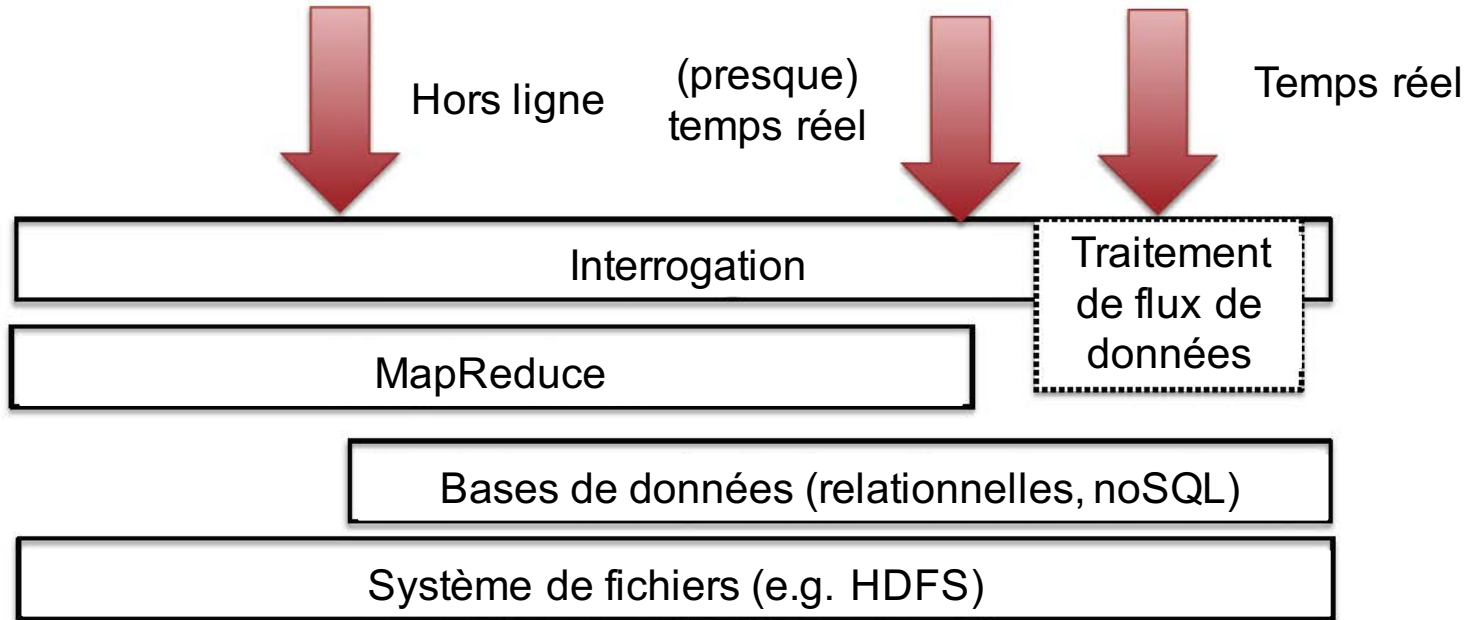


4. Gestion des données urbaines dans les nuages informatiques

- Brève histoire des nuages informatiques
- Modèles de service et de déploiement
- Technologie clé : la virtualisation
- IaaS : les points de vue utilisateur et fournisseur
- PaaS : programmation et déploiement des applications
- Stockage de données
- **Traitement de données**
- Traitement de flux de données

Architecture du système en fonction du type d'analyse



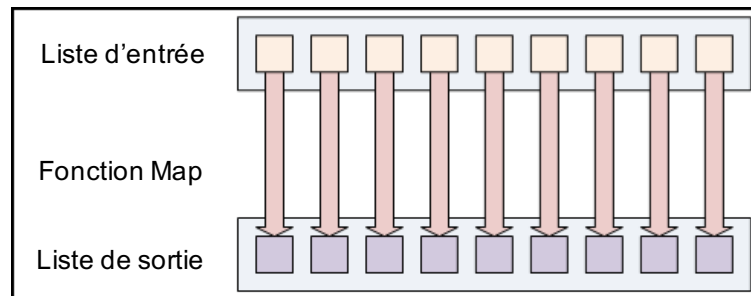
MapReduce : un modèle de programmation pour le traitement de données

- **Programmation fonctionnelle**
 - Deux fonctions : Map et Reduce
- **Modèle maître-esclave**
 - Données traitées en parallèle par différentes tâches
 - Chaque tâche traite un sous-ensemble des données

