

Hadoop & DevOps : better together

Maxime Lanciaux

DevOps System Architect
February 2017





Get your sketch at tinyurl.com/hortonworks



POWERING THE FUTURE OF DATA™

Agenda



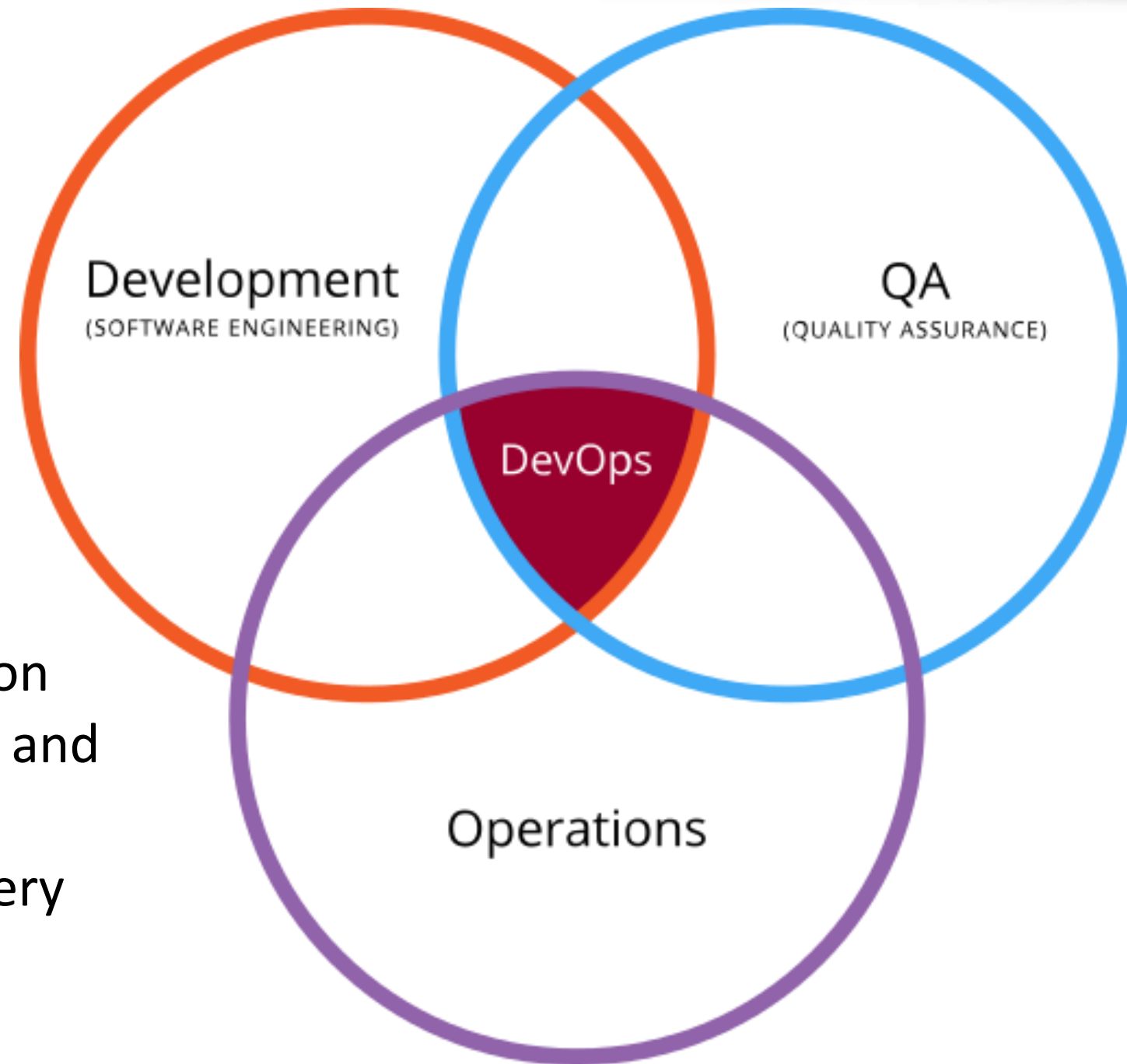
Automate & Speed up deployment

Automate & Speed up deployment

Continuous integration workflow

Dev + Ops

Set of practices that emphasize the collaboration and communication of both software developers and information technology professionals while automating the process of software delivery and infrastructure changes.

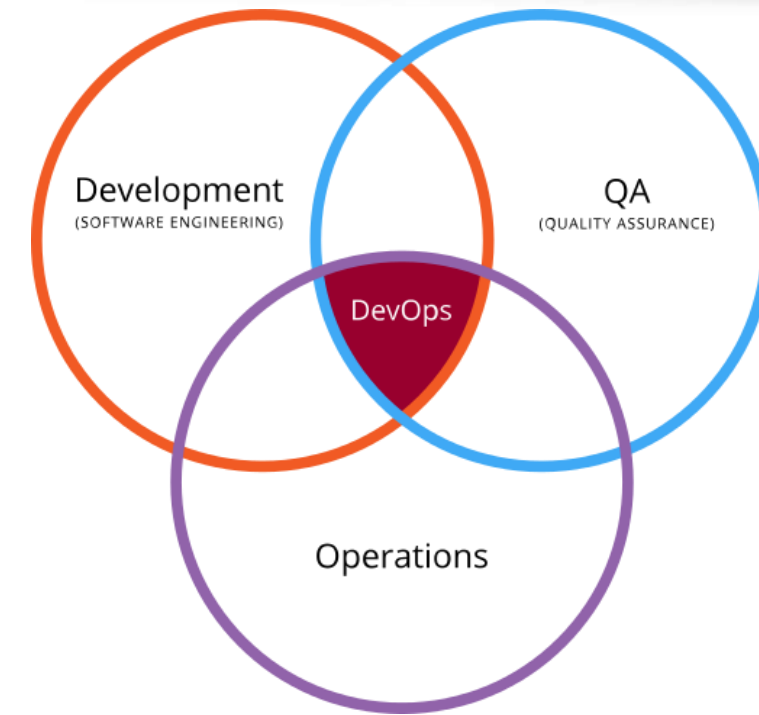


Automate deployment with **tools** to ensure **quality**

Automate & Speed up deployment

Typical set of tool

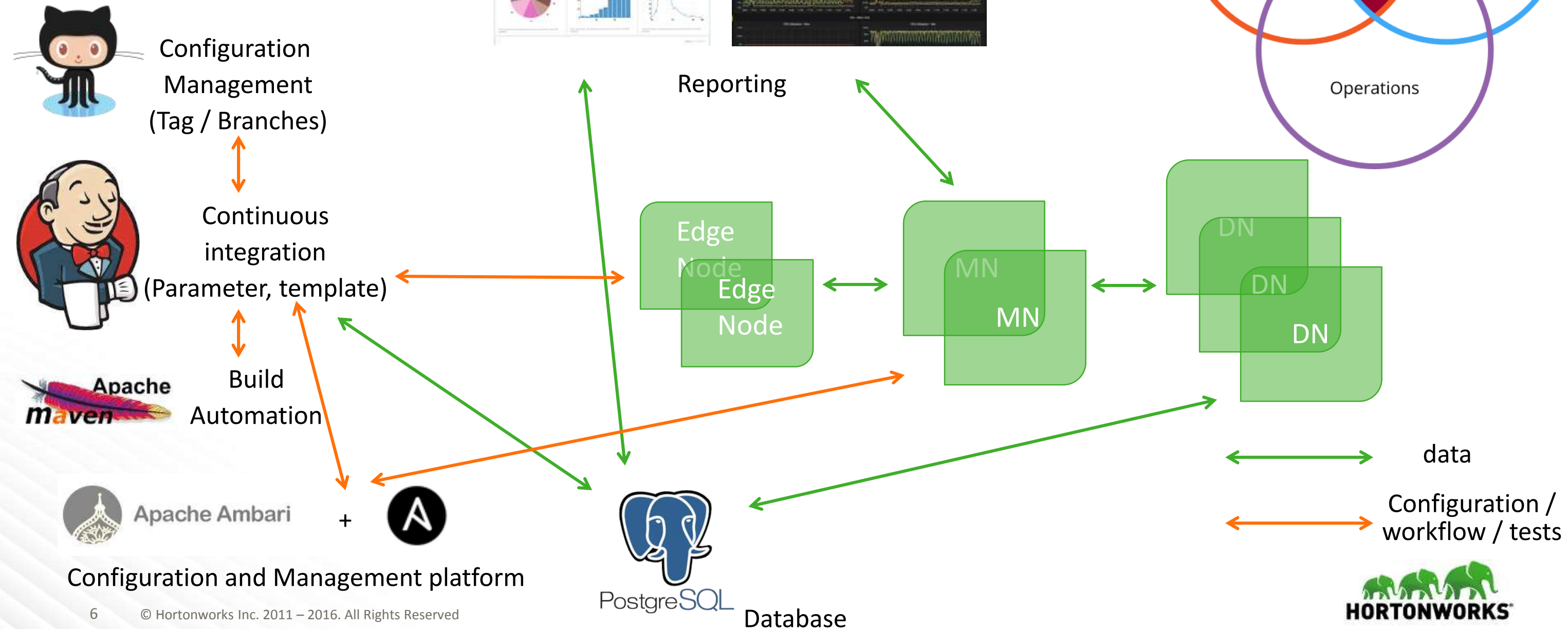
- ◆ Continuous integration + Build machine
 - Jenkins + Maven
- ◆ Configuration management tool (Hadoop + non Hadoop tools)
 - Ambari + Ansible
- ◆ Configuration management (configuration, projects, Jenkins pipeline)
 - Git
- ◆ Big Data Platform + Database + reporting
 - HDP + PostgreSQL + Grafana + Zeppelin
- ◆ Others
 - Custom log4J config, bash template, custom python scripts



