Rabeet Fatmi – COP 5727 - Redo of Exam 2 Q1 and Q2

# Question 1: What is a map-reduce framework? Provide an example that shows how to use it to solve a real-world problem in the context of advanced databases. (15 points)

MapReduce is a java framework started by the Apache Hadoop foundation to process huge amounts of data, in parallel, on large clusters in a reliable manner. The framework is divided into two important parts: a map() function and a reduce() function.

**Example:**  
Real-world problem: Count all entries logged by user “admin” in an arbitrary dataset of 5,000,000 entries.  
Mapper (pseudo-code):  
map() 🡪

if entry.author == ‘admin’

return count = 1

else

return count = 0

Reducer (pseudo-code):

reduce(mappedValues) 🡪

sum = 0

for (each i in mappedValues)

sum = sum + mappedValues[i][count]

return sum

Here the mapper is simply running the isAdmin check on each dataset and returning 0 or 1 to the reducer based on the result. The reducer is then aggregating the counts into one single value “sum” and returning that sum which is the count of all entries logged by the user “admin”. The mapper or reducer can be running on one single node or multiple nodes in the cluster.

# Question 2: What is REST and MIME? How do you use REST? Show an example. (10 points)

**REST:** Short for Representational State Transfer, is an architectural style of Network-based Software that is often used as an approach to develop a web service. It is a guideline for mapping resources to URLs and interacting with them using CRUD verbs.

**MIME:** MIME is a content type identifier for file formats transmitted through the internet. It is composed of a *type*, a *subtype*, and optional parameters. It is mainly used to let the client know what type of file is it downloading so that it can be rendered accordingly.

**Usage of REST:** REST is usually used in the form of a REST API which uses HTTP verbs such as GET, POST, DELETE, etc. to access the underlying data server.

**Example:** To view a list of buckets in a Riak server running on port 8091, we can use the HTTP client program “curl” to speak to Riak server’s HTTP REST interface.