



PS Final Exam Spring 2022 6083B

Principles of Database Systems (New York University)

7th May, 2022

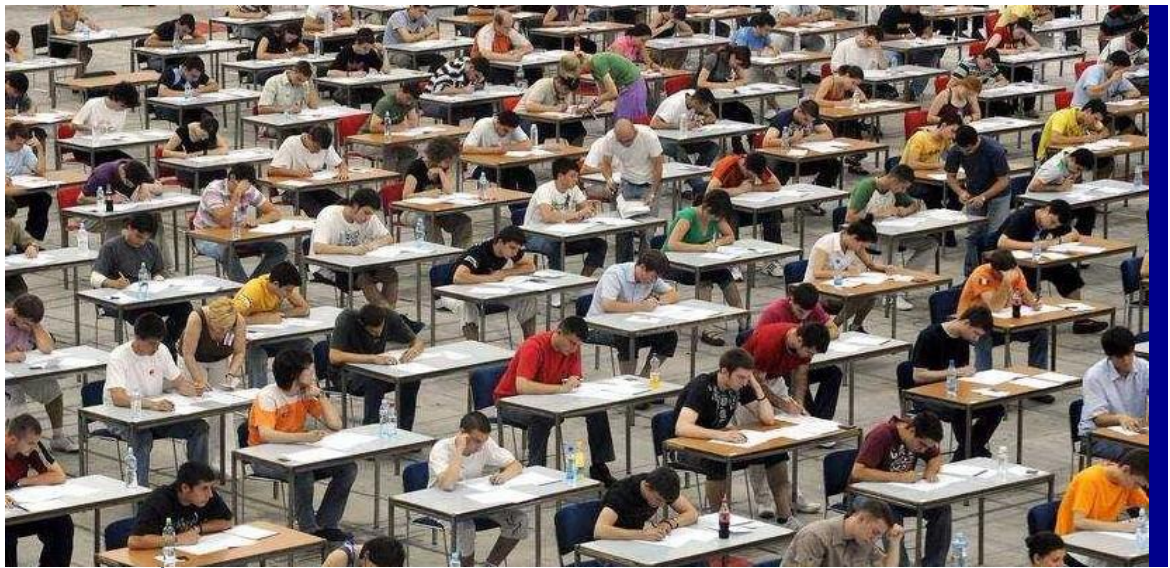
Final Exam



NYU

TANDON SCHOOL
OF ENGINEERING

CS-GY 6083 - B, Spring-2022.
Principles of Database Systems.



FINAL EXAM [100 points with 20% weight]

05/05/2022 11:00 AM to 1:30 PM EST (2 HRS 30 MINS)

Please read instructions carefully before writing exam

Write your name, student id, and net id below

- Last Name:
- Net ID:
- First Name:.
- Student ID:

THIS IS AN ONLINE – OPEN BOOK EXAM. PLEASE LOGIN TO ZOOM MEETING USING YOUR NET ID (DO NOT LOGIN WITH YOUR PERSONAL EMAIL ACCOUNT). Find the Zoom meeting details under ZOOM top bar menu of the course website with title “Final Exam S2022”. Join to the Zoom at least 15 minutes before the exam time.

- WRITE YOUR ANSWERS UNDER EACH QUESTION IN THIS WORD DOCUMENT AND SUBMIT IT ON OR BEFORE 1:30PM TO Course site > ASSIGNMENTS > Final Exam. Save and submit the exam submission document in format <Your Net id>_FinalExam_Spring_2022_6083B. **YOU MAY RESUBMIT YOUR ASSIGNMENT UNLIMITED TIME BEFORE THE SUBMISSION DEADLINE.** The latest submission will be considered for the grading. PLEASE MUTE YOUR MICROPHONE DURING ENTIRE EXAM DURATION and KEEP VIDEO ON SINCE IT IS ONLINE EXAM.
- This exam has 5 sections A, B.C. D. E. All sections and questions have grading points. There is NO negative points for any wrong answers.
- IF YOU HAVE ANY QUESTION DURING THE EXAM, PLES AE SEND YOUR QUESTION PRIVATELY TO PROFESSOR ON ZOOM MEETING CHAT WINDOW. DO NOT SPEAK IN MICROPHONE.
- No hand drawing will be graded. Insert snapshot of database design models in same Word/PDF document. NO ZIP FILE WILL BE ACCEPTED. NO ANY TYPE OF COPY WILL BE GRADED.
- All Objects created should have your initial as prefix, e.g. AP_DEPT
- Keep practice schema (that we used for SQL tutorial) ready to use.

