|  |  |
| --- | --- |
|  | SQL-Mongo Project  Employee Attrition  BUAN 6320  Submitted by:  Tue4PM\_Group 4  Ashu Rawat  Christopher Kupovics  Priyanka Mishra  Rajat Mishra |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Activity** | **Ashu Rawat** | **Chris Kupovics** | **Priyanka Mishra** | **Rajat Mishra** |
| Prepared Data Model and Created Physical DB | x | x |  |  |
| Loaded Data into Database |  |  | x | x |
| Wrote SQL Queries | x | x | x | x |
| Prepared Mongo Database | x | x | x | x |
| Loaded data into Mongo DB | x | x |  |  |
| Wrote Mongo Queries |  |  | x | x |
| Prepared Report | x | x | x | x |
| Reviewed Report | x | x | x | x |

Contents

[Relational Data Model 5](#_Toc39345323)

[**Assumptions/Notes About Data Entities and Relationships** 5](#_Toc39345324)

[**Entity-Relationship Diagram** 6](#_Toc39345325)

[Assumptions/Notes About Data Set 7](#_Toc39345326)

[Screen shot of Physical Database objects 7](#_Toc39345327)

[**Data in the Database** 16](#_Toc39345328)

[SQL Queries 17](#_Toc39345329)

[SQL Query 1 17](#_Toc39345330)

[Question 17](#_Toc39345331)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 17](#_Toc39345332)

[Translation 17](#_Toc39345333)

[Screen Shot of SQL Query and Results 17](#_Toc39345334)

[SQL Query 2 18](#_Toc39345335)

[Question 18](#_Toc39345336)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 18](#_Toc39345337)

[Translation 18](#_Toc39345338)

[Cleanup 18](#_Toc39345339)

[Screen Shot of SQL Query and Results 18](#_Toc39345340)

[SQL Query 3 19](#_Toc39345341)

[Question 19](#_Toc39345342)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 19](#_Toc39345343)

[Translation 19](#_Toc39345344)

[Screen Shot of SQL Query and Results 19](#_Toc39345345)

[SQL Query 4 20](#_Toc39345346)

[Question 20](#_Toc39345347)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 20](#_Toc39345348)

[Translation 20](#_Toc39345349)

[Cleanup 20](#_Toc39345350)

[Screen Shot of SQL Query and Results 20](#_Toc39345351)

[SQL Query 5 21](#_Toc39345352)

[Question 21](#_Toc39345353)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 21](#_Toc39345354)

[Translation 21](#_Toc39345355)

[Cleanup 21](#_Toc39345356)

[Screen Shot of SQL Query and Results 21](#_Toc39345357)

[SQL Query 6 22](#_Toc39345358)

[Question 22](#_Toc39345359)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 22](#_Toc39345360)

[Translation 22](#_Toc39345361)

[Cleanup 22](#_Toc39345362)

[Screen Shot of SQL Query and Results 23](#_Toc39345363)

[Physical Mongo Database 24](#_Toc39345364)

[Assumptions/Notes About Data Set 24](#_Toc39345365)

[Screen shot of Physical Database objects (Database, Collections and Attributes) 25](#_Toc39345366)

[Data in the Database 33](#_Toc39345367)

[MongoDB Queries/Code 34](#_Toc39345368)

[Mongo Query 1 34](#_Toc39345369)

[Question 34](#_Toc39345370)

[Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 34](#_Toc39345371)

[Translation 34](#_Toc39345372)

[Screen Shot of MongoDB Query/Code and Results 34](#_Toc39345373)

[Mongo Query 2 35](#_Toc39345374)

[Question 35](#_Toc39345375)

[Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 35](#_Toc39345376)

[Translation 35](#_Toc39345377)

[Screen Shot of MongoDB Query/Code and Results 35](#_Toc39345378)

[Mongo Query 3 36](#_Toc39345379)

[Question 36](#_Toc39345380)

[Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 36](#_Toc39345381)

[Translation 36](#_Toc39345382)

[Screen Shot of MongoDB Query/Code and Results 36](#_Toc39345383)

# Relational Data Model

**Assumptions/Notes About Data Entities and Relationships**

Include assumptions about data entities and their relationships with each other.

**Relationship:** In this project employee details table is the main table in our model as all other table relate back to the employee details and thus the employee number becomes the primary key for this table. All tables are linked to the employee details tables so their primary key will also be included in the employee details tables as the foreign key. In our model each relationship between table is 1:1.

