

Pràctica 1 SD

Marc Ferré

Miquel Roig

Per a aquesta pràctica primer hem tingut que crear un scrip anomenat registry amb el qual obtenim les instàncies dels host que anirem creant al llarg de tota l'execució de l'exercici.

Després hem creat un scrip anomenat client amb el qual ficant com arguments el identificador de l'ordinador pel qual ha executat el registry, el port pel qual es comunicarà i el nom que tindrà quan creem un host.

Un cop ja hagem creat tots els host correctament i la quantitat desitjada, s'ha d'ejectar el clientb.

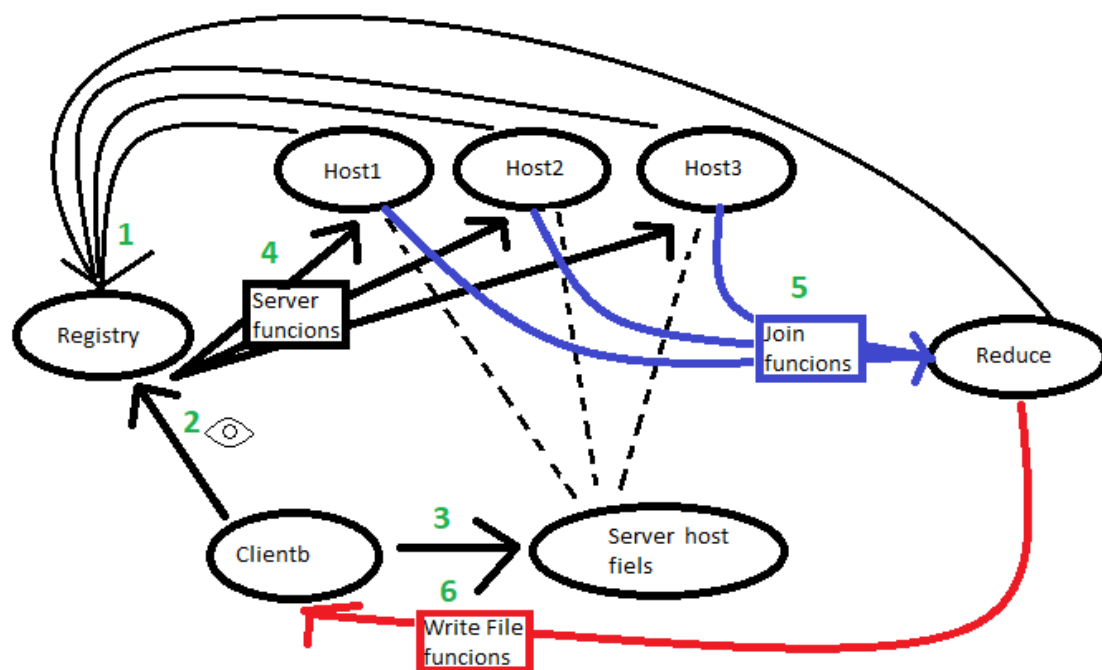
Per a aquest scrip necessitem passar per paràmetre el identificador de l'ordinador pel qual ha executat el registry, el port pel qual es comunicarà, el nom de l'arxiu que es vol analitzar i en quina modalitat es vol executar (0 -> CountWords / 1-> WordsCount).

