MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Sloan School of Management

15.565 – INTEGRATING INFORMATION SYSTEMS:
TECHNOLOGY, STRATEGY, AND ORGANIZATIONAL FACTORS

15.578 – GLOBAL INFORMATION SYSTEMS:
COMMUNICATIONS & CONNECTIVITY AMONG INFORMATION SYSTEMS

Spring 2002

Homework Assignment 4
Due – Lecture #21

Question 1.

The excerpt below represents data obtained from http://quote.yahoo.com when ticker symbol INTC is entered (output is slightly reformated).
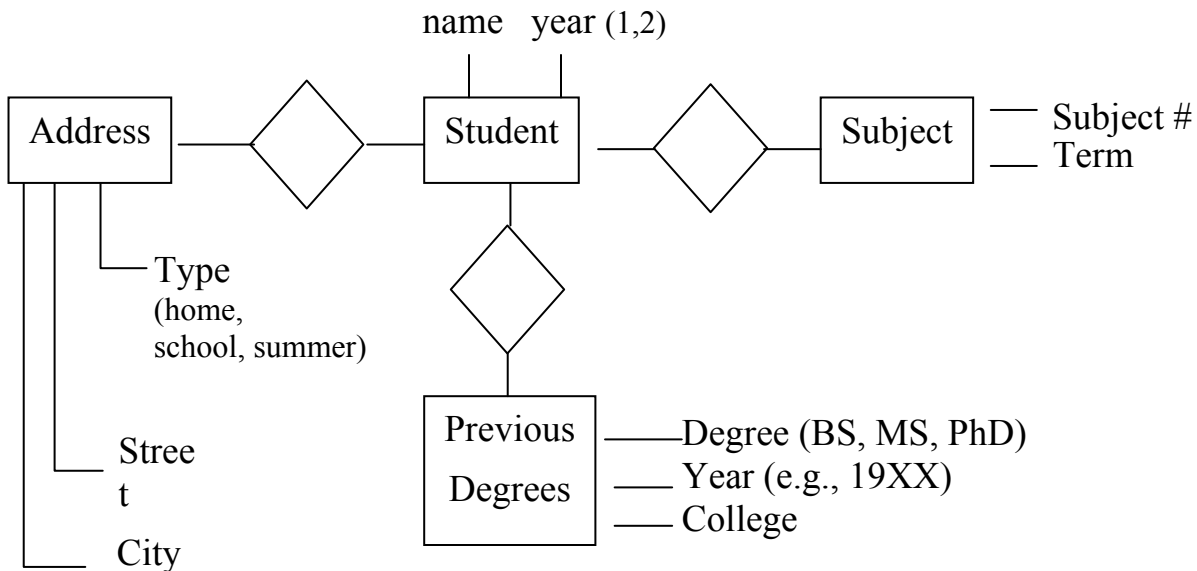
```
INTEL CORPORATION

  Symbol      INTC
  Last trade  30.06        4:00 PM
  Change      -0.04        -0.13%
  Volume      34,154,200
```

a) In class we have referred to such web sites as "semi-structured". Explain what that means.

b) In considering this web site as a database returning database records, what are the 7 field values that are returned by that web page? Give names to each of the fields and briefly (informally) explain the meaning of each (as best you understand them).

c) For each of the 7 fields, explain, in your own terms, what context knowledge is needed and indicate what you believe to be the context definitions for each of the 7 fields in this particular example. Clearly state any assumptions that you feel are appropriate.
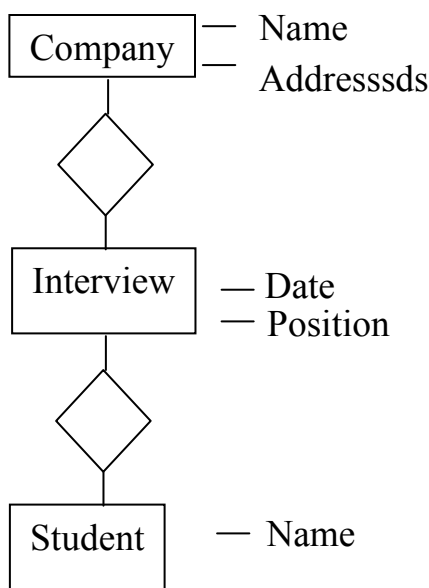
Question 2.

