



# LA MINE D'ALIBABA

MÉTHODES ET PROGRAMMATION POUR LES  
**DONNÉES MASSIVES**



**MATHIEU STEINBACH** HUGO - **ZIMOL** GUILLAUME

**2023 - 2024**

# SOMMAIRE

01

ÉTAPE 1

02

ÉTAPE 2

03

ÉTAPE 3

04

ÉTAPE 4

05

ÉTAPE 5

# SOMMAIRE

01  
**ÉTAPE 1** <

02  
**ÉTAPE 2**

03  
**ÉTAPE 3**

04  
**ÉTAPE 4**

05  
**ÉTAPE 5**

# ÉTAPE 1 – TÂCHES



## MAPPER :

- 1. Obtient la date de début et de fin.
- 2. Émet le nom de la tâche avec sa durée.

## REDUCER :

- 1. Pour chaque tâche, le reducer additionne la durée de toutes ses instances et compte leur nombre total.
- 2. Le reducer calcule la durée moyenne de toutes les instances d'une tâche.
- 3. Identification des Stragglers.
- 4. Tri et écriture des résultats.

## EXEMPLE DE SORTIE

TÂCHES	DURÉE	NOMBRES D'INSTANCES	STRAGGLERS
J10_1_2_9	11650	556	101
J11_4_10	101	9	0
J11_9_10	111	2	0
J12_11_16	1	1	0
J13_3_12	161	1	0
J14_13_22	350	25	0
J16_12_14_15	2	2	0

# ÉTAPE 1 – JOBS



## MAPPER :

- 1. Obtient le job et la tâche.
- 2. Émet le même couple.

## EXEMPLE DE SORTIE

## REDUCER :

- 1. Compte le nombre d'instances et identifie la plus grande tâche pour chaque job.
- 2. Calcule le nombre total de tâches uniques.
- 3. Trie et écrit les résultats (nom du job, nombre de tâches, nombre d'instances).

TÂCHES	DURÉE	NOMBRES D'INSTANCES
j_1008775	21	23531
j_1155151	15	29

# SOMMAIRE

01

ÉTAPE 1

02

ÉTAPE 2 <

03

ÉTAPE 3

04

ÉTAPE 4

05

ÉTAPE 5

# ÉTAPE 2 – Pic de consommation



## MAPPER :

1. Obtient le début et la fin ainsi que l'utilisation max de cœur.
2. Émet la consommation des cœurs pour chaque seconde de la tâche

## REDUCER :

1. Calcule la consommation totale des cœurs pour chaque seconde.
2. Identifie et enregistre les pics de consommation des cœurs.
3. Trie ces pics par consommation moyenne décroissante et les écrit dans le fichier de sortie.

## EXEMPLE DE SORTIE

Horodatage	DURÉE	MOYENNE
Peak from 920 to 932	duration: 12	1026589
Peak from 921 to 932	duration: 11	1028442

