



Discipline(s) : Informatique et télécommunications

BIG DATA STORAGE AND PROCESSING INFRASTRUCTURES

Nature

UE

RESPONSABLES

Gabriel Antoniu

OBJECTIFS

Le but de ce module est de proposer une introduction à la gestion des données massives (Big Data) et à la science des données (Data Science) : principaux concepts, défis, domaines applicatifs, présentation des principaux systèmes représentatifs de l'état de l'art, etc.). Il introduit les principaux modèles de stockage et de traitement des données, notamment le modèle MapReduce et ses dérivés, ainsi que les principales technologies existantes, telles que Hadoop, Spark, Flink, etc. Des descriptions détaillées et des analyses comparatives de ces systèmes seront proposées, afin de permettre une compréhension des objectifs sous-jacents, des domaines applicatifs possibles, des choix architecturaux. Nous étudierons les techniques largement utilisées pour la distribution de traitement de données massives (gossip, flux, vidéo, ...). Nous détaillerons en particulier des protocoles de multicast/streaming vidéo ainsi que des systèmes décentralisés pour l'agrégation des données.

MOTS-CLÉS

Big Data analytics, MapReduce, flux de données, cloud computing, science des données,

PRÉREQUIS

Aucun

CONTENU

Introduction aux Big Data: concepts, enjeux, domaines applicatifs.
Infrastructures physiques et architectures logicielles pour la gestion de de données réparties à très grande échelle: défis, principes de conception, exemples
Modèles et systèmes de traitement de données sur des infrastructures réparties à très grande échelle
 Bref survol de MapReduce et de Hadoop
 Au delà de MapReduce : limitations du modèle, extensions, approches post-Hadoop, études de cas : Spark, Flink, etc.
 Etudes comparatives (architectures, performances, applications)
Techniques pour la distribution de traitement de données massives (gossip, flux, y compris vidéo, ...)
Protocoles de multicast/streaming vidéo
Systèmes d'agrégation décentralisés

COMPÉTENCES ACQUISES

Savoir : défis, motivations, enjeux, domaines applicatifs, modèles de programmation, systèmes de stockage et d'analyse pour les Big Data, architecture et fonctionnement des technologies existantes (Hadoop, Spark, Flonk, etc.).

Savoir faire : modéliser les processus d'analyse de données via le paradigme MapReduce et ses dérivés, proposer des architectures de traitement et de stockage pour un problème donné.

APPARTIENT À

Master 2 informatique parcours Science Informatique

Mise à jour le 17 juillet 2017

CONTACT(S)

Département Informatique et télécommunications

École normale supérieure de Rennes Campus de Ker Lann Avenue Robert Schuman

35170 BRUZ

Tél. : 02 99 05 52 43

E-mail

Site Internet