

# SPARK : Etat de l'Infrastructure

Adrien Ramparison  
Adrien.Ramparison@lal.in2p3.fr

20 Janvier 2016

# PLAN

- 1 L'infrastructure SPARK
- 2 Services
- 3 Utilisation
  - Utilisation
  - Monitoring/Support
- 4 Gestion, Contacts
- 5 TODO List
- 6 Conclusion
- 7 Questions



# INFRASTRUCTURE

- Cluster de production (*experimental*)
  - Openstack liberty : Domain **u-psud**, Project **Spark**, CentOS7
  
- Cluster de développement
  - Openstack liberty : Domain u-psud, Project DI, CentOS7

# INFRASTRUCTURE

- Cluster de production (*expérimental*)
  - Openstack liberty : Domain **u-psud**, Project **Spark**, CentOS7
  - 1 Master : Gabarit 18 cores/36GB RAM (**os.18**)
  
- Cluster de développement
  - Openstack liberty : Domain u-psud, Project DI, CentOS7
  - 1 Master : Gabarit os.1

# INFRASTRUCTURE

- Cluster de production (*experimental*)
  - Openstack liberty : Domain **u-psud**, Project **Spark**, CentOS7
  - 1 Master : Gabarit 18 cores/36GB RAM (**os.18**)
  - 3 Slaves : Gabarit 18 cores/ 36GB RAM, volume cinder HDFS  
**200 Go**
- Cluster de développement
  - Openstack liberty : Domain u-psud, Project DI, CentOS7
  - 1 Master : Gabarit os.1
  - 2 Slaves : Gabarit os.1, pas de volume cinder

# SERVICES

- Standards API : Scala, Python, Java

# SERVICES

- Standards API : Scala, Python, Java
- Built-In Components : SparkR, Spark SQL, MLib, Spark GraphX



# SERVICES

- Standards API : Scala, Python, Java
- Built-In Components : SparkR, Spark SQL, MLib, Spark GraphX
- Système de fichier distribué HDFS

# SERVICES

- Standards API : Scala, Python, Java
- Built-In Components : SparkR, Spark SQL, MLib, Spark GraphX
- Système de fichier distribué HDFS
- Jupyter notebook

# SERVICES

- Standards API : Scala, Python, Java
- Built-In Components : SparkR, Spark SQL, MLib, Spark GraphX
- Système de fichier distribué HDFS
- Jupyter notebook
- HT Condor

# SERVICES

- Standards API : Scala, Python, Java
- Built-In Components : SparkR, Spark SQL, MLib, Spark GraphX
- Système de fichier distribué HDFS
- Jupyter notebook
- HT Condor
- Ganglia

# UTILISATION

- Accès : ssh sur le **MasterNode** (nécessite clé ssh)

- 
1. [www.informatique-scientifique.u-psud.fr/Spark](http://www.informatique-scientifique.u-psud.fr/Spark),
  2. [www.informatique-scientifique.u-psud.fr/Hadoop/TPSpark.tar.gz](http://www.informatique-scientifique.u-psud.fr/Hadoop/TPSpark.tar.gz)
  3. [http://training.databricks.com/workshop/itas\\_workshop.pdf](http://training.databricks.com/workshop/itas_workshop.pdf)

# UTILISATION

- Accès : ssh sur le **MasterNode** (nécessite clé ssh)
- Ouverture de comptes : Compte Adonis (DI). Plus d'infos sur [le site](#)<sup>1</sup> de la DI

---

1. [www.informatique-scientifique.u-psud.fr/Spark](http://www.informatique-scientifique.u-psud.fr/Spark),
2. [www.informatique-scientifique.u-psud.fr/Hadoop/TPSpark.tar.gz](http://www.informatique-scientifique.u-psud.fr/Hadoop/TPSpark.tar.gz)
3. [http://training.databricks.com/workshop/itas\\_workshop.pdf](http://training.databricks.com/workshop/itas_workshop.pdf)

# UTILISATION

- Accès : ssh sur le **MasterNode** (nécessite clé ssh)
- Ouverture de comptes : Compte Adonis (DI). Plus d'infos sur [le site](#)<sup>1</sup> de la DI
- Langages : scala, java, python, R, ...

---

1. [www.informatique-scientifique.u-psud.fr/Spark](http://www.informatique-scientifique.u-psud.fr/Spark),
2. [www.informatique-scientifique.u-psud.fr/Hadoop/TPSpark.tar.gz](http://www.informatique-scientifique.u-psud.fr/Hadoop/TPSpark.tar.gz)
3. [http://training.databricks.com/workshop/itas\\_workshop.pdf](http://training.databricks.com/workshop/itas_workshop.pdf)

# UTILISATION

- Accès : ssh sur le **MasterNode** (nécessite clé ssh)
- Ouverture de comptes : Compte Adonis (DI). Plus d'infos sur [le site](#)<sup>1</sup> de la DI
- Langages : scala, java, python, R, ...
- Comptes, projets :
  - DI : Workflow Spark par J. Nauroy
  - Administrateurs : [exercices](#)<sup>2</sup>, [tests](#)<sup>3</sup>
  - LAL : C. Arnault (LSST)

