

Proiecte Big Data si scalabilitate

Toate proiectele sunt compuse din 2 parti, o parte client cu care interactioneaza utilizatorul (de preferat o interfata WEB) si o parte server care se va implementa folosind componente din AWS.

La punctarea proiectelor se va tine cont de:

- Capacitatea sistemului de scalare (se vor genera mii de requesturi pe secunda pentru a vedea cum se modifica timpul de raspuns).
- Capacitatea sistemului de a autoscala.
- Arhitectura sistemului.
- Lipsa unui "single point of failure".
- Raportarea progresului in aplicatia client.
- Functionarea sistemului :p

Proiect 1: Serviciu de conversie video

Implementati un serviciu scalabil de conversie video care poate executa urmatoarele operatii:

- Modificarea rezolutiei clipului video.
- Transformarea clipului in grayscale (alb-negru).

Proiect 2: Analizator de tcpdump

Implementati un serviciu care poate analiza cantitati mari de tcpdump. Sistemul trebuie sa poata raspunde la urmatoarele intrebari:

- Ce cantitate de date a fost transmisa intre masina gazda si o adresa data ca input din client intr-un interval de timp.
- Cat timp a fost folosit un serviciu intr-un interval de timp.

Proiect 3: Analizator de carti

Implementati un serviciu care poate analiza cantitati mari de carti in format text. Sistemul trebuie sa poata raspunde la urmatoarele intrebari:

- Cate carti a scris un autor dat ca parametru din client.
- De cate ori apare cuvantul X in cartile scrise de autorul Y.

Proiect 4: Motor de cautare in documente PDF

Implementati un serviciu care indexeaza cantitati mari de fisiere PDF si poate cauta text in ele.

Proiect 5: Motor de cautare in poze

Implementati un serviciu care primeste ca input o poza si cauta poze similare intr-o baza de date.

Nu se va pune accentul de algoritmul de similitudine dintre poze!

Proiect 6: Generator de heatzone

Implementati un serviciu care primeste ca input un interval de timp si genereaza o poza cu harta lumii peste care se coloreaza zone in functie de traficul de internet in acea zona. Vor fi disponibile cantitati mari de date de forma:

timestamp latitudine longitudine