# SOMMAIRE

01

ÉTAPE 1

02

ÉTAPE 2

03

ÉTAPE 3 <

04

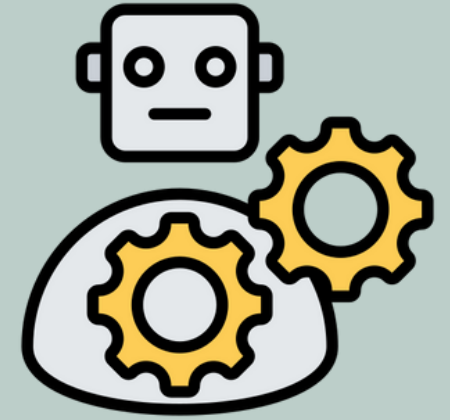
ÉTAPE 4

05

ÉTAPE 5



# ÉTAPE 3 – Puissance machine



## MAPPER :

1. Obtient la consommation max et la machineId associé.
2. Émet la consommation max pour chaque machine.

## REDUCER :

1. Calcule et émet la consommation maximale des cœurs pour chaque machineID fourni.

## EXEMPLE DE SORTIE

Machine ID	DURÉE
m_1	586
m_10	250

# SOMMAIRE

01

ÉTAPE 1

02

ÉTAPE 2

03

ÉTAPE 3

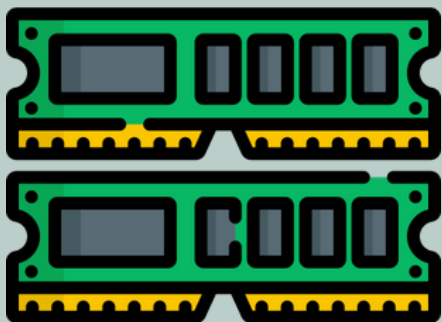
04

ÉTAPE 4 

05

ÉTAPE 5

# ÉTAPE 4 - Capacités mémoire



## Class annexe :

- MachineMemoryWritable

## EXEMPLE DE SORTIE

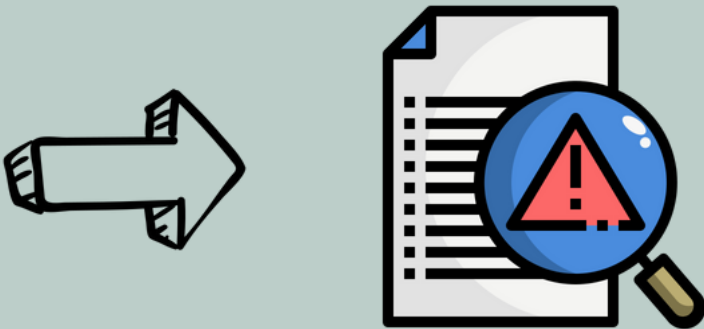
### MAPPER :

1. Obtient la taskID, machineID et l'utilisation maximale de la mémoire.
2. Émet l'utilisation max pour chaque machine via la class MachineMemoryWritable.

Machine ID	DURÉE	RATIO
m_3069	m_2861	0.9767441
m_3069	m_1433	0.7241379
m_3069	m_334	0.76363635
m_3069	m_2828	0.9767441

### REDUCER :

1. Calcule et émet les ratios d'utilisation maximale de la mémoire entre toutes les paires uniques de machines pour une tâche donnée, en excluant les ratios égaux à 1.



# SOMMAIRE

01

ÉTAPE 1

02

ÉTAPE 2

03

ÉTAPE 3

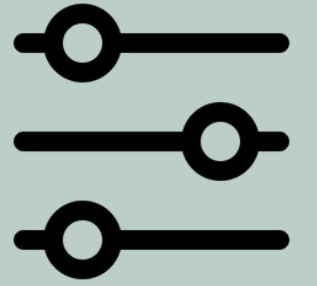
04

ÉTAPE 4

05

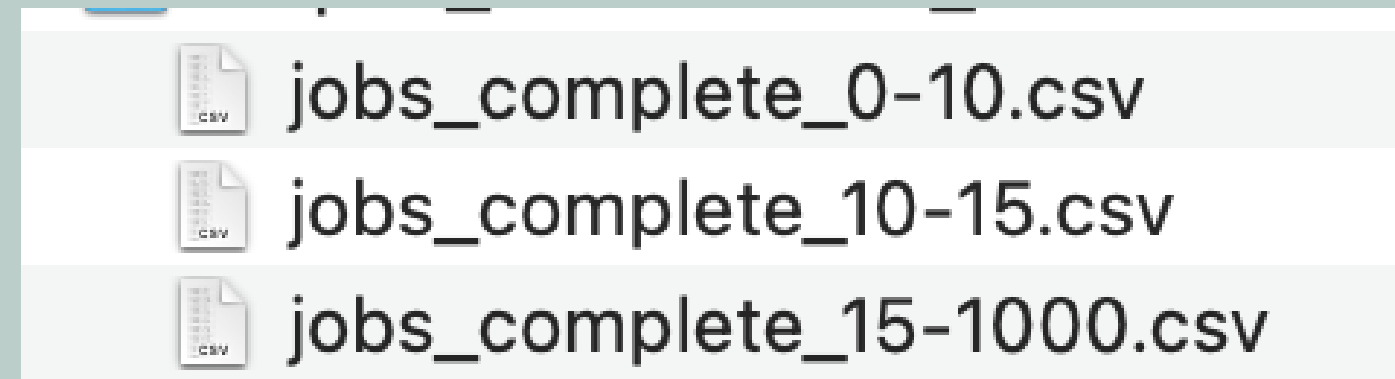
ÉTAPE 5 <

# ÉTAPE 5 - Options



## EXEMPLE DE SORTIE

- **Modification des mappers et reducers**
- **MultipleOutput**
- **Commandes personnalisées**



Pour obtenir cette sortie nous avons exécuté cette commande :

```
hadoop jar MyHadoopApps.jar JobAnalysisDriver -D mapreduce.job.reduces=2 /corpus/selectionCourt.csv /etape5/etape5_2_court_separate -separateFiles 1 3 0-10,10-15,15-1000
```



**MERCI DE VOTRE  
ATTENTION !**

**DES QUESTIONS ?**

MATHIEU STEINBACH HUGO - ZIMOL GUILLAUME