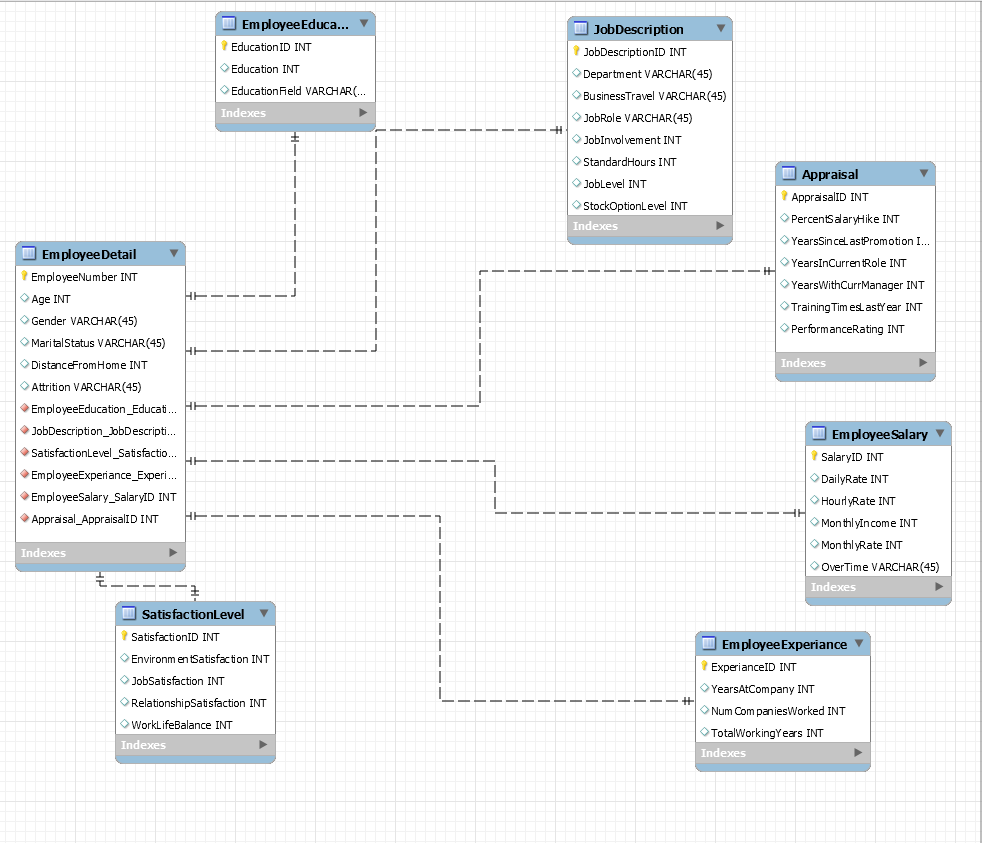
**Include reasons why the data model is in 3NF.**

1. All the rows have same number of columns and all values are singular
2. There are no multipart or multivalued columns and duplicate rows in tables.
3. Partial and transitive dependency is removed wherever it is required.

**Assumptions:**

1. Employee's work experience history would be summed up in the experience table.
2. Satisfaction table has 4 sub-sections, and each can have a rating from 1 to 4.
3. Stock option level is the relative level of stocks the employee owns. Higher the level of stock option higher the number of stocks the employee owns.
4. We are assuming each employee has its own daily and monthly rate which is used to bill client for providing their services to them.
5. Monthly income is the main salary of the employee and hourly rate will be used as an addon if employee works overtime means more than standard hours to calculate his month end payout.
6. We have assumed each employee age is over 18 and that is not going to change in future.
7. There are 6 different education fields in our database and each field has 5 levels of education that an employee can have. Higher the level of education higher the education employee has.
8. There are 9 job roles and 5 job levels. Healthcare Representatives, Sales Executives and Manufacturing Directors are at levels 2-4. Laboratory Technician, HR and Research Scientists are at levels 1-3. Managers and Research Directors are at level 3-5. Sales Representatives are at level 1-2. These levels represent the hierarchy of the chain of command in the specific roles.
9. We are assuming the there are no multiple entries of each employee.
10. We are assuming employee didn’t take the multiple surveys.
11. We have assumed employee count for each employee will be 1 and that is not going to change in futures.
12. All rates are in dollar units.

