер



Балансировка



Кластер



Балансировка



Резервные копии



Кластер



Балансировка



Резервные копии



Мониторинг и логи



Кластер



Балансировка



Резервные копии



Мониторинг и логи



Деплой



Кластер



Балансировка



Резервные копии



Мониторинг и логи



Деплой



CI-CD



Мы здесь



Кластер



Балансировка



Резервные копии



Мониторинг и логи



Деплой



CI-CD



- PostgreSQL 9.4 (9.6)





- PostgreSQL 9.4 (9.6)
- Поточковая репликация



- PostgreSQL 9.4 (9.6)
- Поточковая репликация
- Кластеризация — repmgr от 2ndQuadrant

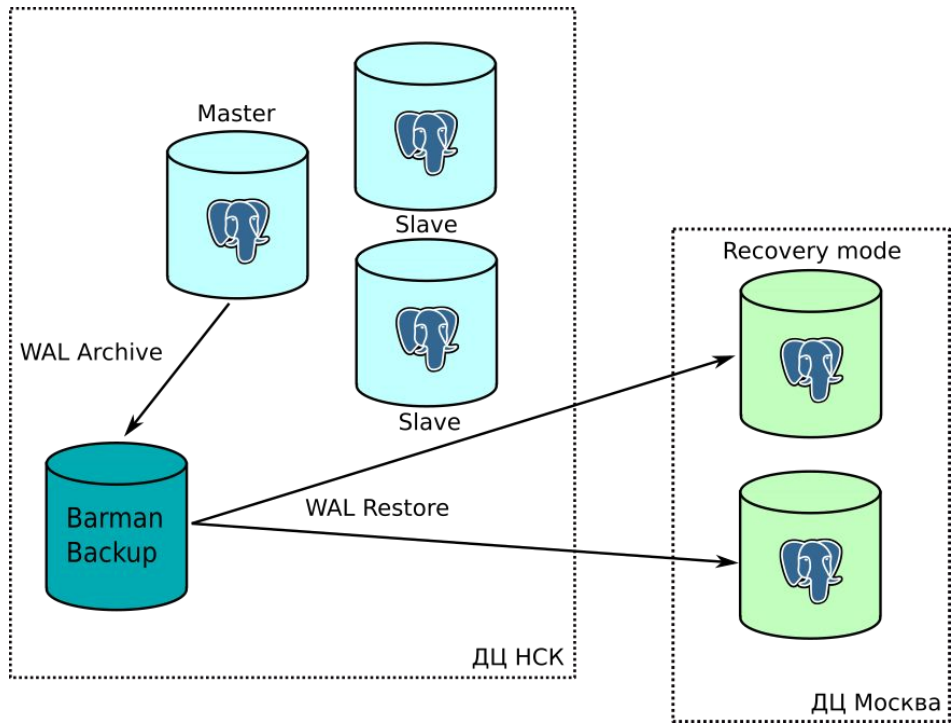




- PostgreSQL 9.4 (9.6)
- Поточковая репликация
- Кластеризация — repmgr от 2ndQuadrant
- Кастомный failover_command



- PostgreSQL 9.4 (9.6)
- Поточковая репликация
- Кластеризация — repmgr от 2ndQuadrant
- Кастомный failover_command
- RO реплика в другом ДЦ





- PostgreSQL 9.4 (9.6)
- Поточковая репликация
- Кластеризация — repmgr от 2ndQuadrant
- Кастомный failover_command
- RO реплика в другом ДЦ
- Проблемы

