

Pràctica 1

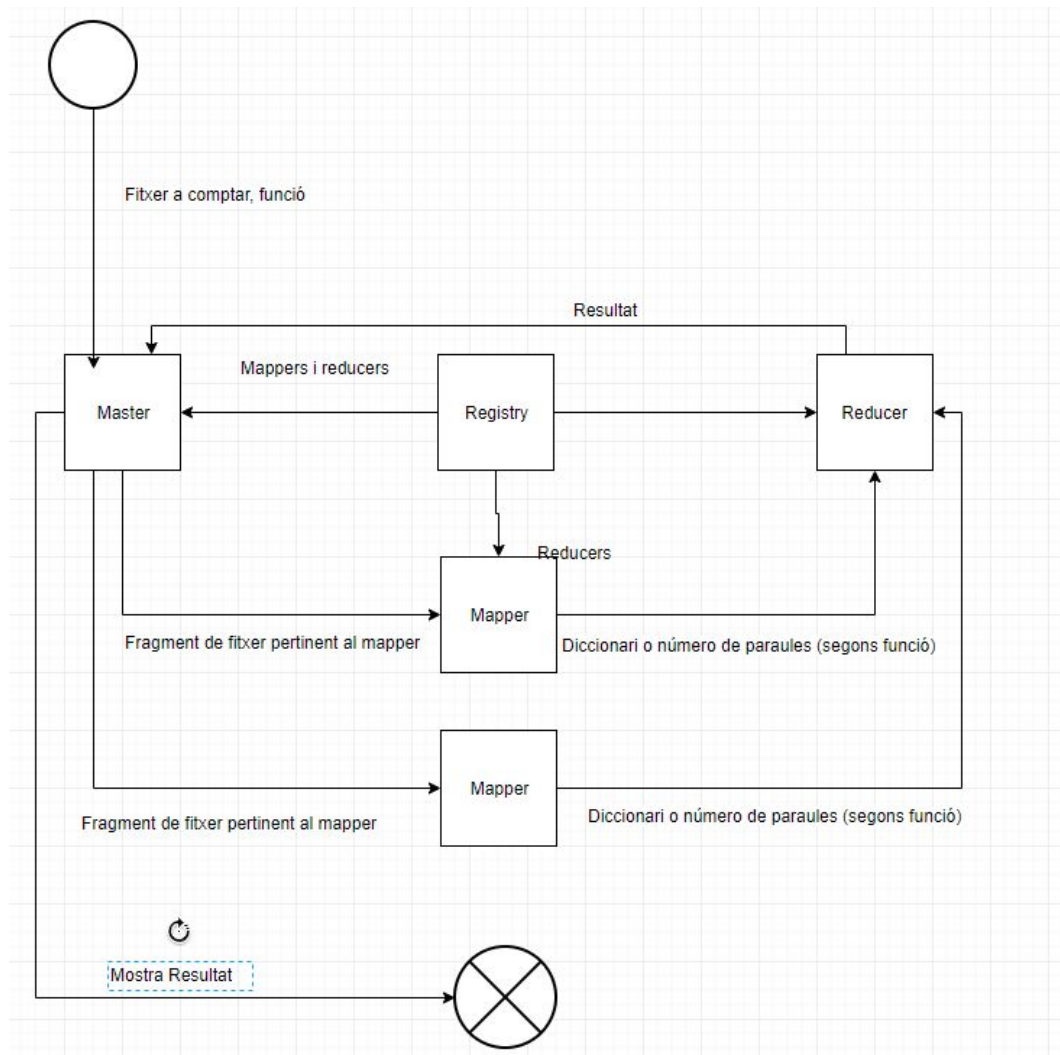
Foundations of distributed systems

Enric Civit
Álvaro Vega

Introducció

Aquesta pràctica consisteix en la implementació de l'algorisme map and reduce utilitzant diverses màquines interconnectades. El sistema consisteix en un master amb el qual es poden realitzar dues funcions, WordCount i CountingWords. WordCounter consistirà en el obtenir el nombre de vegades repetida cada paraula d'un arxiu .txt donat i CountingWords retornarà el nombre de paraules totals d'aquest. La infraestructura inclou un registre que farà d'organitzador, donant a conèixer tots els actors del sistema, tant mappers com reducers i masters. Finalment també existiràn els mappers disponibles per realitzar les tasques de les funcions i el reducer (un en el nostre cas) que realitzarà la tasca d'agrupació i mostreig de resultat.

