

Proposition de plan du groupe 7

Thème 4 : Spark : Traitement de données massive en temps réel

Introduction

- Contexte (Big Data, limites de Hadoop et SQL)
- Transition vers Apache Spark
- Problématique
- Objectifs de l'exposé (présenter les fondements techniques d'Apache Spark, montrer la pertinence de Spark pour le batch et le streaming temps réel, illustrer concrètement l'intégration Spark-Cloud à travers AWS et Databricks)

I. Cadre théorique : Apache Spark

- Présentation générale et concepts de base
- Architecture de Spark (driver, executorn cluster manager, fonctionnement global)
- Modules principaux (Spark SQL, Streaming, MLlib...)
- Méthodologie de traitement : batch et streaming
- Intégration dans les environnements réels (HDFS, AWS, Databricks...)
- Forces et limites

II. Partie pratique : intégration Cloud et Big data

- Utilisation d'AWS et Databricks
- Cas pratique 1 : traitement batch et génération automatique d'un rapport
- Cas pratique 2 : traitement en streaming et création d'un dashboard

Conclusion