Мы здесь



Кластер



Балансировка



Резервные копии



Мониторинг и логи

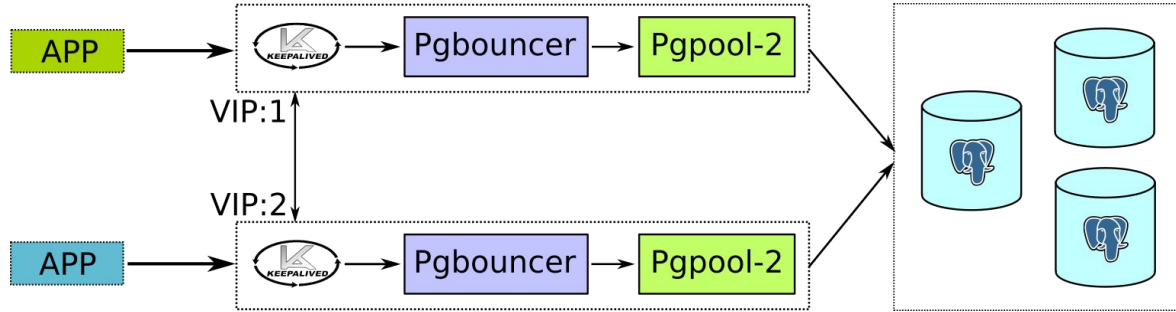


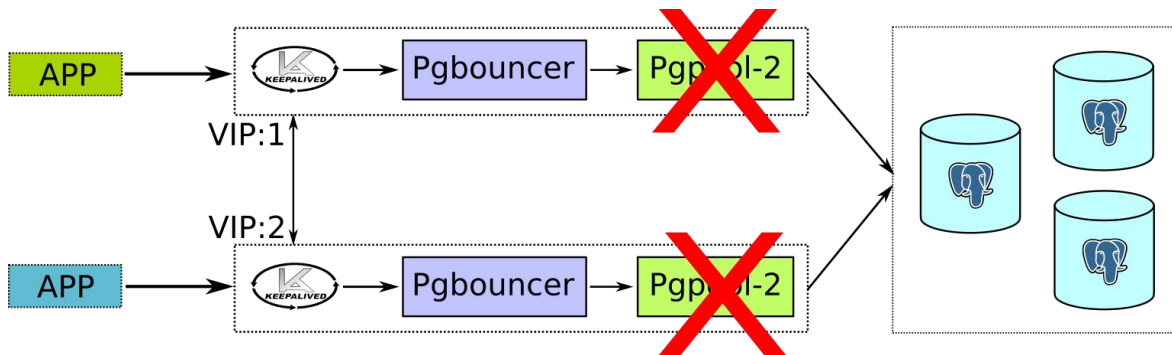
Деплой

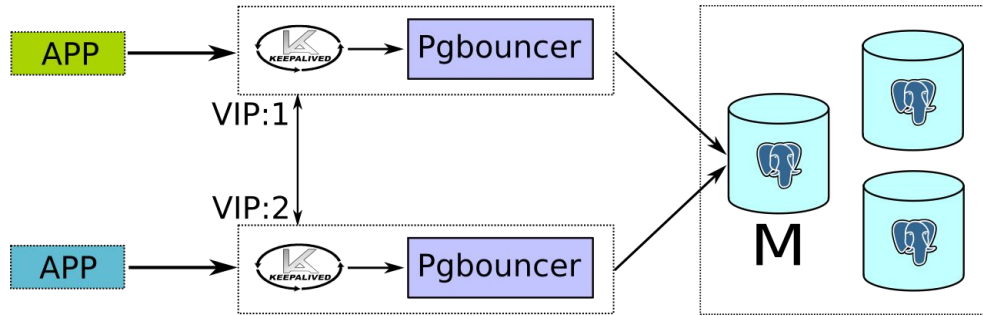


CI-CD

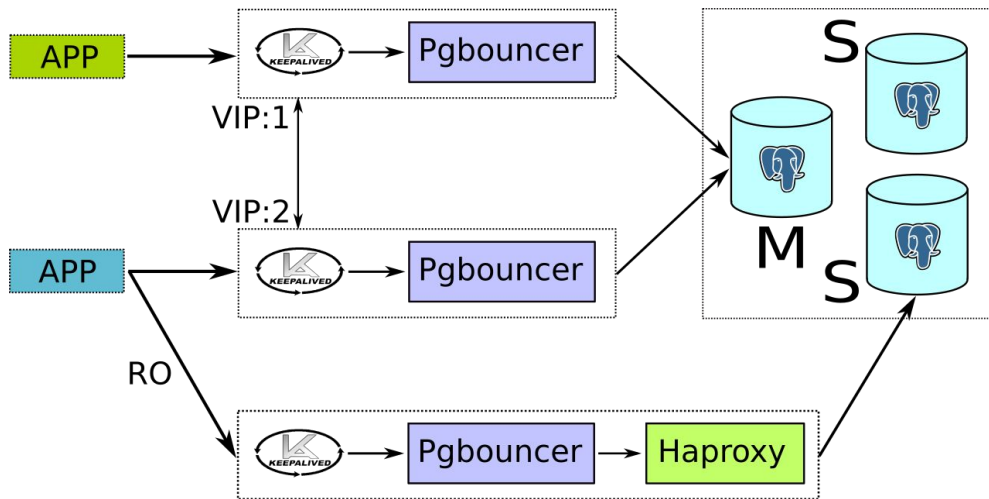








*Failover_command переключает бекенд



Мы здесь



Кластер



Балансировка



Резервные копии



Мониторинг и логи



Деплой



CI-CD



Barman



Barman

- Base backup





Barman

- Base backup
- WAL Archiving для возможности PITR





Barman

- Base backup
- WAL Archiving для возможности PITR

Проверка backup-ов





Barman

- Base backup
- WAL Archiving для возможности PITR

Проверка backup-ов

Недостатки

- Медленный archive_command
- Много места



Кластер



Балансировка



Резервные копии



Мы здесь



Мониторинг и логи



Деплой



CI-CD





- Интеграция с Prometheus



- Интеграция с Prometheus
- Golang Python Bash



- Интеграция с Prometheus
- Golang Python Bash
- Node, Postgres, Pgboncer, Cgroups exporters