**You may want to use
virtualization,
containerization,
continuous inspection, code
collaboration tool as well**

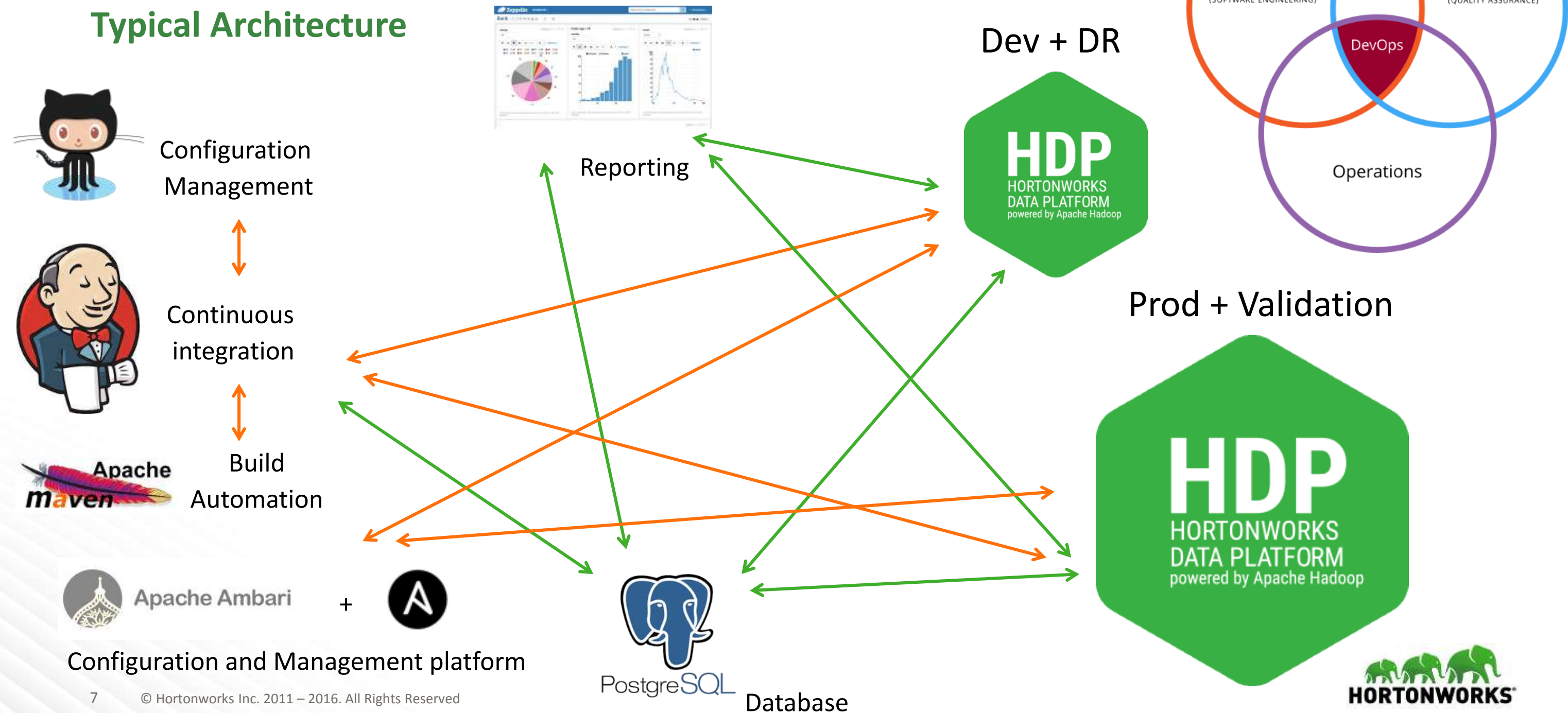
Automate & Speed up deployment

Typical Architecture



Automate & Speed up deployment

Typical Architecture



Automate & Speed up deployment

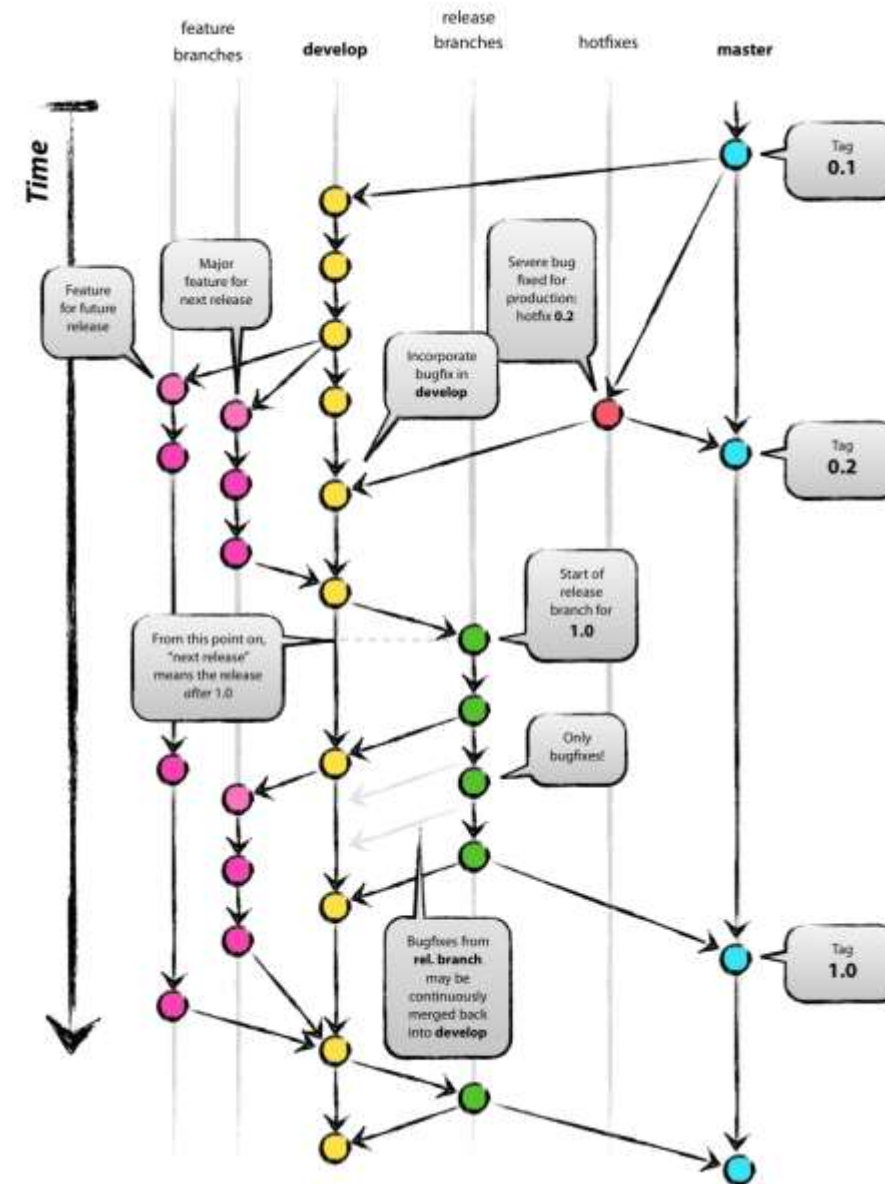
Typical Organization

- ◆ HDFS directory for Dev, Validation, Production, Backup
- ◆ Hive databases and Kafka topics with naming convention
- ◆ HBase dedicated namespace
- ◆ YARN Capacity Scheduler with corresponding queue (Dev, Validation, Production)
- ◆ Ranger rules (Project, team, environment with groups)
- ◆ All automatically created during *init* of project (On Dev by user, on Prod by admin)

Automate & Speed up deployment

On the Git side

- ◆ Codify everything
- ◆ Version everything
- ◆ Organize infrastructure as code



Agenda

Automate & Speed up deployment

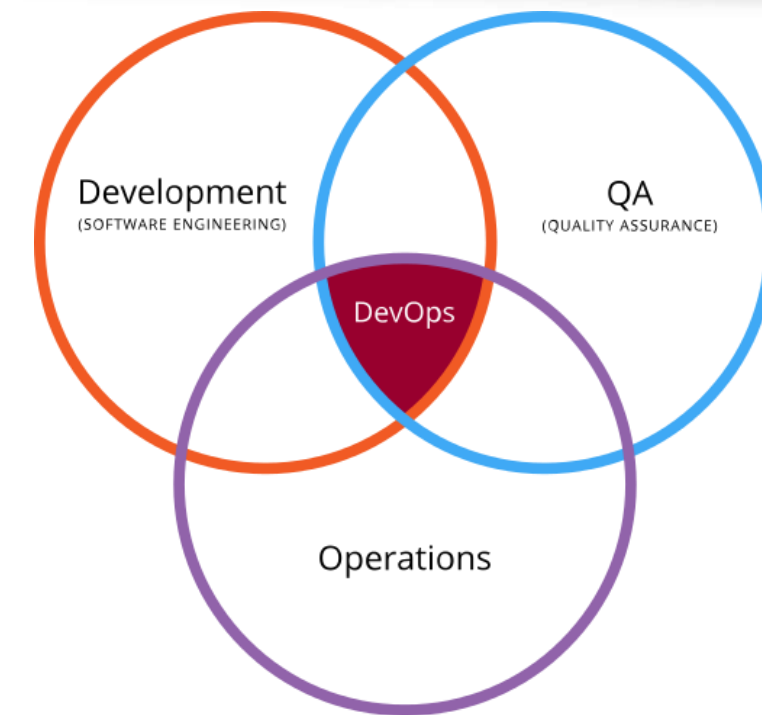


Enforce testing and quality

Enforce testing and quality

Environments / Context

- ◆ Use of multi-tenant (1+n) Hadoop clusters
 - Which means naming convention and parameter / security



Project Biyp project with multi-env support and automatic deployment run

This build requires parameters:

edge_user

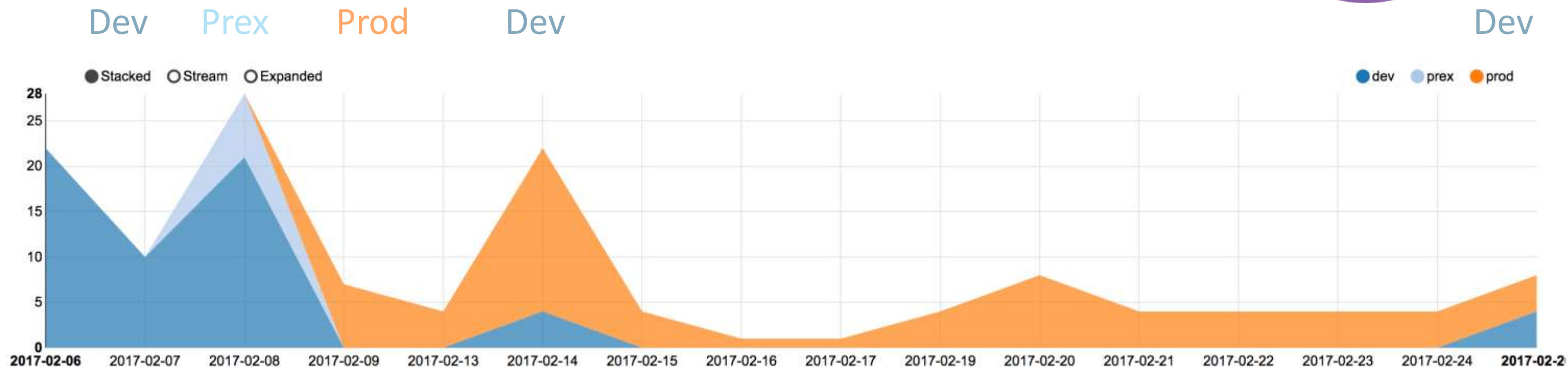
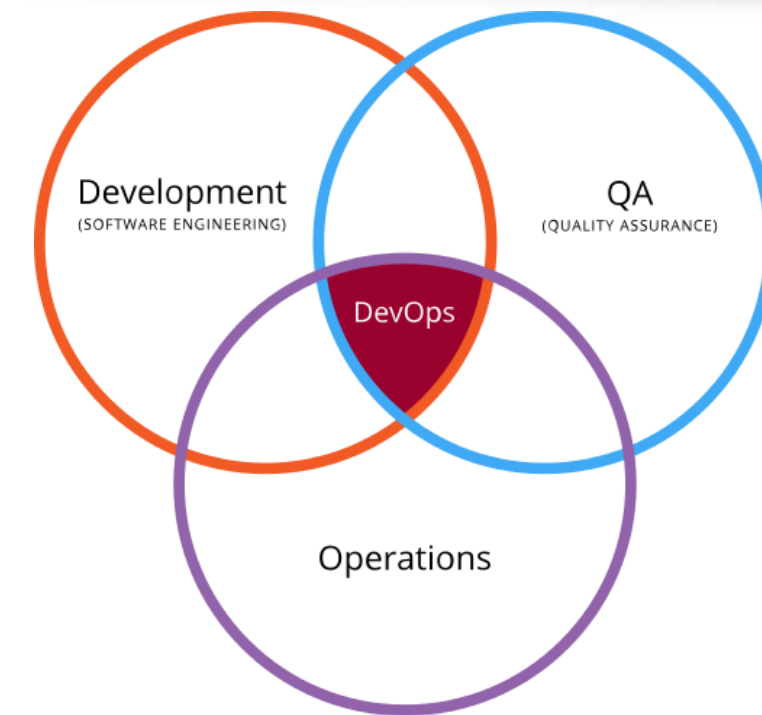
environment