**Entity-Relationship Diagram**

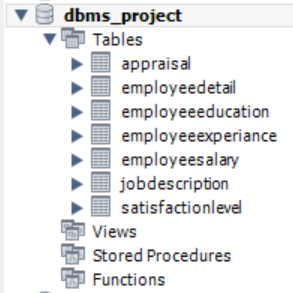


**Physical MySQL Database**

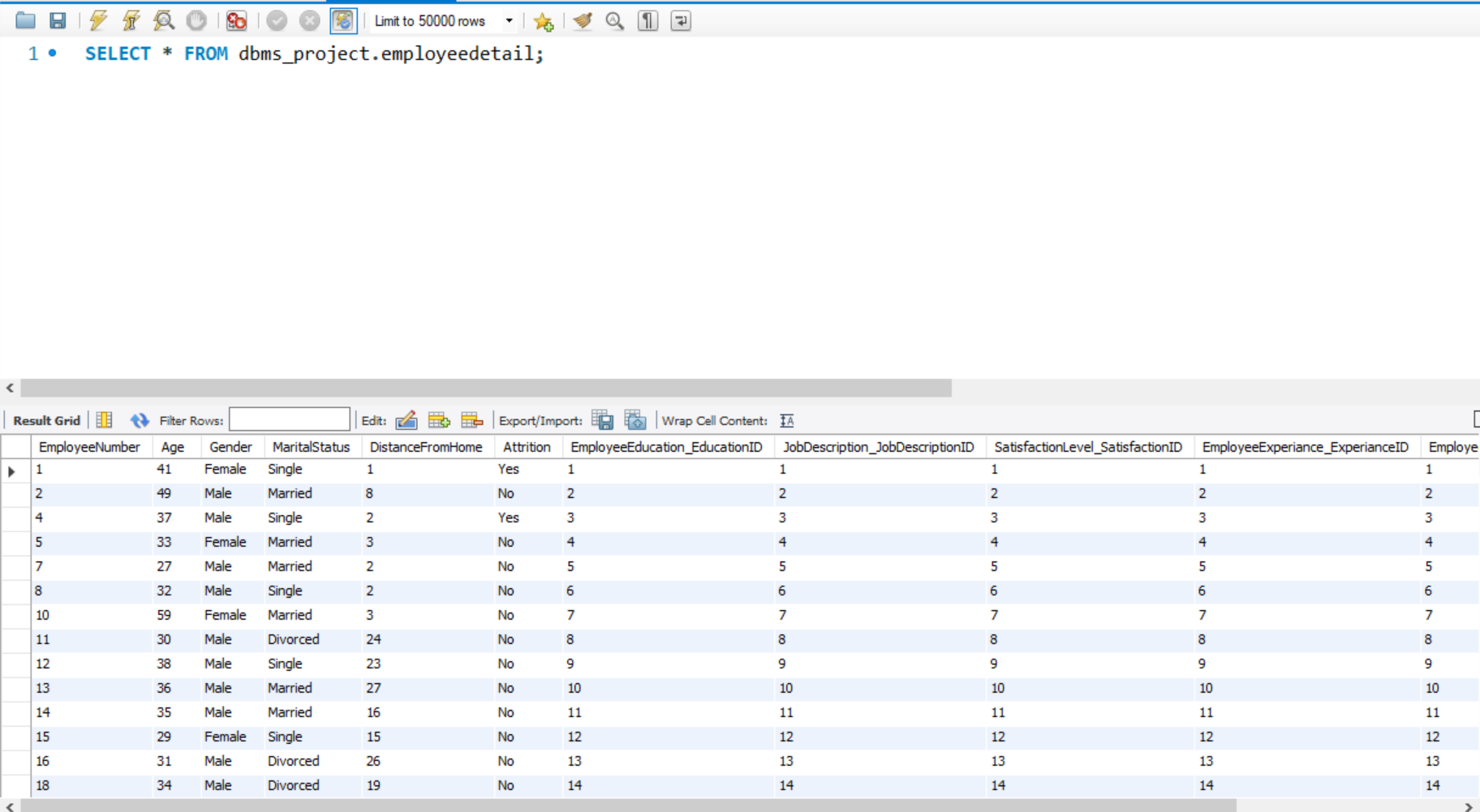
## Assumptions/Notes About Data Set

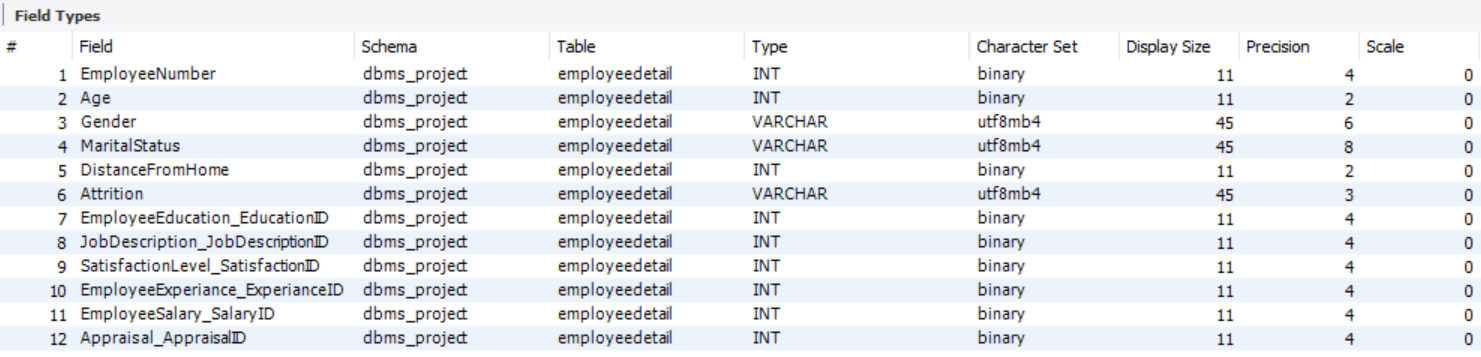
There is no empty fields or sparse data in the data set.

