3/15/2024

SQL DATABASE PROJECT

CDBM-603-001 Enterprise Data Architecture
Artificial Intelligence and Data Analytics
Saskatchewan Polytechnic, Saskatoon

Prepared By-
Jyoti Bisht
Bigyan Sapkota
Ali Azgar Katha

Table of Contents

Introduction-

Database normalization is a critical technique in the creation of successful and efficient database systems. To ensure data integrity and improve overall database speed, it involves organizing data within a database to minimize redundancy, dependency, and inconsistencies. In this report, we will talk about the normalization procedure used on a database that contains information on employees, jobs, payroll, committees, and pay-week end dates. We will discuss why the data has been divided into several tables and how each table complies with the specifications of various normal forms.

Normalization Process-

First Normal Form (1NF)

In the First normal form(1NF), all the columns of the table must have atomic values and all the columns must have unique identifiers. In the actual table, there were rows with no values. So, we fill the empty rows with the values from the previous non-empty values and ensure all the columns have the atomic values.

EmpId	SIN	Last	First	Street	City	Prov	Postal	Job Code	Position	Payrate CAD	Inc. Tax	Birth Date	Hire Date	Job Code Date	Pay Week End Date	Days Available	Hours	OT	Person Hours Worked	Supervisor	Supervisor Cell#
97319	516303417	Novak	Gerry	6803 Park Ave.	Moose Jaw	SK	S6H 1X7	3000	Stockperson	12.99	N	24-Aug-86	07-Jul-03	07-Jul-03						Abu Muktadir	306.304.4545
33982	867481381	Boychuk	Robin	117 East Broadway	Moose Jaw	SK	S6H 3P5	5000	Butcher	18	Y	04-Mar-71	11-Oct-98	11-Oct-98	23-May-13	7	40.00	0.00	322.50	Joseph Herbert	306.304.1212
33982	867481381	Boychuk	Robin	117 East Broadway	Moose Jaw	SK	S6H 3P5	5000	Butcher	18	Y	04-Mar-71	11-Oct-98	11-Oct-98	30-May-13	6	38.25	0.00	236.00	Joseph Herbert	306.304.1212
51537	112893584	Smith	Kim	9745 University Drive	Regina	SK	S4P 7A3	2000	Cashier	11.99	Y	29-Nov-82	02-Dec-01	02-Dec-01	23-May-13	7	27.00	0.00	322.50	Melissa Jones	306.304.8878
41822	717505366	Miller	Chris	72 Railway Ave.	Pense	SK	S0T 1K4	2000	Cashier	11.99	Y	15-Nov-68	19-Feb-85	19-Feb-85	23-May-13	7	40.00	0.00	322.50	Melissa Jones	306.304.8878
41822	717505366	Miller	Chris	72 Railway Ave.	Pense	SK	S0T 1K4	2000	Cashier	11.99	Y	15-Nov-68	19-Feb-85	19-Feb-85	30-May-13	6	38.00	1.25	236.00	Melissa Jones	306.304.8878
3571	374853129	Hashimoto	Jo	386 High Street	Tuxford	SK	S0L 8V6	1000	Greeter	10.25	Y	23-Jun-56	20-Mar-80	30-Aug-99	23-May-13	7	40.00	0.00	322.50	Melissa Jones	306.304.8878
3571	374853129	Hashimoto	Jo	386 High Street	Tuxford	SK	S0L 8V6	1000	Greeter	10.25	Y	23-Jun-56	20-Mar-80	30-Aug-99	30-May-13	6	40.00	0.00	236.00	Melissa Jones	306.304.8878
85833	466128562	Singh	Lindsey	1216 Willow Cres.	Pasqua	SK	S0H 5T8	7000	Pharmacist	30	Y	15-Mar-75	27-Jul-02	27-Jul-02	23-May-13	7	37.50	0.50	322.50	James Snowdale	306.304.9091
85833	466128562	Singh	Lindsey	1216 Willow Cres.	Pasqua	SK	S0H 5T8	7000	Pharmacist	30	Y	15-Mar-75	27-Jul-02	27-Jul-02	30-May-13	6	22.00	0.00	236.00	James Snowdale	306.304.9091
81216	615917448	Hansen	Jaimie	95 Lakeshore Blvd.	Caronport	SK	S0T 3S7	8000	Assistant Baker	15.5	Y	04-Mar-83	21-May-02	21-May-02	23-May-13	7	40.00	0.00	322.50	Joseph Herbert	306.304.1212
32177	306114858	DaSilva	Robbie	4319 Main St.	Moose Jaw	SK	S6H 2M2	4000	Baker	17.5	Y	18-Feb-51	07-Jul-83	15-Sep-92	23-May-13	7	40.00	3.70	322.50	Joseph Herbert	306.304.1212
32177	306114858	DaSilva	Robbie	4319 Main St.	Moose Jaw	SK	S6H 2M2	4000	Baker	17.5	Y	18-Feb-51	07-Jul-83	15-Sep-92	30-May-13	6	40.00	2.25	236.00	Joseph Herbert	306.304.1212
52421	936654021	O'Day	Shelley	27 High St.	Tuxford	SK	S0L 8V6	6000	Cleaner	13.5	Y	31-Jul-63	08-Nov-97	08-Nov-97	23-May-13	7	22.00	0.00	322.50	Abu Muktadir	306.304.4545
52421	936654021	O'Day	Shelley	27 High St.	Tuxford	SK	S0L 8V6	6000	Cleaner	13.5	Y	31-Jul-63	08-Nov-97	08-Nov-97	30-May-13	6	40.00	4.50	236.00	Abu Muktadir	306.304.4545

