Sigui Xi = Consum energètic any-mes i

Es calcula la mitjana del consum mensual com

Avg(X) = Sum (Xi)/#mesos

Solució proposada:

S’utilitzen 2 processos mapreduce:

**1er.- Retorna el consum per a cada any-mes**.

* La rutina map s’encarrega de retornar una tupla (key, value) on la key = mes formatejat %A-%m i el value = Kwh per a cada una de element de la sèrie temporal inicial. Exemple:

2013-01-01 01:00:00;2.3 🡪 (2013-01, 2.3)

* La rutina reduce retorna, el consum mensual per cada mes de la sèrie temporal.

(key, value) = (%A-%m, sum(Kwh)) per a cada mes

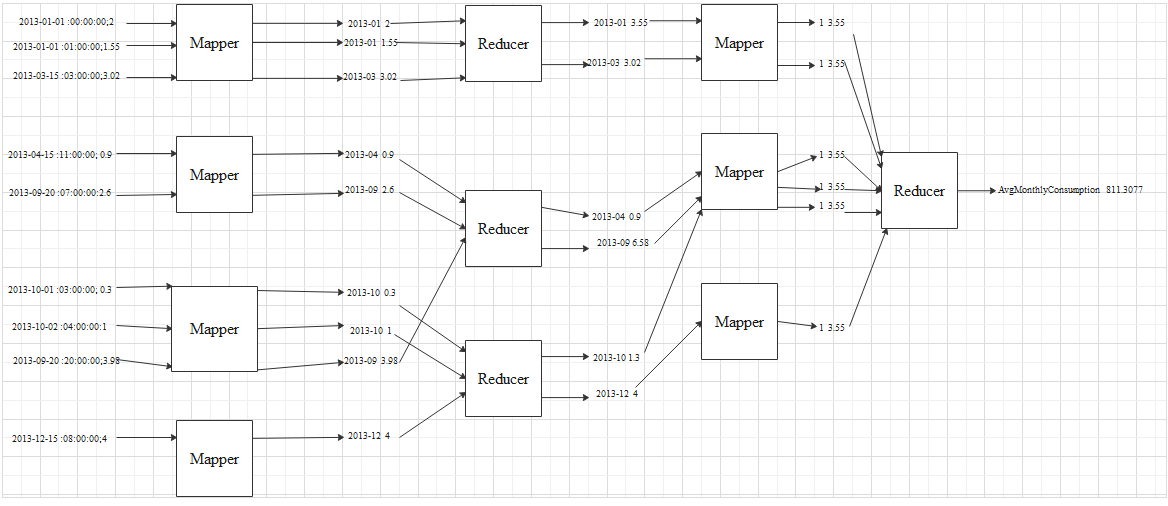
**2on.- Retorna la mitjana del consum mensual**

* Mapper: La entrada d’aquesta rutina es la sortida de l’anterior reduce i per tant son tuples amb el consum de cada mes. La sortida es una tupla (key, value) on la clau es constant i per tant en el seguent reduce agruparà totes les entrades en una 1 tupla de sortida.

(%A-%m consum mensual al mes i) 🡪 (1, consum mensual al mes i)

* Reduce : Aquest reduce s’encarrega de contar el número de mesos presents a la sèrie, que no es res mes que contar el número d’entrades a la rutina i sumar els consums mensuals per acabar fent el càlcul de la mitjana

La sortida (key, value) = (‘avgMonthlyConsumption’, sum(consum mensual i)/#mesos)



**Per executar a un cluster hadoop:**

hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming-\*.jar

-input Input/data.csv

-output Output/monthlyConsumption

-file mapreduce/monthlyConsumption/mapper.py

-file mapreduce/monthlyConsumption/reducer.py

-mapper mapreduce/monthlyConsumption/mapper.py

-combiner mapreduce/monthlyConsumption/reducer.py

-reducer mapreduce/monthlyConsumption/reducer.py

hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming-\*.jar

-input Output/monthlyConsumption/mapper\_output.txt

-output Output/avgMonthlyConsumption

-file mapreduce/avgMonthlyConsumption/mapper.py

-file mapreduce/avgMonthlyConsumption/reducer.py

-mapper mapreduce/avgMonthlyConsumption/mapper.py

-reducer mapreduce/avgMonthlyConsumption/reducer.py

**Observacions:**

* Per a valors de la sèrie temporal inicial no informats, s’ha optat per considerar un consum energètic 0 kwh. Tot i així, existeixen altres possibilitats com per exemple imputar la mitja del mes corresponent o fer mitjanes mòbils.