SPARK/OPENSTACK

pateyron@lal.in2p3.fr, peloton@lal.in2p3.fr,
adrien.ramparison@lal.in2p3.fr

Webinaire IN2P3 - 14 mars 2019

PLAN

Équipement de Recherche Mutualisé

- Programmes ERM U-PSUD 2015-2017, 2017-2019
- Objectif : fournir un environnement d'expérimentation réaliste des traitements dits ***Big Data*** intégré dans la plateforme cloud@VirtualData
- Plusieurs entités : LRI (informatique), LESE (Ecologie),LAL, I2BC, INSERM, SHPEM (Signalisation Hormonale . . .), UUI Lipide (Chimie), DI Cellule calcul scientifique
- $\Rightarrow$ Cluster Apache Spark sur Openstack : Flexibilité, Scalabilité

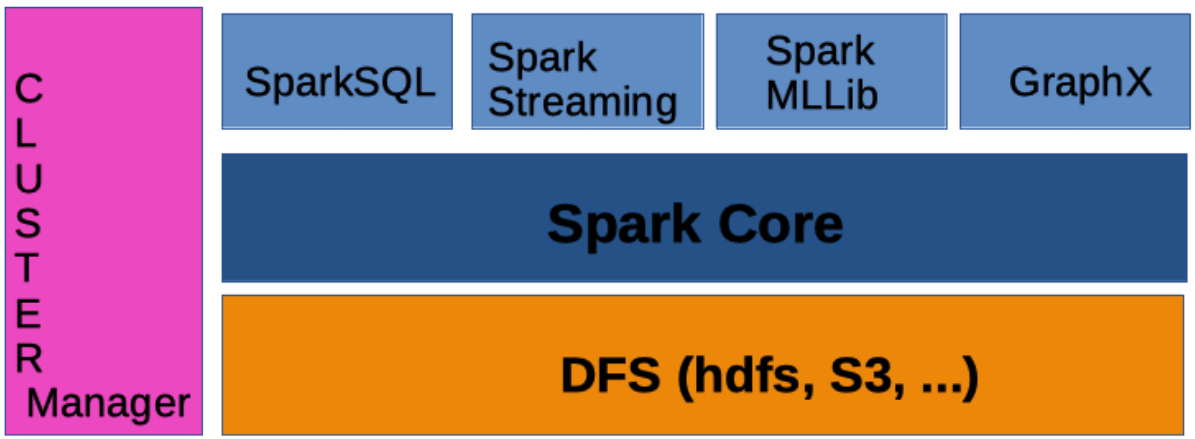
Ecosystème Apache Hadoop

- **Hadoop Map Reduce**
  - $\Rightarrow$ **Système de répartition des calculs**
  - concept **Map** (lecture des données, production Clé/Valeur), **Reduce** (Regroupement des clés, traitement des valeurs)
- **Hadoop HDFS**
  - $\Rightarrow$ **Système de répartition des données**
  - Système de fichiers répartis
  - Non posix
  - Blocs répliqués (réplica 3, configuration globale). Peut-être redéfini par l'utilisateur (variable).
- **Yarn**
  - $\Rightarrow$ **Système de gestion des ressources**

