

**Task 1: Communication models and Middleware**  
**Playing out with MapReduce**

Autors: Antoni Dalmau Diéguez

Oriol Balagué Chorto

Data d'entrega: 16/3/2018

## **Índex:**

1. Decisions de Disseny .....	3
2. Joc de proves .....	5

## 1. Decisions de Disseny

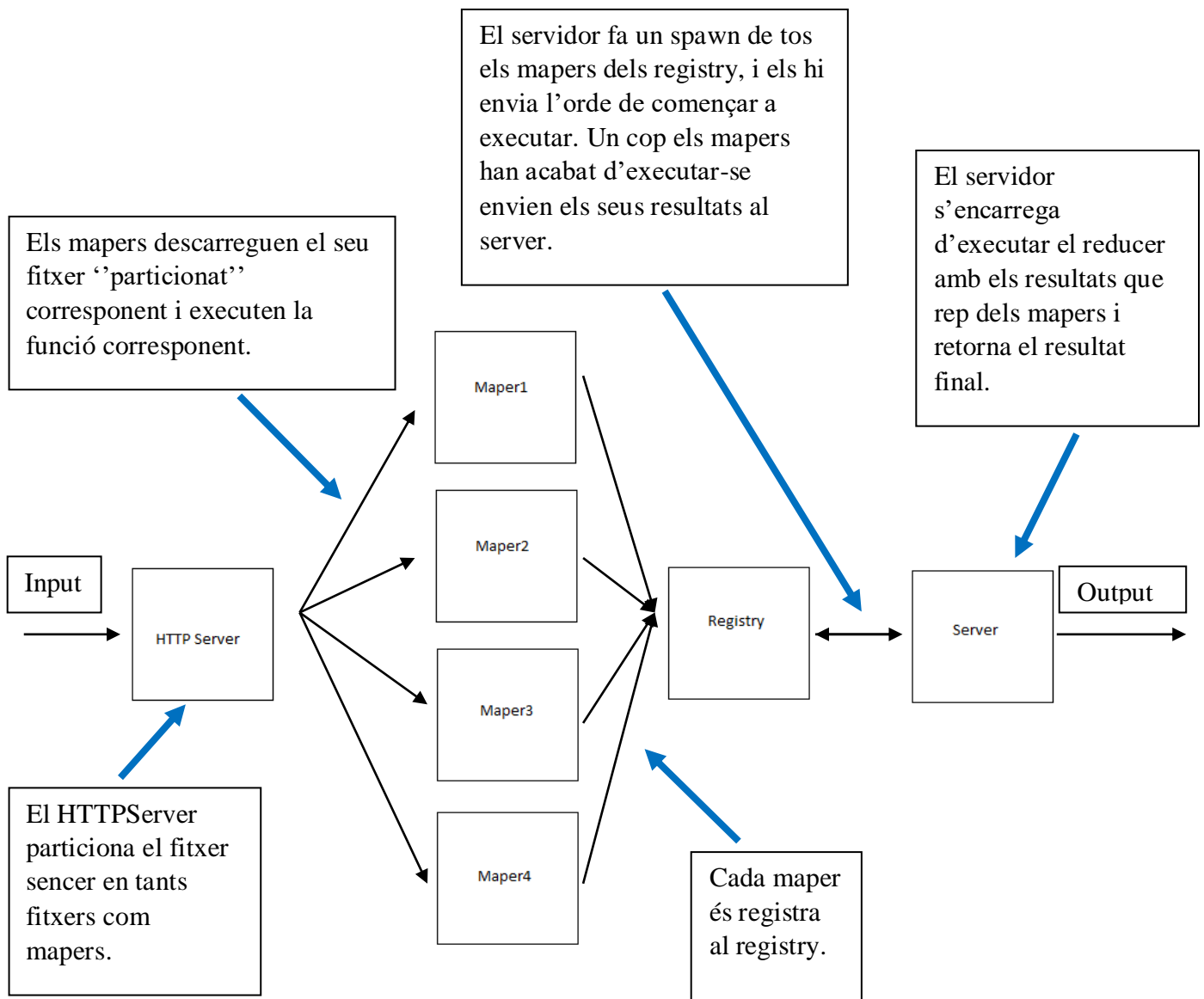
En aquest projecte hem implementat un MapReduce en el qual a partir d'un fitxer amb un tamany considerable executem dues funcions la *CountWords* i la *WordCount*.

La funció *CountWords* s'encarrega de contar el nombre de vegades que apareix una paraula en un string que rep per paràmetre i retorna un diccionari amb clau "paraula" i valor "# vegades que apareix".

La funció *WordCount* s'encarrega de contar el nombre de paraules que conté un string que rep per paràmetre i retorna aquest valor.

En l'esquema que mostrarem a continuació hem representat com funciona el nostre programa seguint els següents passos:

1. HTTPServer particiona el fitxer i el penja en un simple HTTP server local.
2. Registrem els mapers al registry
3. El servidor fa un spawn de tots els mapers del registry, i els hi envia l'orde de començar a executar.
4. Els mapers descarreguen el seu fitxer corresponent del simple HTTP server que hem creat prèviament.
5. Un cop descarregat executen la funció corresponents (*CountWords* o *WordCount*) i envien el resultat al servidor.
6. El servidor executa l'actor reducer i mostra per pantalla el resultat final.