---

1. [www.informatique-scientifique.u-psud.fr/Spark](http://www.informatique-scientifique.u-psud.fr/Spark),

2. [www.informatique-scientifique.u-psud.fr/Hadoop/TPSpark.tar.gz](http://www.informatique-scientifique.u-psud.fr/Hadoop/TPSpark.tar.gz)

3. [http://training.databricks.com/workshop/itas\\_workshop.pdf](http://training.databricks.com/workshop/itas_workshop.pdf)



# MONITORING

- Monitoring
  - Utilisateurs :
    - Ganglia sur <http://vm-75222.lal.in2p3.fr/ganglia>
    - Interface utilisateur  
<http://vm-75222.lal.in2p3.fr:20001/>

# MONITORING

- Monitoring
  - Utilisateurs :
    - Ganglia sur <http://vm-75222.lal.in2p3.fr/ganglia>
    - Interface utilisateur  
<http://vm-75222.lal.in2p3.fr:20001/>
  - Administrateurs : OMD (ToDo List) sur  
<https://monitoring.lal.in2p3.fr>

# MONITORING

- Monitoring
  - Utilisateurs :
    - Ganglia sur <http://vm-75222.lal.in2p3.fr/ganglia>
    - Interface utilisateur  
<http://vm-75222.lal.in2p3.fr:20001/>
  - Administrateurs : OMD (ToDo List) sur  
<https://monitoring.lal.in2p3.fr>
- Suivi des Incidents/Support : GLPI sur  
<https://cloud-support.lal.in2p3.fr>

# CONTACTS

- Administration du cluster Spark (Infrastructure) : Omar Talibi

# CONTACTS

- Administration du cluster Spark (Infrastructure) : Omar Talibi
- Contacts Techniques :

ARNAULT	Christian	LAL	arnault@lal.in2p3.fr
COURET	Denis	DI	Denis.Couret@u-psud.fr
PHILIPPON	Guillaume	LAL	philippo@lal.in2p3.fr
RAMPARISON	Adrien	LAL	Adrien.Ramparison@lal.in2p3.fr
TALIBI	Omar	LAL	Omar.Talibi@lal.in2p3.fr

Table – List of contacts

# CONTACTS

- Administration du cluster Spark (Infrastructure) : Omar Talibi
- Contacts Techniques :

ARNAULT	Christian	LAL	arnault@lal.in2p3.fr
COURET	Denis	DI	Denis.Couret@u-psud.fr
PHILIPPON	Guillaume	LAL	philippo@lal.in2p3.fr
RAMPARISON	Adrien	LAL	Adrien.Ramparison@lal.in2p3.fr
TALIBI	Omar	LAL	Omar.Talibi@lal.in2p3.fr

Table – List of contacts

- Contact Scientifique : Christian ARNAULT, autres ?

# A faire

- 1 MAJ Hadoop **2.6.4** vers **2.6.5/2.7.3**, SPARK **2.0** vers **2.1.0** (correction bugs, sécurités).

# A faire

- 1 MAJ Hadoop **2.6.4** vers **2.6.5/2.7.3**, SPARK **2.0** vers **2.1.0** (correction bugs, sécurités).
- 2 Ajout d'autres fonctionnalités
  - Quota/resources/projet
  - Gestion automatique conf/déploiement
  - Rajout de noeuds + montée en charge
  - Monitoring + Support avec **OMD et GLPI**



# A faire

- 1 MAJ Hadoop **2.6.4** vers **2.6.5/2.7.3**, SPARK **2.0** vers **2.1.0** (correction bugs, sécurités).
- 2 Ajout d'autres fonctionnalités
  - Quota/resources/projet
  - Gestion automatique conf/déploiement
  - Rajout de noeuds + montée en charge
  - Monitoring + Support avec **OMD** et **GLPI**
- 3 Intégration dans Slipstream (Projet Cyclone)

# A faire

- 1 MAJ Hadoop **2.6.4** vers **2.6.5/2.7.3**, SPARK **2.0** vers **2.1.0** (correction bugs, sécurités).
- 2 Ajout d'autres fonctionnalités
  - Quota/resources/projet
  - Gestion automatique conf/déploiement
  - Rajout de noeuds + montée en charge
  - Monitoring + Support avec **OMD et GLPI**
- 3 Intégration dans Slipstream (Projet Cyclone)
- 4 HDFS → CEPH

# A faire

- 1 MAJ Hadoop **2.6.4** vers **2.6.5/2.7.3**, SPARK **2.0** vers **2.1.0** (correction bugs, sécurités).
- 2 Ajout d'autres fonctionnalités
  - Quota/resources/projet
  - Gestion automatique conf/déploiement
  - Rajout de noeuds + montée en charge
  - Monitoring + Support avec **OMD et GLPI**
- 3 Intégration dans Slipstream (Projet Cyclone)
- 4 HDFS → CEPH
- 5 ...

# Conclusion

# Questions ;

