***Compte rendu : BD NoSQL & Big Data***

***TP3 : Hadoop MapReduce***

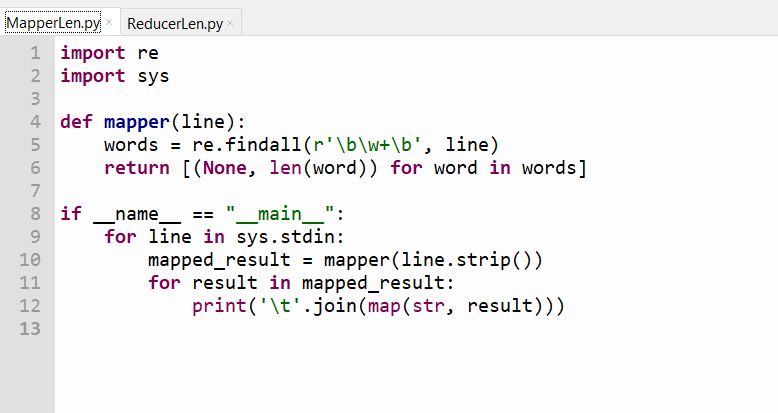
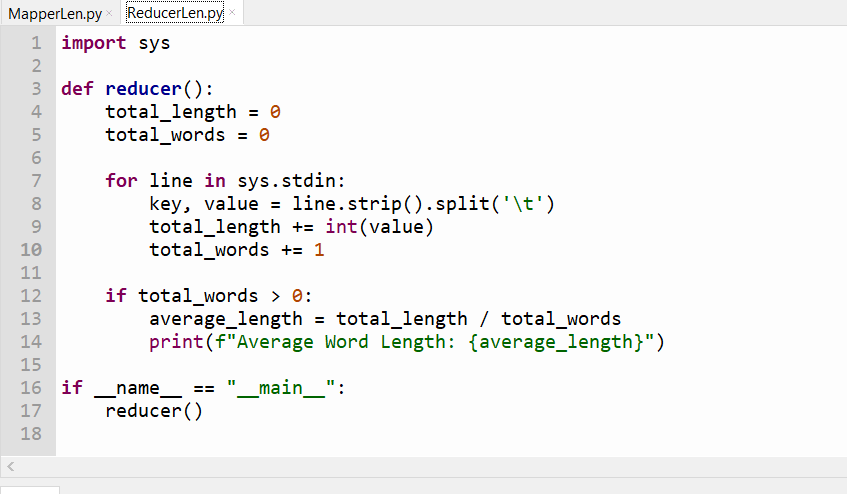
Réalisé par : Riahi Farah

Classe : ING2 INFO

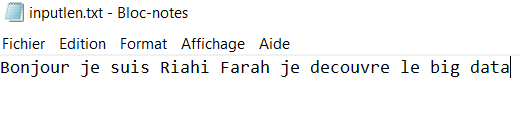
Exercice 1 :

Objectif du programme MapReduce : calculer la longueur moyenne des mots .

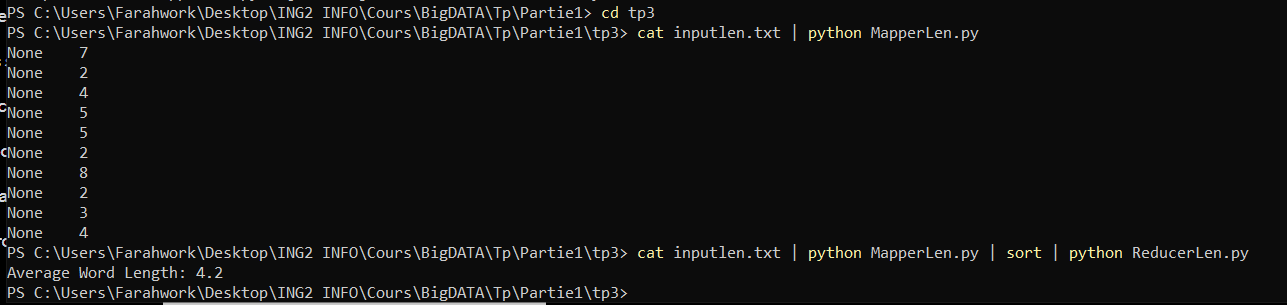
* Aperçu du code :