GOOD LUCK!

A) Answer following questions briefly [20 points]

a) Answer following questions [10 points]

- i) What is RAID and what are major goals of it? Which RAID level will you use for storing entire database and archive log files and why?**

RAID: Different schemes (techniques) of disk organization to achieve reliability (via Redundancy/Mirroring) and Performance (via Parallelism / Stripping)

Major Goals of RAID: Disk performance and reliability.

For storing entire database: raid 5 or raid 6 are used, depending on the trade-off of reliability and cost.

Raid 5 will be used for higher Input output transactions rate as it is faster, but low risk tolerance than Raid 6.

Raid 6 stores extra redundant information to guard against multiple disk failure, so it offers better reliability with higher cost.

Archive log files: Raid Level 1 as archive files has low update rates and large amount of data

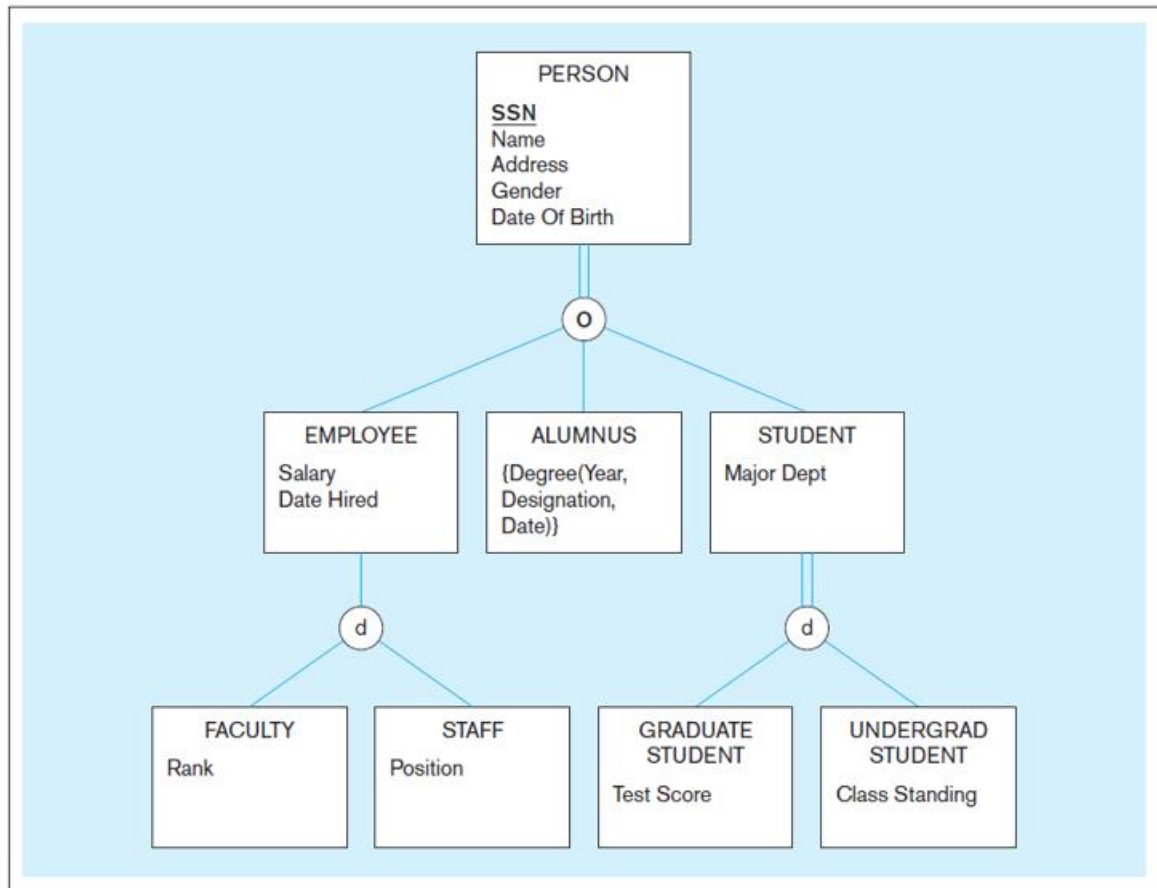
- ii) Explain Shared and Exclusive locks and difference between them.**

