

Semestre : 1 ☒ 2 ☐

Session : Principale ☐ Rattrapage ☒

ETUDIANT(e)

Nom et Prénom :

Code :

Classe:

Module : Big Data

Enseignant(s) : Asma Hamed, Henda Sfaxi, Ines Channoufi, Ines Slimene

Classe(s) : 5 ARCTIC, 5 ERP-BI, 5 GL, 5 SIGMA, 5 TWIN

Documents autorisés : OUI ☐ NON ☒ Nombre de pages : 05

Calculatrice autorisée : OUI ☐ NON ☒ Internet autorisée : OUI ☐ NON ☒

Date : 07/12/2017 Heure : 11h15 Durée : 1h30



Code	Note /20	Nom et Signature du Surveillant	Nom et Signature du Correcteur	Observations

Module :

NB : Les parties réservées aux informations relatives à l'étudiant et à l'administration seront ajoutées au cas où les réponses seront écrites sur les feuilles d'examen.

Exercice : QCM

Cocher la(les) bonne(s) réponse(s).

Q1 : Qui a créé le framework logiciel Hadoop, très populaire dans le monde du Big Data ?

- ☐ Doug Cutting
- ☐ Richard Stallman
- ☐ Alan Cox

Q2 : Soit un bloc avec une taille de 128Mo par défaut et un facteur de réplication de 3. Combien de blocs de données seront créés dans HDFS pour un fichier avec une taille de 1Go.

- ☐ 8
- ☐ 16
- ☐ 24
- ☐ 1024

NE RIEN ECRIRE



Q3 : Laquelle des commandes suivantes va copier le fichier ‘/user/cloudera/Myfile.txt’ dans le répertoire ‘/user/hadoop/hadoopfile’

- ☐ Hadoop cp Myfile.txt /user/hadoop/hadoopfile
- ☐ Hadoop fs -cp Myfile.txt /user/hadoop/hadoopfile
- ☐ Hadoop fs -put Myfile.txt /user/hadoop/hadoopfile
- ☐ Hadoop fs put Myfile.txt /user/hadoop/hadoopfile

Q4 : Hadoop est efficace pour :

- ☐ Les applications ayant de grands ensembles de données
- ☐ Les applications ayant beaucoup de petits fichiers
- ☐ Les applications ayant de grands ensembles de données et/ou beaucoup de petits fichiers

Q5 : HDFS permet :

- ☐ La lecture des données seulement
- ☐ La lecture et l'écriture des données seulement
- ☐ La lecture, l'écriture et la mise à jour des données

Q6 : Peut-on considérer HDFS comme une alternative au datawarehouse ?

- ☐ Vrai
- ☐ Faux

Q7 : Quelle est la caractéristique clé, qui fait la différence entre HDFS1 et HDFS2

- ☐ Accès aléatoire aux données
- ☐ L'accès aux données se fait à travers Hive/Pig
- ☐ Haute disponibilité du NameNode
- ☐ Le TaskTracker est remplacé par Node Manager

NE RIEN ECRIRE



Q8 : Le modèle de programmation MapReduce a été initialement développé par...

- ☐ Apache Software Foundation
- ☐ Google
- ☐ Microsoft Research

Q9 : MapReduce peut être décrit comme un modèle de programmation utilisé pour développer des applications basées sur Hadoop capables de traiter des quantités massives de données non structurées.

- ☐ Vrai
- ☐ Faux

Q10 : Laquelle des affirmations suivantes est correcte par rapport au nombre de mapper et reducer dans un cluster

- ☐ Le nombre de mappers et reducers peut être modifié dans la fichier mapred-site.xml
- ☐ Le nombre de mappers et reducers est fixé par le NameNode
- ☐ Le nombre de mappers et reducers sont toujours égaux
- ☐ Le nombre de mappers doit être égal au nombre de nœuds dans un cluster

Q11 : Laquelle des affirmations suivantes est incorrecte par rapport à l'exécution d'un Job Map Reduce

- ☐ Les fichiers d'entrée doivent être stockés dans le système local
- ☐ Les fichiers d'entrée doivent être stockés sous HDFS
- ☐ Quel que soit l'emplacement des fichiers d'entrée, le programme MapReduce fait automatiquement une copie dans l'emplacement d'exécution du Job Mapreduce.

NE RIEN ECRIRE



Q12 : Dans Hadoop1, si le JobTracker tombe en panne :

- ☐ Le Job Tracker du Secondary Namenode poursuit le Job
- ☐ Le Job s'arrête
- ☐ Le Scheduler poursuit le job

Q13 : Que fait Hadoop lorsqu'un job/une tâche échoue ?

- ☐ Redémarre le job sur le même DataNode
- ☐ Redémarre le job sur un autre DataNode
- ☐ Abandonne le job après un certain nombre de tentatives sur le DataNode de départ

Q14 : Chaque base de données créée dans hive va être stockée en tant que :

- ☐ Répertoire
- ☐ Fichier
- ☐ Bloc HDFS
- ☐ Fichier jar

Q15 : L'avantage de créer des tables hive partitionnées est :

- ☐ Utilisation de stockage en mémoire
- ☐ Vitesse d'exécution des requêtes améliorée
- ☐ Le namemode nécessite moins de RAM
- ☐ Syntaxe de requête plus simple

Q16 : Quelle entreprise a développé Apache Hive ?

- ☐ Linked IN
- ☐ Tweeter
- ☐ facebook
- ☐ yahoo

Semestre : 1 ☒ 2 ☐

Session : Principale ☐ Rattrapage ☒

ETUDIANT(e)

Nom et Prénom :

Code :

Classe:

Module : Big Data

Enseignant(s) : Asma Hamed, Henda Sfaxi, Ines Channoufi, Ines Slimene

Classe(s) : 5 ARCTIC, 5 ERP-BI, 5 GL, 5 SIGMA, 5 TWIN

Documents autorisés : OUI ☐ NON ☒ Nombre de pages : 05

Calculatrice autorisée : OUI ☐ NON ☒ Internet autorisée : OUI ☐ NON ☒

Date : 07/12/2017 Heure : 11h15 Durée : 1h30



Q17 : Quels sont les modes d'exécution proposés par Pig :

- ☐ Local
- ☐ Distribué
- ☐ Hadoop
- ☐ Parallèle

Q18 : Dans le cas des SGBDNR, pour quelle raison les propriétés ACID ne sont plus vérifiées :

- ☐ La scalabilité horizontale constitue une limite par rapport au stockage de données
- ☐ Dans un environnement distribué, les données ne seront plus consistantes.
- ☐ La scalabilité verticale ne garantit pas la durabilité des données.
- ☐ L'accès aux données est aléatoire avec une faible latence.

Q19 : Laquelle des affirmations suivantes n'est pas une raison pour que les SGBDNR deviennent une solution populaire pour certaines organisations ?

- ☐ Meilleure scalabilité
- ☐ Améliore la possibilité de garder les données consistantes
- ☐ Accès à la donnée plus rapide que les SGBDNR
- ☐ Rend plus simple le stockage de données sur plusieurs serveurs de données.

Q20 : Dans HBASE, les données sont stockées physiquement dans :

- ☐ NameNode
- ☐ DataNode
- ☐ Hfile
- ☐ Region Server