



P.T.O.

the same ID, e.g., "CS101") and likewise, studentID's are unique only within a university (different universities may assign the same ID to different students). Suppose we want to answer certain queries approximately from a 1/20th sample of the data. For each of the queries below, indicate how you would construct the sample.

- (a) For each university, estimate the average number of students in a course.
- (b) Estimate the fraction of students who have a GPA of 3.5 or more. [2.5+2.5=5]
5. (a) Describe the process of filtering on data streams. Explain Bloom Filter. [3]
- (b) Write and explain Flajolet-Martin algorithm? [2]
6. Describe the architecture of HDFS in detail with the help of a block diagram. [5]

END