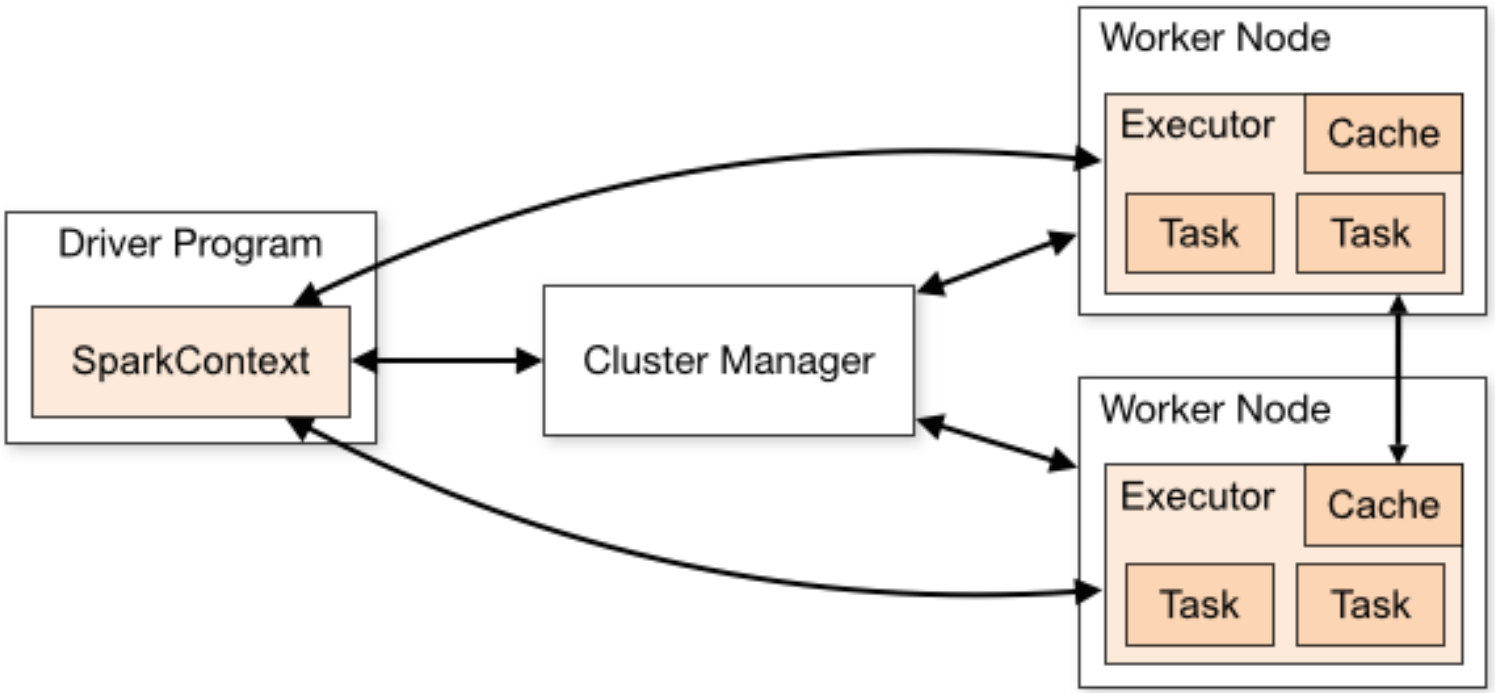
Apache SPARK

Spark en quelques points :

- Apache Spark : Framework pour le calcul distribué
- Similaire à Hadoop MapReduce
- En RAM si possible (Dataset (RDD), résultats). Pas d'écriture intermédiaire sur disque entre opérations comme Hadoop MR.
- Le calcul se fait là où se trouvent les données (idéalement).
- Librairies natives :



Architecture SPARK



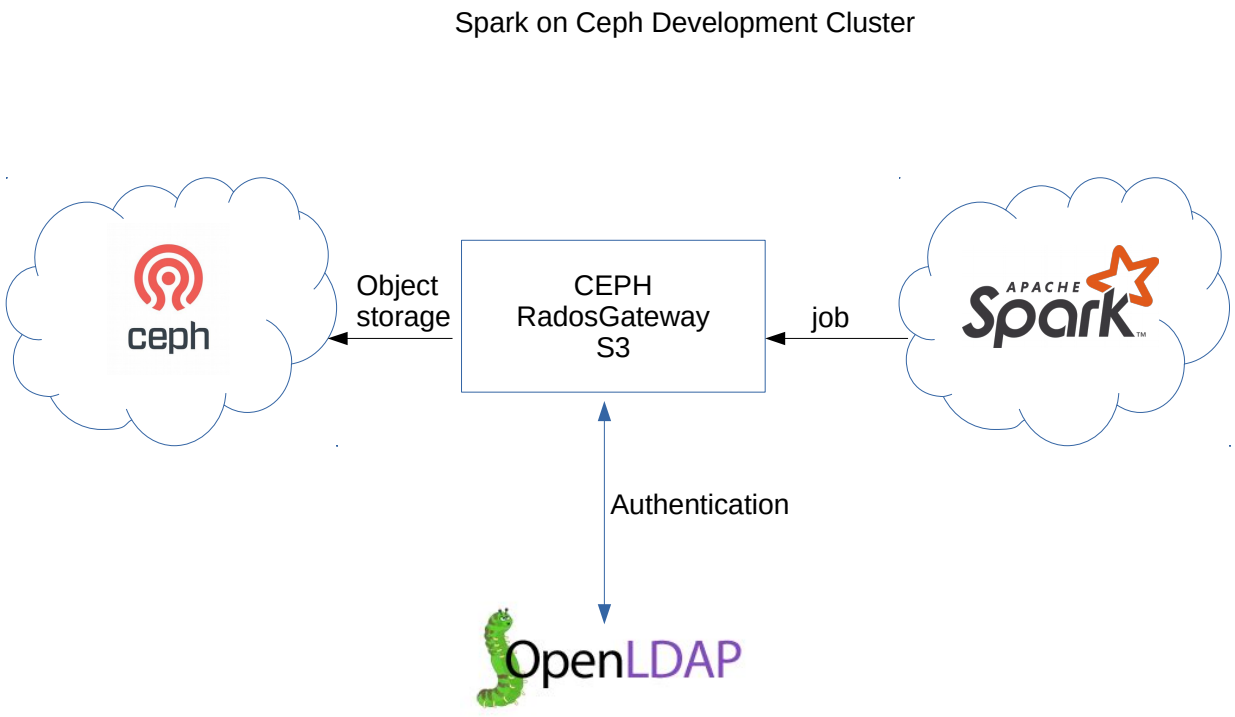
- Cluster Manager
  - Standalone : FIFO
  - Yarn : Capacity Scheduler, Fair Scheduler
  - Mesos : Coarse-grained mode, Dynamic Sharing