## Screen shot of Physical Database objects



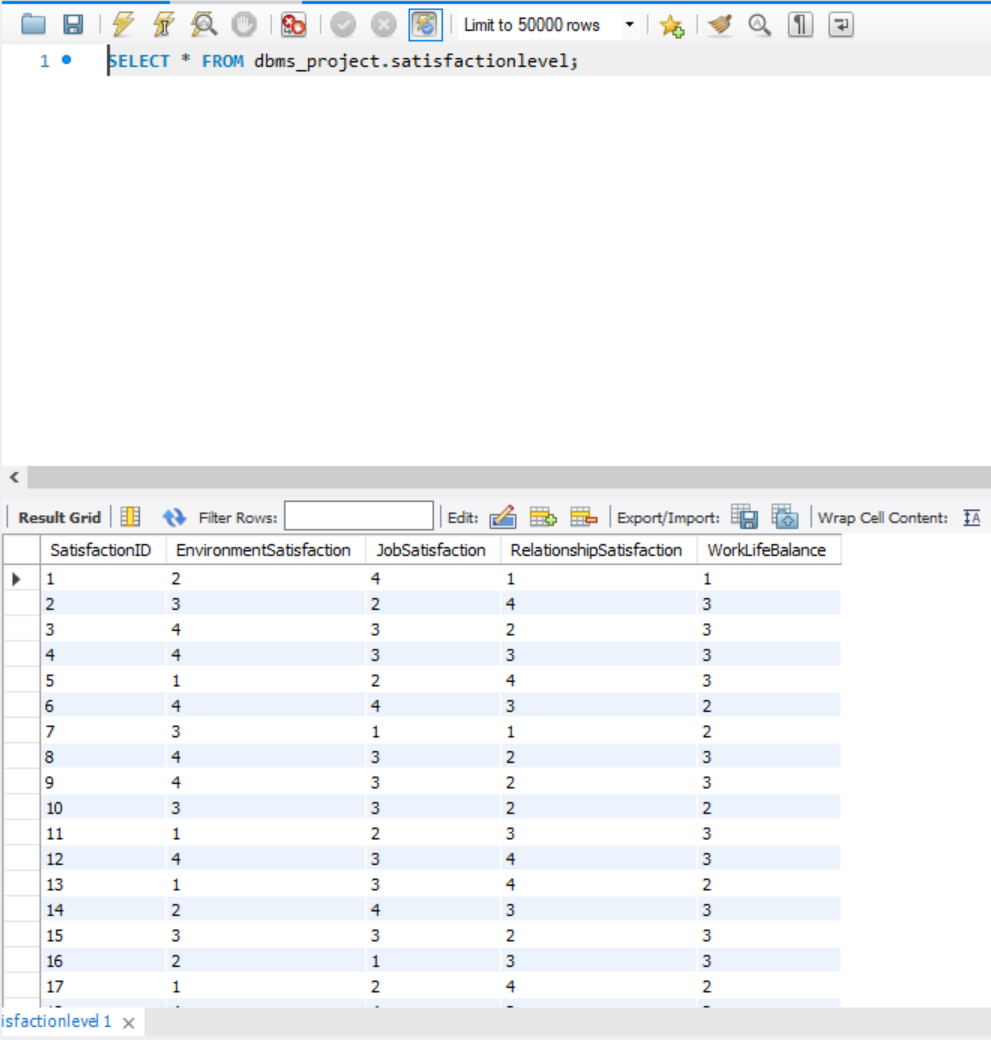
**Table 1: EmployeeDetail**