edge

Enforce testing and quality

Environments / Context

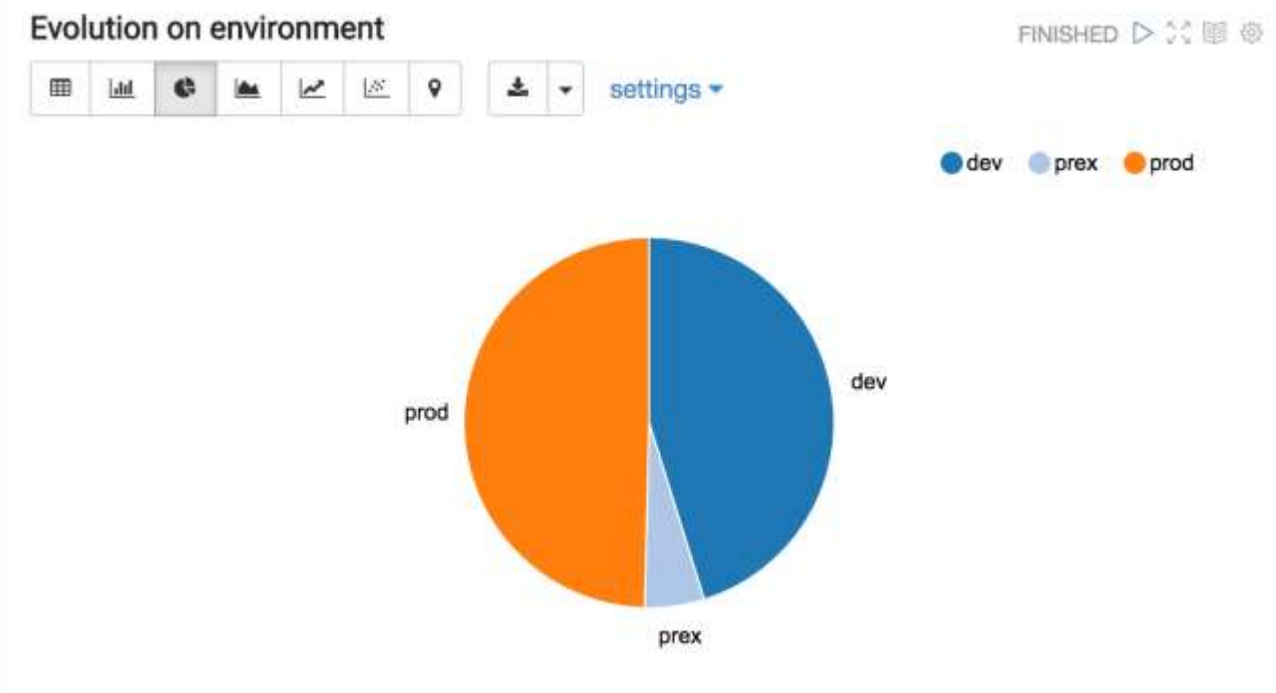
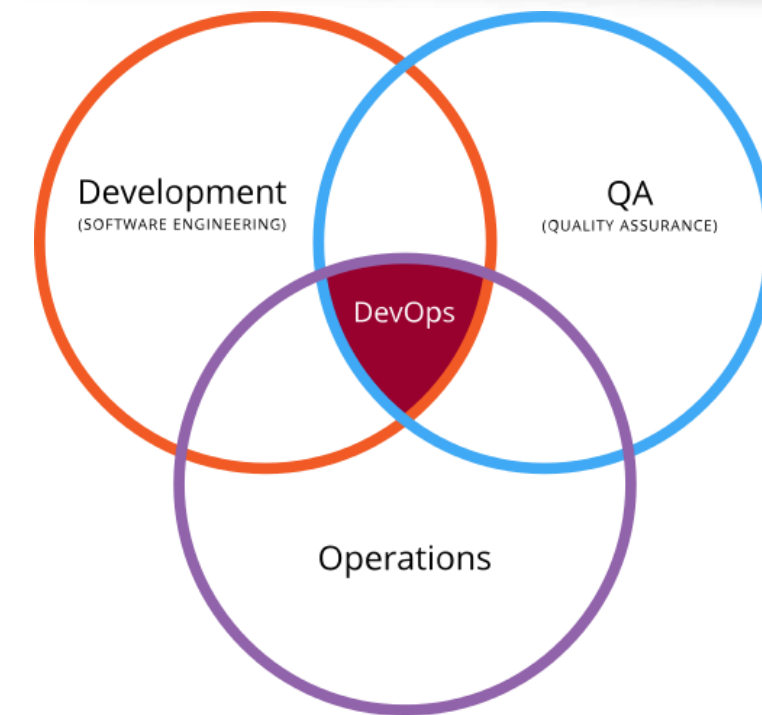
- ◆ Use of multi-tenant Hadoop cluster
 - Which means naming convention and parameter / security



Enforce testing and quality

Environments / Context

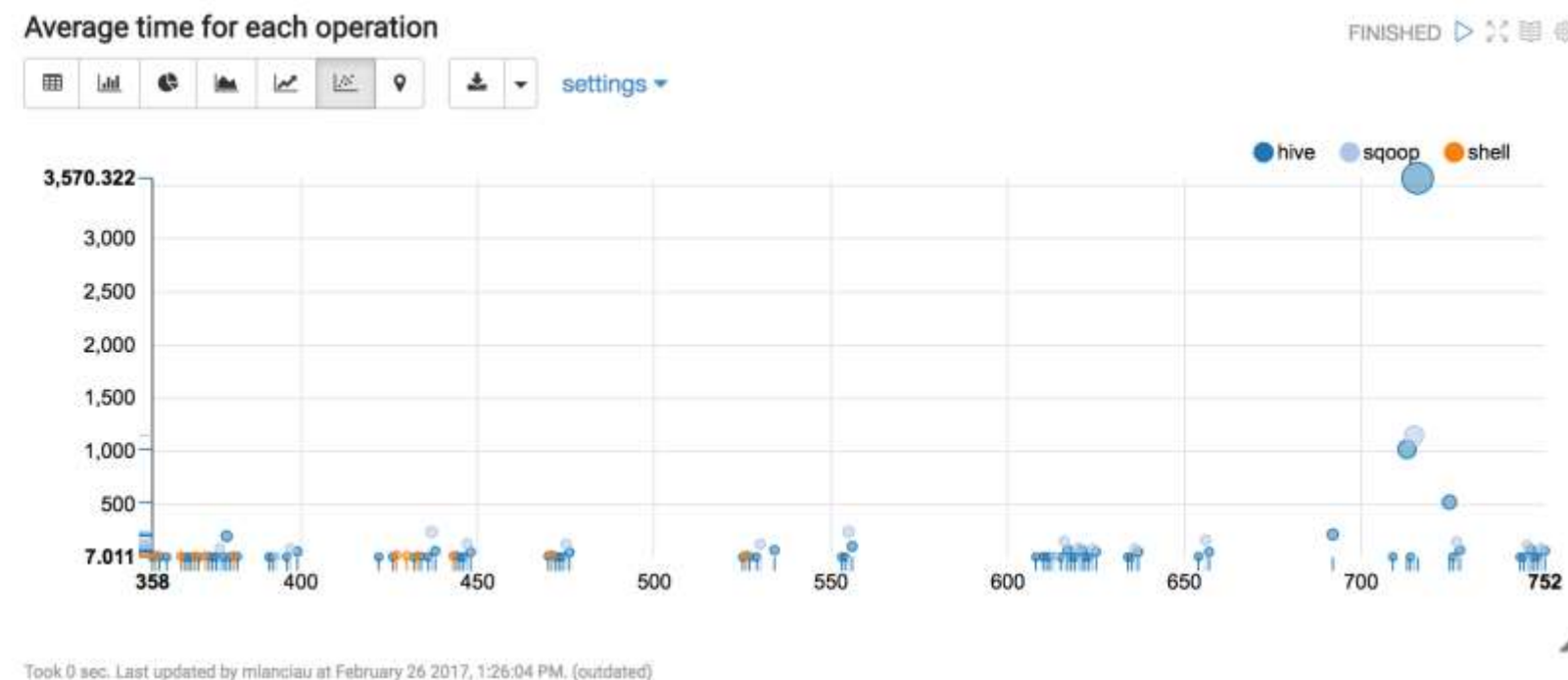
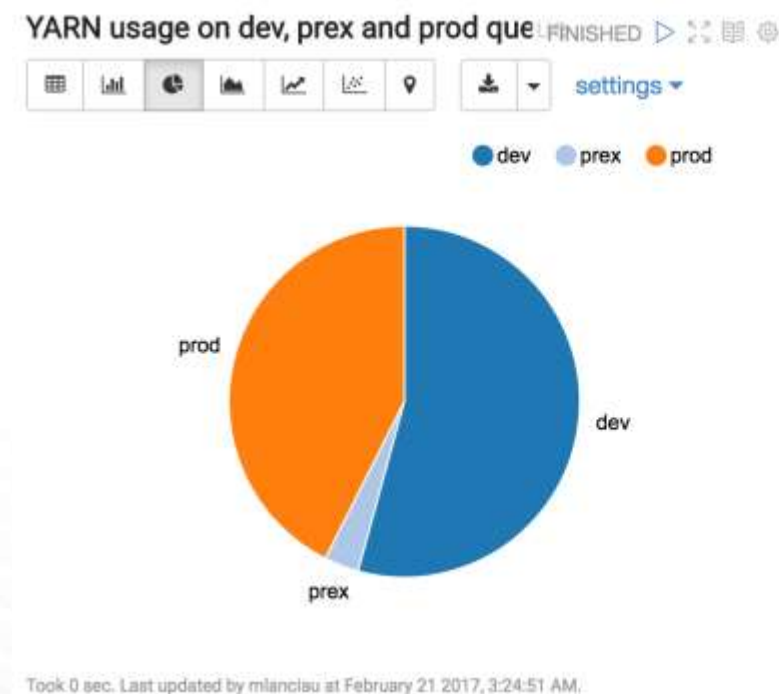
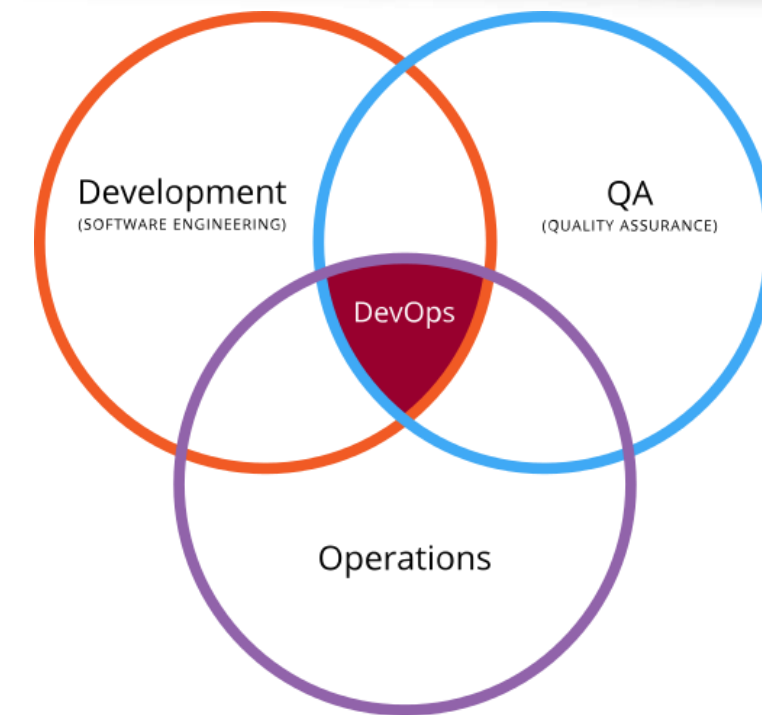
- ◆ Use of multi-tenant Hadoop cluster
 - Which means naming convention and parameter / security



