National University of Computer and Emerging Sciences, Lahore Campus

	Course:	Data Warehousing & Data Mining	**Course Code:**	CS409
	Program:	BS(Computer Science)	**Semester:**	Fall 2017
	Date:	17-Oct-2017	**Total Marks:**	10
	Roll No:		**Weight:**	
	Section:		**Page(s):**	1
	Quiz:	3 (Joining Techniques)		

Instruction/Notes:

Consider the following tables and statistics which are part of a student system:

Student (<u>RollNo</u>, Name, gpa, DeptID, BatchID, DegreeID, …..);   Attendance (<u>RollNo</u>, <u>CourseCode</u>, <u>Semester</u>, AttFlag, …..);

Assume student and attendance tables containing 128,000 and 2,560,000 rows respectively (*Student:Attendance* ratio is *1:20*). Each row and each index entry takes 256 bytes and 16 bytes space respectively. Data block size is 16KB and available memory size is 100 blocks. Suppose degree= 'MS' has a selectivity of 3%, batch= ('2015' or '2005') has a selectivity of (4% + 2%), and dept= ('CS or 'EE') has a selectivity of (10% + 5%).

Query:

> *SELECT AVG(gpa) FROM student JOIN attendance ON student.rollno=attendance.rollno*
> *WHERE DegreeID='MS' AND (BatchID='2015' OR BatchID='2005') AND (DeptID='CS' OR DeptID='EE');*

Calculate the total I/O cost (including the I/O cost to filter the condition on student table) for the above Query using <u>hash join</u> and <u>block nested loop join</u> techniques. You are supposed to filter the condition first and then join. Show all steps clearly.