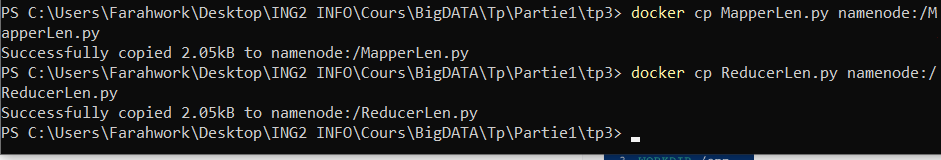
* Contenu inputlen.txt :



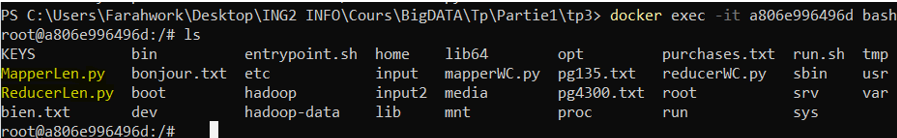
* Exécution du code en local :



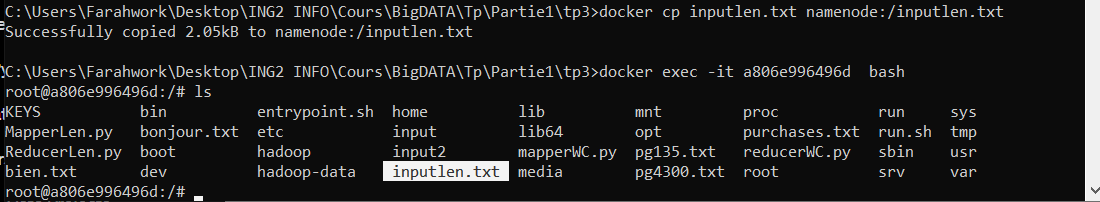
* Le code mapper traite chaque ligne d'entrée et génère des paires clé-valeur : None, longueur du mot lu en input.
* Le code de reducer prend en entrée la sortie du code mapper (paires clé-valeur) et fait retourner l moyenne de tous les valeurs.
* Exécution sur le cluster :



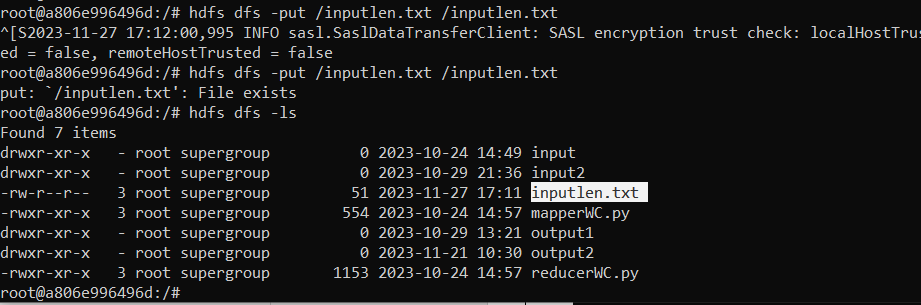
* On copie nos codes dans NameNode.

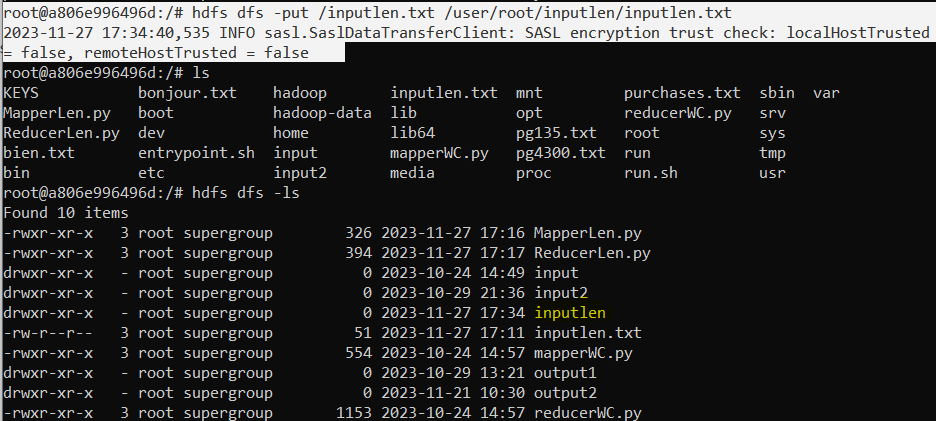


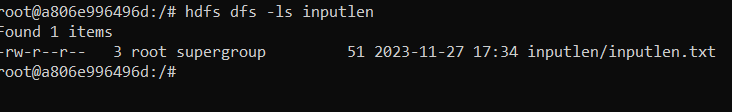
* Codes copiés.

