QCM CURSUS Cycle certifiant Data Scientist

Questionnaire

**« Big Data, méthodes et solutions pratiques pour l'analyse des données »**

**Question n°1 (général)**

D’après vous quelles sont les caractéristiques clés de la Big Data face à la Business intelligence classique ?

* La Big Data traite aussi bien des données structurées et non structurées
* La Big Data traite des volumes de données variables et massifs, qu’elles soient structurés ou non structurées
* La Big Data est le remplaçant de la Business Intelligence car elle permet de réaliser des analyses de données plus précises et plus fréquentes

Bonne Réponse : 2

**Question N°2 *(Big data)***

Lorsque l’on parle de la règle des « 3V », de quoi parle-t-on exactement ?

* Vélocité, Valeur, Variétal
* Vélocité, Volume, Variété
* Vague, Volume, Vélocité
* Veridique, Vélocité, Volume

Bonne Réponse : 2

**Question N°3 *(no sql)***

Le théorème CAP permet de classer les technologies de base de données SQL. Pouvez indiquer ce que permet de recouvre comme terme ce théorème CAP ?

* Consistency, Availability, Partition Tolerance
* Adaptability, Consistency, Persistent
* Convergence, Partition Tolerance, Adaptability
* Conversation, Availability, Proof of concept

Bonne Réponse : 1

**Question N°4 *(no sql)***

Indiquer par les choix ci-dessous les technologies qui ne recouvrent que des bases de données de type « NoSQL »

* MongoDB, Big Table, Hbase, Mysql
* Redis, Cassandra, CouchDB, Dynamo
* RIAK, Postgres, Voldemort, Cassandra
* Neo4J, Oracle, MongoDB, Vertica
* Scalaris, MemCacheDB, Apache Spark, Voldemort

Bonne Réponse : 2

**Question N°5 *(HDFS)***

Le système de fichier HDFS permet de distribuer des données sur plusieurs  machines, pourriez-vous en confirmer le principe de fonctionnement dans le cadre de l’enregistrement d’un fichier ?

* Le client contacte un NameNode qui lui indique ensuite les DataNode où enregistrer les différents blocks de ce fichier
* Le client contacte un NameNode qui lui indique un DataNode où enregistrer le fichier, ce fichier sera ensuite découpé et transféré aux autres Datanode
* Le client contacte chaque DataNode pour y stocker les blocks de fichiers et indique au NameNode le dataNode où sont stockés les différents blocks
* Le client enregistre tous les blocks du fichier sur le NameNode qui se charge ensuite de les répartir sur les différents DataNode

Bonne réponse : 1

**Question N°6 *(HDFS)***

Le RF désigne un paramètre clé dans la technologie HDFS, Pourriez-vous indiquer de quoi il s’agit ?

* RF désigne le Reverse Forward qui permet de déterminer l’ordre dans lequel sont récupérés les différents blocks de données d’un fichier
* RF est le Replication Factor qui permet d’indiquer sur combien de DataNode sont répartis les blocks de données d’un fichier
* RF désigne le Real Factor qui permet au DataNode de gérer son mécanisme de copie de fichier entre les différents DataNode
* RF est le Reverse Forward qui permet au client de fixer le débit de téléchargement des blocks de données d’un fichier

Bonne réponse : 2

**Question N°7 *(Map Reduce)***

La technologie Map Reduce permet de traiter de manière distribuée des traitements sur des données massives, pourriez-vous indiquer les grandes étapes de ce traitement ?

* Map, Shuffle, Reduce, Output
* Shuffle, Map, Reduce, Output
* Output, Shuffle, Map, Reduce
* Input, Map, Reduce, Shuffle, Output

Bonne réponse : 1

**Question N°8 *(hadoop)***

Lorsque l’on parle de l’écosystème Hadoop OSS (open source software), quels logiciels de cet écosystème en font partie ?

* Zookeeper, Oozie, HDFS, MapReduce, Hive, Sqoop, Flume, Avro
* Ambari, HDFS, MapReduce, Hive, Sqoop, Flume, Avro
* MapReduce, Hive, Sqoop, Flume, Avro, Drill, YARN
* Flume, Avro, HUE, HBASE, IMPALA, MapReduce, Hive
* Pig, Oozie, HDFS, Hive, Flume, Impala

Bonne réponse : 1

**Question N°9 *(visualisation)***

Lorsque l’on veut visualiser des données big data, de quelle manière peut-on connecter les outils de visualisation aux données stockées sur HDFS ?

* Via HBASE qui fournit une interface universelle de type SQL d’accès aux fichiers de HDFS
* Via Hive qui fournit un connecteur en pseudo-SQL qui permet de faire des requêtes sur les données du HDFS
* Via Oozie qui fournit un connecteur TCP permettant aux logiciels de visualisation de se connecter et d’extraire les données stockées sur le HDFS
* Si le logiciel de visualisation possède un connecteur Hive, on peut le connecter à HDFS et faire des requêtes pour extraire les données intéressantes pour l’analyse

Bonne réponse : 4

**Question N°10 *(R)***

R est un logiciel intéressant pour la Big Data, pourriez-vous indiquer en quoi ce logiciel est pertinent pour l’analyse de données ?

* R est un logiciel opensource possédant de nombreux connecteurs spécifiques aux technologies Big Data ce qui lui permet de traiter la données directement sur le HDFS
* R est un logiciel permettant de réaliser des analyses poussées de données via des librairies extensibles et customisables en mathématiques et statistiques
* R est un logiciel qui permet de traiter n’importe quel type de données stockées directement sur le HDFS sans avoir besoin de créer des jeux de données spécifiques

Bonne réponse : 2