CS186 Discussion Section 1 (part 2)

Disk I/O

5. Consider two tables:
Students(sid, name, year, department), 200 pages, 1,000 tuples
Enrolled(sid, course, grade), 500 pages, 6,000 tuples
Query: for each student, list all his/her class grades:

```
SELECT name, course, grade
FROM Students, Enrolled
WHERE Students.sid = Enrolled.sid
```

Assume that we only have 1 disk, and that we do **not** have to write the resultant tuples back to disk.
Consider the join of Student and Enrolled in a nested loop (the naïve nested loops algorithm) with Student as the outer.
Also assume that we **don't** cache any pages in our buffer pool.

1. What is the total number of I/Os this join will require?
   200 + 1000 * 500 = 500, 200 I/Os

2. Of the total number of I/Os, how many are **sequential** I/Os? (Assume that the data for each relation is located in a continuous clump, but the two relations are located in different places.)
   Outer: Every access will require a random IO, from the end of the inner.
   Inner: 1000 * 499 = 499,000
   Total: 499,000 sequential I/Os.

3. Of the total number of I/Os, how many are **random** I/Os?
   Outer: 200
   Inner: 1000
   Total: 1200 random IOs