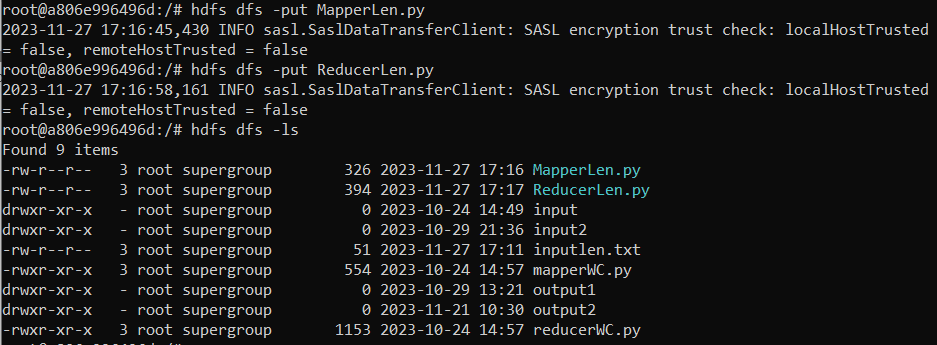


* De même pour intullen.txt pour faciliter la tâche.



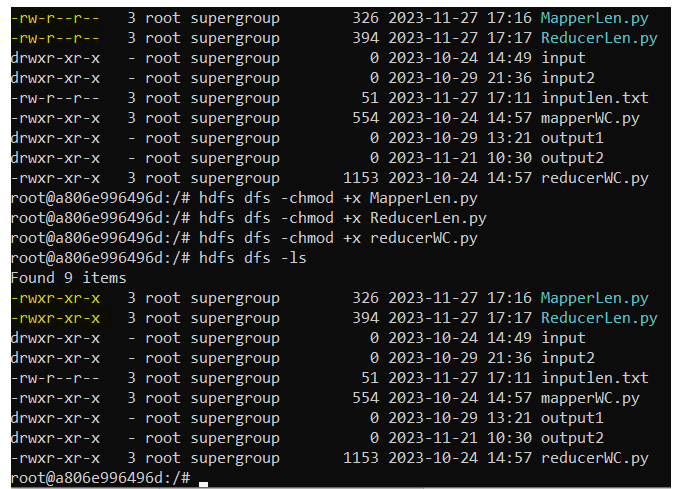








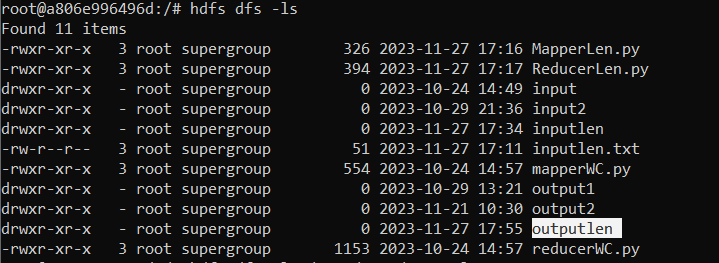
* On change les droits : ajoutant le droit d’exécution des codes mapper et reducer.

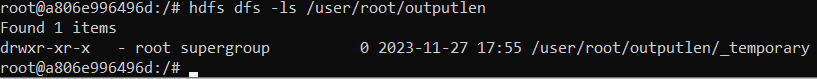


* **La commande suivante est une commande Hadoop Streaming, elle permet d’exécuter des programmes MapReduce :**

**hadoop jar /opt/hadoop-3.2.1/share/hadoop/tools/lib/hadoop-streaming-3.2.1.jar \  
-files MapperLen.py,ReducerLen.py \  
-input inputlen \  
-output outputlen \  
-mapper MapperLen.py \  
-reducer ReducerLen.py**

* Après l’exécution outputlen sera généré : il contient le résultat de l’exécution.

