

Pràctica 2

Foundations of distributed systems

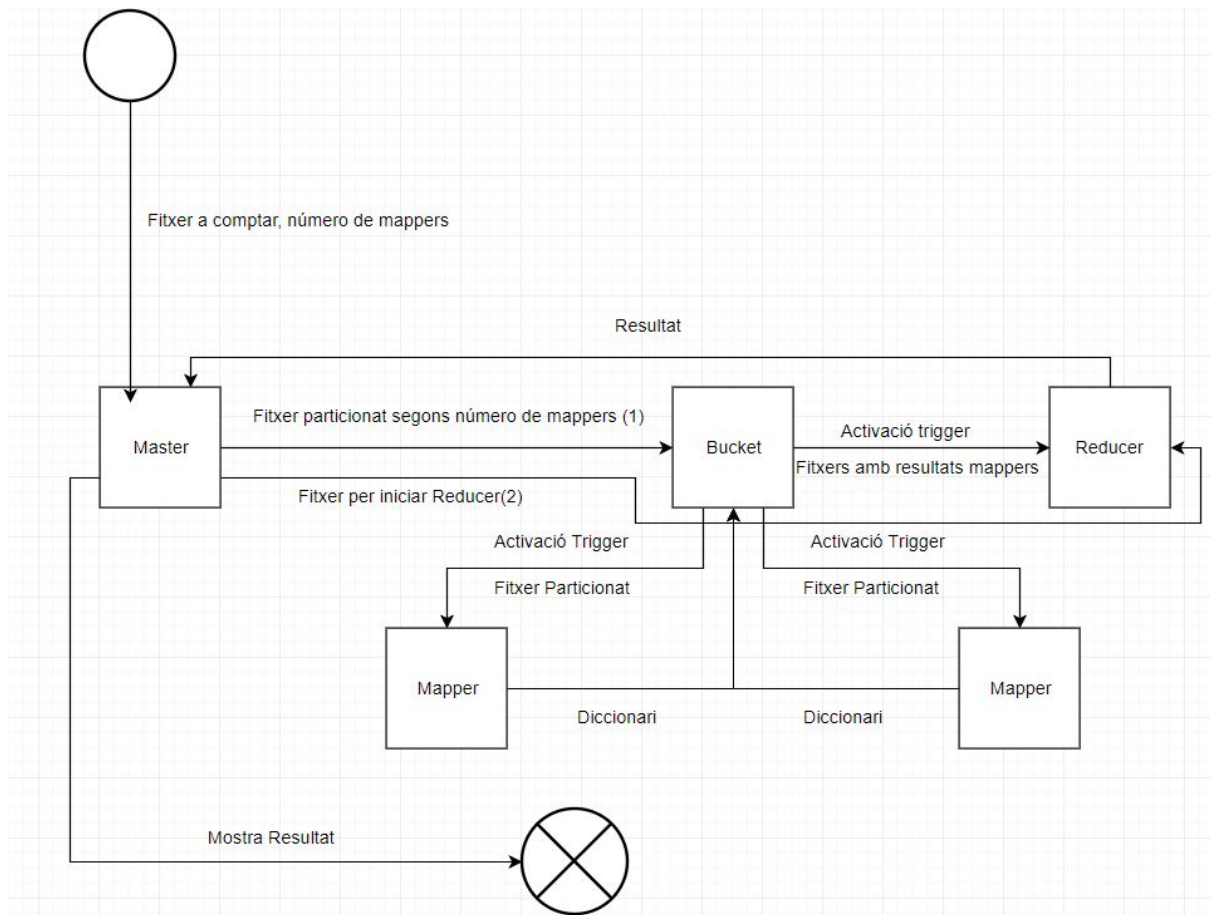
Enric Civit
Álvaro Vega

Introducció

Aquesta pràctica consisteix en la implementació mitjançant cloud services del algorisme Map and Reduce. Utilitzarem la versió gratuïta d'Amazon dels serveis S3 i Lambda per realitzar les tasques d'emmagatzematge i virtualització. El sistema interpretarà un fitxer que guardarem en la carpeta contenedora del codi i aquest mitjançant el nostre cloud se li aplicarà l'algorisme map and reduce per tal d'obtenir un resultat.

En el nostre cas hem fet un WordCount per tal de comptar el nombre de paraules d'un fitxer .txt distribuint la càrrega a través del nombre de mappers que l'introduïm per paràmetre.

Diagrama de l'arquitectura



A nivell més tècnic, el sistema funciona de la següent forma:

- En primer lloc s'executa el master tot introduint el nombre de mappers l'arxiu a comptar i la carpeta contenidora
- Després aquest arxiu es dividirà en tantes parts com mappers existeixin i es penjaran a s3.
- Amb un trigger del mapper a Lambda executarem el codi del mapper amb el que per cada fitxer s'executarà i es calcula el nombre de paraules d'aquest. Quan acaba aquest càlcul es crea un fitxer amb el resultat en s3 per cada mapper existent.
- En paral·lel el master consulta si ja existeix un fitxer de resposta per cada mapper, llavors creem un fitxer a s3 per comunicar a lambda que ja està la resposta llesta, i amb un trigger s'executa el codi del reducer.
- El reducer busca el complet de fitxers de resposta dels mappers i en realitza el càlcul i ho resol en un fitxer amb la resposta.
- Finalment el master sap que ja existeix un fitxer final de resposta del reducer i acaba el procés.

Gràfiques SpeedUp

WordCount de fitxer de 13 MB



La reducció del speedup a partir de 4 mappers es degut a la intercomunicació entre el master i el cloud, no vol dir que el procés sigui més lent sinó que nosaltres estem tenint en compte aquest temps de interconnexió.