72690	655971502	Wong	Jodie	59 Oslo Square	Moose Jaw	SK	S6H 2H9	6000	Cleaner	13.5	N	01-Jan-87	26-Aug-03	26-Aug-03	23-May-13	7	36.00	0.00	322.50	Abu Muktadir	306.304.4545
72201	635111876	Ramirez	Kelly	1015 Brunswick Lane	Moose Jaw	SK	S6H 4T5	3000	Stockperson	12.99	N	29-Sep-86	26-Aug-03	26-Aug-03	30-May-13	6	18.00	0.00	236.00	Abu Muktadir	306.304.4545

Second Normal Form (2NF)

In Second Normal Form(2NF), all non-attributes in the table must be dependent entirely on the Primary key and all the partial dependencies must be removed.

- ✓ In Employee table EmpId is the primary key and all other attributes (SIN, Last, First, Street, City, Prov, Postal, Job Code, Position, Inc. Tax, Birth Date, Hire Date, Supervisor, Supervisor Cell#) are dependent on the EmpId.
- ✓ In Jobs table Job Code is the primary key and all other attributes (Position, Payrate CAD, Inc. Tax) are dependent on the EmpId.
- ✓ In Committee table CommitteeID is the primary key and all other attributes (Committee Name, Meeting Night) are dependent on the EmpId.
- ✓ In Employee to Committee table EmpComID is the primary key and. EmpId and Committee Id together serve as a composite key.By combining these two attributes together, we can uniquely identify each record in the "Employee to Committee" table. This is because an employee can serve on multiple committees, and each committee can have multiple employees associated with it.
- ✓ In Payroll table PayId is the primary key and all other attribites (Hours, OT) are dependent on the PayId. With EmpId in the table, it establishes relationship with Employee Table indicating which employee the payroll entry corresponds to.For EmpId 97319 there are no information in Pay Week End Date, Hours & OT columns so we delete this row.
- ✓ In Pay Week End Date table PayWeekEndId is the primary key and all the other attributes (Pay Week End Date,Days Available) are dependent on the PayWeekEndId.