This question addresses the issue of schema integration. Assume that the schema below are used by the Sloan Student database, the Sloan Alumni database, and the Career Development Office (CDO) Recruit database. We want to produce an integrated schema. Perform a step-by-step process, similar to Lecture 16. Show the results of each step: (1) identify equivalent entities and attributes, (2) convert attributes to entities, and (3) integrate the two schema. Briefly explain the reasons, justifications, and assumptions behind each step.
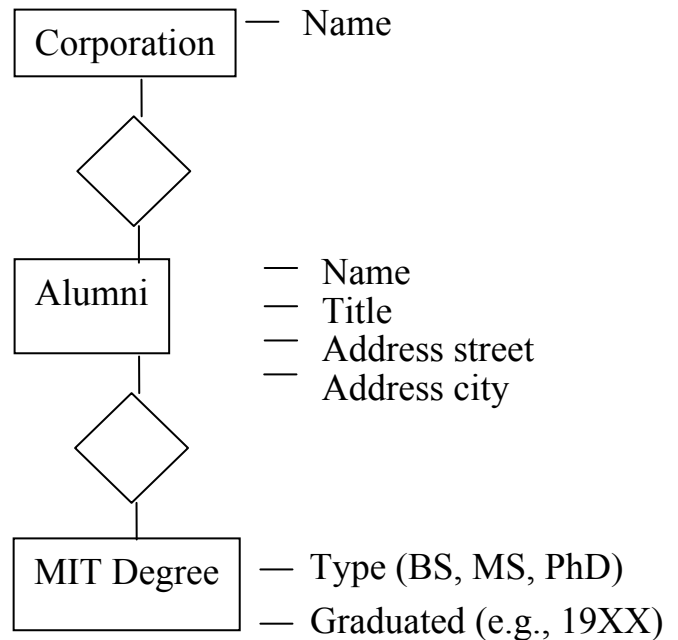
Sloan Students Database

name   year (1,2)

Address ── ◇ ── Student ── ◇ ── Subject ── Subject #
                                            ── Term

Address:
── Type (home, school, summer)
── Street
── City

Student ── ◇ ── Previous Degrees
── Degree (BS, MS, PhD)
── Year (e.g., 19XX)
── College

CDO Recruit Database

Company ── Name
        ── Addresssds

Company ── ◇ ── Interview ── Date
                          ── Position

Interview ── ◇ ── Student ── Name

Sloan Alumni Database

Corporation ── Name

Corporation ── ◇ ── Alumni
── Name
── Title
── Address street
── Address city

Alumni ── ◇ ── MIT Degree ── Type (BS, MS, PhD)
                          ── Graduated (e.g., 19XX)

For this question, let us consider the distributed homogenous SQL database depicted below.

