

MashReduce

- 1 vuosi tehty
 - 1. vaihe: OpenCL-laskentaa etäpalvelimella
 - 2. vaihe: Yhdistetään monta serveriä gridiksi
 - 3. vaihe: MapReduce + Mashup ohjelmointimalli

OpenCL

- Näytönohjain rinnakkaisena prosessorina
- Paljon rinnakkaisia prosessoreita (stream)
- EC2:n tuli syksyllä GPU-instance tuki
 - Maksaa 2\$/tunnilta
- Ohjelmointi C/C++, Python, Ruby, Java...
- Tianhe-A1 pohjautuu GPU-tekniikkaan

Agentti

- MapAgent
 - Agentti joka hajautetaan workereille
- ReduceAgent
 - Agentti joka kokoaa MapAgenttien tulokset masterilla
- Rajapintana toimii komentoriviparametrit sekä tiedostot ja hakemistot
 - Ei kirjastoriippuvuuksia
- C++, Python, jne..

MashReduce

- Master
 - Koordinoi hajautusta
- Worker
 - Voi olla useita, suorittaa suurimman osan työstä
- Molemmat on Webbspalveluita
 - REST API ja käyttöliittymä
 - Django
- Kommunikoidaan tiedostoilla, muttei ole käytössä hajautettua tiedostojärjestelmää

Worker

- Voi ajaa tietyn määrän taskeja kerrallaan
- Käynnistää ja suorittaa MapAgentin taskin
- Osa ladata map-agentille tiedostoja valmiiksi
 - Http://
 - Flickr
 - Paikallisia tiedostoja

Master

- Lähettää map-agentit workereille
- Antaa map-agenteille taskeja, sitä mukaa kun workereillä on aikaa
- Lukee Map-Agenttien ulostulotiedostot ja käynnistää ReduceAgentin joka tuottaa lopulliset tulostiedostot

MapReduce

- Google kehitti 2003
- Luotettavuus siirretty softaan
 - Kotikone hajoaa keskimäärin kerran 3 vuodessa
 - Jos on 10 000 konetta, niin joka päivä hajoaa 3
- Ohjelmointimalli
 - Map+Reduce funktiot on helppo käsittää ja toteuttaa
 - Map+Reduce funktiot voi tehdä yleiskäyttöiseksi, jolloin niitä voidaan yhdistää

MapReduce Esimerkki 1

Map(tiedosto, tiedoston sisältö):

Word = readWord()

EmitIntermediate(word, 1)

Reduce(word, occurrences)

Emit(count(occurrences))

MapReduce Esimerkki 2

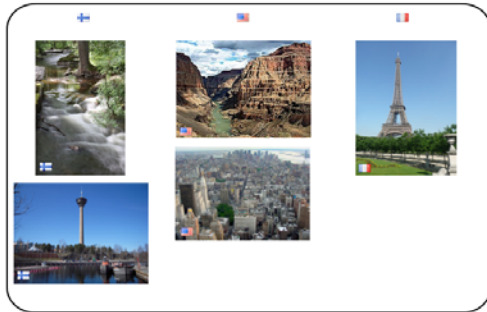
- Map
 - Selataan kuvia verkkohakemistoissa
 - Järjestetään kuvat lokaation perusteella
- Sort
 - Sovelluskehys järjestää kaikkien mappien tuottamat välituokset lokaation perusteella
- Reduce
 - Lasketaan montako kuvaa eri lokaatioista on otettu



MAP



MAP

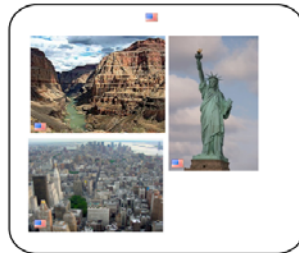


SORT



REDUCE

↓
4



REDUCE

↓
3



REDUCE

↓
2

ID hajautetussa järjestelmässä

- http://localhost/user/joonas/agent/hello_world
- <http://localhost/user/joonas/agent/23>
 - <http://localhost/agent/23>
- <http://localhost/user/joonas/agent/4dfj321i44>
 - <http://amazon.com/user/joonas/agent/4dj321i44>

Demo

- MapAgentti
 - Ajetaan workereillä
 - Tunnistaa kuvista kasvot ja luo pieniä kasvokuvia
 - C++/OpenCV
- ReduceAgentti
 - Ajetaan masterilla
 - Yhdistää kasvokuvat kuvakollaasiksi
 - Python/PIL