Un exemple : comptage de mots dans un texte

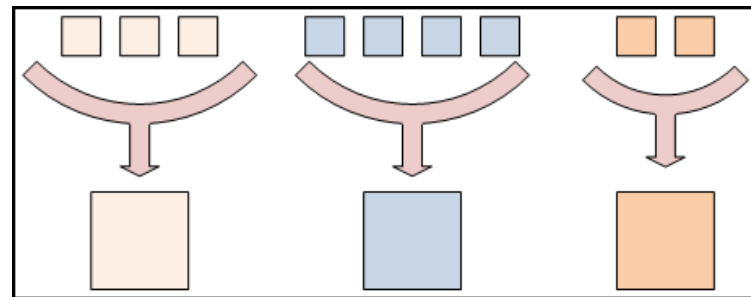
Map (fichier)

pour chaque ligne du fichier :
 pour chaque mot de la ligne :
 émettre (mot, 1) ;



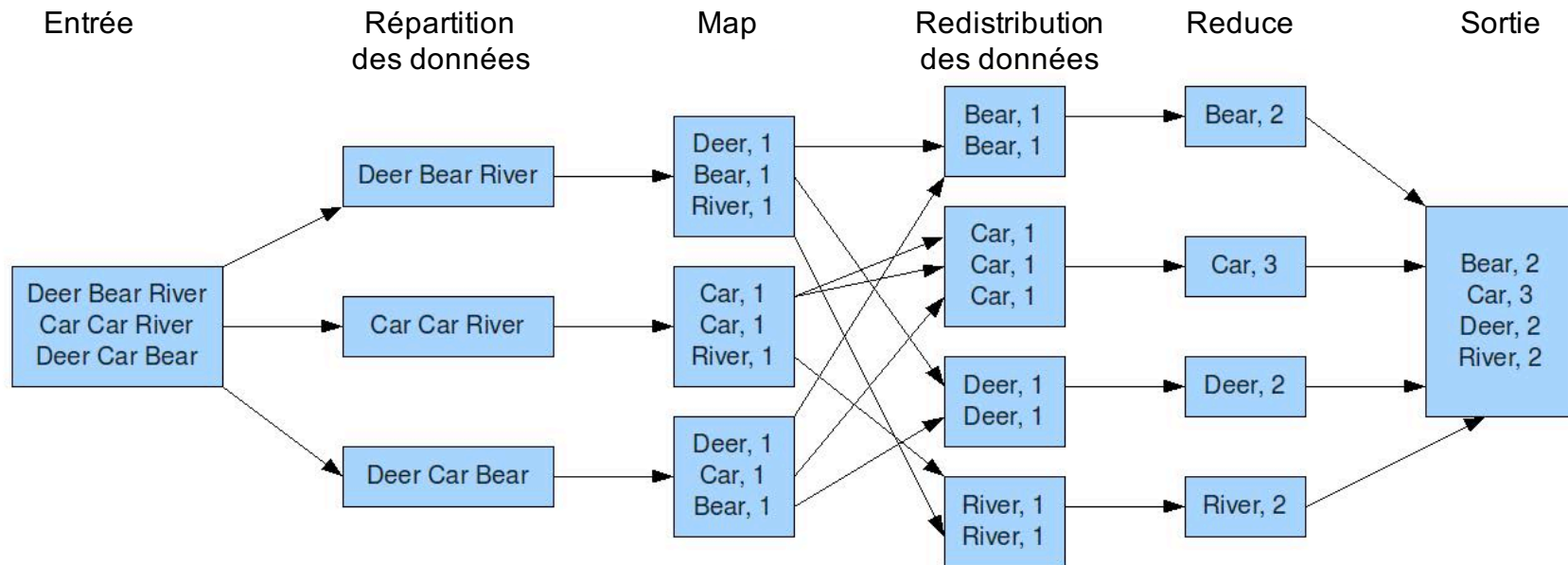
Reduce (mot, valeurs)

somme = 0 ;
pour chaque valeur dans valeurs :
 somme = somme + valeur ;
émettre (mot, somme) ;

