



# Jenkins User Conference

## Gerrit and Jenkins for Big Data Continuous Delivery

London, UK, June 2015





Jenkins  
User Conference

# About GerritForge



#jenkinsconf

- Founded in 2009 in London
- Committed to OpenSource





# The Team



## Luca Milanesio

- Co-founder and Director of GerritForge
- over 20 years in Agile Development and ALM
- OpenSource contributor to many projects (BigData, Continuous Integration, Git/Gerrit)



## Antonios Chalkiopoulos

- Author of Programming MapReduce with Scalding
- Open source contributor to many BigData projects
- Working on the "land-of-Hadoop" (ladoop.com)

# The Team (2)

## Tiago Palma

- Data Warehouse & Big Data Development
- Senior Data Modeler
- Big Data infrastructure specialist



## Stefano Galarraga

- 20 years of Agile Development
- Middleware, Big Data, Reactive Distributed Systems.
- Open Source contributor to many BigData projects.

# Agenda

- Why continuous deployment on BigData?
- Our Development Lifecycle ingredients
  - Gerrit, Jenkins, Mesos, Marathon, CDH / Spark
- Topics to address in BigData development
  - Type of tests (Unit vs. Integration)
  - Testing the "real thing" (aka the Cluster)
- Our BigData virtualised infrastructure
  - Marathon, Mesos and Dockers all around
- Live (minimised) Demo





# WHY?

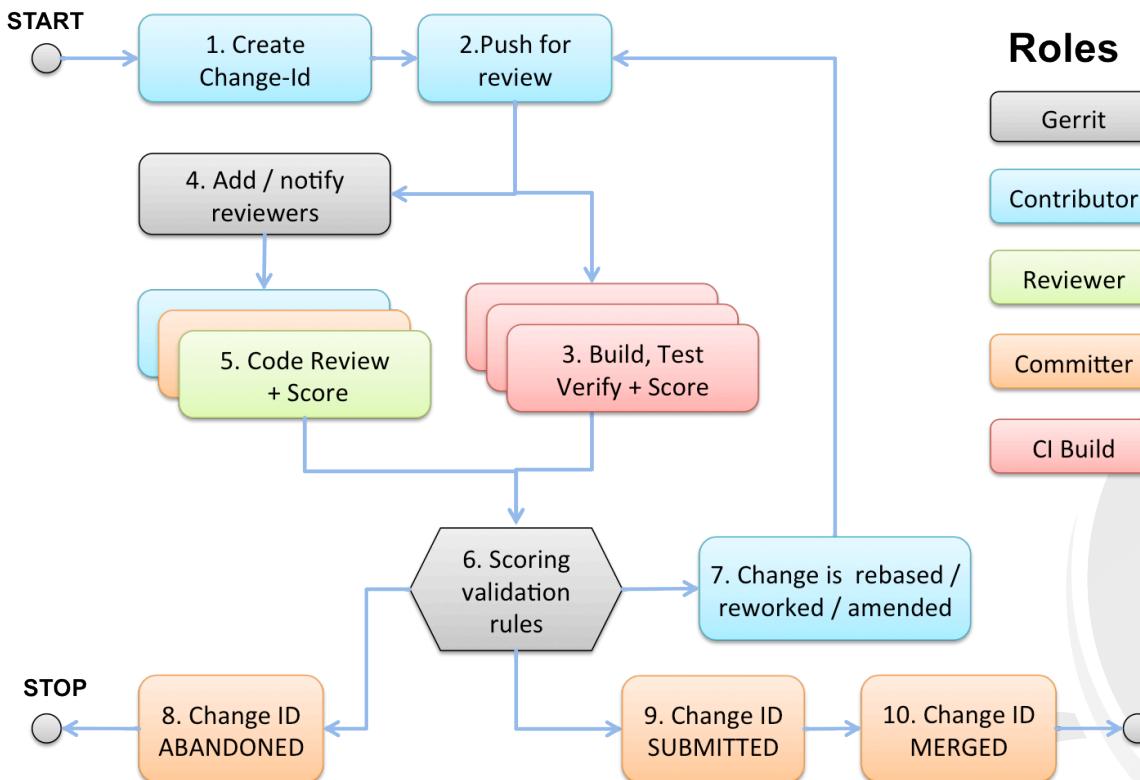


- Early BigData had no process at all = may fail at any time
- Mature BigData is mission critical decision maker
- Need for more stable sw-engineering methodologies:
  - Test-Driven Development (Stefano's ScaldingUnit)
  - Continuous Integration with Jenkins
  - Integration & Performance testing
  - Code review and validation

