CS585 Midterm

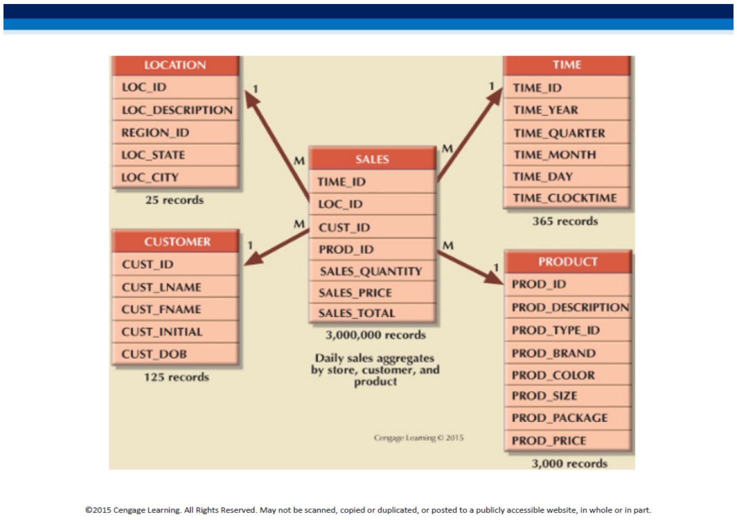
Fall term, 10/11/19
Duration: 1 hour

Instructions/notes

- the exam is closed books/notes/devices/neighbors, and open mind :)
- there are 6 questions, a 'non-data-related' bonus, for a total of **35** points
- there are no 'trick' questions, or ones with long calculations or formulae
- you can write on the two blank sheets (that are at the end) if you like
- **please DO NOT CHEAT; you will get a 0 if you are found to have cheated**
- **when time is up, please STOP WRITING**; **you will get a 0 if you continue**

Q	Your score	Max possible score
1		6
2		6
3		6
4		5
5		6
6		6
Bonus		1
Total		**35 (NOT 36)**

Q1 (4+2=6 points).
In a snowflake schema, the fact table (at the center) is related to several dimension tables (on the periphery, ie. ‘outside’):



LOCATION: LOC_ID, LOC_DESCRIPTION, REGION_ID, LOC_STATE, LOC_CITY — 25 records

TIME: TIME_ID, TIME_YEAR, TIME_QUARTER, TIME_MONTH, TIME_DAY, TIME_CLOCKTIME — 365 records

SALES: TIME_ID, LOC_ID, CUST_ID, PROD_ID, SALES_QUANTITY, SALES_PRICE, SALES_TOTAL — 3,000,000 records

Daily sales aggregates by store, customer, and product

CUSTOMER: CUST_ID, CUST_LNAME, CUST_FNAME, CUST_INITIAL, CUST_DOB — 125 records

PRODUCT: PROD_ID, PROD_DESCRIPTION, PROD_TYPE_ID, PROD_BRAND, PROD_COLOR, PROD_SIZE, PROD_PACKAGE, PROD_PRICE — 3,000 records

Often, a dimension table contains columns that form a hierarchy (eg. TIME and LOCATION, above) - **what is the purpose of this?**

A. To be able to drill down or roll up (zoom out) along that dimension.

Also, you have come across the equivalent of a fact table before - what is it? Explain briefly, using a diagram.

