

# Evaluation des dossiers sur la programmation en R

Les « 12 travaux d'AstéRix »

Daif Hakim

23/01/2021

#HADOOP ET SPARK

## 1. Critères d'évaluation

Dans ce dossier, nous allons parcourir et faire une évaluation des travaux de mes camarades sur la Programmation en R:

Nos critères seront :

- 1. Visuel et l'organisation du dossier:** Es ce que la mise en page est propre ? Le travail est-il organisé (chapitres, sections, sous-sections, Parties) ?
- 2. Qualité de la rédaction du dossier:** le document est-il agréable à lire? es ce que l'auteur a respecté les consignes de rédaction d'un article scientifique ou d'une thèse... ?
- 3. Compréhension de l'idée générale:**Les Auteurs arrivent-ils à transmettre l'idée générale ?
- 4. lisibilité et fonctionnement du code (explication et commentaires):** Es ce que le code fonctionne parfaitement sans erreurs ? Es ce que l'auteur a mis des commentaires pour expliquer chaque ligne de commande ?
- 5. L'utilisation d'exemples et de cas concrets (cas pratique):** l'Auteur a t'il utilisé des exemples et des cas pratique lors de sa rédaction ?

## 2. Lien vers le document commenté

Pour ce dossier, nous avons étudié le travail de Florine COMLAN et Ramya HOUNTONDJI qui porte sur HADOOP ET SPARK

GitHub : [https://github.com/fcom-stack/PSBX/tree/main/Hadoop\\_et\\_spark](https://github.com/fcom-stack/PSBX/tree/main/Hadoop_et_spark)

## 3. Synthèse de travail :

Dans un travail réparti sur 6 parties Florine et Ryma ont abordé deux outils de gestion et traitement de données massives à savoir Hadoop et Spark en commençant par l'explication des anciennes pratiques de traitement et stockage des données à savoir l'architecture client/serveur classique pour arriver à la nécessité de créer une nouvelle stratégie appelée architecture distribuée – cluster computing.

Après cette introduction, elles ont exposé les composants de base de cette dernière stratégie et son écosystème pour entamer, ensuite, une comparaison entre deux outils phares de cette nouvelle technologie qui sont Hadoop et Spark et expliquer enfin la méthode d'installation de Hadoop avec Docker.

#### 4. Extrait commenté :

La première partie était assez courte mais très bien détaillée pour illustrer le besoin de développement d'une telle technologie ce qui a attiré toute notre attention pour suivre l'enchaînement du travail. La partie de présentation de l'écosystème de Hadoop est bien concise par le schéma ce qui donne une vision globale simplifiée. Enfin, la partie de comparaison entre les deux concurrents Hadoop et Spark a évoqué les caractéristiques, différences et inconvénients de chacun ce qui nous a aidé à mieux comprendre leurs fonctionnements afin de nous faciliter le choix d'un modèle plus adéquat pour nos projets futurs de ce type.

#### 5. Evaluation du travail suivant les 5 critères précités

**1. Visuel et l'organisation du :** Le visuel du document est bien travaillé, les textes sont bien alignés avec un interligne respecté.

**2. Qualité de la rédaction du dossier :** La qualité rédactionnelle est à la hauteur du travail, le document est agréable à lire.

**3. Compréhension de l'idée générale :** L'enchaînement logique des éléments du dossier a facilité la compréhension de l'idée générale, appuyé par de multiples exemples et comparaison des outils ainsi que l'utilisation d'images illustratives et schémas récapitulatives.

**4. Lisibilité et fonctionnement du code :** Les étapes d'installation et d'exécution sont très bien détaillées et organisées d'une façon lisible avec un tableau récapitulatif des différentes instructions et leurs fonctionnalités.

**5. Qualité des applications permettant d'illustrer le package :** Quant aux applications d'exécution de ces outils, ils restent très volumineux et compliqués à installer à cause de la multitude d'erreurs signalées tout au long du processus d'installation et qui ont exigé un matériel très performant en termes de RAM et qualité du processeur.

#### 6. Conclusion

La qualité de travail fournie tout au long de ce dossier est irréprochable malgré la complexité du sujet traité, cependant, il est tout à fait possible de l'enrichir encore davantage. Par ailleurs, la partie 6 traitant les Packages introduite dans le plan de travail du début mais demeure inexistante dans le corps du rapport.