

# Big Data et MapReduce

## Démonstration avec Hadoop

Marion Dalle

- Explosion de la quantité de données
  - 1969 : 32KB pour marcher sur la lune
  - 2013 : 8GB de RAM dans mon ordinateur
- Enjeu économique sur leur traitement
- Traitement en base de données dépassé

# Qu'est ce que MapReduce ?

⇒ Patron d'architecture de développement informatique pour la calcul parallèle et distribué de données importante

## Grands utilisateurs :

- Google, qui l'a rendu célèbre
- Amazon
- Facebook

- Grep distribué
- Tri distribué
- Inversion de graphe des liens webs
- Classification automatique
- Traduction automatique

## Map

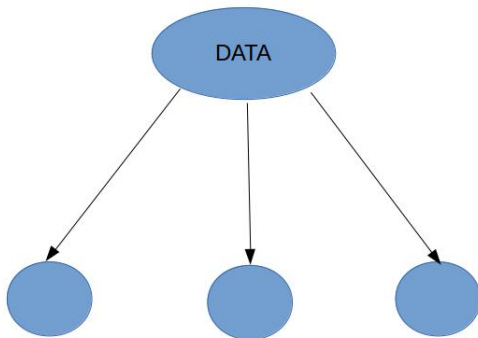
$\langle \text{clé}, \text{valeur} \rangle \rightarrow \text{liste} \langle \text{clé}, \text{valeur} \rangle$

## Reduce

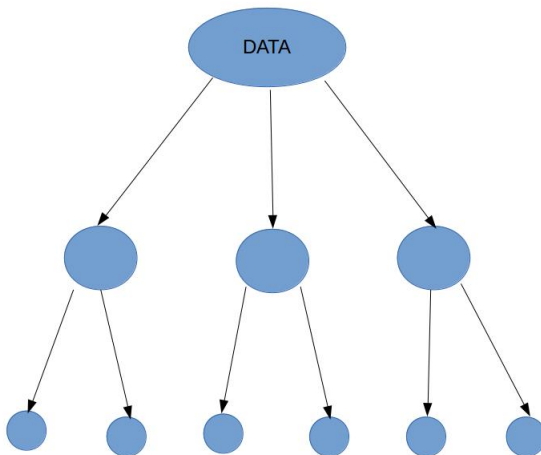
$\langle \text{clé}, \text{liste}(\text{valeur}) \rangle \rightarrow \text{liste}(\text{valeur})$



DATA

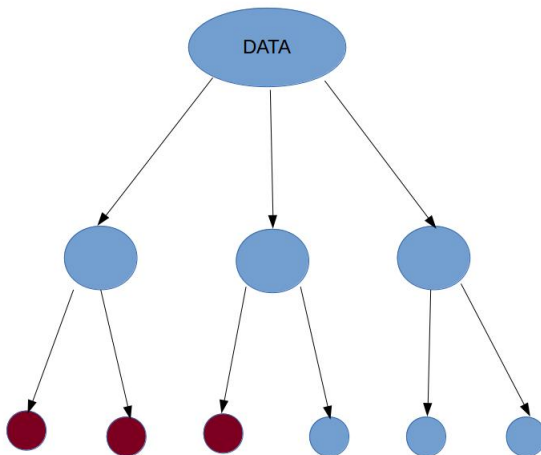


# Déroulement de l'algorithme

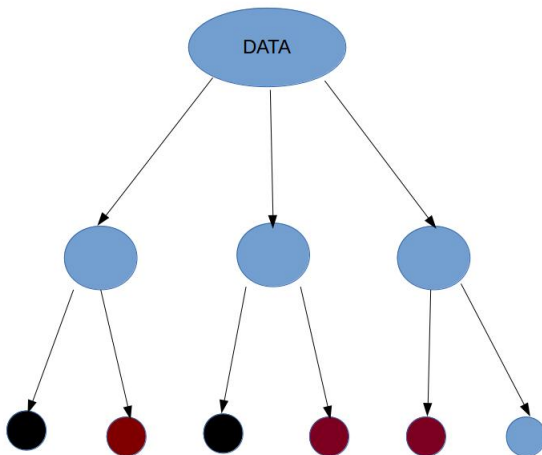




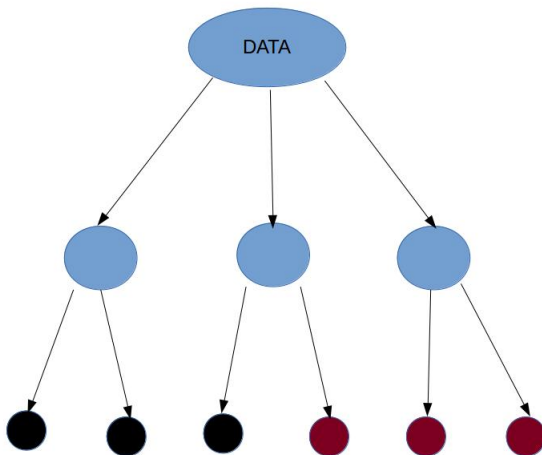
# Déroulement de l'algorithme



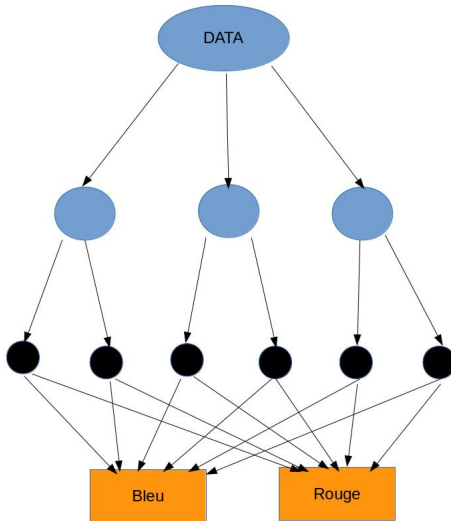
# Déroulement de l'algorithme

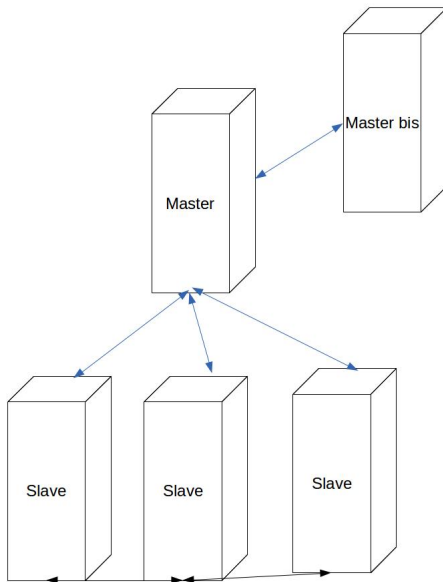


# Déroulement de l'algorithme



# Déroulement de l'algorithme





- Pas de SQL
- Mémoriser et parser les données pour les transformer en objet
- Un seul flot de données
- Méthode bloquante
- Aucune optimisation du transfert
- Performances

# Questions ?

