**Final Exam**

**Prof. Prem Nair**

**09/29/2016**

**Question 1(10 points):** Write pseudo-code for Mapper and Reducer class of **Inverted Indexing (using revised implementation)**

**Question 2(12 points):** Given a posting list (I cannot remember correctly, so let take a posting list below as example)

[(5, 2), (7, 3), (12, 1), (49, 1), (51, 2), …]

Do index compression with

1. Using VarInt
2. Using Simpe-9
3. Using Bit-Aligned

**Question 3(8 points):** (This is theoretical questions)

1. Given a 5000-server cluster, with a mean-time between failures (MTBF) of 2000 days. What is the failure rate?
2. Read 1 TB data. 1 machine having 2 I/O channels (or 2 hard drives) such that each can read 200 MB/sec. How long does this machine read all data?
3. List four distinct features in which you can conclude that it’s not suitable for MapReduce Algorithm.
4. What is the terms of RDD and YARN?

**Question 4 (20 points):**

Given an input text. Let compute average, min, max of length of words begin with the same letter, output following **(letter, average, min, max)**

1. Using MapReduce **(Please don’t using In-Mapper Combining).**   
   For example, we have

[{dome apple bat Day}, {Zero do Be}

Output: following (letter, average, min, max)

(A, 5, 5, 5)

(D, 3, 2, 4)

(B, 2.5, 2, 3)

…..

So on

Explain: We have only “apple” (length of 5) begin with letter “a”, so output will be (A, 5/1, 5, 5). Likewise, we have “dome” (length of 4), “Day” (3), “do” (2), so output will be (D, (4+3+2)/3, 2, 4) = (D, 3, 2, 4). The logic is the same as the others.

1. Modify the MapReduce in **a)** to use **In-Mapper Combining**

**Question 5 (6 points):** Regarding Scala code (I cannot remember, but it’s very easy. Just review the slides, it’s enough.