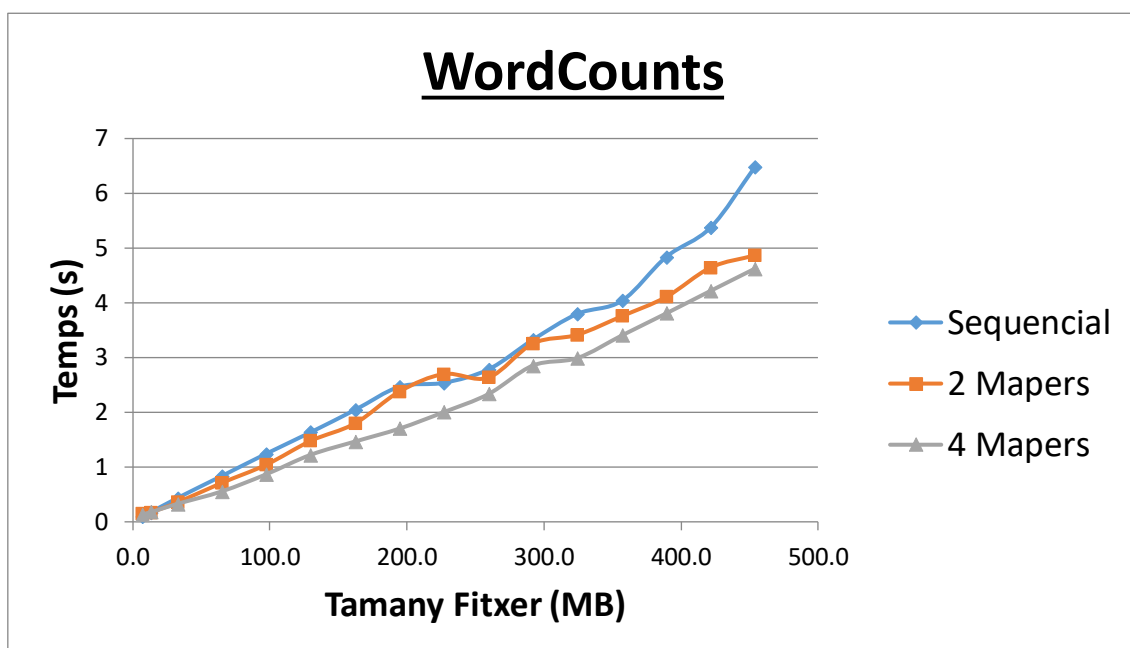
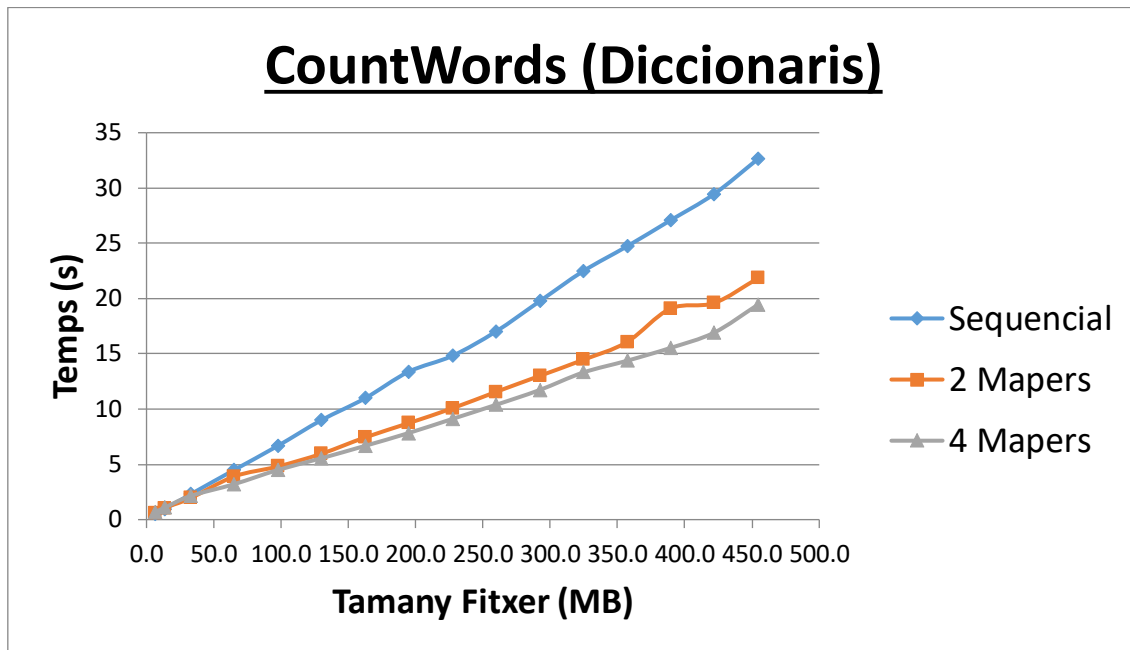


## 2. Joc de proves

Per realitzar el joc de proves, hem utilitzat una maquina virtual amb SO Linux amb les següents especificacions:

- 4 CPUs
- 7GB memoria RAM

Per poder comparar els speedup's hem realitzat dues implementacions per cadascuna de les funcions: una en seqüencial i l'altra en distribuïda, on variant el nombre de mapers i el tamany del fitxer hem calculat la mitjana de temps d'execució. Els resultats obtinguts son els següents:



Les dades amb les quals hem realitzat les gràfiques anteriors, es troben en el fitxer: "JocProvesSD.xlsx".