Employees

EmpId	SIN	Last	First	Street	City	Prov	Postal	Job Code	Position	Inc. Tax	Birth Date	Hire Date	Supervisor	Supervisor Cell#
97319	516303417	Novak	Gerry	6803 Park Ave.	Moose Jaw	SK	S6H 1X7	3000	Stockperson	N	24-Aug-86	07-Jul-03	Abu Muktadir	306.304.4545
33982	867481381	Boychuk	Robin	117 East Broadway	Moose Jaw	SK	S6H 3P5	5000	Butcher	Y	04-Mar-71	11-Oct-98	Joseph Herbert	306.304.1212
51537	112893584	Smith	Kim	9745 University Drive	Regina	SK	S4P 7A3	2000	Cashier	Y	29-Nov-82	02-Dec-01	Melissa Jones	306.304.8878
41822	717505366	Miller	Chris	72 Railway Ave.	Pense	SK	S0T 1K4	2000	Cashier	Y	15-Nov-68	19-Feb-85	Melissa Jones	306.304.8878
3571	374853129	Hashimoto	Jo	386 High Street	Tuxford	SK	S0L 8V6	1000	Greeter	Y	23-Jun-56	20-Mar-80	Melissa Jones	306.304.8878
85833	466128562	Singh	Lindsey	1216 Willow Cres.	Pasqua	SK	S0H 5T8	7000	Pharmacist	Y	15-Mar-75	27-Jul-02	James Snowdale	306.304.9091
81216	615917448	Hansen	Jaimie	95 Lakeshore Blvd.	Caronport	SK	S0T 3S7	8000	Assistant Baker	Y	04-Mar-83	21-May-02	Joseph Herbert	306.304.1212
32177	306114858	DaSilva	Robbie	4319 Main St.	Moose Jaw	SK	S6H 2M2	4000	Baker	Y	18-Feb-51	07-Jul-83	Joseph Herbert	306.304.1212
52421	936654021	O'Day	Shelley	27 High St.	Tuxford	SK	S0L 8V6	6000	Cleaner	Y	31-Jul-63	08-Nov-97	Abu Muktadir	306.304.4545
72690	655971502	Wong	Jodie	59 Oslo Square	Moose Jaw	SK	S6H 2H9	6000	Cleaner	N	01-Jan-87	26-Aug-03	Abu Muktadir	306.304.4545
72201	635111876	Ramirez	Kelly	1015 Brunswick Lane	Moose Jaw	SK	S6H 4T5	3000	Stockperson	N	29-Sep-86	26-Aug-03	Abu Muktadir	306.304.4545

Jobs

Job Code	Position	Payrate CAD	Inc. Tax
3000	Stockperson	12.99	N
5000	Butcher	18	Y
2000	Cashier	11.99	Y
1000	Greeter	10.25	Y
7000	Pharmacist	30	Y
8000	Assistant Baker	15.5	Y
4000	Baker	17.5	Y
6000	Cleaner	13.5	Y
6000	Cleaner	13.5	N

Committee

Committee Id	Committee Name	Meeting Night
1	OH&S	Fri
2	Party Committee	Wed
3	Social Res. Com.	Mon

Employee to Committee

EmpComID	EmpID	Committee Id
1	97319	1
2	72201	1
3	33982	1
4	32177	1
5	72690	2
6	51537	2
7	81216	2
8	41822	2
9	41822	3
11	72690	3
12	97319	3

Payroll

PayId	EmpId	Hours	OT
1	33982	40.00	0.00
2	33982	38.25	0.00
3	51537	27.00	0.00
4	41822	40.00	0.00
5	41822	38.00	1.25
6	3571	40.00	0.00
7	3571	40.00	0.00
8	85833	37.50	0.50
9	85833	22.00	0.00
10	81216	40.00	0.00
11	32177	40.00	3.70
12	32177	40.00	2.25
13	52421	22.00	0.00
14	52421	40.00	4.50
15	72690	36.00	0.00
16	72201	18.00	0.00

Pay WeekendDate

PayWeekendID	Pay Week End Date	Days Available
1	23-May-13	7
2	30-May-13	6

Third Normal Form (3NF)

- ✓ In Committee table CommitteeID is the primary key and all other attributes (Committee Name, Meeting Night) are dependent on the EmpId.
- ✓ In Employee to Committee table EmpComID is the primary key and. EmpId and Committee Id together serve as a composite key.By combining these two attributes together, we can uniquely identify each record in the "Employee to Committee" table. This is because an employee can serve on multiple committees, and each committee can have multiple employees associated with it.
- ✓ In Payroll table PayId is the primary key and all other attribites (Hours, OT) are dependent on the PayId. With PayWeekDayId in the table, it indicates the specific pay week end date for each payroll entry and each employee.