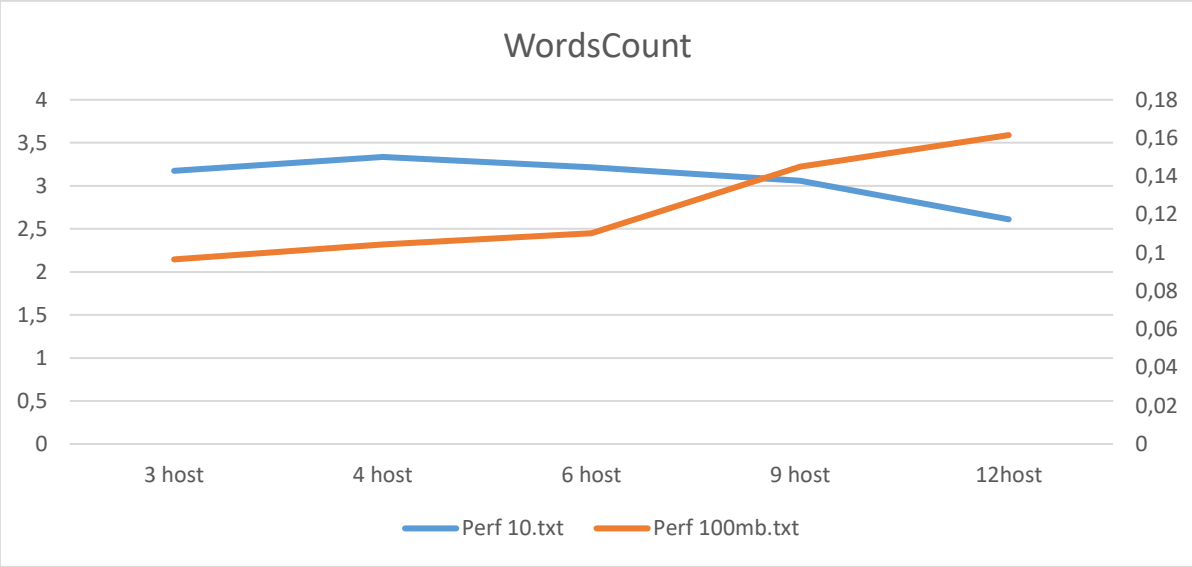
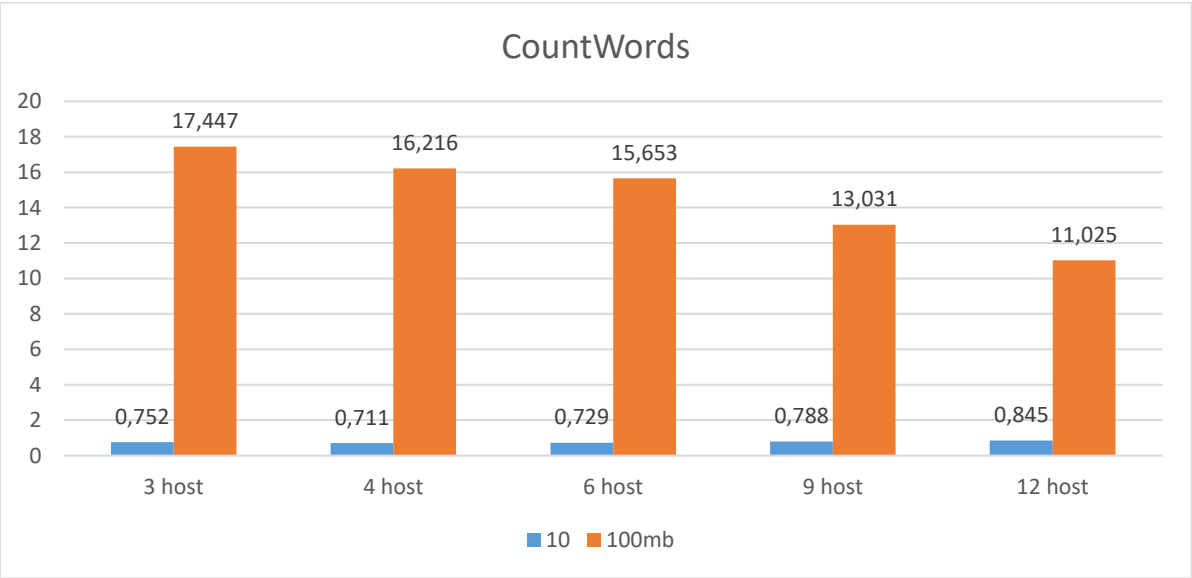
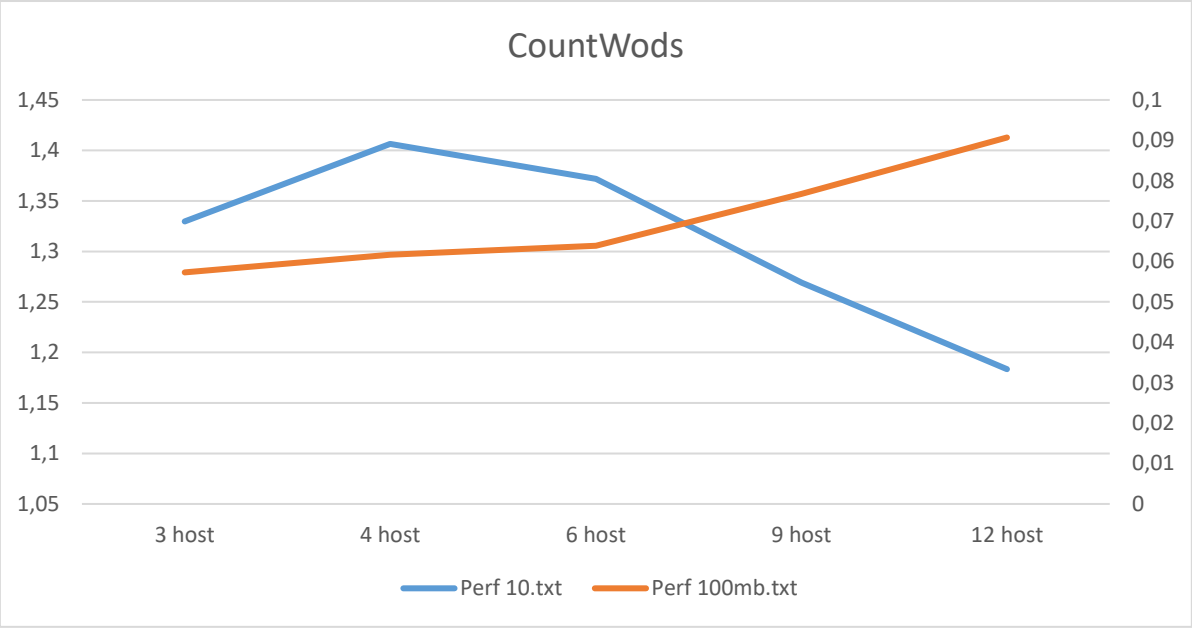
Un cop s'aguin passat els arguments correctament, el primer que fa l'scrip és obtindrà la pròpia IP i crear-se. Després cridem a la funció sub_arxius per tal de distribuir el fitxer en altres de més petits i ens guardem el llistat dels noms dels nous arxius que s'han creat en la carpeta "Arxius".