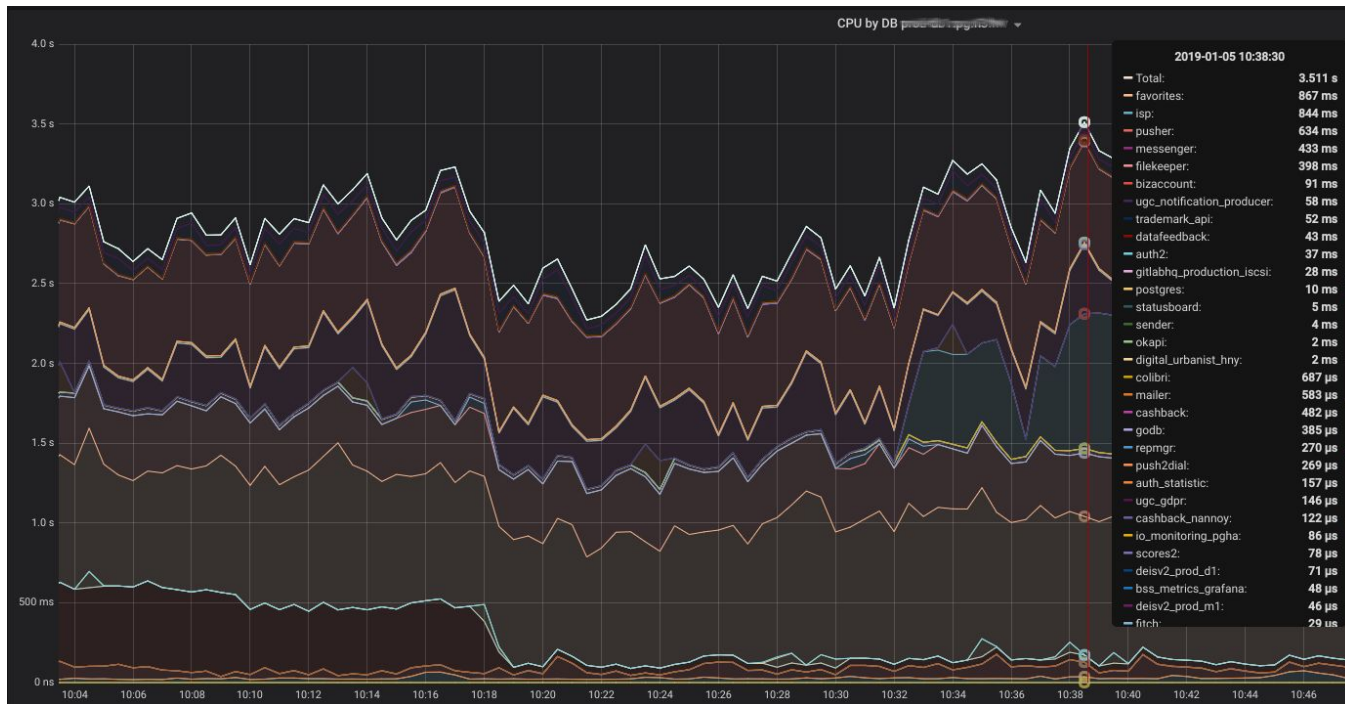


- Интеграция с Prometheus
- Golang Python Bash
- Node, Postgres, Pgrounder, Cgroups exporters
- Системные метрики (CPU, RAM, IO)



- Интеграция с Prometheus
- Golang Python Bash
- Node, Postgres, Pgrounder, Cgroups exporters
- Системные метрики (CPU, RAM, IO)
- Метрики всех компонентов

Метрики в Grafana



Alerts



Alerts



AlertManager APP 6:39 PM

[FIRING:2] PghaMasterIsNotWritableOne pgsq1-ha-stage-ng postgresql-ha (io-slack n3 loadbalancer critical node)



AlertManager APP 6:59 PM

[RESOLVED] PghaMasterIsNotWritableOne pgsq1-ha-stage-ng postgresql-ha (io-slack n3 loadbalancer critical node)

Alerts



AlertManager APP 6:39 PM

[FIRING:2] PghaMasterIsNotWritableOne pgsq1-ha-stage-ng postgresql-ha (io-slack n3 loadbalancer critical node)



AlertManager APP 6:59 PM

[RESOLVED] PghaMasterIsNotWritableOne pgsq1-ha-stage-ng postgresql-ha (io-slack n3 loadbalancer critical node)

- Failover

Alerts



AlertManager APP 6:39 PM

[FIRING:2] PghaMasterIsNotWritableOne pgsq1-ha-stage-ng postgresql-ha (io-slack n3 loadbalancer critical node)



AlertManager APP 6:59 PM

[RESOLVED] PghaMasterIsNotWritableOne pgsq1-ha-stage-ng postgresql-ha (io-slack n3 loadbalancer critical node)

- Failover
- Pg_settings

Alerts



AlertManager APP 6:39 PM

[FIRING:2] PghaMasterIsNotWritableOne pgsql-ha-stage-ng postgresql-ha (io-slack n3 loadbalancer critical node)



AlertManager APP 6:59 PM

[RESOLVED] PghaMasterIsNotWritableOne pgsql-ha-stage-ng postgresql-ha (io-slack n3 loadbalancer critical node)

- Failover
- Pg_settings
- Лаг репликации

Alerts



AlertManager APP 6:39 PM

[FIRING:2] PghaMasterIsNotWritableOne pgsql-ha-stage-ng postgresql-ha (io-slack n3 loadbalancer critical node)



AlertManager APP 6:59 PM

[RESOLVED] PghaMasterIsNotWritableOne pgsql-ha-stage-ng postgresql-ha (io-slack n3 loadbalancer critical node)