A. A ‘bridge’ table, similar to the one in the lecture:

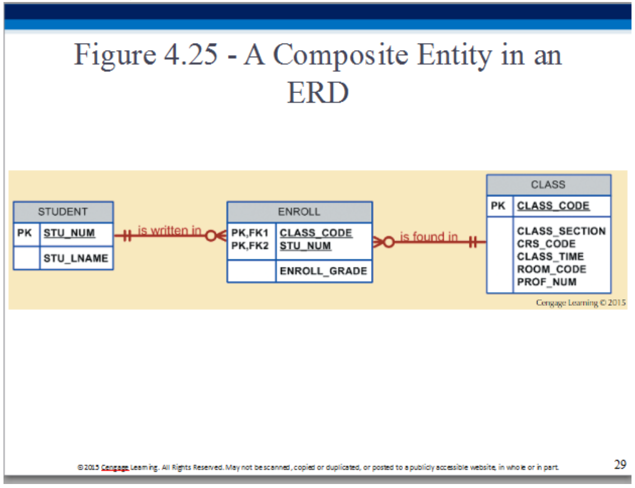


Figure 4.25 - A Composite Entity in an ERD

STUDENT: PK STU_NUM; STU_LNAME

ENROLL: PK,FK1 CLASS_CODE; PK,FK2 STU_NUM; ENROLL_GRADE

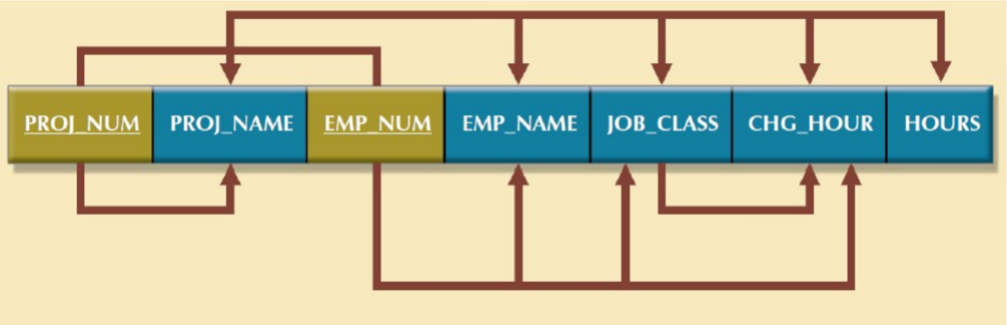
CLASS: PK CLASS_CODE; CLASS_SECTION, CRS_CODE, CLASS_TIME, ROOM_CODE, PROF_NUM

Q2 (4*1.5=6 points). For your HW1, you were asked to design an E-R diagram for a 'STEM' organization. The next step would be, to use the diagram to create tables, connect them appropriately, and deploy the resulting database. **What are important principles and practices that would result in a 'good' relational database?** Think 'across' all the relevant material you learned. Describe each item (principle or practice) using a sentence or two. Provide at least 4 items.

A.

* choose a 'blind' (non-intelligent) (and numeric) primary key
* create normalized entities
* choose 'good' names for entities and attributes
* create indices for non-PK columns frequently used in queries
* create a bridge entity for M:N related entities - limit redundancies to these
* ...

Q3 (2+4=6 points). A 1NF table, such as the one shown below (we covered this in class on great detail), is analyzed to detect problems (related to unwanted dependencies), which are then systematically eliminated (the table is converted to 2NF, then 3NF).



a. **What is the diagram (shown above) called?**

A. Dependency diagram.

b. **How does the diagram aid in normalization?** Explain briefly, using the above diagram (you can mark it up (draw on it) if you want).

A. It helps identify partial and transitive dependencies, thereby allowing us to create 2NF, then 3NF normal forms that systematically eliminate such unwanted dependencies.

Q4 (3+(2*1)=5 points).
a. In the context of database performance tuning, what is an 'access plan'?

A. An access plan is a sequence of optimized I/O operations for data fetching and storage, that result from parsing and optimizing a SQL query - loosely, it is an ‘assembly language’ version of higher-level code statements, after ‘compilation’.

b. What are a couple of ways using which a SQL programmer can enhance her queries (make them be executed efficiently)?

A.
* in expressions, use literals where possible
* in an OR compound expression, place the subexpression most likely to succeed first
* ...

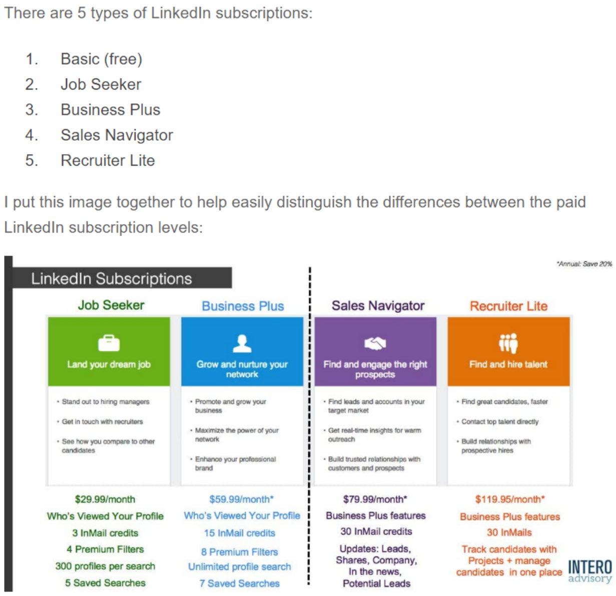
Q5 (2+4=6 points).
Given an EMP table of the form (EMP_ID,EMP_NAME,EMP_DEPT,EMP_SALARY,EMP_MGR), where the column names have 'usual' meanings, **what would the following SQL query output? You need to explain your answer** (ie. how the query produces the result).

```
SELECT DISTINCT salary
FROM EMP E1
WHERE 2 = (SELECT COUNT(DISTINCT EMP_SALARY)
           FROM EMP E2
           WHERE E1.EMP_SALARY <= E2.EMP_SALARY);
```