Share Lock: share locks can be placed on objects that do not have an exclusive lock already placed on them. It prevents other from updating the data. But still others can read the data (others can place S-locks on it.). More than one share lock can be placed on the same object at the same time.

Exclusive lock: can only be placed on the rows that do not have any other kind of lock (not even S-lock) on it. Once an exclusive lock is placed on a row, no other locks can be placed on the same row anymore. It prevents others from reading or updating the date.

b) Consider following EERD. [10 points]

Explain Supertype-Subtype business rules (completeness/disjoint/overlap) for each of the hierarchy of supertype-subtype.



i) Super type -> Person

subtypes -> Employee, Alumnus, student

Total Completeness: Each record of the person must also be a member of either Employee, Alumnus or student. That is, a person has to be either an employee, student or alumnus.

Overlap: An instance of person could be in more than one supertype, that is a person can be student as well alumnus as employee.

ii) Supertype -> Employee

Subtype -> Faculty, staff

Partial completeness: Employee can be either faculty or staff or can be neither. That is, it is not necessary an instance of employee have to be present in the subtypes.

Disjoint: An instance of employee can be either faculty or staff, but not both. That is can be present in only ONE of the subtypes.

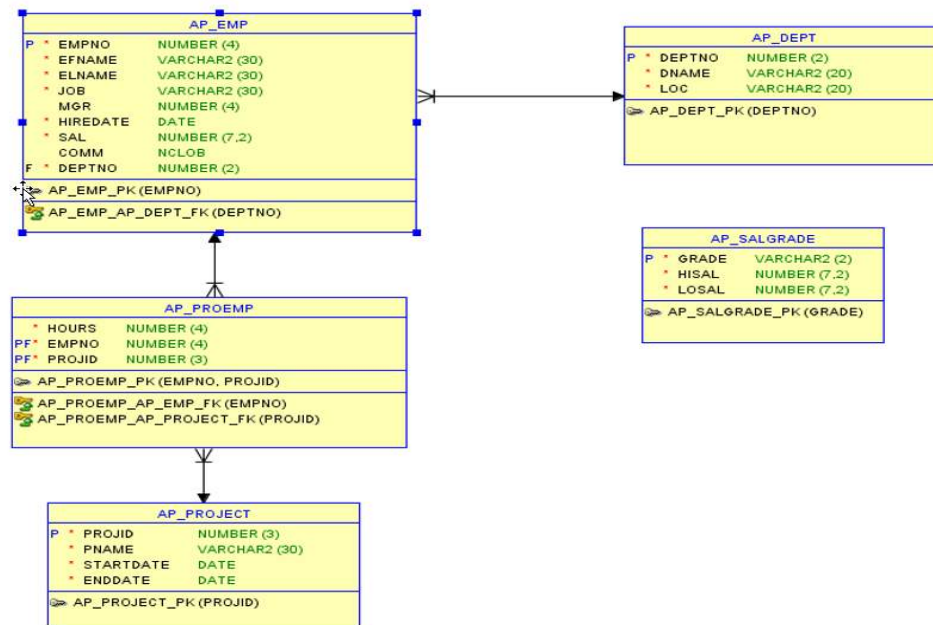
ii) Supertype-> Student

subtype-> Graduate student, Undergrad student

Total completeness: Each instance of student must also be a member of either graduate or undergraduate subtype. That is all students must be categorized in undergraduate or graduate student.

Disjoint: An instance of student can be either graduate or undergraduate but not both at the same time.