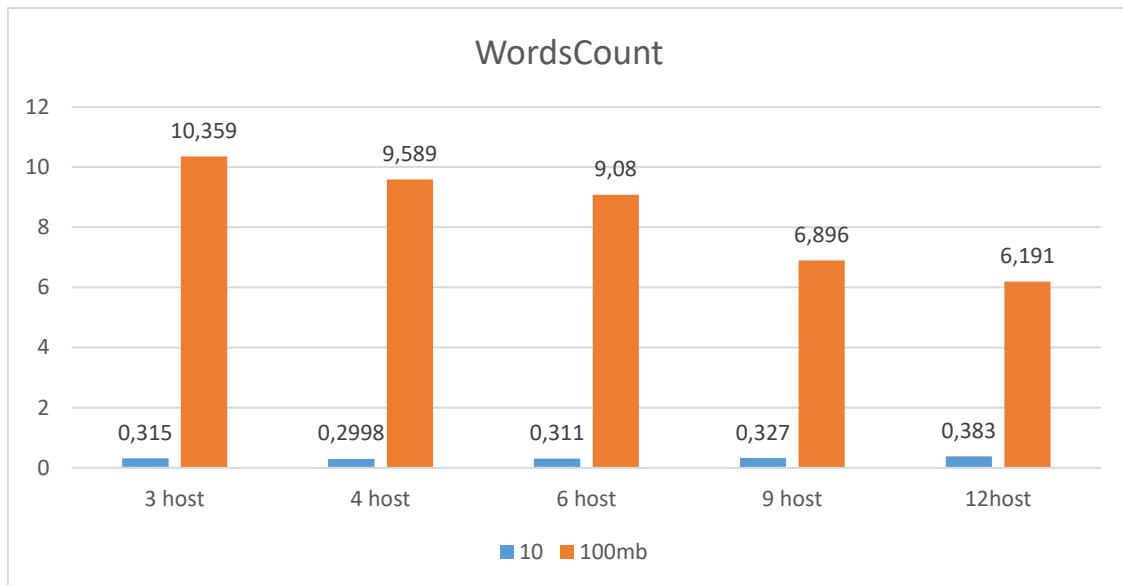
Un cop que ja tenim creats els diferents fitxers, adjudiquem a cada host creat un fitxer de la llista i els executem de forma síncrona. Capa host al finalitzar anirà al reduce i allí tindrem una condició que ens permetrà analitzar les dades rebudes un cop tots els hosts hagin acabat d'analitzar el seu fitxer.

El host que s'encarregui de fer el reduce, al final retornarà les dades obtingudes al clientb i des d'allí crearem un fitxer out.txt o imprimirem el resultat per pantalla.

El joc de proves que hem dut ha terme han estat executades des d'un mateix ordinador i per tan pot ser que les dades obtingudes no siguin les mateixes que podríem obtindre si ho fem des de diferents ordinadors.







El Perf 10.txt i el Perf 100mb.txt són les gràfica de rendiment que hem obtingut. En la gràfica observem dos corbes que descriuen el rendiment es de la funció coutwords en funció del numero de host disponibles. De la gràfica arribem a la conclusió que **en el nostre sistema** el màxim rendiment l'obtenim quan la mida del fitxer que correspon a cada host té una mida aproximada de 1mb amb la capacitat de carga del meu ordinador. Per això, el arxiu que pesa 100mb encara podria augmentar de rendiment al augmentar el nombre de host. És a dir, segons el sistema sempre hi haurà una mida ideal o un nombre de hort ideal en funció de les dades a analitzar i la manera en que es processen les dades.