Immutable Database Infrastructure with PXC

Satoshi Mitani | @mita2
Yahoo Japan Corporation
Agenda

- Yahoo! JAPAN Introduction
- Demo
- What is Immutable Infrastructure
- Architecture
- Why Percona XtraDB Cluster
- Disadvantages of our method
- Q&A
Yahoo! JAPAN Introduction
Yahoo! JAPAN Introduction
Yahoo! JAPAN Introduction

- **Monthly Page Views**: 70+ Billion
- **Number of services**: 100+

- **Daily Unique Browser**: 90+ Million
- **Daily Unique Browser (Only Smartphone)**: 60+ Million
Demo
Demo

- Our steps to release new software
  1. Take the node offline
  2. Rebuild the node with image including new software
  3. Bring the node back online
What is Immutable Infrastructure?
Legacy infrastructure (Mutable)

- Accumulated changes
- Long life-span

- Advantage
  - Existing Infrastructure
  - Persistent

- Disadvantage
  - Need to track states
  - Need to upgrade perfectly
  - Difficult to test all combinations
Immutable Infrastructure

- Does not change after creation
- Disposable
- Replace servers to release new features
- Short life-span

- Advantage.
  - Always fresh
  - Less combinations
- Disadvantage
  - Volatile
Why do we need Immutable Infrastructure?

- huge number of DBs
- hard to track state
- hard to test all combination
Architecture
Architecture overview

Image factory:
- GitHub Enterprise
  - Chef recipes
- Screwdriver.cd (CI System)
- IaaS API
- VM

Golden Image

Databases on IaaS:
- Image Repo
- IaaS API
- Config Backup Storage
- XtraDB Cluster
- my.cnf etc...

Image factory

Databases on IaaS
Architecture – Image factory

- Golden Image
  - Include all software
  - PXC
  - Prometheus
  - Fluentd
  - etc..

Image factory

- Chef recipes
- GitHub Enterprise
- Screwdriver.cd (CI System)

IaaS API

VM

Golden Image
1. Update Chef recipe

```ruby
  version [pxc_version, pxc_version]
  action [:install, :lock]
  options '--enablerepo="percona-release"'
end
```

```ruby
cookbook_file
"/etc/systemd/system/mysql.service.d/override.conf" do
  source 'etc/systemd/system/mysqld.service.d/override.conf'
  mode 00444
  owner 'root'
  group 'root'
end
```
Architecture – Image factory

2. Boot new VM
3. Run chef-client
   - chef-client local mode
   - No workstation
   - No server

$ sudo chef-client -z -r "role[some-role]"

Image factory
Architecture – Image factory

- Snapshot VM

4. Create Snapshot

Image factory
Architecture – Image factory

- Tests
  - Based on new Golden Image
  - Creating new Database Cluster
  - Monitoring Process
  - Load Balancing
  - etc...

- Tests are covered by our own python scripts
  - Fabric

5. Tests
Architecture - Database

- Re-imaging clears all data
  - MySQL configuration
  - OS configuration
  - MySQL data
  - etc..

- MySQL configuration files

- Other OS configuration files
  - network-scripts/if-cfg, /etc/hosts etc..
  - Generated automatically by IaaS

Database
Architecture - Database

- Database consists of 3 nodes
- Re-imaging the node one by one
  - To avoid downtime
- Pass the backed up config file to rebuild OpenStack IaaS API

6. Rebuild
Why Percona XtraDB Cluster
Our maintenance requirements

• No downtime

• Anytime, without scheduling
Percona XtraDB Cluster (PXC)

• MySQL compatible High-availability solution

• Multi-writer
  • Galera replication

• Automatic data recovery
  • State Snapshot Transfer (SST)
Zero-downtime maintenance

- Taking node offline before re-imaging
- Wait for all client connections move to others
- Possible write across the nodes
- PXC supports multi-writer
SST - Automatic data recovery

• All data cleared by re-imaging

• State Snapshot Transfer
  • full data copy from one node to the joining node
Disadvantages of our method
SST Problem (1)

1. SST compatibility issue between 5.7.22 and before 5.7.21
   • If you have TDE tables (ENCRIPTIONION=Y)
   • Need to upgrade all node before SST

2. SST failed with TDE and Compressed Table
   • ENCRYPTION=Y, ROW_FORMAT=COMPRESSED
   • Will be fixed in next Percona XtraBackup release 2.4.15
   • [https://jira.percona.com/browse/PXB-1867](https://jira.percona.com/browse/PXB-1867)
3. SST blocks DDL
   • Not a bug 😨
   • xtrabackup runs with –lock-ddl for safety
   • App with frequent DDL faces this problem
Disadvantage of our method

• PXC has some limitations

• Deploy takes much time
  • Emergency release by manual operation

• Limited volume
  • Large data cause long SST
  • We limited < 500GB
Thank you