LABORATOIRE DE L'ACCÉLÉRATEUR LINÉAIRE

Séminaire LAL

Stéphane Plaszczynski
(LAL)

Mardi 13 novembre 2018 à 11h00

Spark pour les physiciens

Apache-Spark est une technologie issue du monde du big data très utilisée dans l'industrie mais assez peu dans celui de la recherche scientifique. Le but de ce séminaire interne est de présenter cet outil et ses potentialités en particulier pour l'analyse des gros volumes de données tels que ceux attendus par les prochains relevés de galaxies. Après une introduction pédagogique concernant le calcul distribue avec Spark et ses avantages, je présenterai les méthodes et performances obtenues sur un cas d'utilisation d'analyse d'une simulation de 10 ans de données de type LSST (6 milliards de galaxies). Puis je présenterai des développements récents obtenus au LAL en particulier dans le domaine de l'identification des clusters et de la visualisation. Enfin j'esquisserai l'intérêt de combiner du calcul haute-performance a ce type d'approche. Ce travail s'inscrit dans le cadre de l'organisation AstroLab (https://astrolabsoftware.github.io/) qui vise à insuffler de la complexité scientifique dans le traitement des larges volumes de données.

Salle 101 - Bât. 200, Orsay

Organisation : Reisaburo Tanaka (LAL) - seminaires@lal.in2p3.fr
LAL web : http://www.lal.in2p3.fr
Indico: https://indico.lal.in2p3.fr/category/31/

cnrs

UNIVERSITÉ PARIS SUD
université PARIS-SACLAY