HP UNIX	IBM 390	Sun Solaris	IBM NT
SQL	SQL	SQL	SQL

POS Code	Positio n

POSITIONTB

Nam e	Company	POS Cod e

ALUMNIT B

Company	Dat e

COMPANYTB

DATABASES
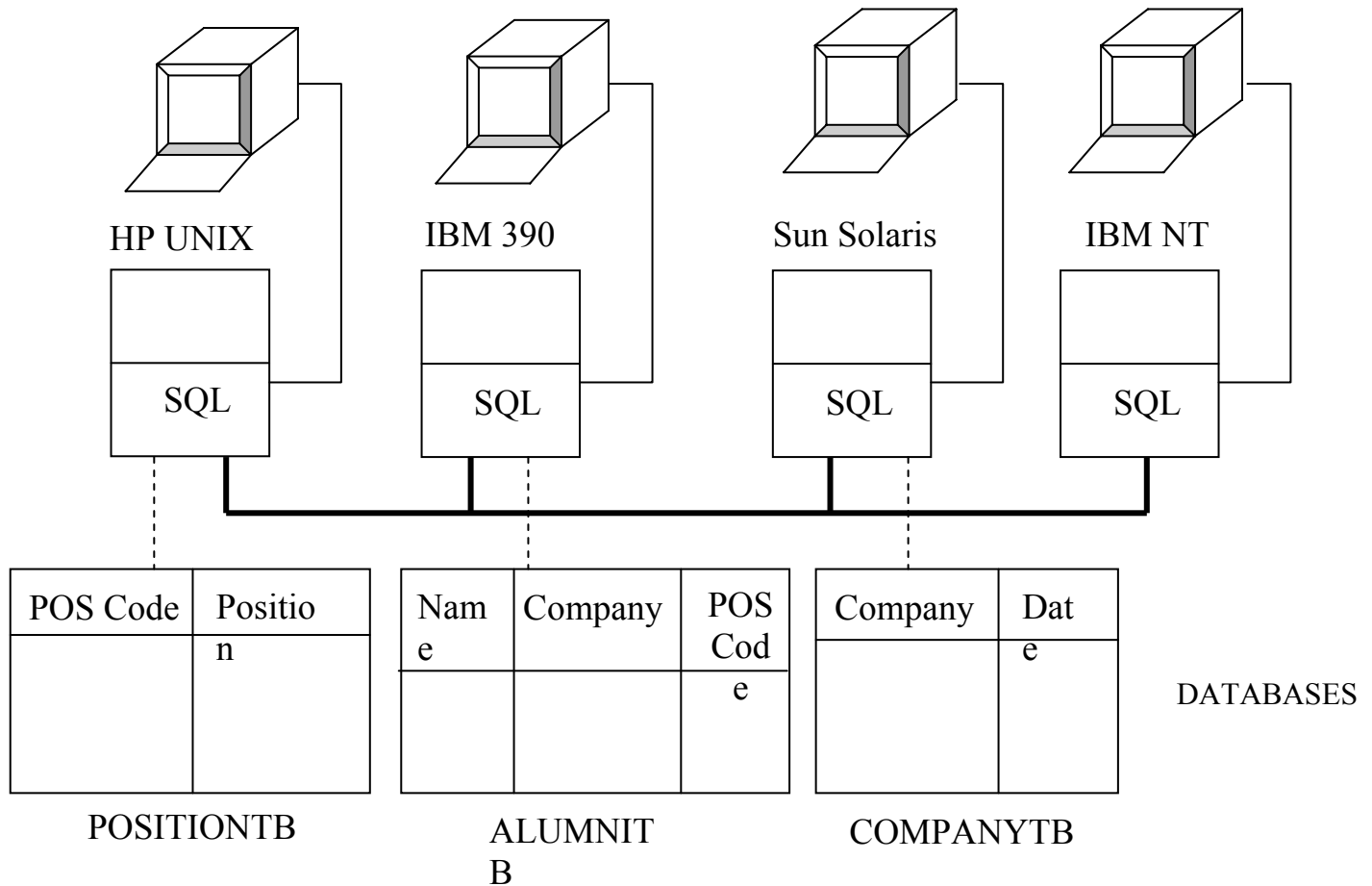
Let us consider the SQL query to print the names of all alumni that are "vice presidents" of companies that are interviewing on "December 14". The query might be:

```
select a.Name
from alumnitb a, positiontb p, companytb c
where p.Position = "vice president"
and p.Poscode = a.Poscode
and c.Date = "December 14"
and c.Company = a.Company
```

a)   Explain what "retrieval transparency" means in the context of the above situation.

b)   Explain what "performance transparency" means in the context of the above situation.

c)   Describe at least two different ways to accomplish "performance transparency" for the above situation (i.e., two different sequences for processing the query.) Make maximum use of concurrency and explain which steps can be performed concurrently.

d)   Of the two approaches described above, which one is better? Explain your reasoning and any assumptions made (or information that you would need.) If helpful, make hypothetical assumptions about table sizes, etc.