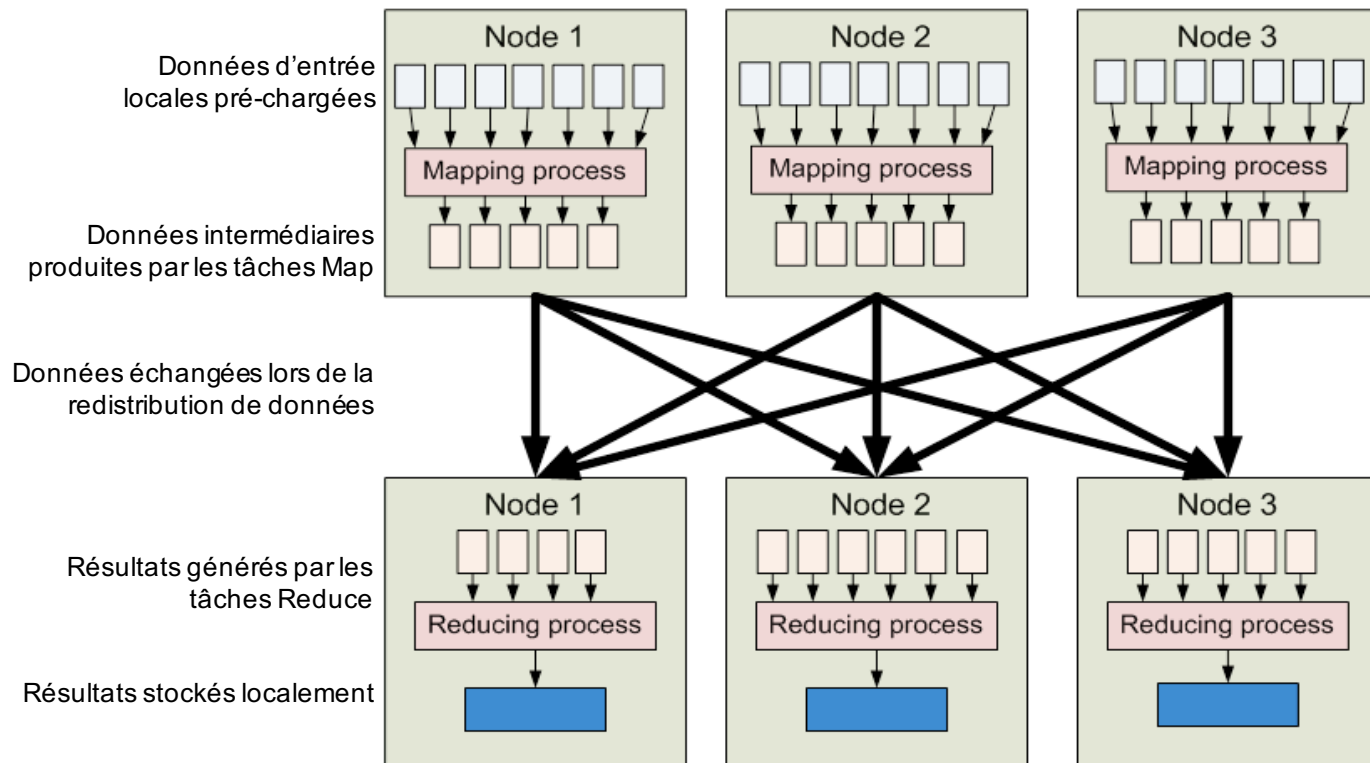


Un exemple d'application MapReduce

MapReduce – Application de comptage de mots dans un texte



Environnement d'exécution MapReduce

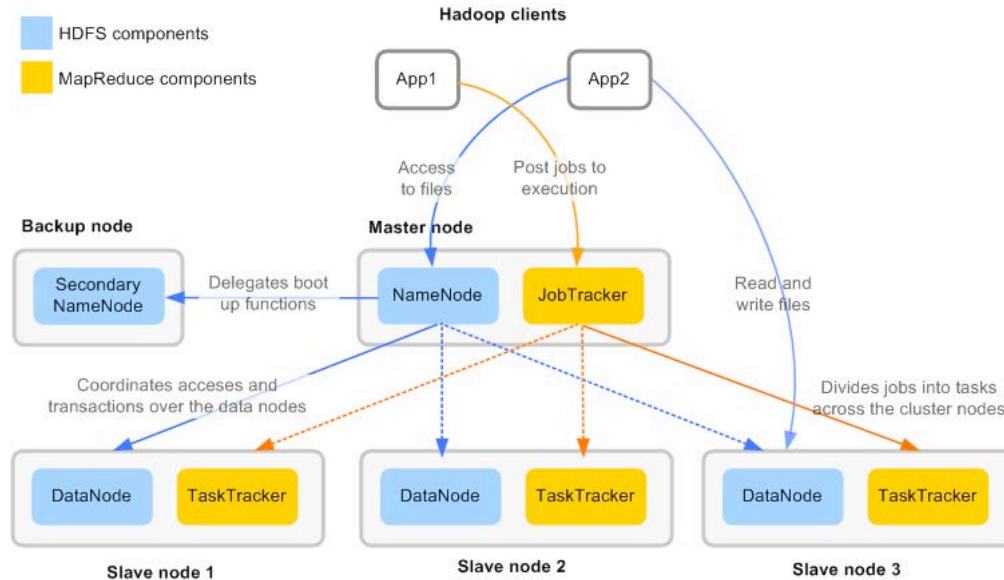


Hadoop MapReduce



For more information visit me at
www.hadooper.blogspot.com

Hadoop base platform brief



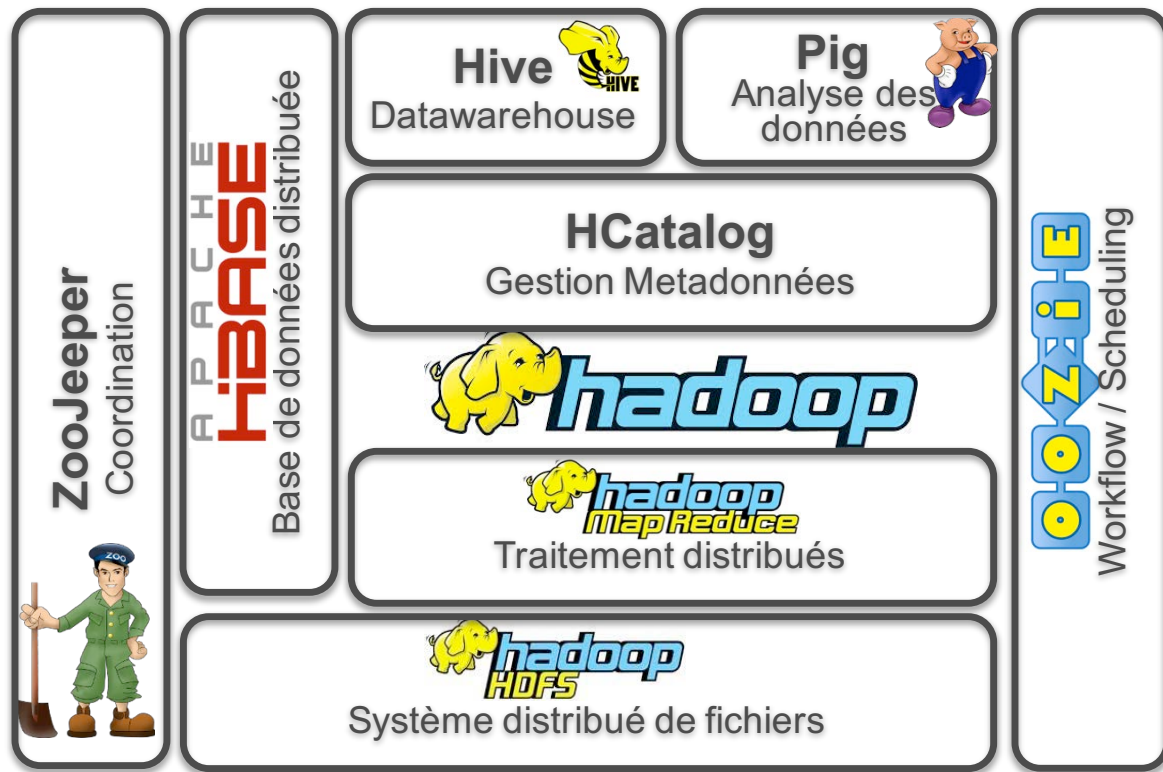
De nombreux paramètres qui impactent les performances

- Caractéristiques des nœuds (nombre de cœurs, capacité mémoire, taille disque)
- Latence et débit réseau
- Taille des blocs
- Degré de réplication des blocs
- Nombre de tâches Map et Reduce par nœud

Hadoop MapReduce dans le cloud

- **Choix des caractéristiques des machines virtuelles du cluster Hadoop**
 - Capacité disque importante
- **Impact du placement des machines virtuelles sur les performances et la disponibilité des données**
 - Exécution de plusieurs VMs sur une même machine physique
- **Lourdeur de la migration des machines virtuelles contenant les nœuds de données HDFS**
- **Elasticité du cluster Hadoop MapReduce**
 - Exemple : Amazon EMR, Sahara, Resilin

Hadoop : un écosystème en constante évolution



Illustrations & photos : crédits

p. 2 : © Julien Eichinger, Fotolia ; By Camelia.boban CC BY-SA 3.0, via Wikimedia Commons ; Domaine public, Pixabay.

p. 3 : d'après Lavanya Ramakrishnan, *Cloud Computing and Big Data : Overview*, LBNL, 2012 workshop Discovery Workshops covering High Performance Computing, Cloud Computing and Big Data topics.

p. 5 : "Hadoop Tutorial from Yahoo!" by Yahoo! Inc. – CC BY - <https://developer.yahoo.com/hadoop/tutorial/module4.html#basics>

p. 6 : <http://www.rabidgremlin.com/data20>

p. 7 : "Hadoop Tutorial from Yahoo!" by Yahoo! Inc. – CC BY - <https://developer.yahoo.com/hadoop/tutorial/module4.html#basics>

p. 8 : droits réservés