

Homework 4 for CS542 - Fall 2023

Assigned: Wednesday September 20, 2023

Due: Wednesday September 27, 2023 10 am ET (via CANVAS).

Maximum: 100pts.

Note: This homework is to be done by each student individually.

Purpose of homework: To become familiar with the foundations of relational database querying.

Problem 1: SQL Queries About Employment. [100 pts]

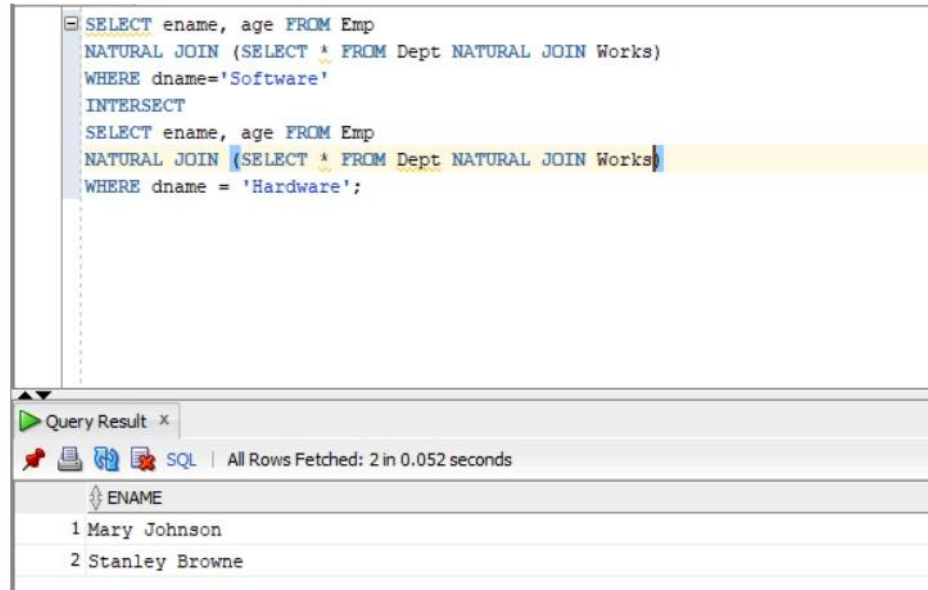
```
Relation Emp
eid Integer,
ename Char Width 255,
age Integer,
salary Float,
PRIMARY KEY (eid);
```

```
relation Works
eid Integer,
did Integer,
pct_time Integer,
PRIMARY KEY (eid,did),
FOREIGN KEY (eid) REFERENCES Emp (eid),
FOREIGN KEY (did) REFERENCES Dept (did);
```

```
relation Dept
did Integer,
dname Char Width 255,
budget Float,
managerid Integer,
PRIMARY KEY (did),
FOREIGN KEY (managerid) REFERENCES Emp (eid);
```

Part 1:

1. `SELECT` `ename`, `age` `FROM` `Emp`
`NATURAL JOIN` (`SELECT` `*` `FROM` `Dept` `NATURAL JOIN` `Works`)
`WHERE` `dname` = 'Software'
`INTERSECT`
`SELECT` `ename`, `age` `FROM` `Emp`
`NATURAL JOIN` (`SELECT` `*` `FROM` `Dept` `NATURAL JOIN` `Works`)
`WHERE` `dname` = 'Hardware';



The screenshot shows a SQL query editor with the following text:

