**SPARK**

Table des matières

[Chapitre 1 INTRODCTION : 2](#_Toc103158879)

[Chapitre 2 : ENVIRONNEMENT DE DEVELOPPEMENT UNIFIE 2](#_Toc103158880)

[Chapitre 3 2](#_Toc103158881)

[Titre 1 2](#_Toc103158882)

[Chapitre 4 2](#_Toc103158883)

[Titre 1 2](#_Toc103158884)

[Chapitre 5 2](#_Toc103158885)

[Titre 1 2](#_Toc103158886)

[Chapitre 6 2](#_Toc103158887)

[Titre 1 2](#_Toc103158888)

# 

# INTRODCTION :

* Apach spark est un framework de calcul distribué en memory (parce qu’il écrit sur le disque rarement.) => traitement très rapide. (Plus optimisé par rapport à map reduce qui sollicite l’écriture à chaque étape : les workers font des vas et viens entre la ram et le disk)
* Un framework c’est lui qui dessine qui encadre le programmeur
* Il est écrit en langage scala.
* Spark est facile à utiliser. (Commande : spark submit)
* Possibilité de développement en local (mode standalone, émulateur spark sur Eclipse)
* On peut effectuer différentes tâches avec spark : etl ; sql, Machine Learning, analyses de données,
* Tolérant aux pannes
* Beaucoup plus performant et beaucoup plus complet que map reduce.
* Il est basé sur les RDD (Resilient Distributed Data set) : Une API similaire à l’api collection : consiste à charger dans la RAM de chaque nœud d’un cluster spark la même méthode (variable distribuée) et de traiter les données de chaque nœud avec la même méthode en même temps (en parallèle)
* Librairie très riche : JDBC, SQL, traitement des chaines de caractères…

# : ENVIRONNEMENT DE DEVELOPPEMENT UNIFIE

* Il est interactif avec spark shell
* Il supporte plusieurs types de données
* Il supporte plusieursmanagers de ressources (YARN, Mesos, Standalone)

# : LA SUITE DE APACH SPARK

## Titre 1

# 

## Titre 1

# 

## Titre 1

# 

## Titre 1