- Failover
- Pg_settings
- Лаг репликации
- Тестовое восстановление

Alerts & Playbooks



```
1 - alert: PghaPostgresServicePendingRestart
2   expr: pg_service_pending_restart{cluster=~"(?cluster='postgres|postgres1|postgres2|postgres3|postgres4|postgres5|postgres6|postgres7|postgres8|postgres9|postgres10|postgres11|postgres12|postgres13|postgres14|postgres15|postgres16|postgres17|postgres18|postgres19|postgres20|postgres21|postgres22|postgres23|postgres24|postgres25|postgres26|postgres27|postgres28|postgres29|postgres30|postgres31|postgres32|postgres33|postgres34|postgres35|postgres36|postgres37|postgres38|postgres39|postgres40|postgres41|postgres42|postgres43|postgres44|postgres45|postgres46|postgres47|postgres48|postgres49|postgres50|postgres51|postgres52|postgres53|postgres54|postgres55|postgres56|postgres57|postgres58|postgres59|postgres60|postgres61|postgres62|postgres63|postgres64|postgres65|postgres66|postgres67|postgres68|postgres69|postgres70|postgres71|postgres72|postgres73|postgres74|postgres75|postgres76|postgres77|postgres78|postgres79|postgres80|postgres81|postgres82|postgres83|postgres84|postgres85|postgres86|postgres87|postgres88|postgres89|postgres90|postgres91|postgres92|postgres93|postgres94|postgres95|postgres96|postgres97|postgres98|postgres99|postgres100')"} > 0
3   for: 3h
4   labels:
5     severity: major
6   annotations:
7     description: Сервис postgresql требует перезапуска в связи с тем, что в конфигурационный
8     файл были внесены изменения, требующие перезапуска. Данная операция является
9     ОПАСНОЙ в случае применения на МАСТЕРЕ, поэтому необходимо точно убедиться,
10    что перезапуск необходим. Особенно, если данный алерт пришел _неожиданно_.
11    Для этого, надо на сервере сделать выборку из view pg_settings where pending_restart='t',
12    таме будет указывать на имя директивы, которая была изменена.
13    summary: Postgresql service is pending restart on {{ $labels.instance }}.
```

Кластер



Балансировка



Резервные копии



Мы здесь



Мониторинг и логи



Деплой

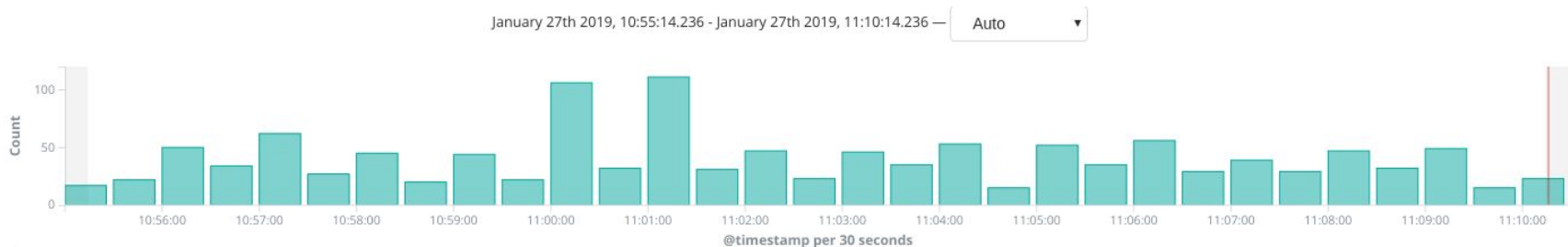


CI-CD





- PostgreSQL log CSV -> Python Beaver -> Logstash -> Elasticsearch
- Pgbadger



Time

_source

```
▶ January 27th 2019, 10:55:15.321 project: pgha detail: - log_time: 2019-01-27 03:55:15.321 UTC user_name: - database_name: - query: - session_line_num: 383,958 location: -
query_pos: - internal_query: - cluster_name: pgha-prod @timestamp: January 27th 2019, 10:55:15.321 tags: command_tag: - hint: -
error_severity: LOG type: pgsq-log application_name: sql_state_code: 00000 team: io process_id: 1,353 session_id: 5c3576a1.549
transaction_id: 0 @version: 1 context: - host: prod-db2-msk file: /var/log/postgresql/postgresql-9.6-main.csv internal_query_pos: -
session_start_time: 2019-01-09 04:20:49 UTC connection_from: - virtual_transaction_id: 1/0 text_message: restored log file "000000200003E6C000000E
```

Кластер



Балансировка



Резервные копии



Мы здесь



Мониторинг и логи



Деплой



CI-CD





- Ansible. 25 ролей. 20 самописных. 12 вырожденных



- Ansible. 25 ролей. 20 самописных. 12 вырожденных
- Environments



- Ansible. 25 ролей. 20 самописных. 12 вырожденных
- Environments
- Секреты



- Ansible. 25 ролей. 20 самописных. 12 вырожденных
- Environments
- Секреты
- Установка из apt



- Ansible. 25 ролей. 20 самописных. 12 вырожденных
- Environments
- Секреты
- Установка из apt
- Рутинные операции



- Ansible. 25 ролей. 20 самописных. 12 вырожденных
- Environments
- Секреты
- Установка из apt
- Рутинные операции
- Git