Enforce testing and quality

Environments / Context

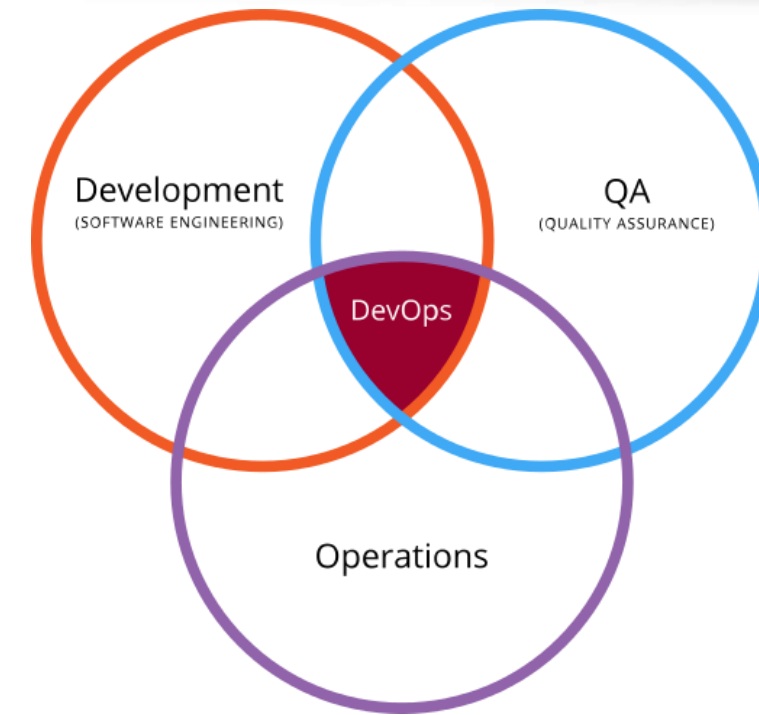
- ◆ Use of multi-tenant Hadoop cluster
 - Which means naming convention and parameter / security



Enforce testing and quality

Massive use of Pareto

- ◆ Hadoop logs
- ◆ Monitoring of full stacks component
- ◆ Application logs
- ◆ Evolution of performance
- ◆ Code review with or without tools
- ◆ Test are include in the bundle (for non-reg)



Agenda

Automate & Speed up deployment

Enforce testing and quality

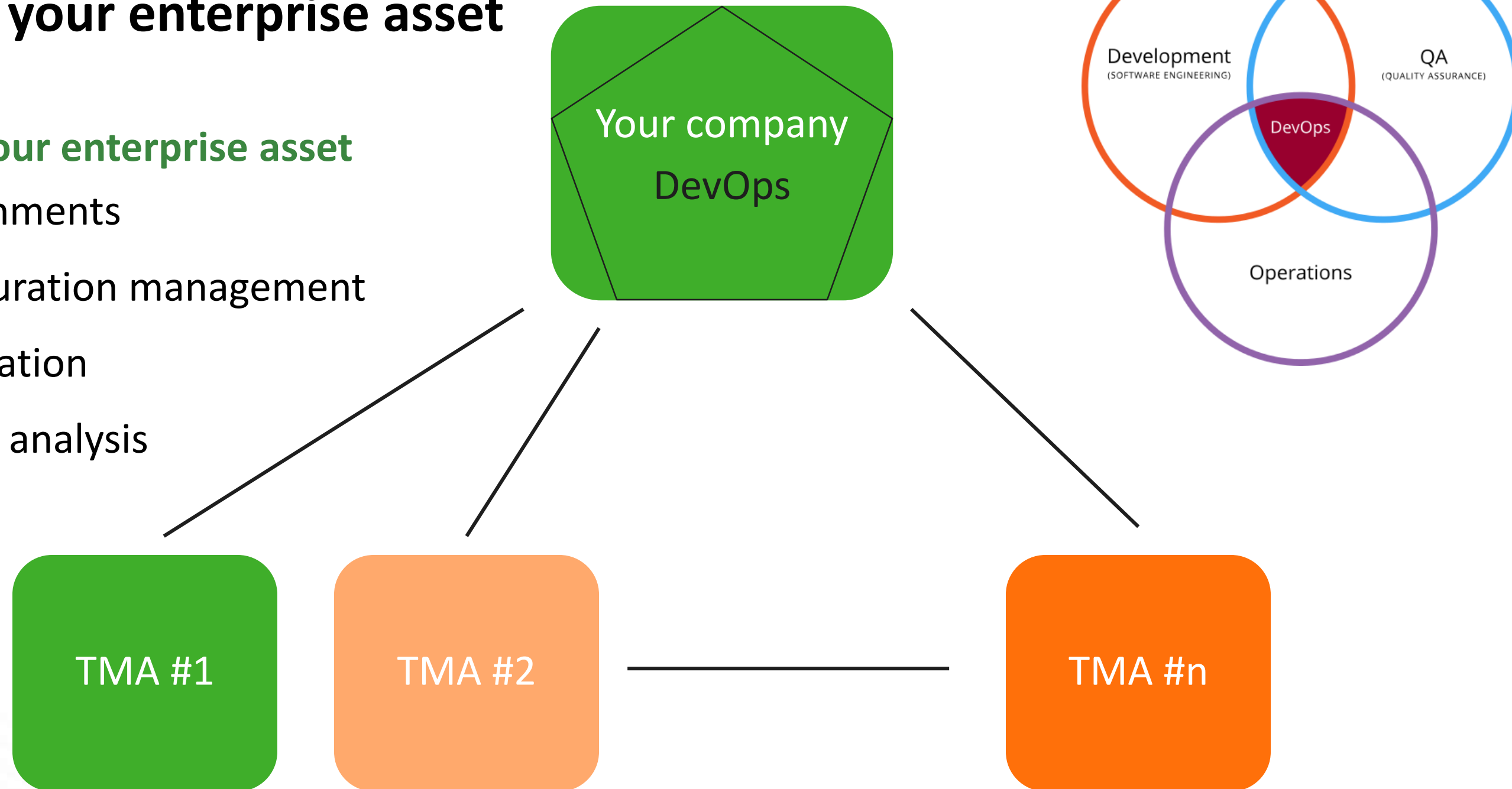


Protect your enterprise asset

Protect your enterprise asset

Protect your enterprise asset

- ◆ Environments
- ◆ Configuration management
- ◆ Automation
- ◆ Impact analysis
- ◆ 5W