Diagrama de l'arquitectura



A nivell tècnic el sistema funciona de la següent forma:

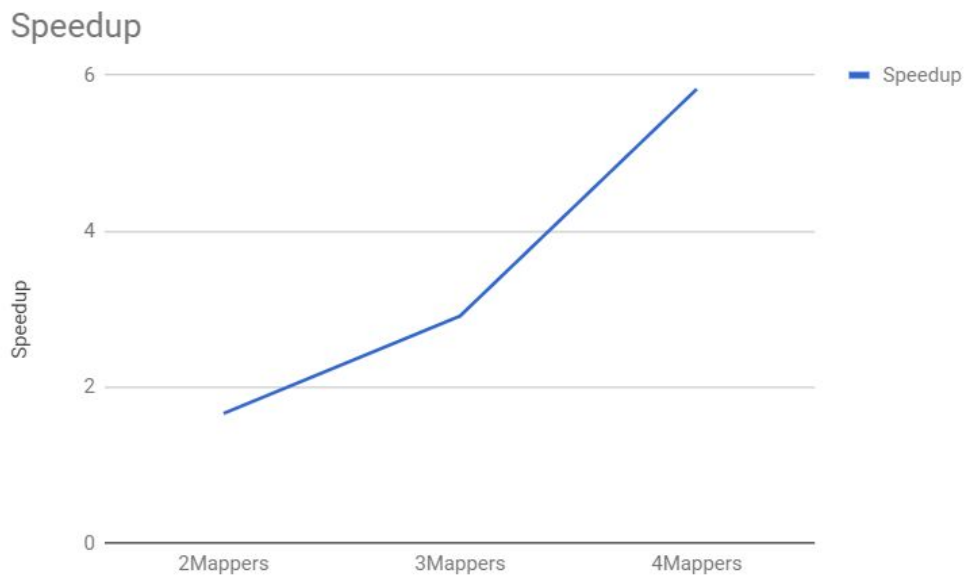
- En primer lloc es donen d'alta en el registre tots els actors participants, els mappers, el reducer i el mapper, en aquest ordre.
- Després el master separa l'arxiu en tants mappers com es vulgui.
- A continuació el master li envia l'arxiu dividit al mapper que realitzarà la seva tasca de càlcul en relació a la funció que se li demani.
- El mapper li envia al reducer la seva resposta.
- El reducer espera a tenir tots els resultats esperats dels mappers i executa la funció reduce per tal d'obtenir el resultat de la funció.
- Finalment el master, recull el resultat i el mostreja.

Gràfiques SpeedUp

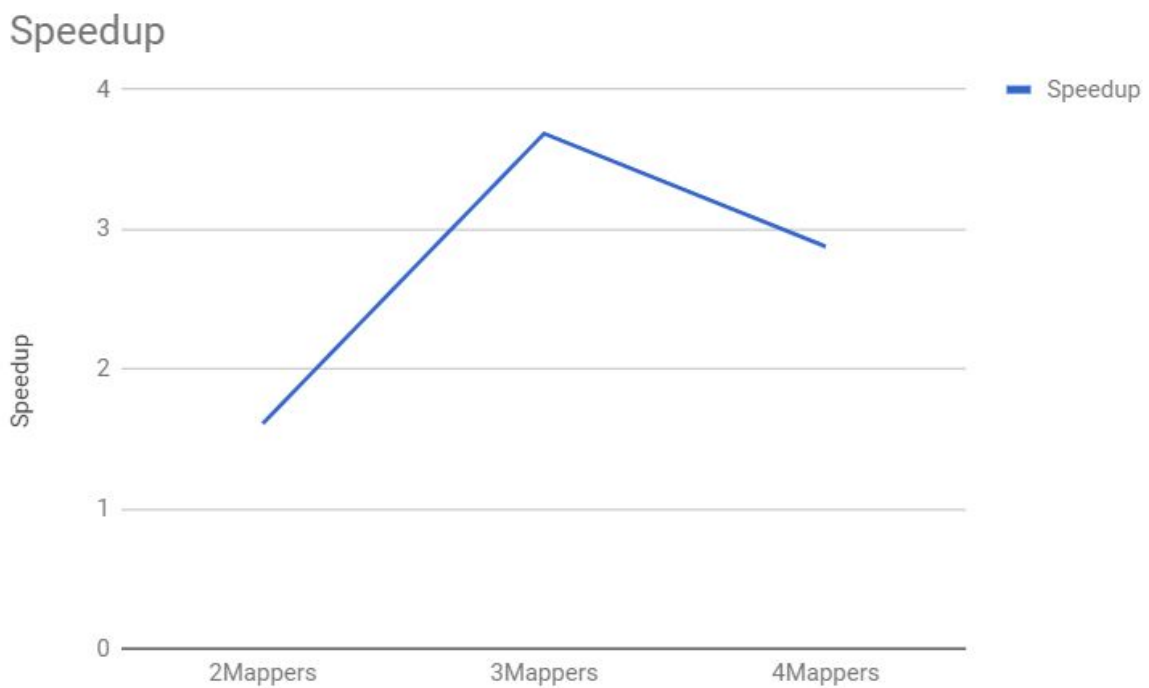
WordCount de fitxer de 13 MB



CountingWords de fitxer de 13 MB



WordCount de fitxer de 6 MB



CountingWords de fitxer de 6 MB