- Ansible. 25 ролей. 20 самописных. 12 вырожденных
- Environments
- Секреты
- Установка из apt
- Рутинные операции
- Git
- MR



- Ansible. 25 ролей. 20 самописных. 12 вырожденных
- Environments
- Секреты
- Установка из apt
- Рутинные операции
- Git
- MR
- Bootstrap для bare-metal, Openstack

Тесты

- Ansible




Тесты



- Ansible 
ANSIBLE
- Запуск на тестовом окружении

Тесты



- Ansible 
ANSIBLE
- Запуск на тестовом окружении
- Проверяем
 - ⚡ Запуск кластера
 - ⚡ Failover
 - ⚡ Бэкап, восстановление

Кластер



Балансировка



Резервные копии



Мониторинг и логи



Деплой



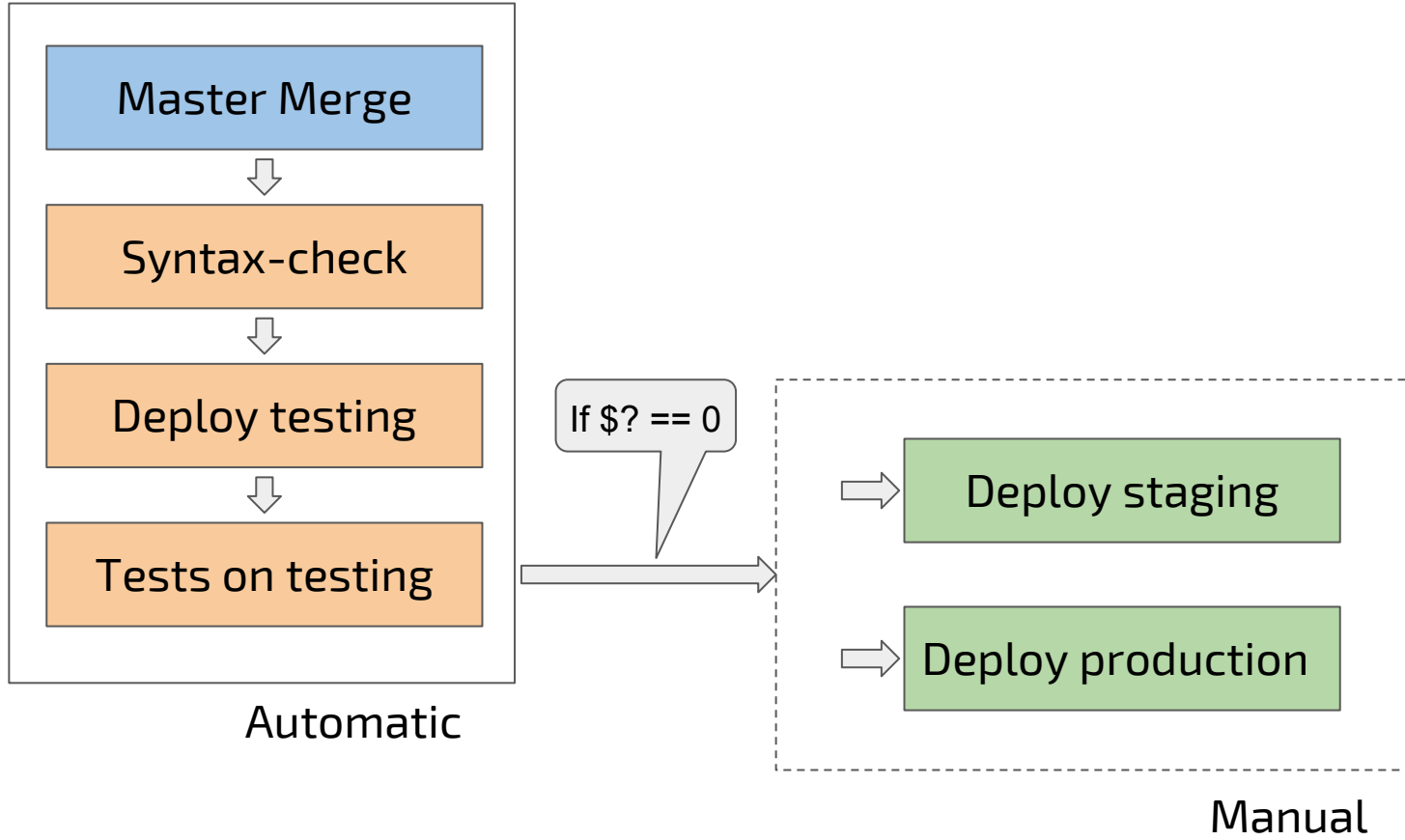
Мы здесь



CI-CD



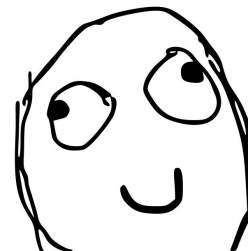
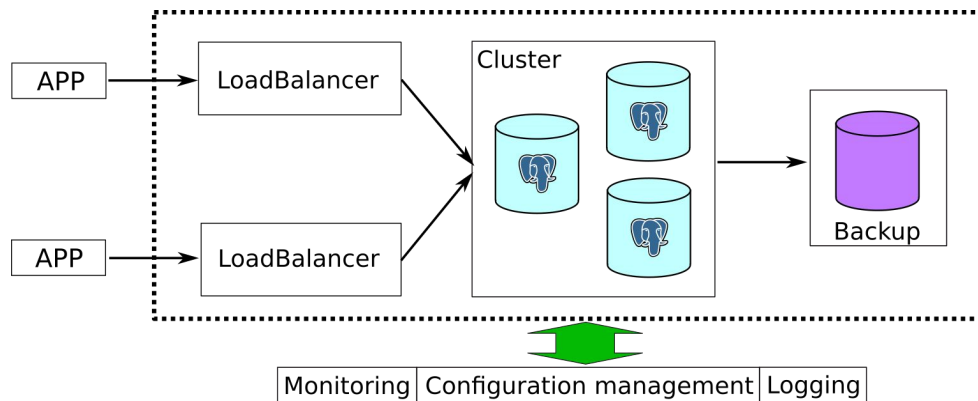




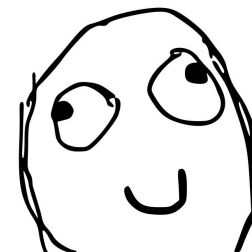
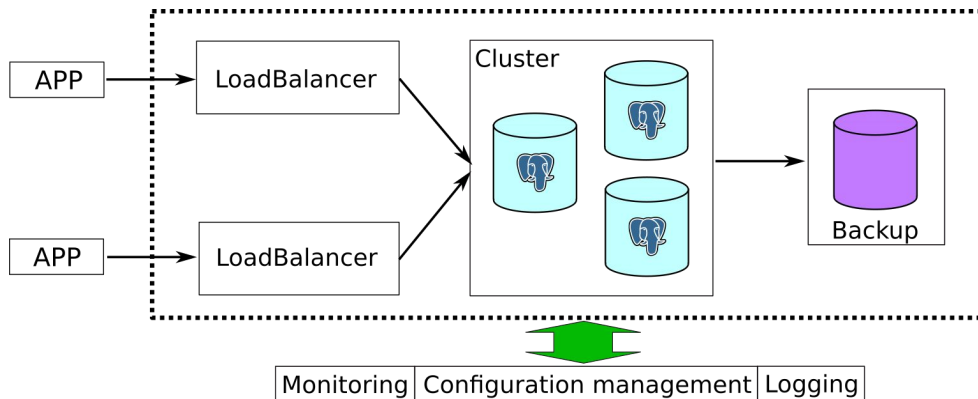
Что удалось