Spark@VirtualData

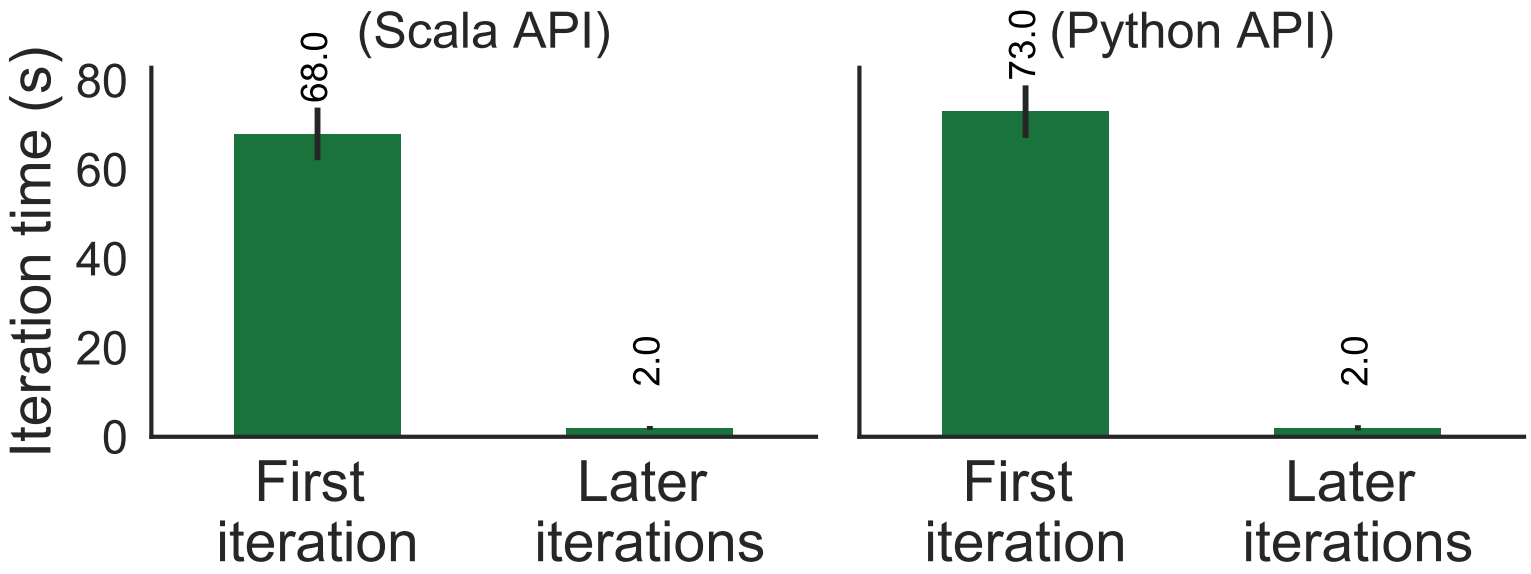
- Cluster de production (calcul distribué)
  - Openstack : Domain **u-psud**, Project **Spark**, CentOs7
  - 1 Master : Gabarit 18 cores/36GB RAM **os.18**
  - 1 Yarn Cluster Manager : Gabarit 18 cores/36GB RAM **os.18**
  - 9 Slaves : Gabarit 18 cores/ 36GB RAM, volume cinder HDFS **4To**
- Cluster Kafka (streaming distribué)
  - Openstack : Domain **u-psud**, Project **Spark**, CentOs7
  - 5 Vms : Gabarit **os.4**
  - volume cinder : **5 x 1To**
- Installation, administration
  - Manuellement : Création des VMs, scripts d'installation du master et des slaves
  - Evolution 1 : Terraform pour le déploiement des VMs, provisioning : user-data $\Rightarrow$ scripts
  - Evolution 2 : Full Ansible pour le déploiement, le provisioning et l'administration