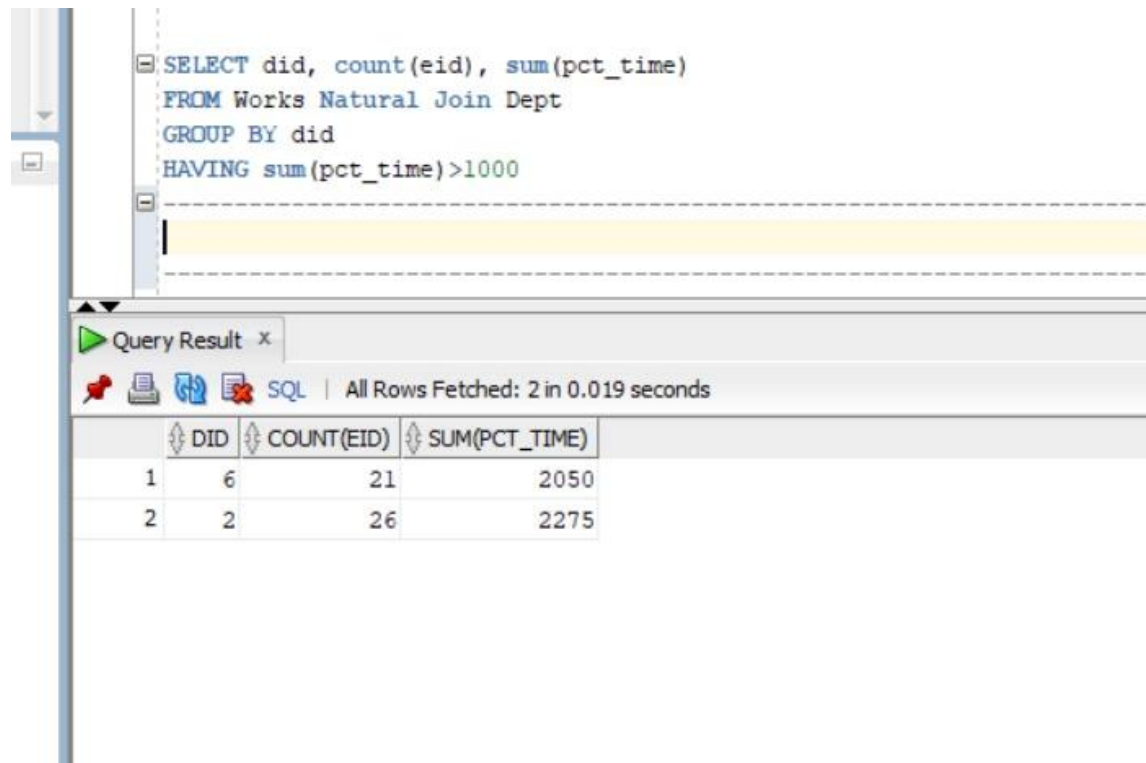
```
SELECT ename, age FROM Emp
NATURAL JOIN (SELECT * FROM Dept NATURAL JOIN Works)
WHERE dname='Software'
INTERSECT
SELECT ename, age FROM Emp
NATURAL JOIN (SELECT * FROM Dept NATURAL JOIN Works)
WHERE dname = 'Hardware';
```

Below the editor is a 'Query Result' window. It shows the execution status: 'All Rows Fetched: 2 in 0.052 seconds'. The results are displayed in a table with one column, 'ENAME', and two rows:

	ENAME
1	Mary Johnson
2	Stanley Browne

2.

```
SELECT did, count(eid), sum(pct_time)
FROM Works Natural Join Dept
GROUP BY did
HAVING sum(pct_time)>1000;
```



The screenshot shows a database query editor with the following SQL query:

```
SELECT did, count(eid), sum(pct_time)
FROM Works Natural Join Dept
GROUP BY did
HAVING sum(pct_time)>1000
```

Below the query editor is a "Query Result" window. It displays the results of the query in a table format. The table has four columns: "DID", "COUNT(EID)", and "SUM(PCT_TIME)". The results are as follows:

	DID	COUNT(EID)	SUM(PCT_TIME)
1	6	21	2050
2	2	26	2275

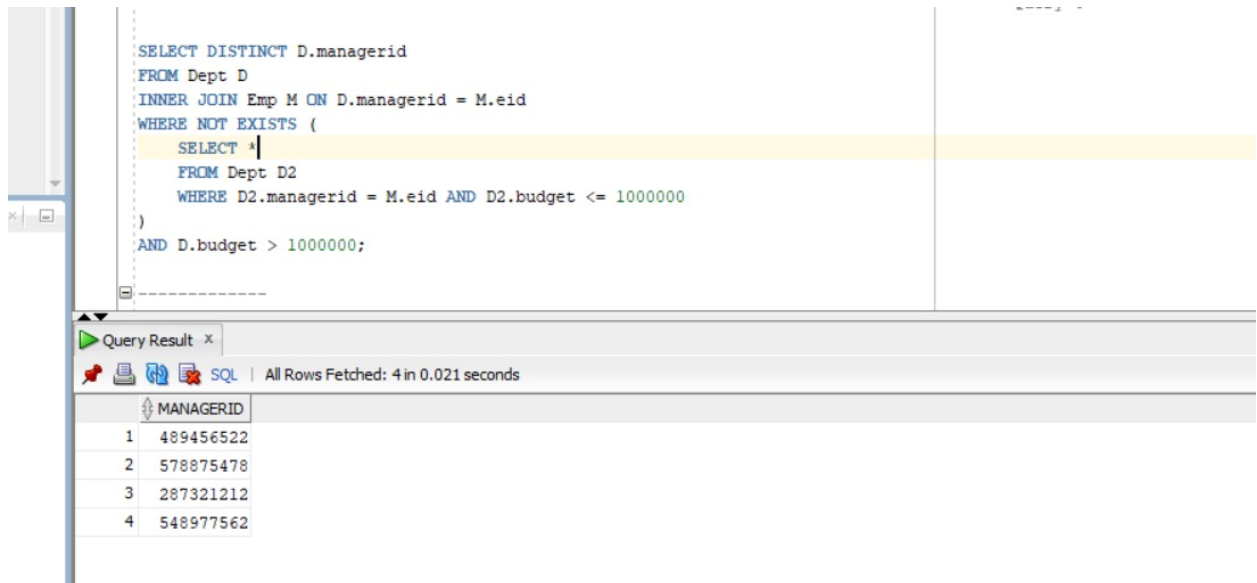
The "Query Result" window also includes a status bar that reads "All Rows Fetched: 2 in 0.019 seconds".