- ✓ In Pay Week End Date table PayWeekEndId is the primary key and all the other attributes (Pay Week End Date,Days Available) are dependent on the PayWeekEndId. In Third Normal Form(3NF), all non-attributes in the table must be dependent only on the Primary key and there should be no functional dependency between non-key attributes.
- ✓ In Employee table EmpId is the primary key and all other attributes (SIN, Last, First, Street, City, Prov, Postal, Job Code, Position, Inc. Tax, Birth Date, Hire Date, Supervisor, Supervisor Cell#,SupervisorId) are dependent on the EmpId.Here SupervisorId column connects the Employee table to Supervisor Table(1 to many relation).
- ✓ In Supervisor table SupervisorId is the primary key and all other attributes (Supervisor, Supervisor Cell#) are dependent on the SupervisorId.
- ✓ In Jobs table Job Code is the primary key and all other attributes (Position, Payrate CAD, Inc. Tax) are dependent on the EmpId.
- ✓ In Committee table CommitteeID is the primary key and all other attributes (Committee Name, Meeting Night) are dependent on the EmpId.
- ✓ In Employee to Committee table EmpComID is the primary key and. EmpId and Committee Id together serve as a composite key.By combining these two attributes together, we can uniquely identify each record in the "Employee to Committee" table. This is because an employee can serve on multiple committees, and each committee can have multiple employees associated with it.
- ✓ In Payroll table PayId is the primary key and all other attribites (Hours, OT) are dependent on the PayId. With PayWeekDayId in the table, it indicates the specific pay week end date for each payroll entry and each employee.
- ✓ In Pay Week End Date table PayWeekEndId is the primary key and all the other attributes (Pay Week End Date,Days Available) are dependent on the PayWeekEndId.

Employees **Primary key (EmpId), foreign key (Job Code,SupervisorId)**

EmpId	SIN	Last	First	Street	City	Prov	Postal	Job Code	Inc. Tax	Birth Date	Hire Date	SupervisorID
97319	516303417	Novak	Gerry	6803 Park Ave.	Moose Jaw	SK	S6H 1X7	3000	N	24-Aug-86	07-Jul-03	1
33982	867481381	Boychuk	Robin	117 East Broadway	Moose Jaw	SK	S6H 3P5	5000	Y	04-Mar-71	11-Oct-98	2
51537	112893584	Smith	Kim	9745 University Drive	Regina	SK	S4P 7A3	2000	Y	29-Nov-82	02-Dec-01	3
41822	717505366	Miller	Chris	72 Railway Ave.	Pense	SK	S0T 1K4	2000	Y	15-Nov-68	19-Feb-85	3
3571	374853129	Hashimoto	Jo	386 High Street	Tuxford	SK	S0L 8V6	1000	Y	23-Jun-56	20-Mar-80	3
85833	466128562	Singh	Lindsey	1216 Willow Cres.	Pasqua	SK	S0H 5T8	7000	Y	15-Mar-75	27-Jul-02	4
81216	615917448	Hansen	Jaimie	95 Lakeshore Blvd.	Caronport	SK	S0T 3S7	8000	Y	04-Mar-83	21-May-02	2
32177	306114858	DaSilva	Robbie	4319 Main St.	Moose Jaw	SK	S6H 2M2	4000	Y	18-Feb-51	07-Jul-83	2
52421	936654021	O'Day	Shelley	27 High St.	Tuxford	SK	S0L 8V6	6000	Y	31-Jul-63	08-Nov-97	1
72690	655971502	Wong	Jodie	59 Oslo Square	Moose Jaw	SK	S6H 2H9	6000	N	01-Jan-87	26-Aug-03	1
72201	635111876	Ramirez	Kelly	1015 Brunswick Lane	Moose Jaw	SK	S6H 4T5	3000	N	29-Sep-86	26-Aug-03	1