Consider following model of HRD database for Questions B, C, and D



RELATIONAL MODEL

B) For HRD database in relational model above, consider following transaction and answer the question [10 points]

- i) CONNECT apatel/NYUTand0n22@HRD
- ii) UPDATE AP EMP SET COMM= COMM + SAL*0.08 WHERE DEPTNO=60;
- iii) SELECT FROM AP_EMP WHERE DEPTNO=60
- iv) COMMIT;
- v) ROLLBACK;
- vi) DELETE FROM EMP WEHRE COMM IS NULL and DEPTNO=60;
- vii) ALTER TABLE DEPT MODIFY LOC NULL;
- viii) DELETE FROM DEPT WEHRE DEPTNO=60;
- ix) ROLLBACK;
- x) COMMIT;

For above set of database changes, list all transactions with transaction number e.g. TX1, TX2 etc., along with when each transaction starts and when it ends. [5 points]

Transaction Name	TX1	TX2	TX3	TX4	TX5
Tx start activity (i to x)	i	iv	v	vii	ix
Tx end activity (i to x)	iv	v	vii	ix	x

At the end of these transactions what changes will take place in HRD database and why? [5 points]

At the end of these transactions, the commission is set to $\text{comm} + \text{sal} * 0.08$ where deptno is 60, FOR THOSE EMPLOYEEES WHO ARE ELIGIBLE TO RECEIVE COMMISSION. Because there is no effect of ROLLBACK after commit. And rows are deleted where COMM IS NULL and DEPTNO=60. Also, LOC is set to NULL in the table DEPT. Because ALTER is self-commit. Delete from dept where deptno=60 will not be done as DELETE is not self-commit and ROLBACK will restore DELETE command. Since TX2 has no DML activities, there will not be no effect of ROLLBACK. Similarly since TX5 has no DML activities there will not be effect of commit after the rollback.

C) For HRD database in relational model above, correct each of SQL [30 points]

Following are incorrect SQLs.

For each of these SQL, identify and list all mistakes and then write corrected SQL. State the purpose of the corrected query (what business question that query answers to)

i) **ALTER TABLE AP_DEPT ADD CONSTRAINT ck_dept_loc
CHECK (LOC = (NEW YORK, BOSTON, CHICGO, DALLAS)
WHERE DEPTNO=20;**

Mistakes: LOC= is incorrect. The correct format is LOC IN
As we have used ALTER TABLE, we cannot use the WHERE clause

CORRECT SQL:
ALTER TABLE AP_DEPT ADD CONSTRAINT ck_dept_loc
CHECK (LOC in ('NEW YORK', 'BOSTON', 'CHICAGO', 'DALLAS'));

Purpose: Apply a constraint such that the location has one of the values as New York, Boston, Chicago, or Dallas

ii) **SELECT JOB, LOC, AVG(COMM) 'AVG COMMISON'**

```
FROM AP_EMP A JOIN AP_DEPT B ON  
A.DEPTNO=B.DEPTNO  
WHERE COMM<>NULL AND AVG(COMM)>=300  
ORDER BY LOC;
```

Mistakes: AVG COMMISSION should be in " " instead of ' '
COMM<>NULL is incorrect. It should be COMM IS NOT NULL. Null values cannot be compared with comparison operator.

Group by clause should be used as we used aggregate function AVG

AVG(COMM)>=300 cannot be used in the where clause, it should be used in the HAVING clause

CORRECT SQL:

```
SELECT JOB, LOC, AVG(COMM) "AVG COMMISSION"  
FROM AP_EMP A JOIN AP_DEPT B ON  
A.DEPTNO=B.DEPTNO  
WHERE COMM IS NOT NULL  
GROUP BY JOB, LOC  
HAVING AVG(COMM)>=300  
ORDER BY LOC;
```

Purpose: It should retrieve all the job, location, and location average commission location wise

except those whose commission is null and except those whose average commission is greater than or equal to 300

**iii) SELECT EMPNO, EFNAME, DEPTNO, SAL
FROM AP_EMP WHERE SAL>=(SELECT JOB, AVG(SAL) FROM
AP_EMP GROUP BY LOC);**

iii. Mistakes: It should have ANY/ANY/IN operator in where clause as the single-row subquery is returning more than one record.

