



Topic ANALYSIS

Le mapper

+

Le reducer

Problème de taille

<u>Taille</u>	<u>Problème</u>	<u>Solution</u>
• 100M mots	• Pas de problèmes	• Naïve avec 1 seul ordinateur
• 1000M mots	• Mémoire insuffisante	• Utiliser le disque, Fenêtre glissante
• 100MM mots	• Processeur insuffisant	• Multithreading, éliminer $< N$
• 1000MM mots	• 1 ordinateur insuffisant	• Distribuer le calcul
• Plus encore!	• Réseau insuffisant, contrôleur surchargé	• MapReduce (ou solution du même ordre comme MPI, etc.)

MAPER

- **Les biblio**
- import sys
- import re
- import string
- from itertools import chain
- from collections import defaultdict
- **Les fonctions:**
- **#donner une liste qui rassemble au couple (cle valeur)**
- develop_dict(par1)
- **#inverser un dict**
- invert_dict(par2)
- **#enlever les apotrophe**
- simplify(par3)
- **Apres**
- **# build base dict from file and invert it**

REDUCER

#on prend le max de nombre de fois que le topic
est apparait dans le mapper

2) Préparer les scripts
mapper/reducer pythons

Choisir les mapper/reducer

Rendre les scripts exécutables

**chmod u+x mapper.py
chmod u+x reducer.py**

Simuler
l'exécution à
l'aide de
l'émulateur
(Hadoop
emulater from
Scratch)

Les biblio:

import argparse

Import subprocess

import glob

Ou

Visualiser le
résultats sur
HDFS with
Hadopp

```
hadoop fs -ls output-dir
```

```
hadoop fs -cat output-dir/output.txt
```

A high-angle, dark photograph of three people sitting on a wooden deck. The text "Thank you!" is overlaid in white. The people are sitting on a dark, circular mat or blanket. One person is wearing a light blue shirt and dark pants, another is wearing a patterned shirt and dark pants, and the third is wearing a striped shirt and dark pants. There are some items on the ground, including a green flip-flop and a light blue bag. The background is a wooden deck with horizontal planks.

Thank you!