Supervisor — Primary key (SupervisorId)

SupervisorID	Supervisor	Supervisor Cell#
1	Abu Muktadir	306.304.4545
2	Joseph Herbert	306.304.1212
3	Melissa Jones	306.304.8878
4	James Snowdale	306.304.9091

Committee — Primary key (CommitteeId)

Committee Id	Committee Name	Meeting Night
1	OH&S	Fri
2	Party Committee	Wed
3	Social Res. Com.	Mon

Employee to Committee — Primary key (EmpComId)

EmpComID	EmpID	Committee Id
1	97319	1
2	72201	1
3	33982	1
4	32177	1
5	72690	2
6	51537	2
7	81216	2
8	41822	2
9	41822	3
10	32177	3
11	72690	3
12	97319	3

Jobs — Primary key (Job Code)

Job Code	Position	Payrate CAD
3000	Stockperson	12.99
5000	Butcher	18
2000	Cashier	11.99
1000	Greeter	10.25
7000	Pharmacist	30
8000	Assistant Baker	15.5
4000	Baker	17.5
6000	Cleaner	13.5

Payroll — Primary key (PayId), foreign key (EmpId,PayWeekEndId)

PayId	EmpId	PayWeekEndID	Hours	OT
1	33982	1	40.00	0.00
2	33982	2	38.25	0.00
3	51537	1	27.00	0.00
4	41822	1	40.00	0.00
5	41822	2	38.00	1.25
6	3571	1	40.00	0.00
7	3571	2	40.00	0.00
8	85833	1	37.50	0.50
9	85833	2	22.00	0.00
10	81216	1	40.00	0.00
11	32177	1	40.00	3.70
12	32177	2	40.00	2.25
13	52421	1	22.00	0.00
14	52421	2	40.00	4.50
15	72690	1	36.00	0.00
16	72201	2	18.00	0.00

Pay Week End Date — Primary key (PayWeekEndId)

PayWeekEndID	Pay Week End Date	Days Available
1	23-May-13	7
2	30-May-13	6

Q & A –

1. A Normalization section with second and third normal forms diagrammed out or using set notation should answer the following four questions/concerns:

- There is a concern that the spreadsheet will get out of hand as the number of pay periods increases; how does your 3NF solve this issue?

  -In our 3NF, we have made PayWeekEndDate table separately with PayWeekEndId, Pay week end date and Days available attributes. If the number of periods increases, we can add new rows into this table. Additionally, this table has linked with Payroll table which has EmdId and PayWeekEndId as its foreign keys. These two tables have one- to- many relationship as one PayWeekEndDate can be similar for many Employees. So, by these 2 tables we can get the information regarding the Employees if the number of pay periods increases.

- Supervisors often manage several departments; for example, Joseph Herbert manages butchers and bakers. How did you address this problem?

  -In our Employees table we have Job Code feature as foreign key and in Job table Job code is the primary key. So these 2 tables have one-to-many relationship as one employee can be assigned to one job but one job can be assigned to many employees. Similarly, In our Employees table we have SupervisorId feature as foreign key and in Supervisor table SupervisorId is the primary key. So these 2 tables have one-to-many relationship as one employee can be assigned to one supervisor but one supervisor can have many employees under him. Therefore, in case we need information about which supervisor managing several departments we can get from these 3 tables.

- Employees often serve on multiple committees at once. How did you address this?

