

## Examen Final :

Durée : 02h45

### I. QCM : (3.75 points)

Choisir la bonne réponse (Bonne réponse : 0.75 pt ; Fausse réponse : -0,25 pt ; Aucune réponse : 0 pt)

1) HBase est un SGBD orienté :

Colonne	Document	Graphe	Clé/Valeur
---------	----------	--------	------------

2) Pig est construit au-dessus de ...

HBase	Hadoop	HDFS	Giraph
-------	--------	------	--------

3) HBase est construit au-dessus de ...

Pig	HDFS	RDBMS	Giraph
-----	------	-------	--------

4) La commande pour afficher un enregistrement HBase est :

Get	Scan	Print	Out
-----	------	-------	-----

5) Permet de gérer les tâches d'une même application dans un Cluster YARN

NameNode	JobTracker	Ressource Manager	Application Master
----------	------------	-------------------	--------------------

### II. Questions de Cours : (5 points)

1. Quels sont les différents fichiers de configuration de Hadoop ?
2. Définir le **SPOF**, ses caractéristiques et solutions?
3. Vous rencontrez un problème de calcul **itératif intensif** et on vous propose de choisir, entre Hadoop et Spark pour résoudre le problème. Vous allez choisir quel Framework ? pourquoi ?
4. Quels sont les avantages des Base de données NoSQL par rapport aux bases de données traditionnelles?
5. Pour les problèmes ci-dessous, proposer le type des Base de données NoSQL le plus adéquat pour le stockage/traitement des données.
  - a. Le vote
  - b. Sondage
  - c. Réseaux sociaux
  - d. Indexation des documents

### III. NoSQL: (3.25points)

1. Dans le cas de gestion des achats et paniers des utilisateurs dans Amazon considérons le cas suivant :  
Quand un utilisateur clique sur le bouton « acheter » que faut-il faire ?
  - a. Assurer la cohérence entre les serveurs
  - b. Ajouter l'article dans le panier et assurer la vente

Justifier votre réponse en se basant sur le théorème de CAP en rappelant son principe.

2. « A l'arrivée des bases de données NoSQL, les bases de données relationnelles n'auront plus de place dans les systèmes d'information des entreprises. » Vous êtes d'accord avec cet énoncé ? Justifier votre réponse
3. Lier chaque cas d'application avec la base de données convenable

1- Réseau social
2- Moteur de recherche
3- Compteur de vitesse instantanée

A- Cassandra
B- Neo4j
C- MongoDB

### IV. Algorithmique Map-Reduce : (8 points)

- 1) On considère deux matrices A et B, carrées et creuses, c'est-à-dire comportant beaucoup de valeurs nulles (0.0). Elles sont donc stockées sous la forme d'ensembles de triplets  $(i, j, val_{i,j})$  où  $val_{i,j} \neq 0$ . Ce format est plus économique qu'un stockage sous forme de tableau dès qu'il y a principalement des valeurs nulles dans les matrices, et reste générique (adapté à tout type de matrices creuses).

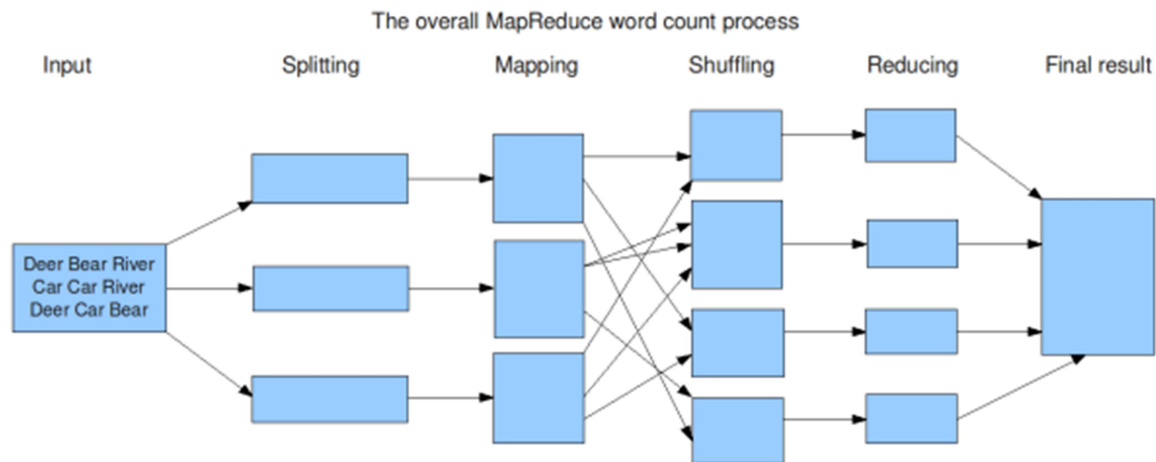
**Question :** On suppose que la longueur du côté des matrices est connue, et notée n. Proposez un algorithme Map-Reduce pour calculer la matrice carrée  $C = A \times B$ , de  $n \times n$  éléments.

- 2) Soit le bloc suivant :

A B R  
C C R  
A C B

Tracer un schéma expliquant le principe MapReduce qui consiste à calculer le nombre d'occurrence de chaque caractère sur un cluster de 3 nœuds.

- 3) Remplissez les cases vides :



4) Soit la matrice M suivante :

**1 2 3**

**4 5 6**

**7 8 9**

Tracer un schéma expliquant le principe MapReduce qui consiste à calculer la somme des lignes de M sur un cluster de 3 nœuds.