3. SELECT ename FROM Emp
MINUS
SELECT ename
FROM Emp NATURAL JOIN (SELECT * FROM Dept NATURAL JOIN Works)
WHERE salary <= budget;

```
SELECT ename
FROM Emp
MINUS
SELECT ename
FROM Emp NATURAL JOIN (SELECT * FROM Dept NATURAL JOIN Works)
WHERE salary <= budget;
```

Query Result x	
All Rows Fetched: 2 in 0.05 seconds	
ENAME	
1 Linda Davis	
2 Patricia Jones	

4. `SELECT DISTINCT D.managerid
FROM Dept D
INNER JOIN Emp M ON D.managerid = M.eid
WHERE NOT EXISTS (
SELECT * FROM Dept D2
WHERE D2.managerid = M.eid AND D2.budget <= 1000000)
AND D.budget > 1000000;`



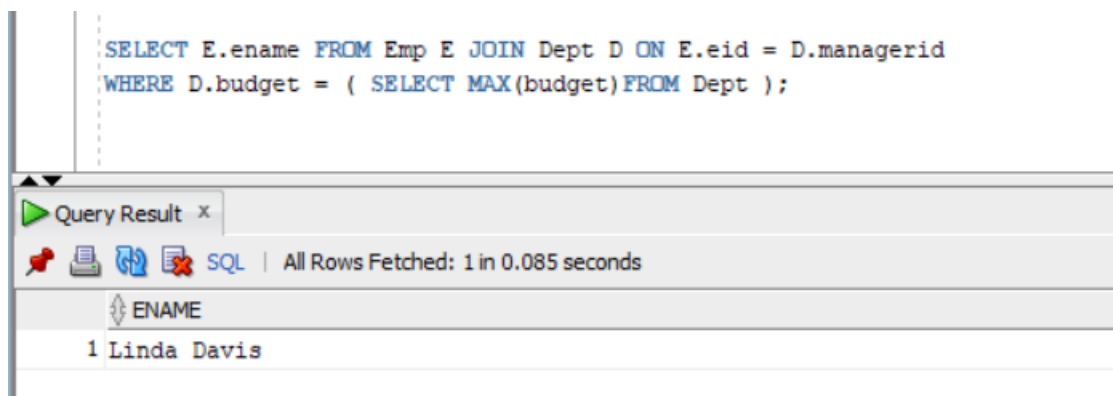
```
SELECT DISTINCT D.managerid  
FROM Dept D  
INNER JOIN Emp M ON D.managerid = M.eid  
WHERE NOT EXISTS (  
  SELECT *  
  FROM Dept D2  
  WHERE D2.managerid = M.eid AND D2.budget <= 1000000  
)  
AND D.budget > 1000000;
```

Query Result x

All Rows Fetched: 4 in 0.021 seconds

	MANAGERID
1	489456522
2	578875478
3	287321212
4	548977562

5. `SELECT E.ename FROM Emp E JOIN Dept D ON E.eid = D.managerid
WHERE D.budget = (SELECT MAX(budget) FROM Dept);`



```
SELECT E.ename FROM Emp E JOIN Dept D ON E.eid = D.managerid  
WHERE D.budget = ( SELECT MAX(budget) FROM Dept );
```

Query Result x

All Rows Fetched: 1 in 0.085 seconds

	ENAME
1	Linda Davis

Part 2:

We tried following queries with Emp table empty and tried inserting data into Dept table

```
INSERT INTO DEPT1 VALUES(111111, 'FIN', 12340, 456)
```

SQL developer will give error as it is integrity constraint

We tried following queries with Emp and Dept table empty, and tried inserting data into Works table

SQL developer will give error for: 'Parent key not found'

We tried for the bigger dataset using following query instead of using the above Query no 4.

We noticed that when it comes to bigger dataset following query generates same result but in faster time. Because we havent used DISTINCT keyword.

```
SELECT D.managerid
FROM Dept D
INNER JOIN Emp M ON D.managerid = M.eid
WHERE NOT EXISTS (
  SELECT * FROM Dept D2
  WHERE D2.managerid = M.eid AND D2.budget <= 1000000
)
AND D.budget > 1000000
GROUP BY D.managerid
HAVING COUNT(*) = 1;
```