



ID de Contribution: 33

Type: Non spécifié

Big Data cosmology with Spark

mercredi 20 novembre 2019 12:30 (10 minutes)

L'environnement de travail Apache Spark est très utilisé dans l'industrie pour organiser et analyser les grands volumes de données. Je montrerai que cette "technologie Big Data" peut également être appliquée en science sans trop de souffrances, en particulier en astrophysique où les volumes de données des prochains grands relevés de galaxies augmentent drastiquement.

A partir d'une simulation simple mais réaliste de 10 ans de données du télescope LSST, je montrerai les divers intérêts d'une telle approche pour l'analyse interactive d'un relevé de 6 milliards de galaxies.

Je développerai l'aspect utilisateur et montrerai par des exemples simples et pratiques en Python comment obtenir d'excellentes performances sur un cluster de taille modeste.

Auteur principal: Dr PLASZCZYNSKI, stephane (CNRS)

Orateur: Dr PLASZCZYNSKI, stephane (CNRS)