Agenda

Automate & Speed up deployment

Enforce testing and quality

Protect your enterprise asset

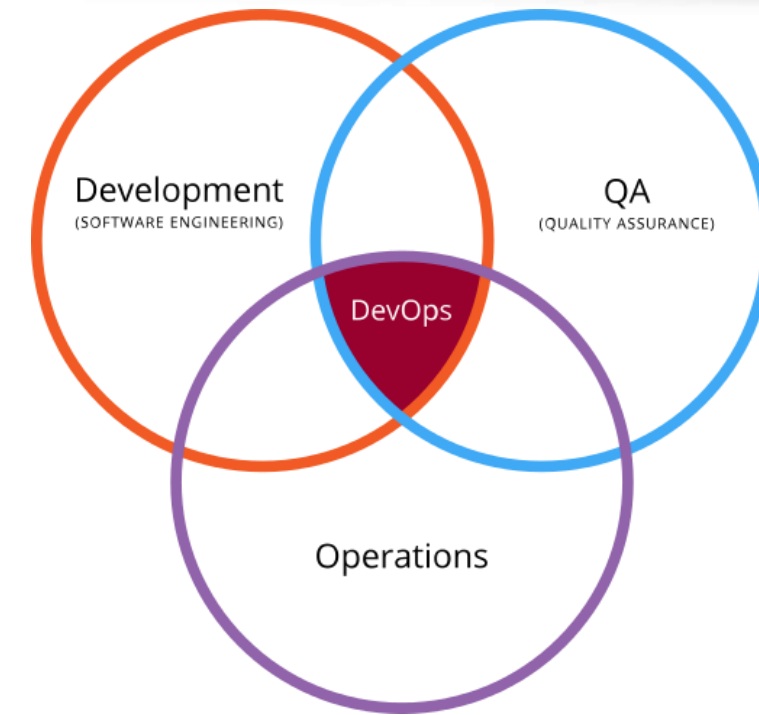


Ease Hadoop migration

Ease Hadoop migration

A lovely hand

- ◆ Install, configure and deploy
- ◆ All test to be schedule / run
- ◆ Follow performance evolution
- ◆ Can be as easy as configuring access to a new edge node + configuration / password / SSH / certificates / security
- ◆ Multi-component environments



Agenda

Automate & Speed up deployment

Enforce testing and quality

Protect your enterprise asset

Ease Hadoop migration

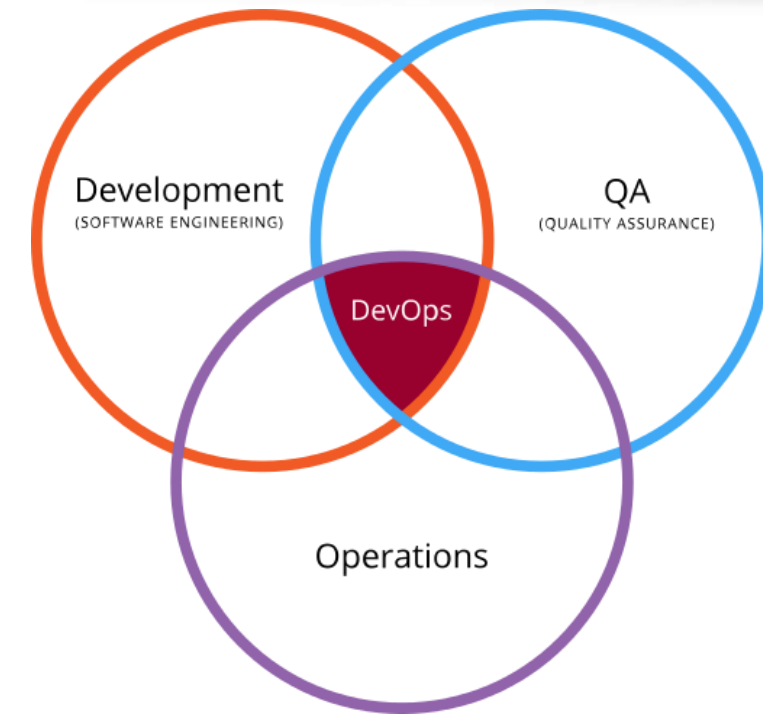


Empower Administrator and Developer

Empower Administrator and Developer

Same tools, same discussion

- Developer to get a better feedback on their application
- Administrator to check / push back best practices
- Numerous action to be completely automated (ssh)
- Focus more high level task, with direct access to information
- Get directly access to template (no need to reinvent the wheels)