If using the WHERE clause, salary cannot be compared with two columns in a subquery, it should only have AVG(SAL)

LOC column in GROUP BY is invalid since EMP table does not have this column. It should be group by JOB

CORRECT SQL:

```
SELECT EMPNO, EFNAME, DEPTNO, SAL  
FROM AP_EMP  
WHERE SAL>= ANY (SELECT AVG(SAL)  
FROM AP_EMP GROUP BY JOB);
```


Purpose:

Retrieve employee number, first name, department number, and salary of employees whose salary is greater than or equal to the minimum average salary of employees in grouping of their JOB

iv) **TRUNCATE AP_PROEMP
WHERE PROJID=100, HOURS>100;**

Mistakes: If Truncate clause is used then we cannot use the WHERE clause

CORRECT SQL:
TRUNCATE TABLE AP_PROEMP;

Purpose:

Is to remove all rows of AP_PROEMP table

v) **SELECT EMPNO, ENAME, DEPTNO, SAL
FROM AP_EMP
WHERE SAL>=4000
ORDER BY SAL
INTERSECT
SELECT ENAME, EMPNO, DEPTNO, SAL
FROM AP_EMP
WHERE DEPTNO=20
AND COMM=NULL;**

Mistakes: The position of Order by clause is incorrect. It should be used as the last clause. Column names in both select statements should match the data types. COMM=NULL is incorrect. It should be COMM IS NULL

CORRECT SQL:
SELECT EMPNO, EFNAME, DEPTNO, SAL FROM AP_EMP
WHERE SAL>=4000
INTERSECT
SELECT EMPNO, EFNAME, DEPTNO, SAL FROM AP_EMP
WHERE DEPTNO=20 AND COMM IS NULL
ORDER BY SAL;

Purpose:

Retrieve employee number, employee name, department number, and salary of employees whose salary is greater than or equal to 4000 and employees who work in department 20 and those who are not eligible for commission

D) Consider following business scenario and answer questions [20 points]

Assume that, there are 9,000 employees, 30 managers, 6 jobs, and total 90,000 combination of projects and employee.

The company frequently reassigns manager to employees with employees work location or job changes. Data Analytics Department runs a report daily based upon following query

```
SELECT a.empno,  
       a.ename || ' ' || a.ename  "Employee",  
       a.job,  
       d.dname,  
       b.ename || ' ' || b.ename  "Manager",  
       a.sal,  
       b.sal,  
       c.hours  
FROM ap_emp a  
     JOIN ap_emp b ON a.mgr = b.empno  
     JOIN ap_proemp c ON a.empno = c.empno  
     JOIN ap_dept d ON a.deptno = d.deptno  
WHERE a.JOB = 'SALESMAN' OR a.JOB = 'ANALYST'  
ORDER BY a.sal, UPPER (a.ename || ' ' || a.ename);
```

Answer following questions.

- i) Which attributes are most suitable for creating index to improve the performance of this report query?

For AP_EMP table, index on UPPER (ename || ' ' || ename) [there should not be table alias when creating index, and this index will help for Order By clause]

For AP_EMP table, index on DEPTNO. [this will help table joins]

For AP_EMP table, index on JOB [this will help for WHERE clause filter]

- ii) What type of index is appropriate for each column that you have intended to create index and why?

For AP_EMP: JOB(Bitmap index is suitable since few distinct values) ; UPPER(a.ename || ' ' || a.ename) (since function is used along with column in order by clause, function based index is suitable); DEPTNO (secondary (non-unique) index)