# Code-Review BigData Lifecycle (1)

- GIT used by distributed teams (UK, Israel, India)
- Topics and Code Review
- Jenkins build on every patch-set
- Commits reviewed / approved via Gerrit Submit

# Code-Review BigData Lifecycle (2)

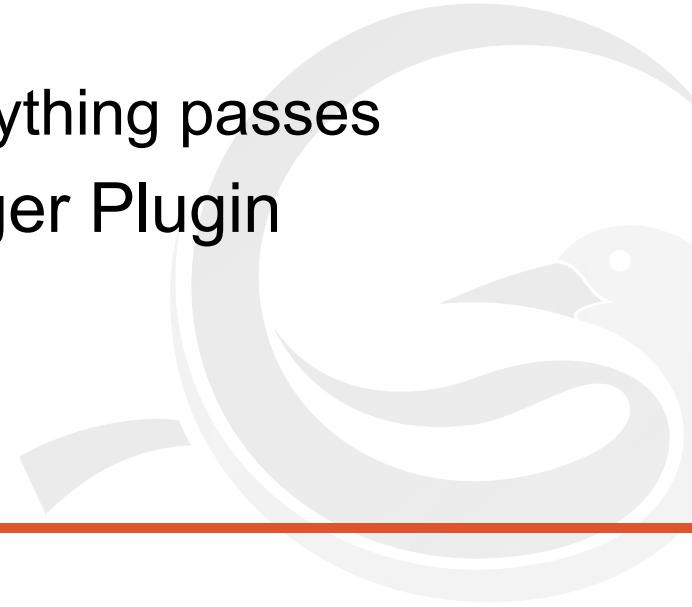


## Roles

- Gerrit
- Contributor
- Reviewer
- Committer
- CI Build

 #jenkinsconf

- Submitting a Topic automatically does:
  - all patch-sets merged (semi-atomically)
  - trigger a longer chain of CI steps
  - automatically promote a RC if everything passes
- Jenkins automation via Gerrit Trigger Plugin



# Ingredients: Gerrit

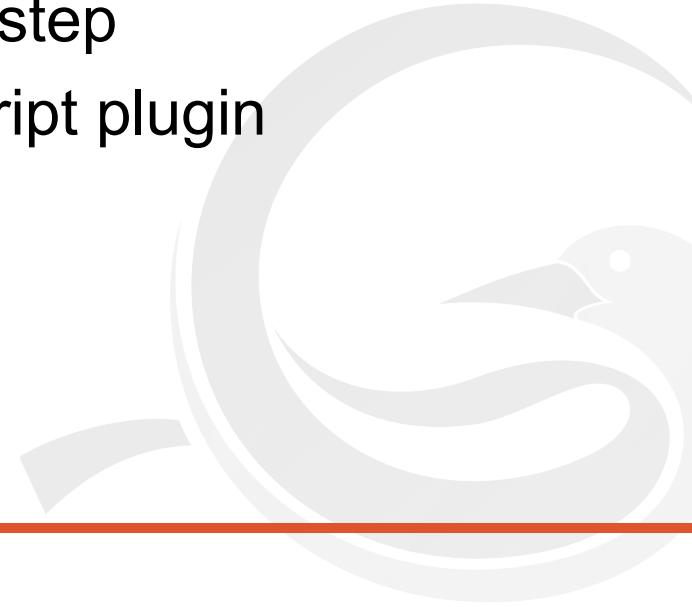


- Git-based Code Review system
- Pre-commit review
- Allows multiple validation steps (pipeline)
- Validation + Integration flags

# Ingredients: Jenkins



- Plugins:
  - Gerrit trigger
  - Docker build step
  - Post-build script plugin



# Fitting CDH Into this Picture

- Integration Test
  - Running integration tests into an CDH-enabled docker container
  - Hadoop/local and Spark/standalone is not enough
  - Need to test classes serialisation
  - Validate package fat-jars (libs conflicts with CDH)
  - Performance on a real cluster