Soumission de jobs en mode cluster

En cours de test de performance

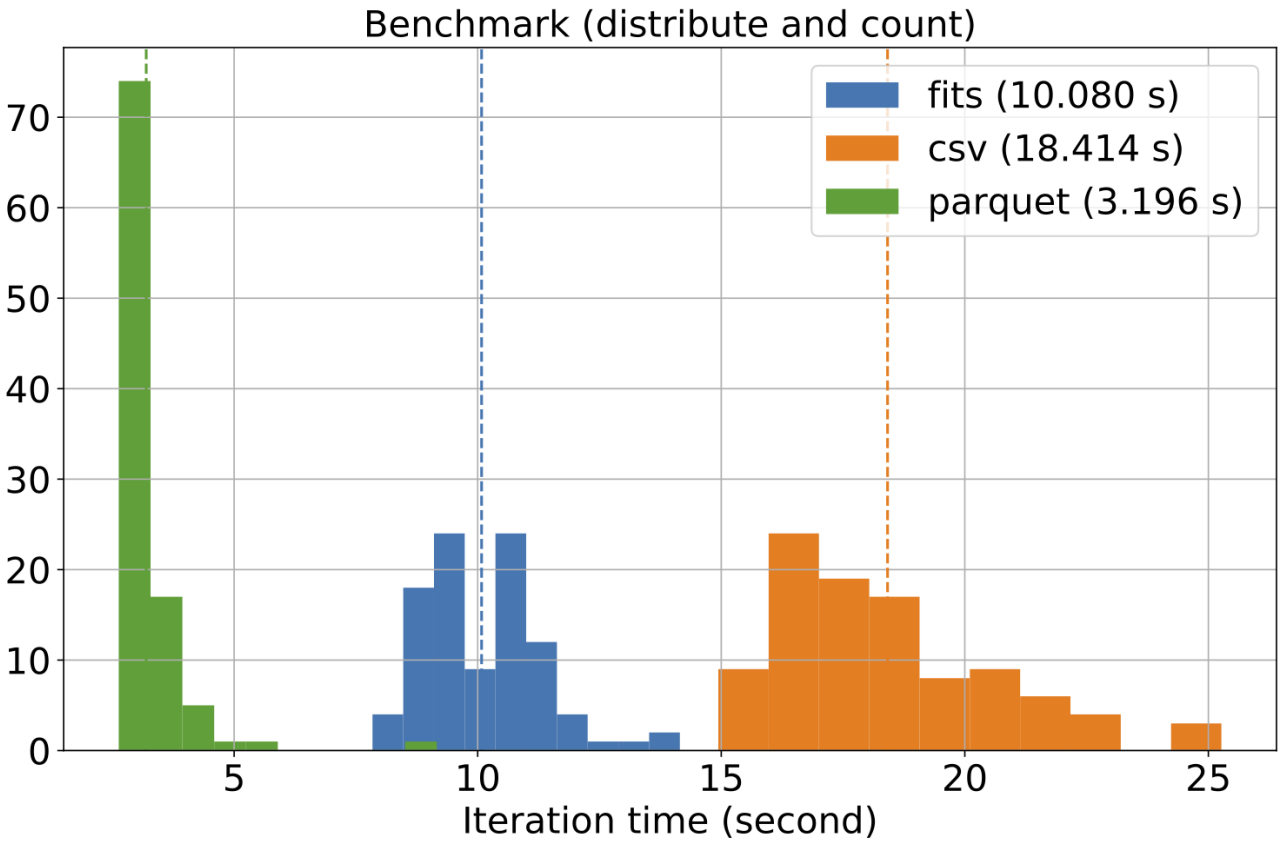


Connecter FITS et Apache Spark



- *FITS Data Source for Apache Spark* (spark-fits).
- Benchmarks avec 110 GB sur disque, 153 coeurs.
- Tests effectués jusqu'a 1.1 TB sur le cluster Spark, avec bon passage à l'echelle.