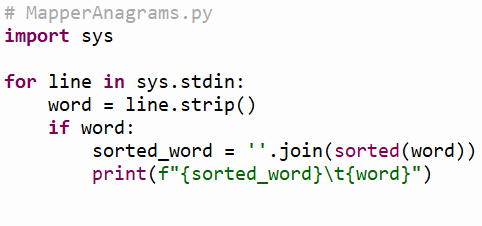




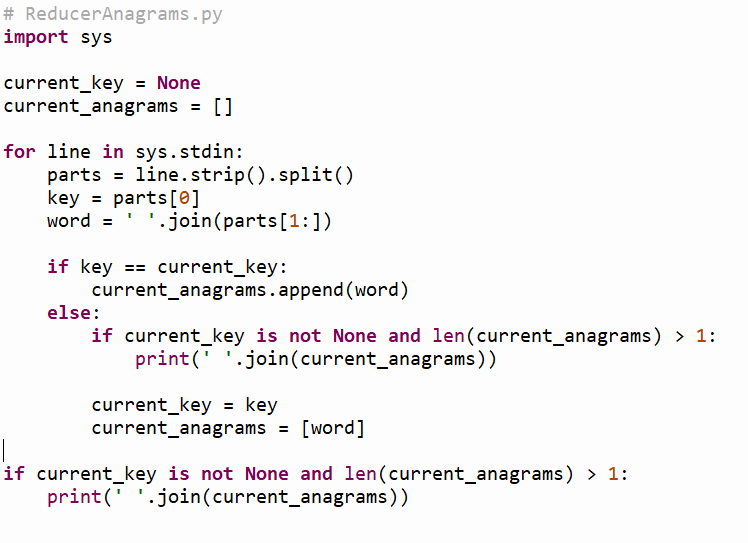
Exercice 2 :

Objectif du programme MapReduce : rechercher des anagrammes dans un fichier texte.

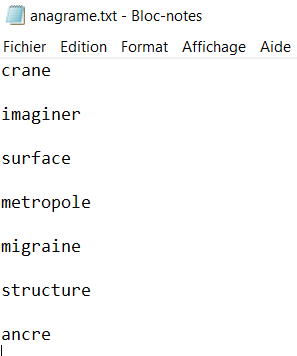
* Aperçu du code :
* Mapper :



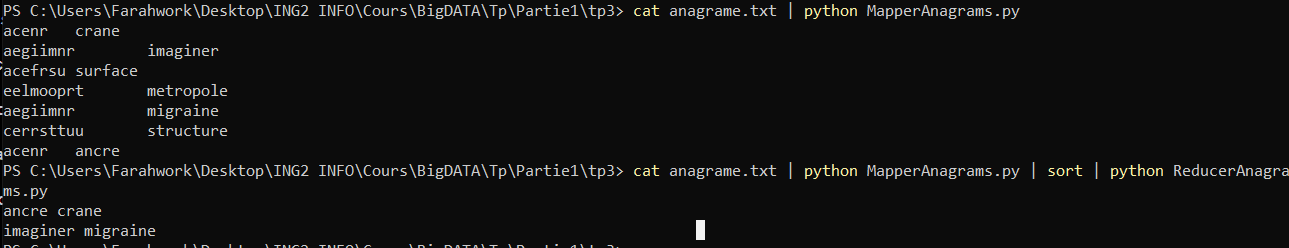
* Reducer :



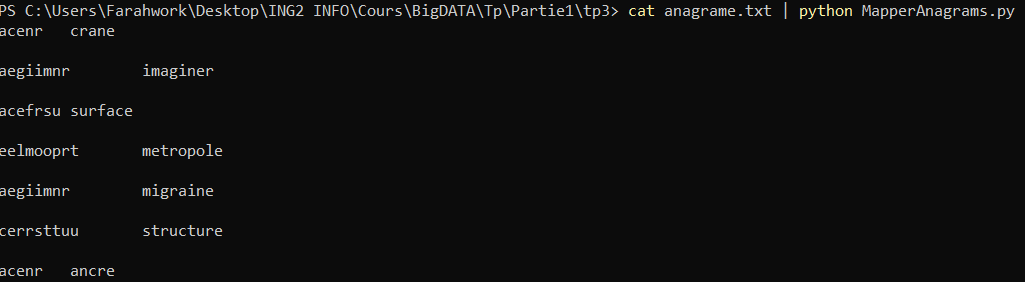
* Contenu anagrame.txt :

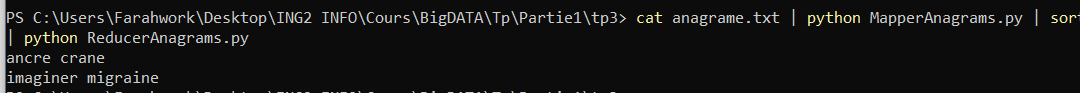


* Exécution du code en local :



* Ce script mapper prend une liste de mots en entrée, trie les lettres de chaque mot, et émet des paires clé-valeur où la clé est la représentation triée du mot et la valeur est le mot d'origine.



