
Utilisation d'une plateforme Hadoop/Spark pour des données astrophysiques

Christian Arnault*†¹

¹Laboratoire de l'Accélérateur Linéaire (LAL) – CNRS : UMR8607, IN2P3, Université Paris XI - Paris Sud – Centre Scientifique d'Orsay B.P. 34 91898 ORSAY Cedex, France

Résumé

Une expérimentation est menée dans l'environnement VirtualData/OpenStack au LAL. Plusieurs caractéristiques de cette technologie sont en cours d'évaluation:

- mise en place d'une infrastructure Hadoop/Spark/scipy sur une cloud OpenStack
- mesures de performance (configuration du cluster, suivi des facteurs d'échelle)
- apprentissage des modules fonctionnels de la plateforme Spark (Scala, Python)
- exploitation des données binaires structurées dans Spark à travers le module Hadoop/avro
- application aux fichiers FITS
- quelques exemples de cas d'usage dans le contexte de l'expérience LSST

Christian Arnault

Axel Chevarin

Réza Ansari

Johann Cohen Tanoudji

*Intervenant

†Auteur correspondant: arnault@lal.in2p3.fr