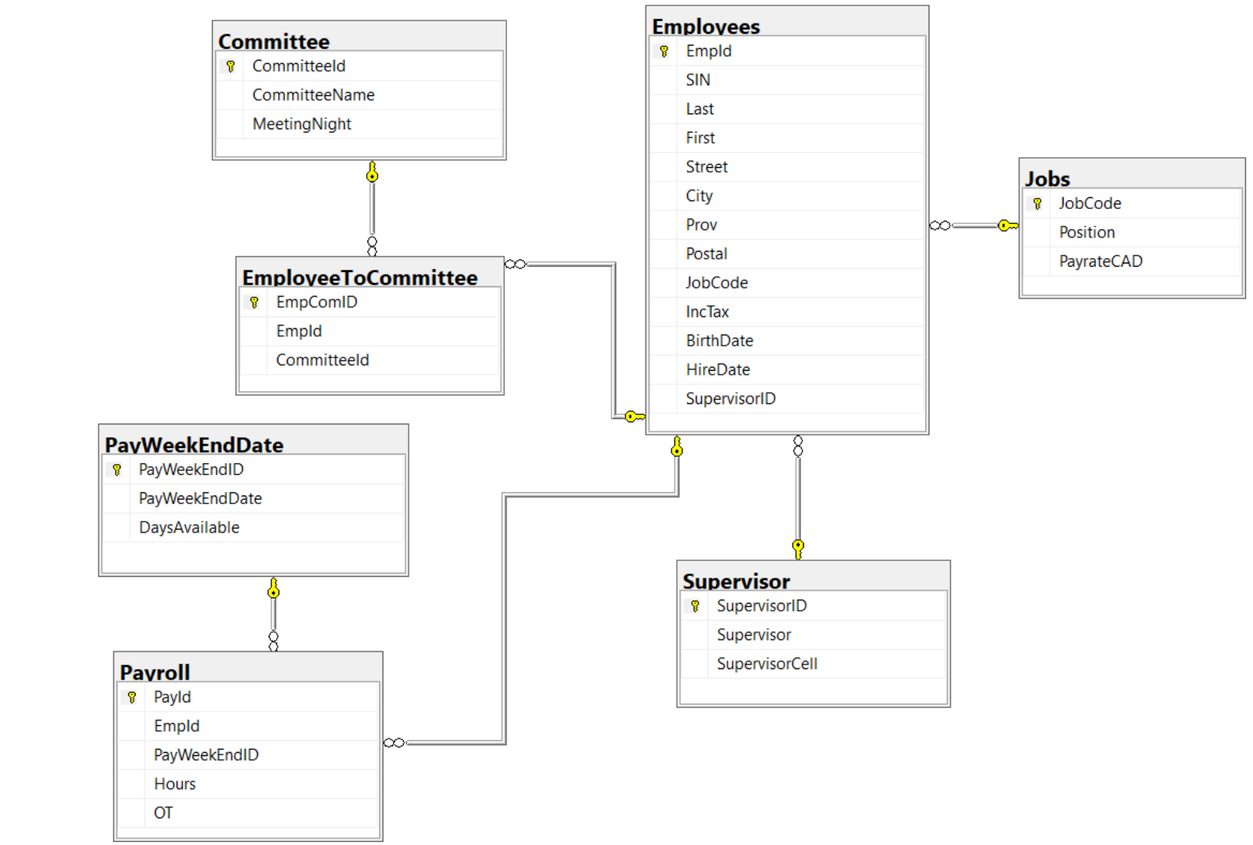
  -In our EmployeesTocommitee table we have EmpId, CommitteeId which connects the Employees and the Committee tables. easily know which employee is serving which committee. In case of Employees serving multiple committees, we can get these information easily from this table. These tables have many-to-many relationships as each employee can be associated with multiple committees, and each committee can have multiple employees. Similarly, each employee may participate in multiple committees, and each committee may have several employees involved. So, from these tables we can easily know Employees who are working on multiple committees at once.

- 'Person hours worked' is calculated (it represents information, not data) for all hours worked by all employees for that week including overtimes. Is there a better way to handle this rather than recording it in the database? Implement your solution in the queries section.

  -'Person hours worked' is a data which calculates the total amount of time worked by the employee including the overtime within a specific period. In our 3NF, we have Payroll and PayWeekEnd tables linked by the PayWeekEndId column. If we need to find out the information for those pay periods, we can use SUM () function and calculate the total hours worked as we have payroll table which contains (Hours+OT) for each employee.

