Hadoop 16/11/18

# Hadoop componenten

## HDFS (hadoop distribute file system)

Client spreekt name node aan -> stuurt client naar een bepaalede node.

Client schrijft de block direct naar een data node

Data node kopieert de block

## Apache Hadoop 2.0 and YARN

Resource manager

## Hadoop MapReduce

1 taak opsplitsen in meerdere taken -> map gedeelte

Alles terug samenvoegen -> reduce

## Zookeeper

Clusterbeheer

# Spark

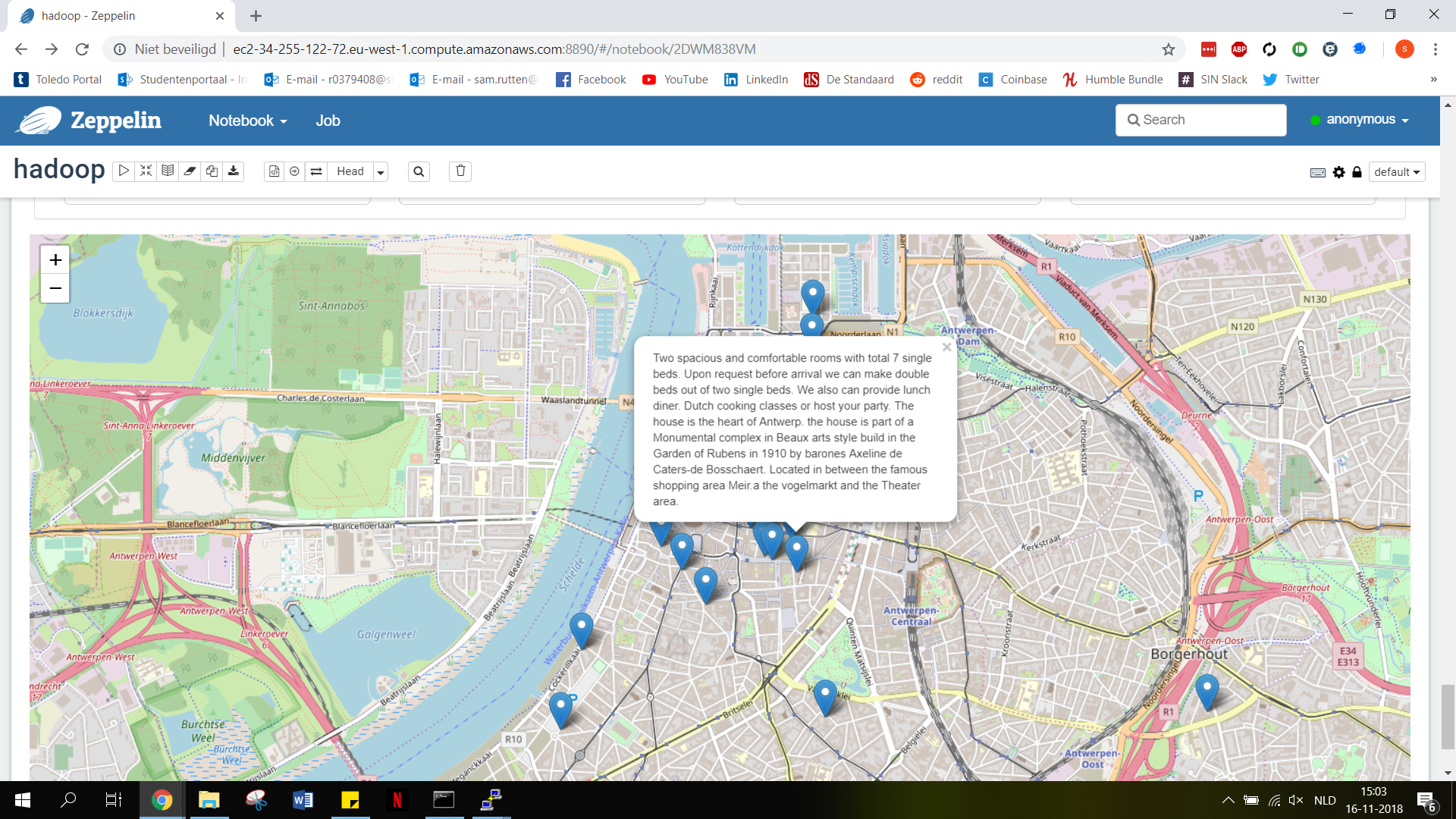
Framework voor cluster computing

# EMR

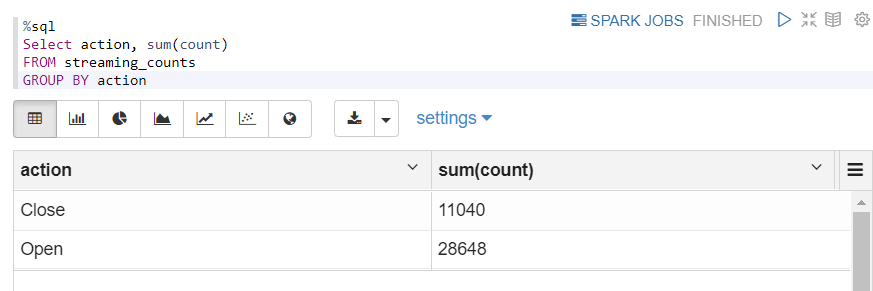
Elastic Map Reduce

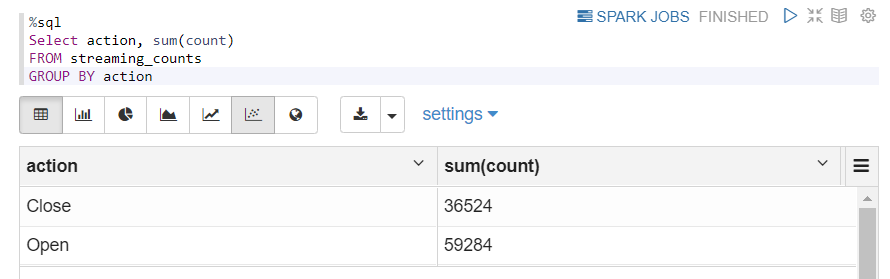
* Zelf niks voor opzetten

# Opdracht 1



# Opdracht 2





# Quicksight

