



ID de Contribution: 25

Type: **Présentation orale 15 min.**

AstroLab: naviguer au milieu de milliards d'astres

mercredi 3 octobre 2018 11:00 (15 minutes)

AstroLab est une organisation pour le développement d'outils logiciels de pointe pour aider à résoudre des problématiques big data dans la recherche.

Partager les efforts de R&D entre groupes, améliorer l'inter-opérabilité entre le milieu industriel et celui de la recherche dans les projets open-sources, et développer de nouveaux outils collaboratifs permettra aux équipes de recherche d'exploiter au mieux les outils de l'écosystème du big data.

Dans cette présentation, je me concentrerai sur Apache Spark, un ensemble d'outils pour le calcul distribué principalement utilisé dans le milieu industriel, pour analyser des données de futur telescopes tels que LSST qui enregistrera des téraoctets de données par nuits pendant plusieurs années.

Plus en détails, je décrirai les mécanismes mis en place pour connecter le format de données scientifique FITS dans Spark, et j'étudierai les performances pour distribuer et manipuler des téraoctets de données distribuées sur un ensemble de machines. Pour finir, j'introduirai spark3D, une extension d'Apache Spark pour manipuler efficacement des jeux de données tri-dimensionnels venant de l'astrophysique, de la physique des hautes énergies, ou encore de la météorologie.

Thèmes associés

Calcul distribué
Programmation fonctionnelle
Analyse de données
Big data
Apache Spark

Auteur principal: PELOTON, Julien (CNRS)

Co-auteurs: ARNAULT, Christian (CNRS); PLASZCZYNSKI, stephane (CNRS)

Orateur: ARNAULT, Christian (CNRS)

Classification de Session: Présentations plénières

Classification de thématique: Calcul haute-performances