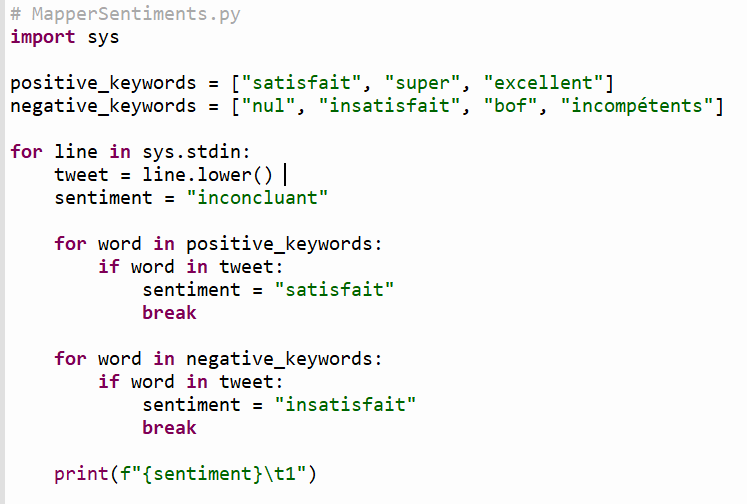


* Les données d'entrée du reducer sont des paires clé-valeur, où la clé est la représentation triée alphabétiquement d'un mot et la valeur est le mot d'origine.
* Le script regroupe les mots ayant la même représentation triée (anagrammes) et les imprime ensemble s'il y a au moins deux mots dans le groupe.

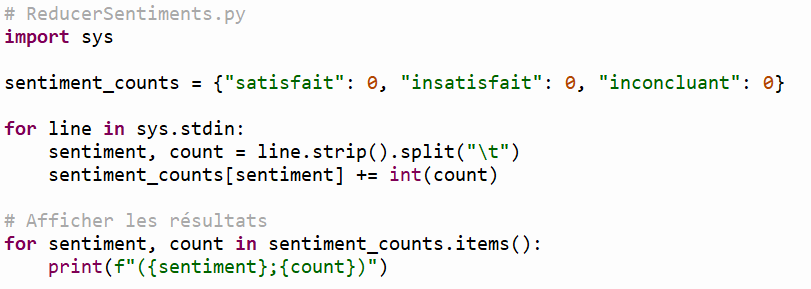
Exercice 3 :

Objectif du programme MapReduce : Analyse des sentiments des clients sur Twitter.

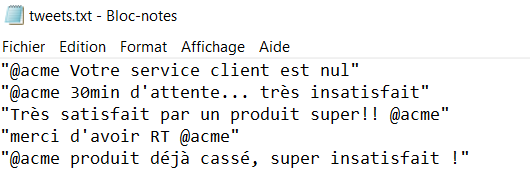
* Aperçu du code :
* Mapper :



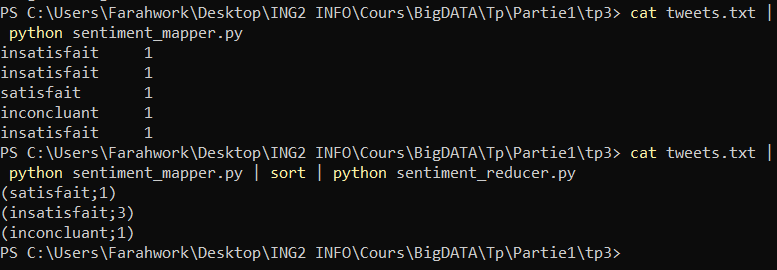
* Reducer :



* Contenu de fichier :



* Exécution du code en local :

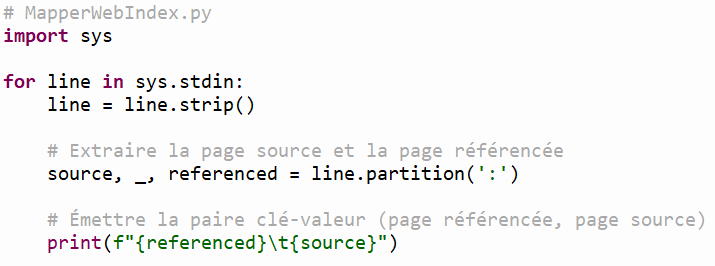


* Le script mapper va analyser des tweets et déterminer le sentiment associé à chacun. Il utilise des mots-clés positifs et négatifs pour attribuer un sentiment à chaque tweet. Il émet des paires clé-valeur où la clé est le sentiment du tweet et la valeur est 1, indiquant qu'un tweet avec ce sentiment a été trouvé.
* Le script reducer agrège les résultats du mappage en comptant le nombre de tweets pour chaque sentiment.

