

Managing configuration drifts in large computing infrastructures: an experimental approach at CERN

Student: Andrea Giardini

Supervisor: Prof. Anna Ciampolini

Mentor: Ben Dylan Jones

December 19, 2016

University of Bologna

Outline

Introduction

CERN

Configuration management

Package Inventory

Project Structure

Results

Continuous Integration

Project Structure

Results

Conclusions

Introduction

- European Organization for Nuclear Research
- Situated in the border between Switzerland and France
- 21 Member states
- Big challenges



Data Centres

Two data centers:

- Budapest
- Geneva

Two dedicated links:

- 2 x 100Gbps

The number of resources is growing year by year. As today:

- 18k servers
- 180PB on tape
- 260PB on disk

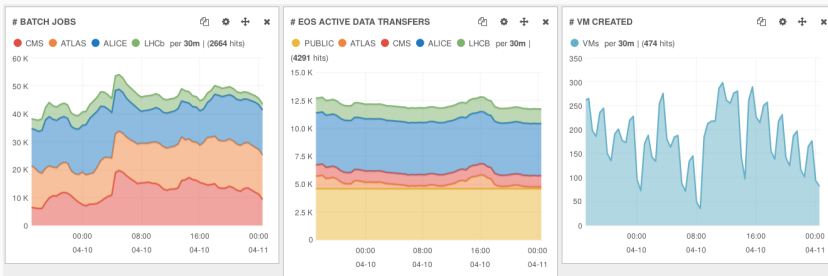


Going Agile

Requirements started to grow

- Agile approach was needed

Since a few years we started using Openstack to deploy virtual machines for our users and Puppet to configure the services



Configuration Management with Puppet

Puppet is an open-source configuration management tool. It is designed to manage the configuration of Unix-like and Microsoft Windows systems declaratively.

- Service configuration
- Users/Groups management
- Automates repetitive tasks



puppet

Problems

- Multiple service managers
- Deployment interval
- Configuration and package drifting

Package Inventory

Introduction

Configuration drifts started to be a problem:

- Servers with outdated packages
- Difficult to spot
- Users forcing their servers not to update

It is not easy to keep all the configuration in sync and guarantee security



The configuration team needed a tool to query packages over a large number of hosts in a timely manner.

Package Inventory is made by three components:

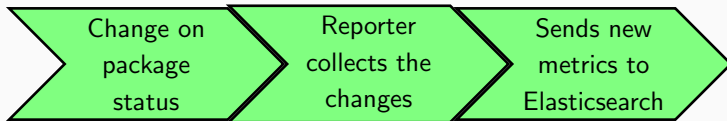
- Reporter
- Cli
- Elasticsearch Cluster

Reporter

Software installed on every server, integrated with Yum using a plugin.

- Reports to Elasticsearch every package change
- Fault-tolerant
- Real-time updates

Elasticsearch stores and indexes all the metrics to make them searchable.



The Cli is used by the users to query the metrics that are stored in the Elasticsearch cluster.

- Possibility to compare clusters of machines
- Package history
- Package version across a set of servers

Optimized to scale for large queries:
Query time to compare over twenty hundreds hosts is two minutes.

```
pkginv -H bi/batch/gridworker/aishare
       -e production
       compare
```

Processing hosts with:

- Hostgroup: bi/batch/gridworker/aishare
- Environment: production

Group 0	Group 1	Group 2
b6b3d71dfa	b6b1576a51	b60ba703db
b636f67ca8	b604a66b3b	b6e493a1b6
b62b21e394	b6c6f66a36	
b65573753a	b67d439ea5	
b6a09855b8		
b6a2bb83c8		
b67af0481a		
b64f0f53c1		
b69b92afd1		
b67f818ae3		
b67cc79825		
b622b1cfdc		

Now that we know which hosts are drifting we can get more details.

Packages can have:

- Different versions
- Different status

Comparing two hosts gives us details about the installed packages.

```
pkginv -m 'b6b3d71dfa b6b1576a51' compare
```

Package	Field	b6b3d71dfa	b6b1576a51
httpd		Present	Not present
httpd-tools		Present	Not present
gridsite		Present	Not present
mod_ssl		Present	Not present
castor-lib	epoch	8.slc6	9.slc6
castor-rfio-client	epoch	8.slc6	9.slc6
castor-ns-client	epoch	8.slc6	9.slc6
castor-devel	epoch	8.slc6	9.slc6

- Installed in more than **55 hundreds of servers**
- Spotted over **two hundreds of servers** out of sync
- Used extensively to monitor the deployment of security updates
- Debugging performance issues

Continuous Integration

Shipping changes to production used to be a manual procedure:

- Multiple actions involved
- Human error

The process could had to be automated:

- Less time wasted
- More organization

Implementation of a Continuous
Integration platform using Jenkins

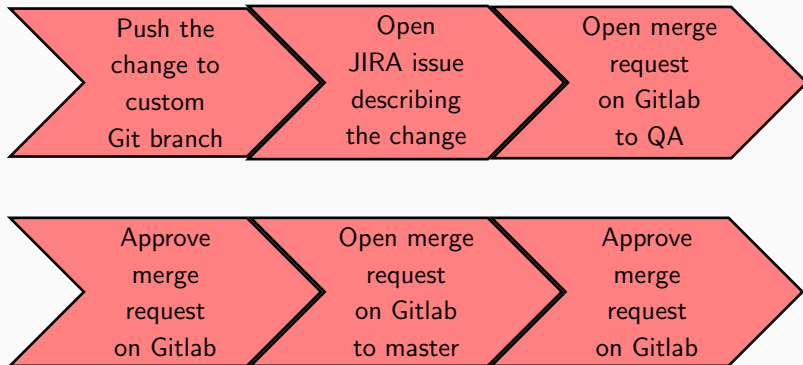
- Open source software
- Highly customizable
- Active community

blabla

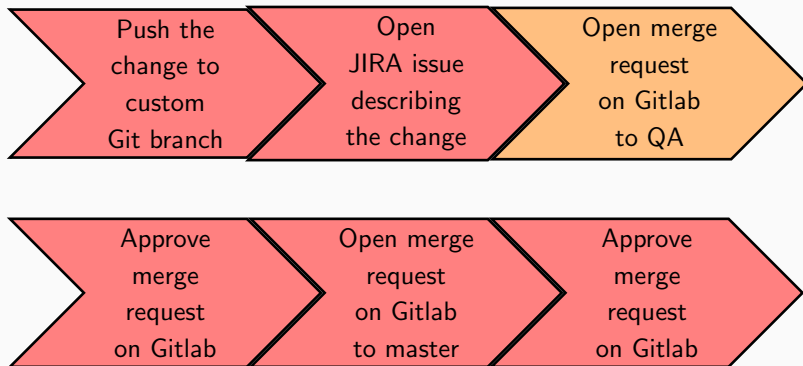


Jenkins

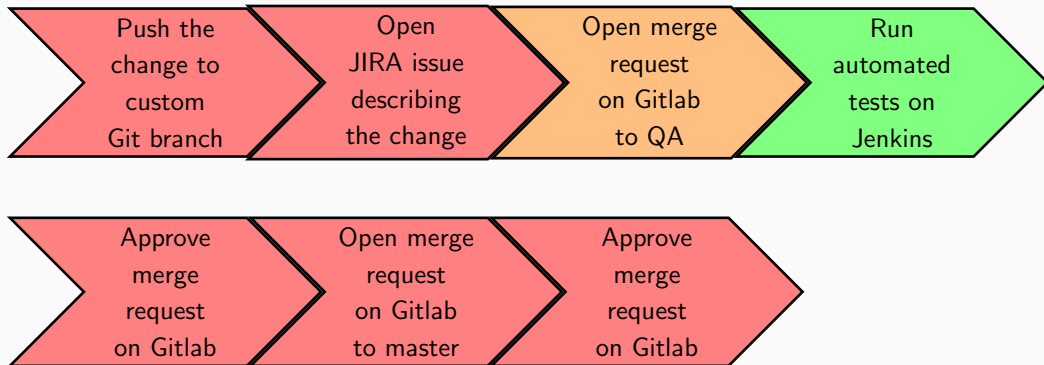
Manual CRM process



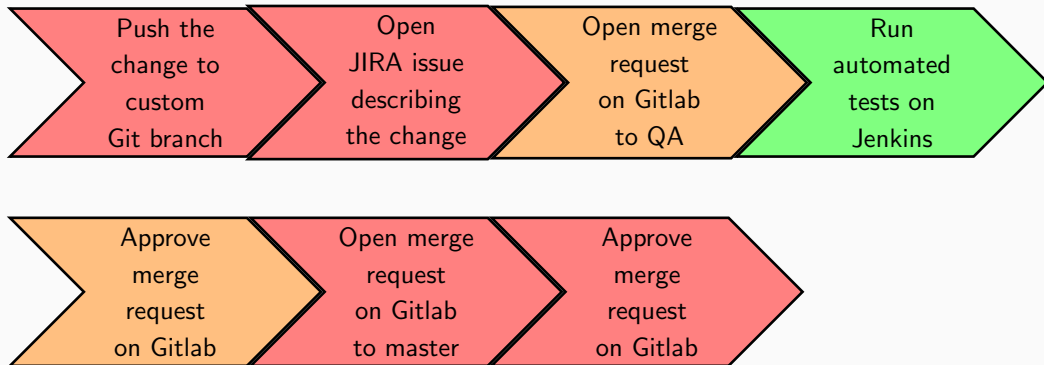
Automated CRM Process



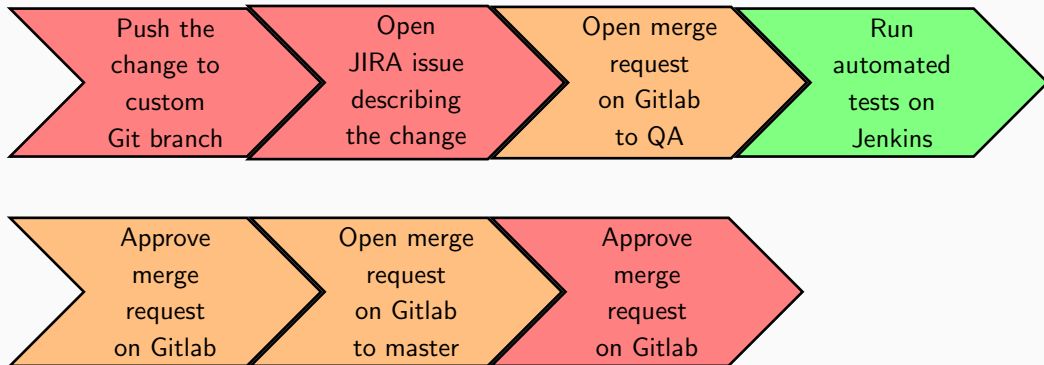
Automated CRM Process