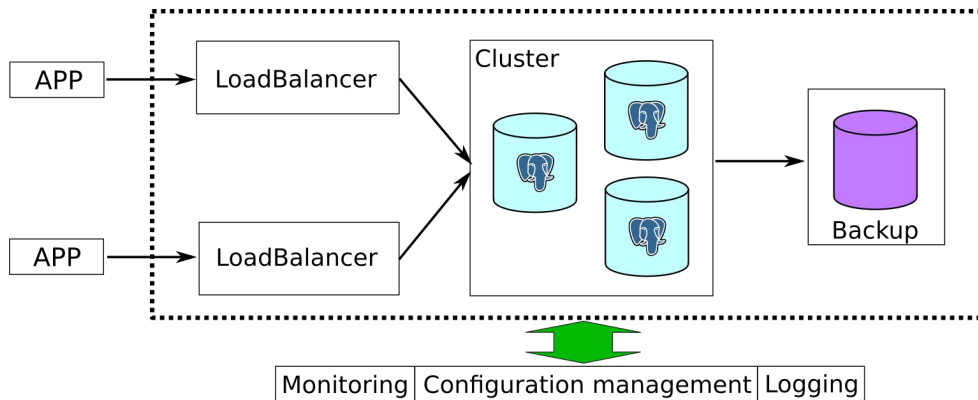
Что удалось



Что удалось



Что удалось



15 мин.



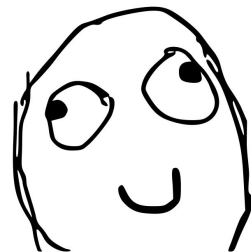
Что удалось

✓ Несколько кластеров



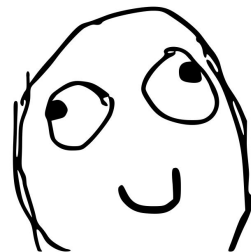
Что удалось

- ✓ Несколько кластеров
- ✓ --Поддержка командных СУБД



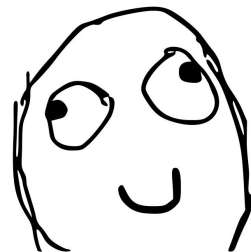
Что удалось

- ✓ Несколько кластеров
- ✓ --Поддержка командных СУБД
- ✓ XР++



Что удалось

- ✓ Несколько кластеров
- ✓ --Поддержка командных СУБД
- ✓ ХР++
- ✓ Failover работает, мы проверяли, и спим спокойно по ночам :-)



Что не удалось



Что не удалось

X Сделать решение простым



Что не удалось

X Сделать решение простым

Fork 0



Что не удалось

- ✗ Сделать решение простым
- ✗ Приходится дорабатывать

Fork 0



Что не удалось

- ✗ Сделать решение простым
- ✗ Приходится дорабатывать
- ✗ Полностью избавиться от ручных операций

Fork 0



Что не удалось

- ✗ Сделать решение простым
- ✗ Приходится дорабатывать
- ✗ Полностью избавиться от ручных операций
- ✗ Обновление — отдельная боль

Fork 0



Что не удалось

- ✗ Сделать решение простым
- ✗ Приходится дорабатывать
- ✗ Полностью избавиться от ручных операций
- ✗ Обновление — отдельная боль
- ✗ Приходится самим собирать пакеты

Fork 0



Что не удалось

- ✗ Сделать решение простым
- ✗ Приходится дорабатывать
- ✗ Полностью избавиться от ручных операций
- ✗ Обновление — отдельная боль
- ✗ Приходится самим собирать пакеты
- ✗ Cattle vs. pet

Fork 0



Куда дальше?


Куда дальше?

- Stolon, Patroni 


Куда дальше?

- Stolon, Patroni 
- Решение без кластера

Куда дальше?

- Stolon, Patroni 
- Решение без кластера
- Апгрейд на PostgreSQL 10-11

Куда дальше?

- Stolon, Patroni 
- Решение без кластера
- Апгрейд на PostgreSQL 10-11

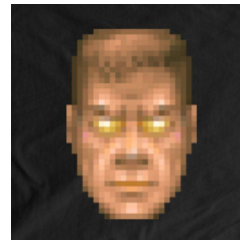




Берегите свой Postgres!

p.molyavin@2gis.ru

Секретный уровень



- 3 кластера
- ~80 баз
- ~1,2 TB данных
- ~3895 хacts/s, 6310 queries/s
- ~3,5 млн аутентифицированных пользователей у сервисов в production
- ~10 млн анонимных пользователей

Типичный конфиг для сервера БД:

- 2 x XEON Gold 6134
- 128 GB RAM
- 4 x 900 GB SSD