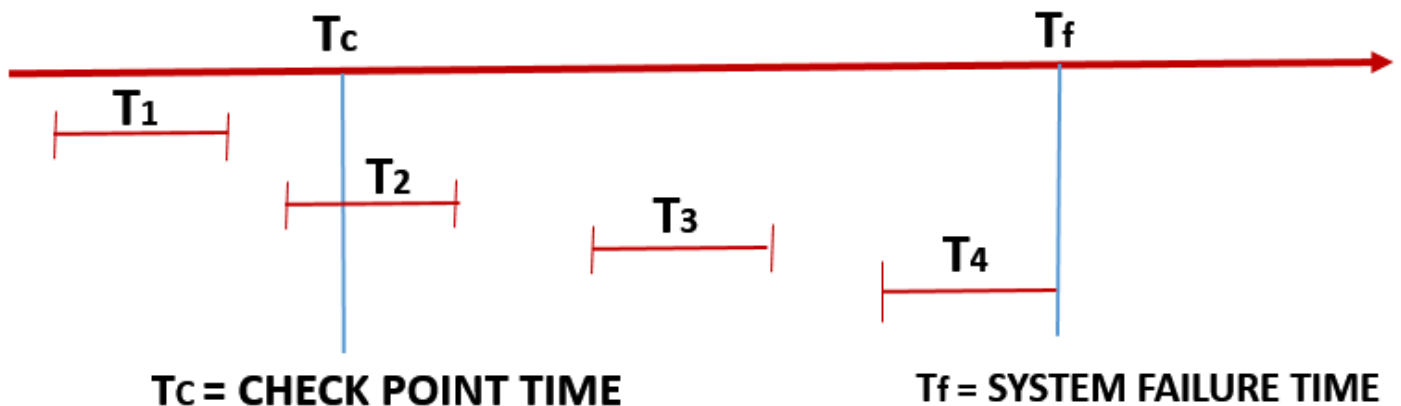
- iii) **Write DDL command to create the index for each column that you have intended to create index.**

```
CREATE BITMAP INDEX idx_job on AP_EMP(JOB);  
CREATE INDEX idx_ename on AP_EMP(UPPER(ename || ' ' || ename));  
CREATE INDEX idx_deptno on AP_EMP(DEPTNO);
```

- iv) **Which attribute(s) in each table has index created automatically, and what is the type of index they are.**

EMPNO attribute of AP_EMP table, DEPTNO attribute of AP_DEPT table, EMPNO and PROJID attributes of AP_PROEMP table will have index created automatically as they are PKs of the respective entities.

E) Consider following scenario of transactions in HRD database. [20 points]



T_c is the time when database has checkpoint.

T_f is the time when database has system failure

T₁, T₂, T₃, and T₄ are database transactions

Transactions T₁, T₂, T₃, and T₄ occurred in chronological order. The checkpoint in database happened at given time T_c and later on time at T_f, the system crashed on power failure.

Assume that, the employee JASON with employee id 7788 is working in department number 20 with monthly salary of \$9000. JASON is currently not eligible to earn the commission.

Following is the details of work done in each transaction.

T1:

```
CONNECT apatel/N3wY0rk21@HRD
```

```
SELECT deptno, sal, comm FROM ap_emp WHERE EMPNO='7788';
```

```
UPDATE ap_emp SET sal=sal+700, comm=sal*0.1 WHERE EMPNO='7788'
```

```
CREATE TABLE EMP_TEST AS SELECT * FROM ap_emp WHERE 'Tf'='Tc';
```

T2:

```
DELETE FROM ap_emp WHERE SAL BETWEEN 8000 and 9000;
```

```
ALTER TABLE EMP_TEST ADD CONSTRAINT  
pk_emp_test PRIMARY KEY(EMPNO);
```

T3:

```
UPDATE ap_emp SET deptno=30 WHERE empno='7788';
```

```
UPDATE ap_emp SET sal=sal+sal*0.1 WHERE deptno=30;
```

```
COMMIT;
```

T4:

```
UPDATE ap_emp  
SET comm=nvl(comm,0)+100 , sal=sal+300  
WHERE empno=7788;
```

- i) Upon system recovery, which transaction(s) will undergo REDO operations and which transactions will undergo UNDO operations and why? [8 points]**

Upon system recovery, transactions T2 and T3 will undergo REDO operation since T2 started but didn't complete at checkpoint Tc. Also, T3 started after checkpoint and was complete before system crash at Tf. Transaction T4 will undergo UNDO operation since, T4 started after the checkpoint Tc but was never completed.

- ii) **For transaction(s) that will undergo UNDO, what will be written out in transaction log? [8 points]**

<T4, COMM, 900>, <T4, SAL 10670>, T4, ABORT>

- iii) **What will the department number, salary, and commission of JASON after the system is recovered. [4 points]**

After the system is recovered JASON's department number will be 30, salary will be 10670 and commission will be 900.

===== End of Exam Paper=====