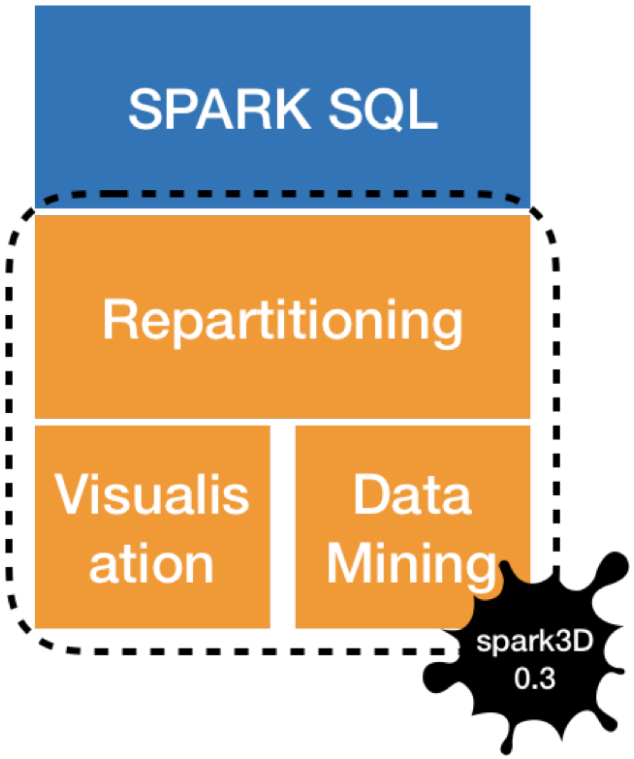
Connecter FITS et Apache Spark



- *FITS Data Source for Apache Spark* (spark-fits).
- Benchmarks sur cluster entre différents formats de données : resultats comparables.

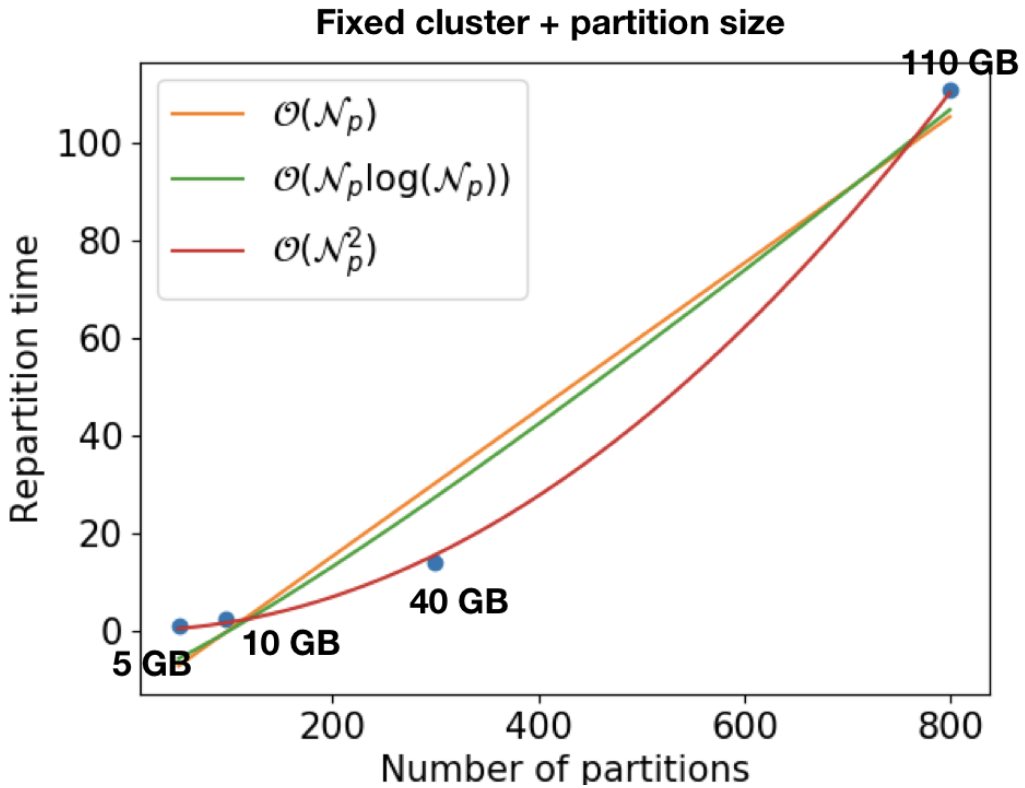
Manipuler des données 3D : spark3D

- Repartitionement distribué de données en 3D, requêtes spatiales, visualisation.
- Exemple : Recherche des plus proches voisins (KNN) : 6 milliards de galaxies, K=1000 pour une galaxie en O(10) sec.
- Projets liés : visualisation (inexlib) & machine learning.