# Fitting CDH Into this Picture

- Acceptance / performance test with short-lived CDHs
- Solution: Mesos, Marathon and Docker:
  - Ephemeral clusters with defined capacity
  - Automatic cluster-config
  - All controlled via Docker/Mesos



# Mesos + Marathon



# Marathon

- Apache Mesos
  - Abstracts CPU, memory, storage, other compute resources away from machines
- Marathon Framework
  - Runs on top of Mesos
  - Guarantees that long-running applications never stop
  - REST API for managing and scaling services

# CDH Components

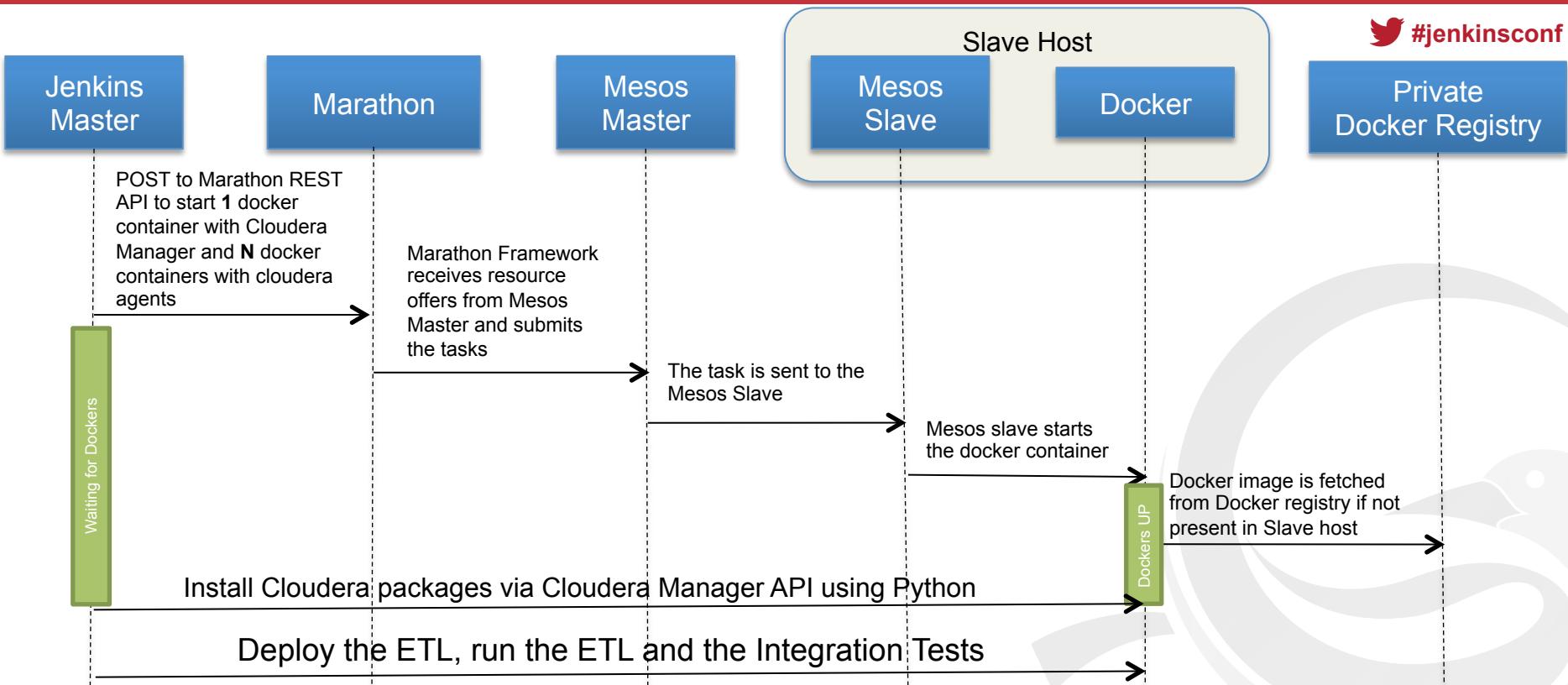


- CDH 5.4.1 distribution
  - Apache Spark
  - Hadoop HDFS
  - YARN





# Integration Test Flow on CDH Cluster

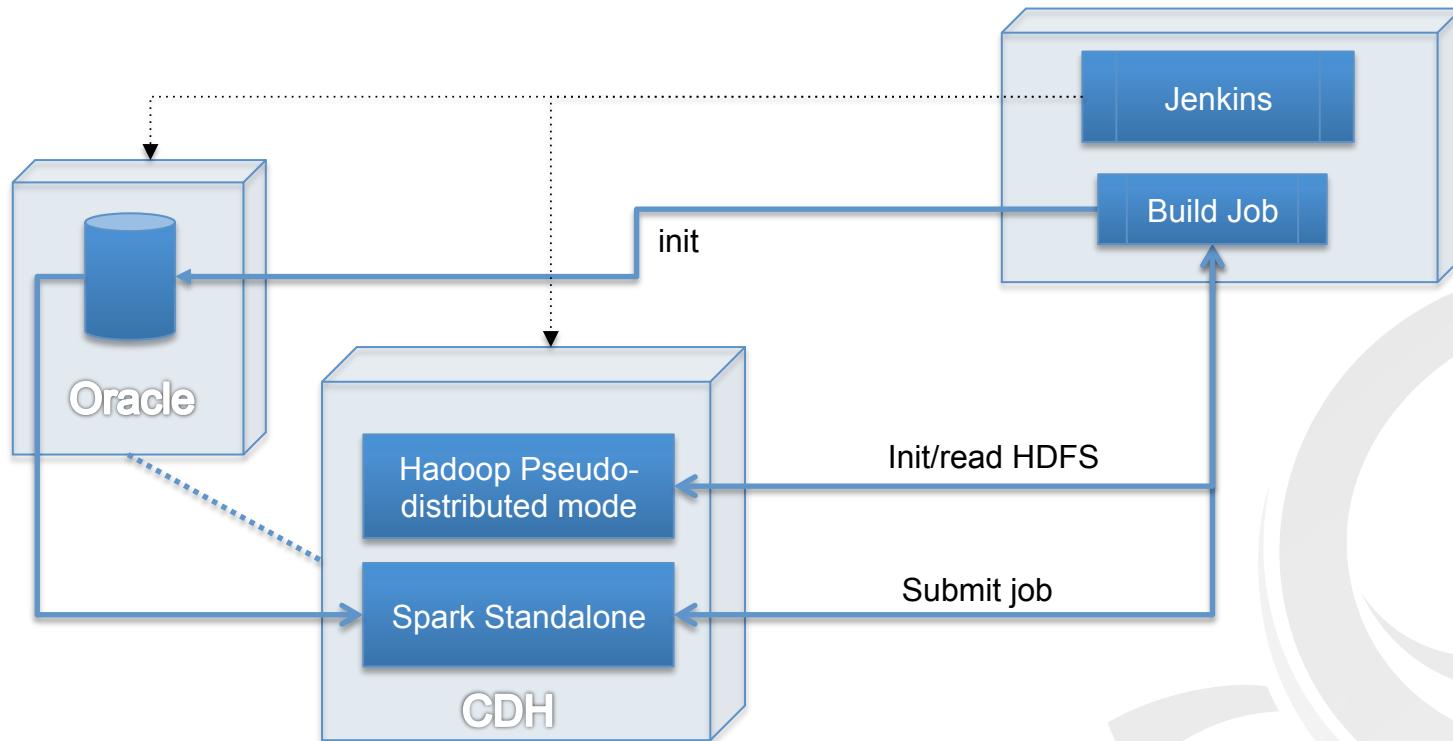


# Unit and Integration Tests sample

- Test project:
  - Test Spark project
  - ETL from Oracle to HDFS
- Unit-test directly on Spark logic
- Integration tests for every patch-set:
  - VERY small dataset just for this demo
  - CDH and Oracle Docker Images



# Unit and Integration Tests





**Jenkins**  
**User Conference**

 #jenkinsconf

# DEMO

Small-scale of BigData Delivery Pipeline



# References

- Demo sources  
<https://github.com/GerritForge>
- Blog:  
<https://gitenterprise.me>
- Twitter:  
[@GerritReview](#) [@GitEnterprise](#) [@GerritForge](#)
- Learn Gerrit Code Review book:  
[GerritHub.io/book](http://GerritHub.io/book)
- Get in touch with GerritForge:  
[info@GerritForge.com](mailto:info@GerritForge.com)