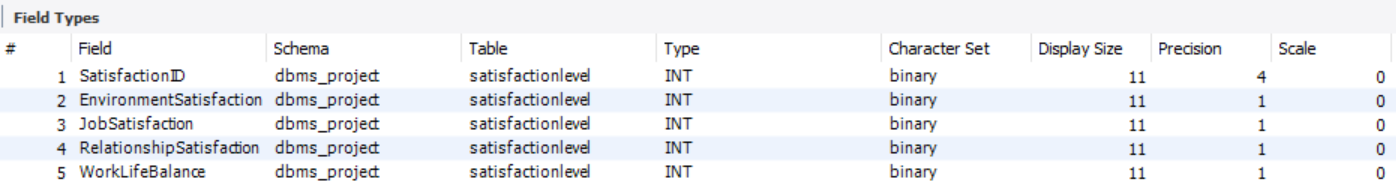






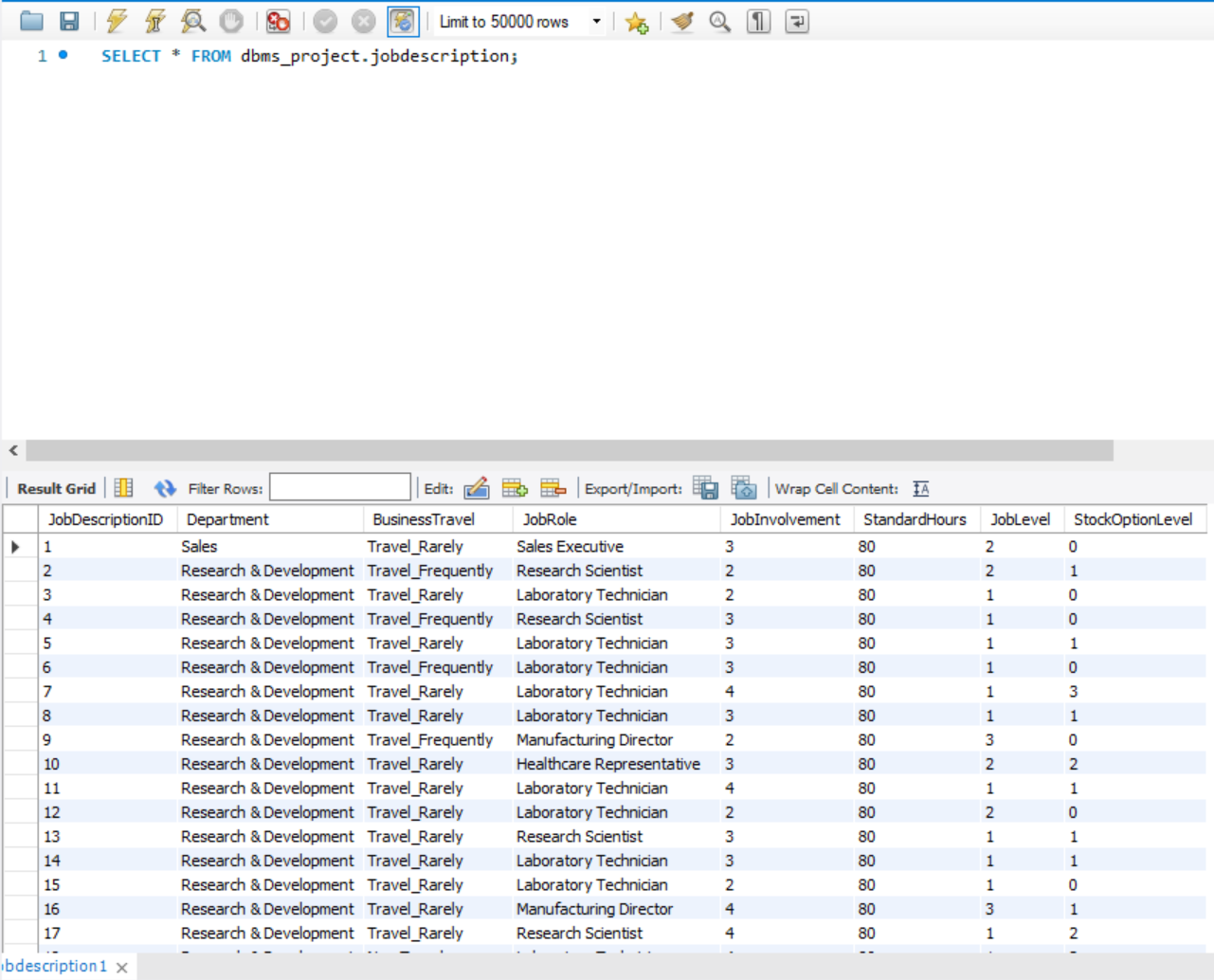