Agenda

Automate & Speed up deployment

Enforce testing and quality

Protect your enterprise asset

Ease Hadoop migration

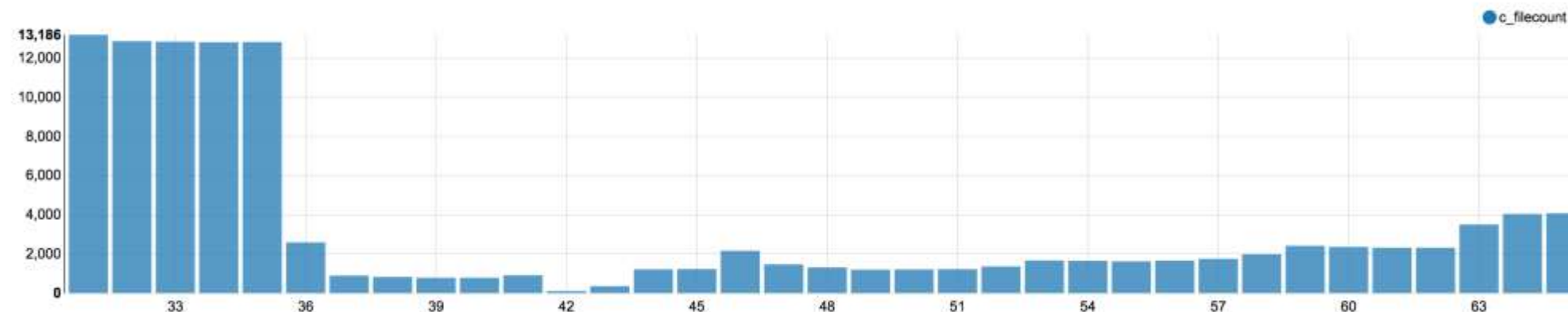
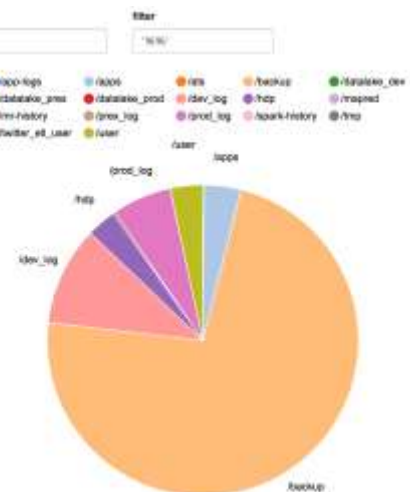
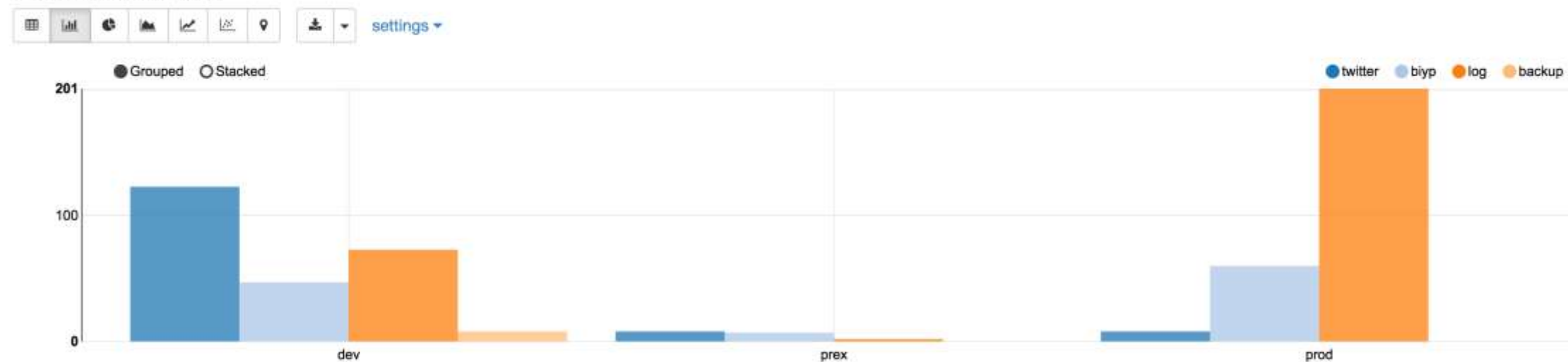
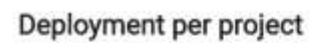