A. The query will produce the second largest salary value in the EMP table.

Because EMP_SALARY is used in the sub query, to compare E1's value with E2, it is a correlated subquery - for each row's value of E1.EMP_SALARY, we count how many distinct values in all of E2 are greater than or equal to it. **When we get to an E1 row that contains the second highest salary**, our count will be 2, because in E2, that E1 value and a higher value (the overall highest salary value) are the only 2 values that will satisfy the <= condition.

Q6 (6 points). The diagram below (made by 'INTERO advisory'), lists the 5 types of LinkedIn subscriptions a user can sign up for. **Represent them using a small EER diagram using appropriate notations; for the entity supertype and each entity subtype, list a few relevant attributes.**



There are 5 types of LinkedIn subscriptions:

1. Basic (free)
2. Job Seeker
3. Business Plus
4. Sales Navigator
5. Recruiter Lite

I put this image together to help easily distinguish the differences between the paid LinkedIn subscription levels:

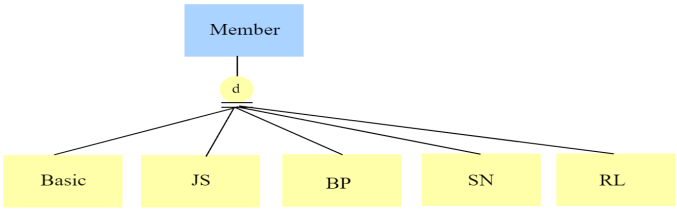
LinkedIn Subscriptions

*Annual: Save 20%

Job Seeker	Business Plus	Sales Navigator	Recruiter Lite
Land your dream job	Grow and nurture your network	Find and engage the right prospects	Find and hire talent
• Stand out to hiring managers • Get in touch with recruiters • See how you compare to other candidates	• Promote and grow your business • Maximize the power of your network • Enhance your professional brand	• Find leads and accounts in your target market • Get real-time insights for warm outreach • Build trusted relationships with customers and prospects	• Find great candidates, faster • Contact top talent directly • Build relationships with prospective hires
$29.99/month	$59.99/month*	$79.99/month*	$119.95/month*
Who's Viewed Your Profile	Who's Viewed Your Profile	Business Plus features	Business Plus features
3 InMail credits	15 InMail credits	30 InMail credits	30 InMails
4 Premium Filters	8 Premium Filters	Updates: Leads, Shares, Company, In the news, Potential Leads	Track candidates with Projects + manage candidates in one place
300 profiles per search	Unlimited profile search		
5 Saved Searches	7 Saved Searches		

INTERO advisory

A.



Member:
* ID
* subscription type (subtype discriminator)
* name
* number of contacts
* date joined
* number of endorsements
* ...

Basic:
* ID

JobSeeker:
* ID
* number of profile viewers
* InMail credits used
* Premium Filters used
* Saved Searches used

BusinessPlus:
* ID
* number of profile viewers
* InMail credits used
* Premium Filters used
* Saved Searches used

SalesNavigator:
* ID
* number of leads used

RecruiterLite
* ID
* number of candidates tracked

Other possibilities for the diagram (two level hierarchies, with disjoint and total constraints at both levels):

* a 2-level hierarchy, with Basic and Paid(JS,BP,SN,RL):

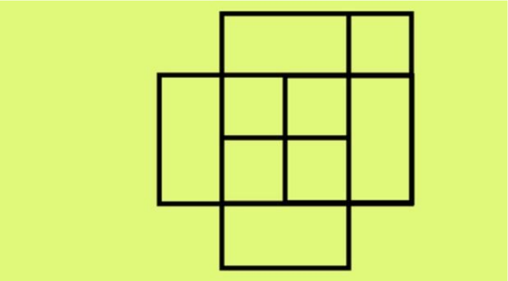
```
Member
  Basic
  Paid
    JS BP SN RL
```

* or a 2-level hierarchy with (Basic,JS,BP) and (SN,RL):

```
Member
  JSBP
    JS BP
  SNRL
    SN RL
```

Bonus (1 point).

How many squares are in the figure below?



11.