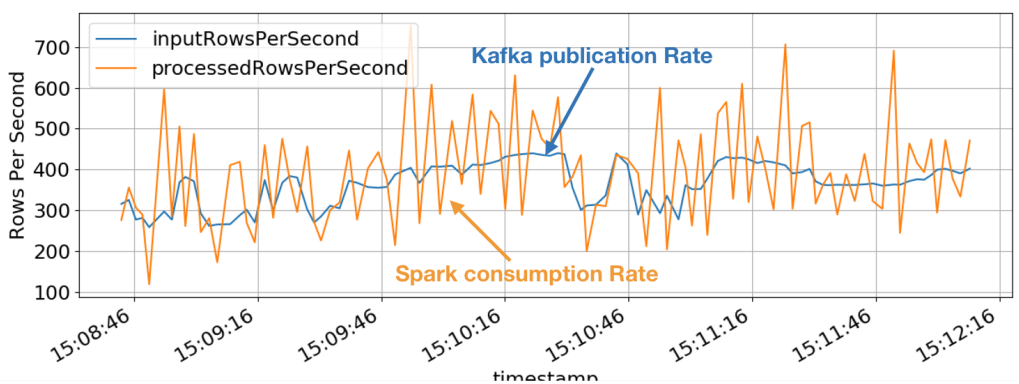
Manipuler des données 3D : spark3D

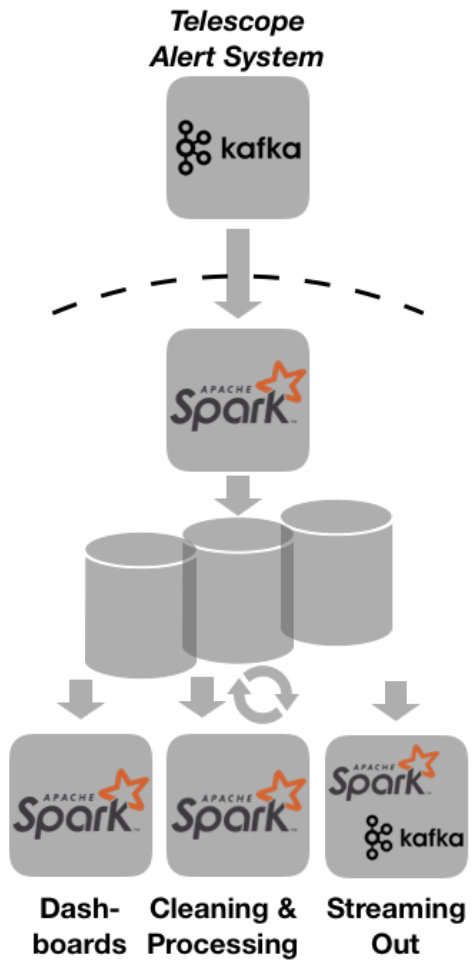
- Repartitioner les données est une operation coûteuse.
- Communication + délai de scheduler.
- Spark n'a pas la connaissance de la "physique" a priori pour optimiser les flux.



Capter les millions d'alertes du ciel

- Apache Kafka (producer) & Apache Spark Structured Streaming module (consumer/producer).
- Combinaison de modes streaming & batch. Archivage des données entrantes, et post-processing fait sur place.
- Tests fait jusqu'à 1000 alerts / seconde (100 MB/s).





Questions ¿

Merci pour votre attention.

Spark-Team (LSST/SI) :
C. Arnault <arnault@lal.in2p3.fr>, G. Barrand <barrand@lal.in2p3.fr>,
JE. Campagne <campagne@lal.in2p3.fr>, V. Givaudan <Valerie.Givaudan@lal.in2p3.fr>,
J. Hrivnac <hrivnac@lal.in2p3.fr>, M. Jouvin <jouvin@lal.in2p3.fr>,
S. Pateyron <pateyron@lal.in2p3.fr> , J. Peloton <peloton@lal.in2p3.fr>,
G. Philippon <philippo@lal.in2p3.fr>, S. Plaszczynski <plaszczy@lal.in2p3.fr>,
A. Ramparison <Adrien.Ramparison@lal.in2p3.fr>