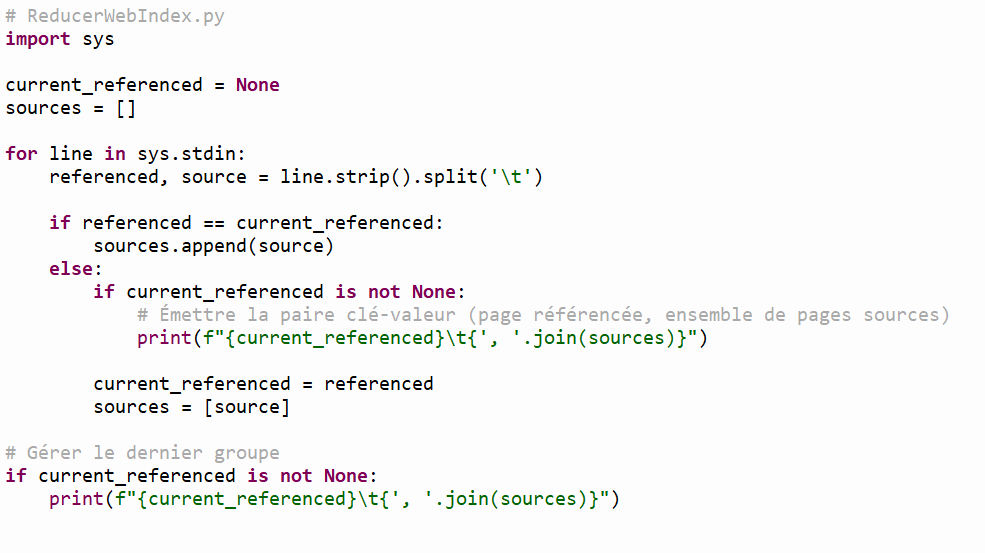
Exercice 4 :

Objectif du programme MapReduce : Index web inversé.

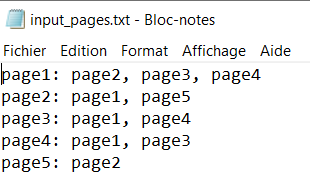
* Aperçu du code :
* Mapper :



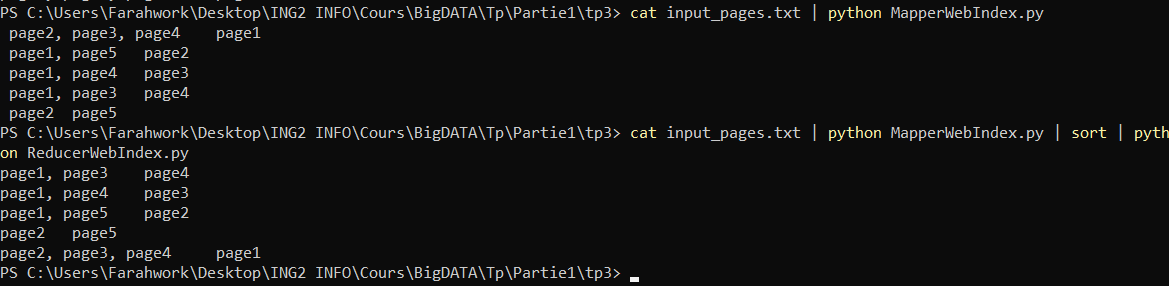
* Reducer :



* Contenu fichier input\_pages.txt :



* Exécution du code en local :



* Le script mapper prend des lignes en entrée, où chaque ligne représente une relation de lien entre deux pages web. Le script extrait la page source et la page référencée à partir de chaque ligne et émet des paires clé-valeur où la clé est la page référencée et la valeur est la page source.
* Le script reducer agrège les résultats du mappage en regroupant les pages sources associées à chaque page référencée.