Topic:

MapReduce Framework and Hadoop. Embarrassingly parallel processes and

other design patterns for big data processing. Cloudera virtual machine. HDFS -

Hadoop Distributed Filesystem, YARN - Yet Another Resource Negotiator.

Questions to be asked to professor:

where will be email correspondence/any other IM?

TA ?

Fri lecture:

**Intro to R**

R language

Is object oriented

R studio – IDE

Interpreted

Can add c c++ routines

Microsoft bought the company that provides support of R

Statistics = mathematics of higher numbers

Eclipse plugin available for R ide needs

Functions can return only one value

Vectors matrices lists

Vectors: elements of same type use c() to create vector

Lists: different types

Last statement value is returned by default in a function.

Matrices are made by combining vectors (not lists)

Class(object) can give list/vector/matrix/factor

Sat lecture:

Statistics is basically to find correlation between 2 or more variables