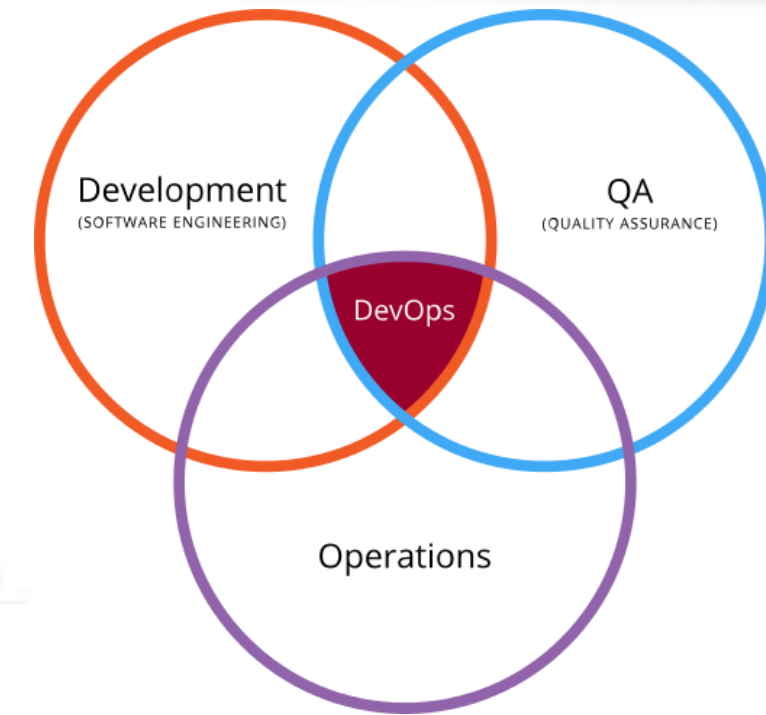
Empower Administrator and Developer



With full monitoring for Human

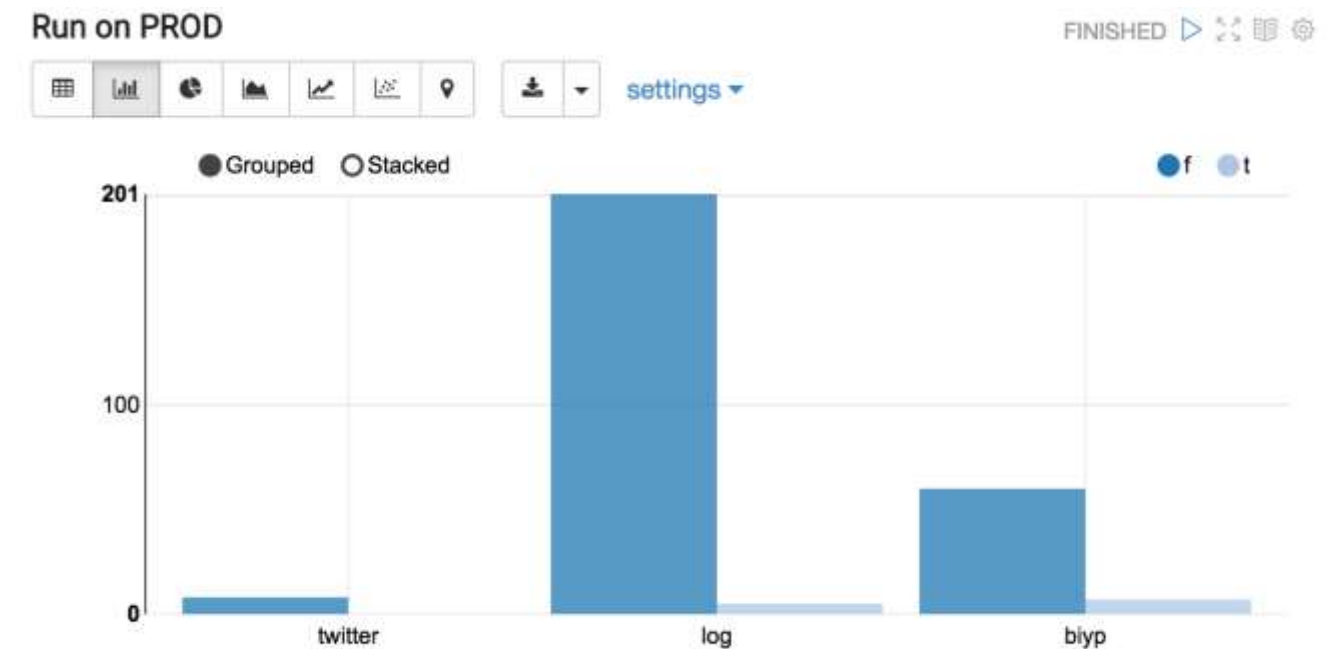
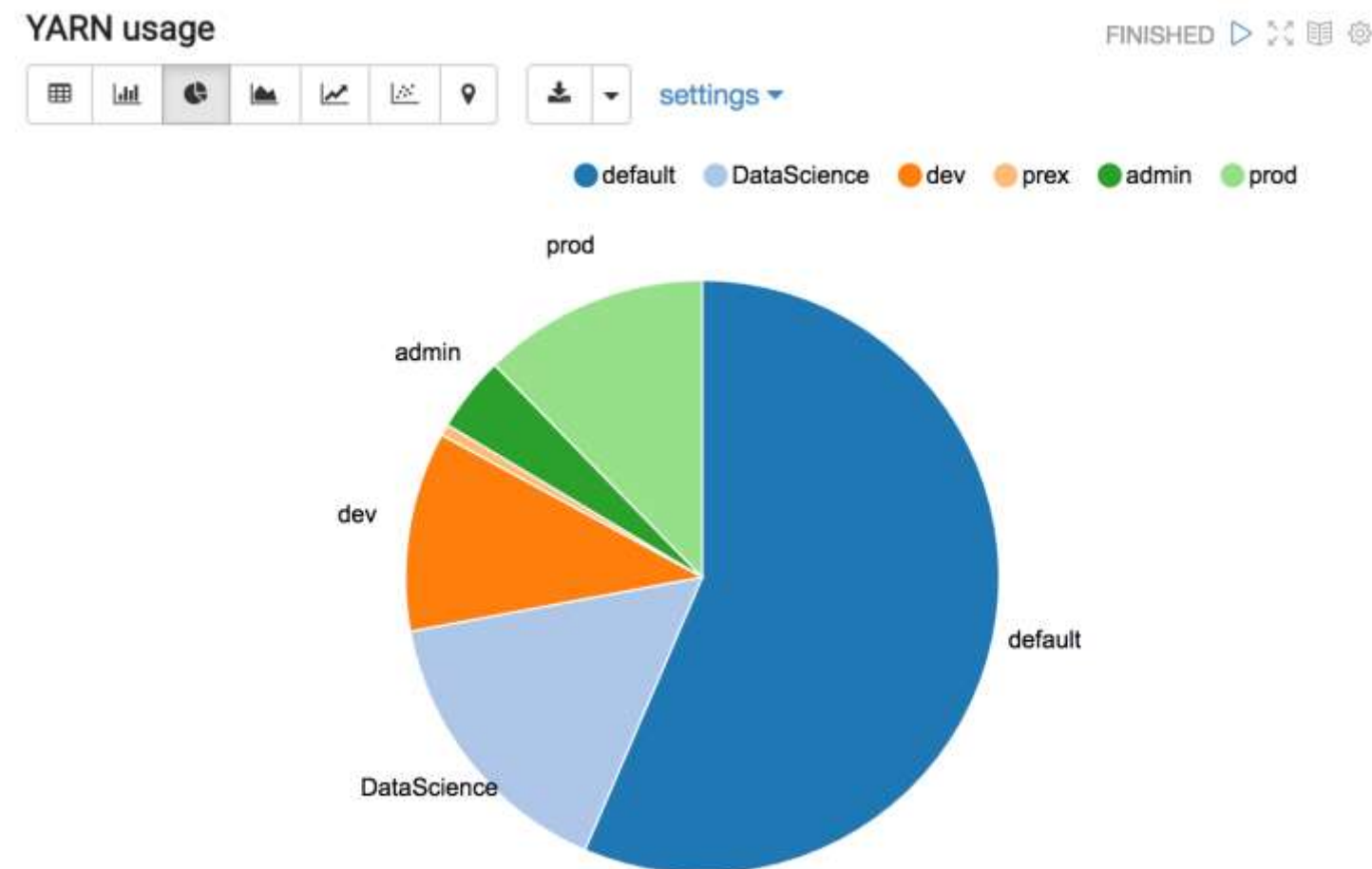
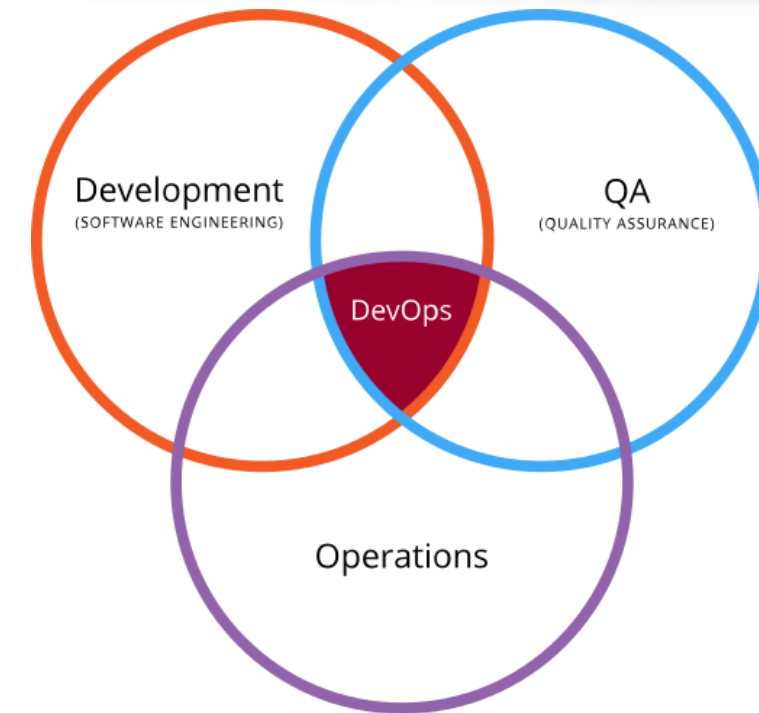
With full monitoring for human