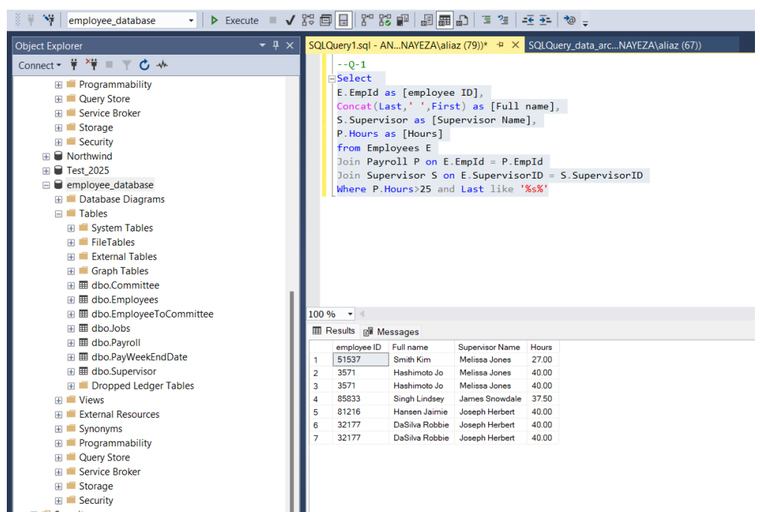
2. An ERD section that will include a full physical ERD of your database. The ERD diagram should clearly describe PK, FK, and 1-to-1, 1-to-many, and many-to-many relationships using Crowfoot notations (and/or) text comments.

-Here is ERD diagram

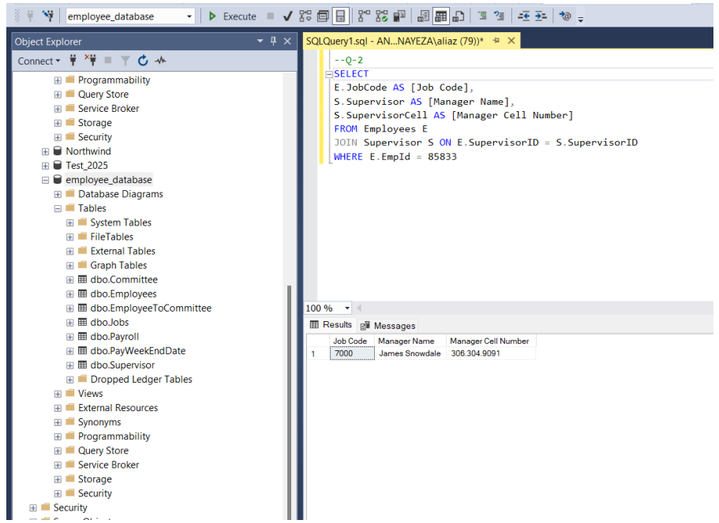


3. A queries section that will include a screen snippet (an example of a screen snippet is shown below) with the answer and the SQL code to answer the following questions:

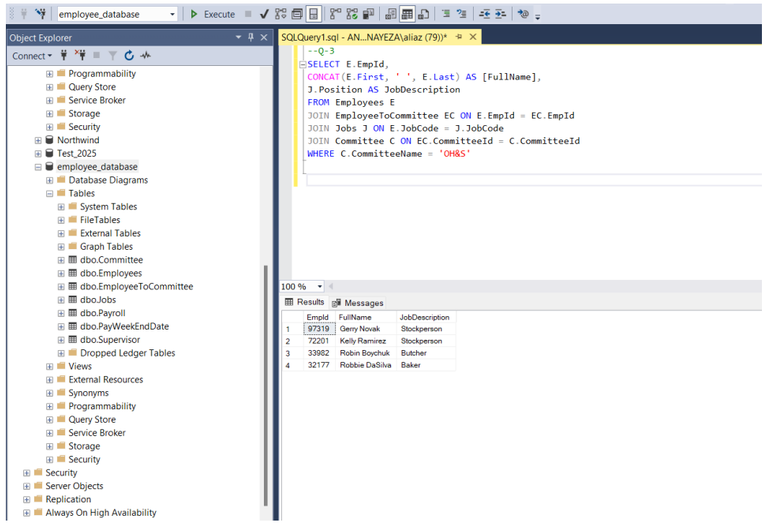
Q1. Who worked more than 25 hours in any pay period and the employee's last name include the character 's'? Please print the employee's [employee ID], [Full name] (Last name + First Name) as 'Full Name', [Supervisor Name], [Hours].



