## Spark dans le cloud

Mohamed Zayed: <u>zayed@lal.in2p3.fr</u>
Guillaume Philippon: <u>guillaume.philippon@lal.in2p3.fr</u>
09 Novembre 2016
Laboratoire de l'accélérateur linéaire





## **Overview**

#### Infrastructure Cloud du LAL

Le projet cyclone

#### Spark sur le cloud

- Structure actuelle
- Objectif

#### Développement en cours

• Développement d'une application avec Slipstream ou Ansible

### Infrastructure du cloud au LAL



2200 coeurs



350 TOctets

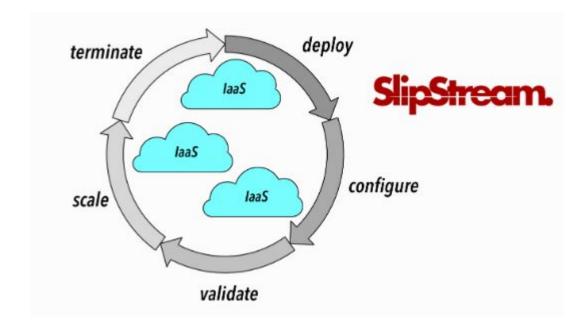


- Orchestrateur: permet de déployer des applications automatiquement sur le cloud (possibilité de multi-cloud).
- Lien: https://slipstream-v3.lal.in 2p3.fr

## SlipStream.

# SlipStream Permet de déployer une infrastructure de machines virtuelles:

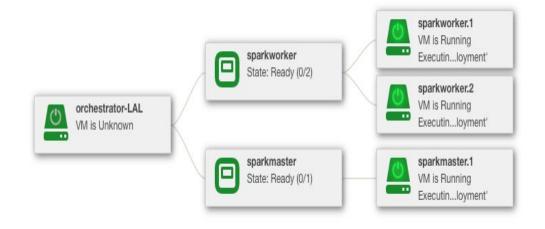
 Exemple: Lancement d'une infrastructure de machines master / Workers, sur un cloud ou en multicloud



## SlipStream.

SlipStream Permet de déployer une infrastructure de machines virtuelles:

 Exemple: Lancement d'une infrastructure de machines master / Workers, sur un cloud ou en multicloud





#### Services

- Déploiement multi-cloud d'applications fédérées.
- CNSMO services (VPN, FW, Load Balancer)

#### **Exemple**

Ftablir une connection VPN entre une VM lancée sur le cloud du LAL et une VM lancée sur le cloud de l'IPHC.

## Spark dans le cloud

#### **Problèmes:**

- Dans les infrastructures Spark, le calcul est très lié au placement de données
  - Difficile d'ajouter ou retirer du stockage
  - Peu adapté au cloud où l'infrastructure de stockage est indépendante de l'infrastructure de calcul
- Deployer une infrastructure Spark multi-cloud

## Spark dans le cloud

#### **Solution actuelle:**

- Démarrer une infrastructure minimal (calcul + stockage)
  - Volume géré par cinder (Ceph)
- Possibilité de déployer du calcul seul à la demande

## Développement en cours

## Infrastructure de stockage HDFS

#### **Utiliser le cluster Ceph comme backend HDFS**

- Permet de déconnecter le stockage et le calcul
- Nécessaire pour rendre le cluster Spark dynamique

### Les services CNSMO

#### Services de networking as a service (NaaS)

- VPN
  - Créer des clients VPN entre différents cloud (par exemple: LAL -IPHC)
- Firewall
  - Établir des règles de firewall différentes sur les machines

### Utiliser un orchestrateur

#### Déploiement d'un cluster Spark avec SlipStream

Permet un déploiement multi-cloud

### Comparaison entre plusieurs Orchestrateur

Slipstream, Ansible

## Merci!