**Student Name:**

**Paper 1 [Hadoop++]**

Q1: In Hadoop++, the Trojan index is created at which granularity? Is it file-level, block-level, tuple-level?

Q2: Describe how the Trojan Join works between two files R and S? That is, how the data is pre-processed, organized on HDFS, and then join at query time.

Q3: If we have a Trojan index on column X of a file consisting of 200 HDFS blocks, and we have a selection query that select all records where column X = *100.* Assume that value 100 in column X appear only in 10 blocks of the file. Answer the following:

1. Do we need a map-only job or map-reduce job to process the above query?
2. How many mappers will get started in Hadoop++ as part of the above query?
3. Describe how Hadoop++ will make use of the index

**Paper 2 [HaLoop]**

Q1: What are the types of caches that HaLoop proposes? For each cache type, give a brief description on how it can save computations.

Q2: How the Task scheduling in HaLoop works? Does it enforce a task to be at the same node with its cache? If yes, then how Haloop manages busy nodes? If No, then how it works?