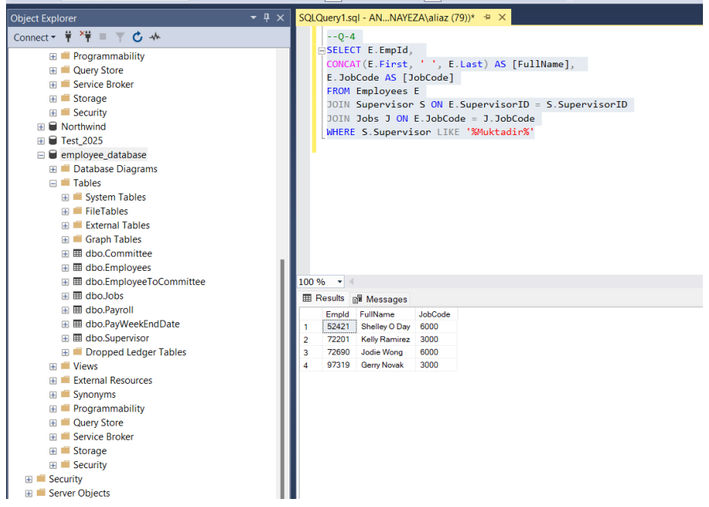
Q2. I want to enter an employee's ID and find their job code, manager's name, and cell number. (No duplicate records)



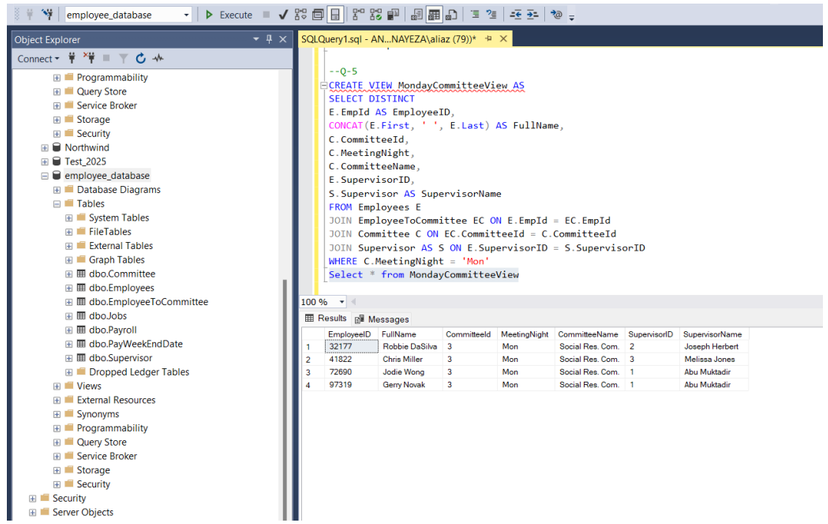
Q3. Given a committee name 'OH&S', I'd like a list of people who work on it and their job description.



Q4. Given a supervisor's last name, e.g., 'Muktadir', provide a list of employees being supervised and their job codes.



Q5. Create a view to show the employee ID, employee's full name, committee ID, meeting night, committee name, supervisor id and supervisor's name, where the condition is the meeting night is 'Tues'. No duplicate records.



Conclusion-

To sum up, the process of normalization has produced an organized and optimized database schema that can effectively handle information on employees, jobs, payroll, committees, supervisors, and pay weeks end. By splitting the data into several tables and applying normalization principles to its organization, we have reduced redundancy, maintained data integrity, and enhanced database performance as a whole. The database is flexible and extensible for future updates and additions due to the development of suitable connections and the use of dynamic calculations for specific data. This project demonstrates the importance of database normalization in designing robust and efficient database systems.