Per project, per environment, per Hadoop cluster or on all of them



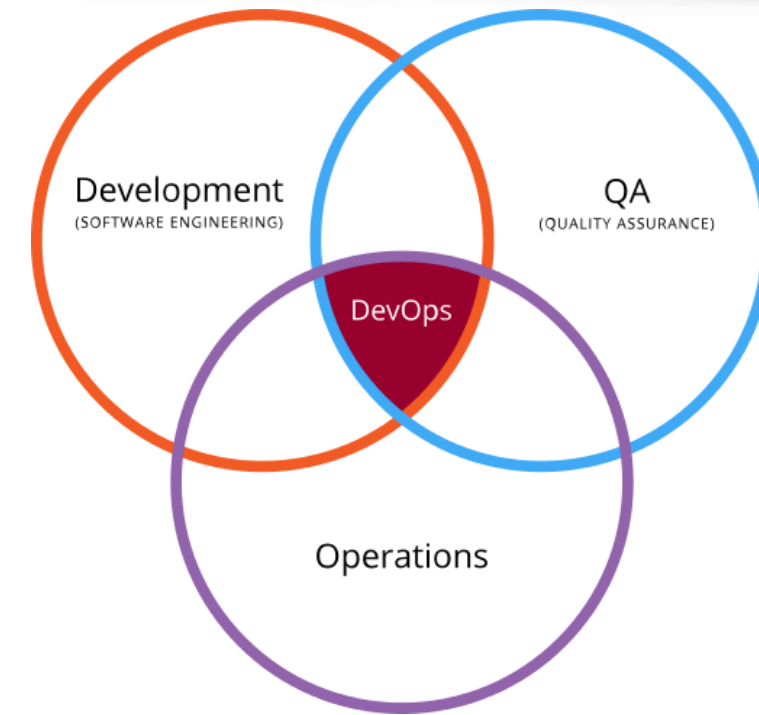
With full monitoring for human

Per project, per environment, per Hadoop cluster or on all of them

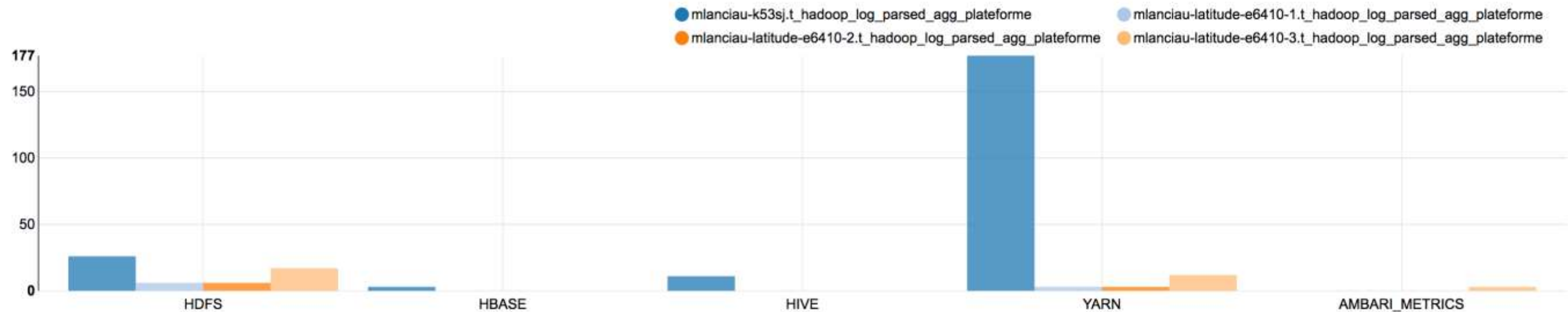


With full monitoring for human

Per project, per environment, per Hadoop cluster or on all of them



ERROR par Service sur les 3 derniers jours



Agenda

Automate & Speed up deployment

Enforce testing and quality

Protect your enterprise asset

Ease Hadoop migration

Empower Administrator and Developer

With full monitoring for Human



Demo

We are Recruiting !

Email me :

mlanciaux@hortonworks.com

Thank You