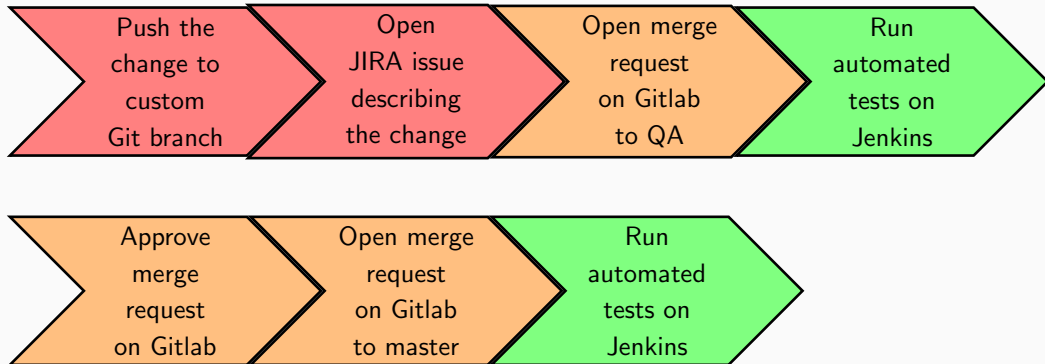
Automated CRM Process



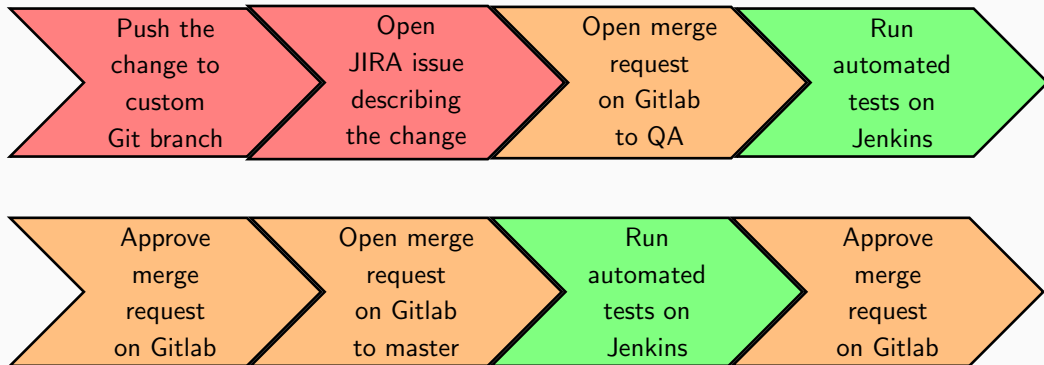
Automated CRM Process



Automated CRM Process



Automated CRM Process



- Extremely customizable infrastructure
 - Users can define their tests
 - Service managers specify how the test server needs to be built
 - Tests can be re-triggered with a comment on JIRA or pushing new commits
- Every service manager is responsible for its Puppet module
- Procedure completely automated
- No need for users to learn how to use Jenkins

Conclusions