**Table 2: SatisfactionLevel**

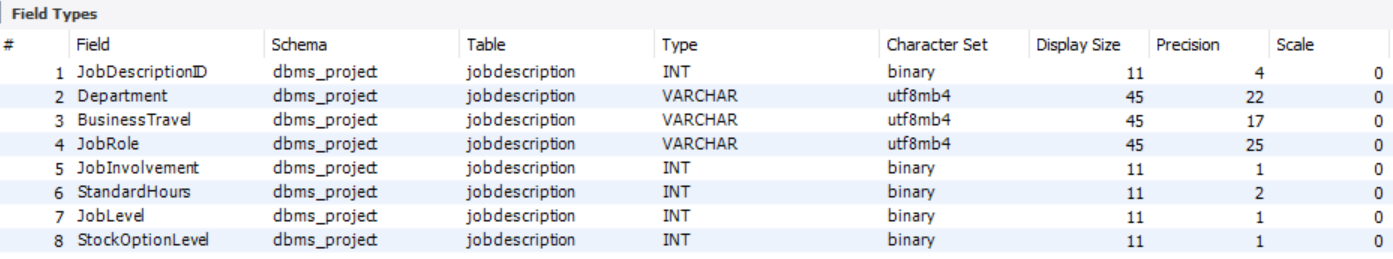


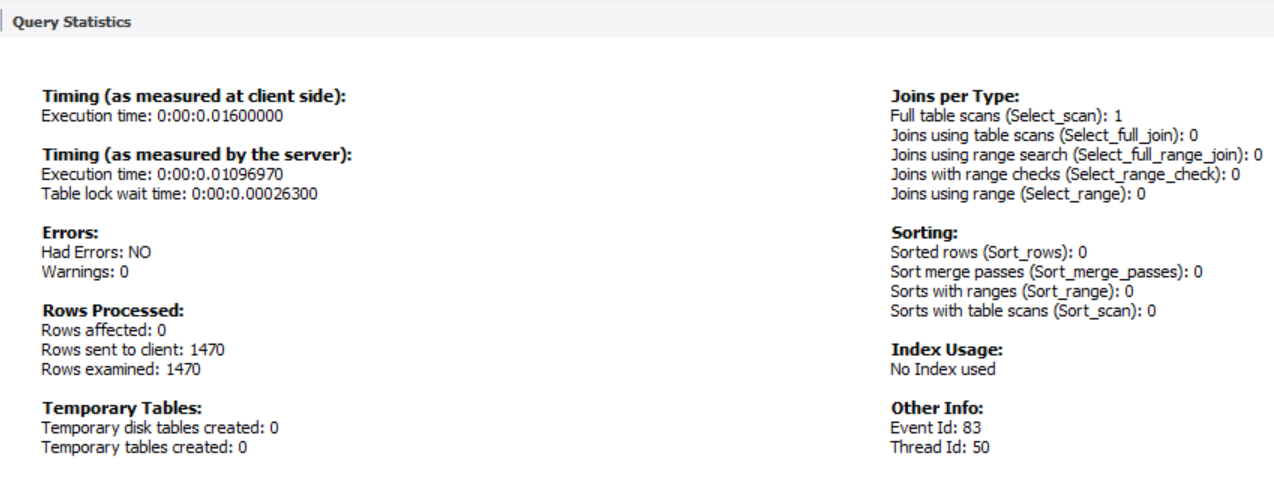




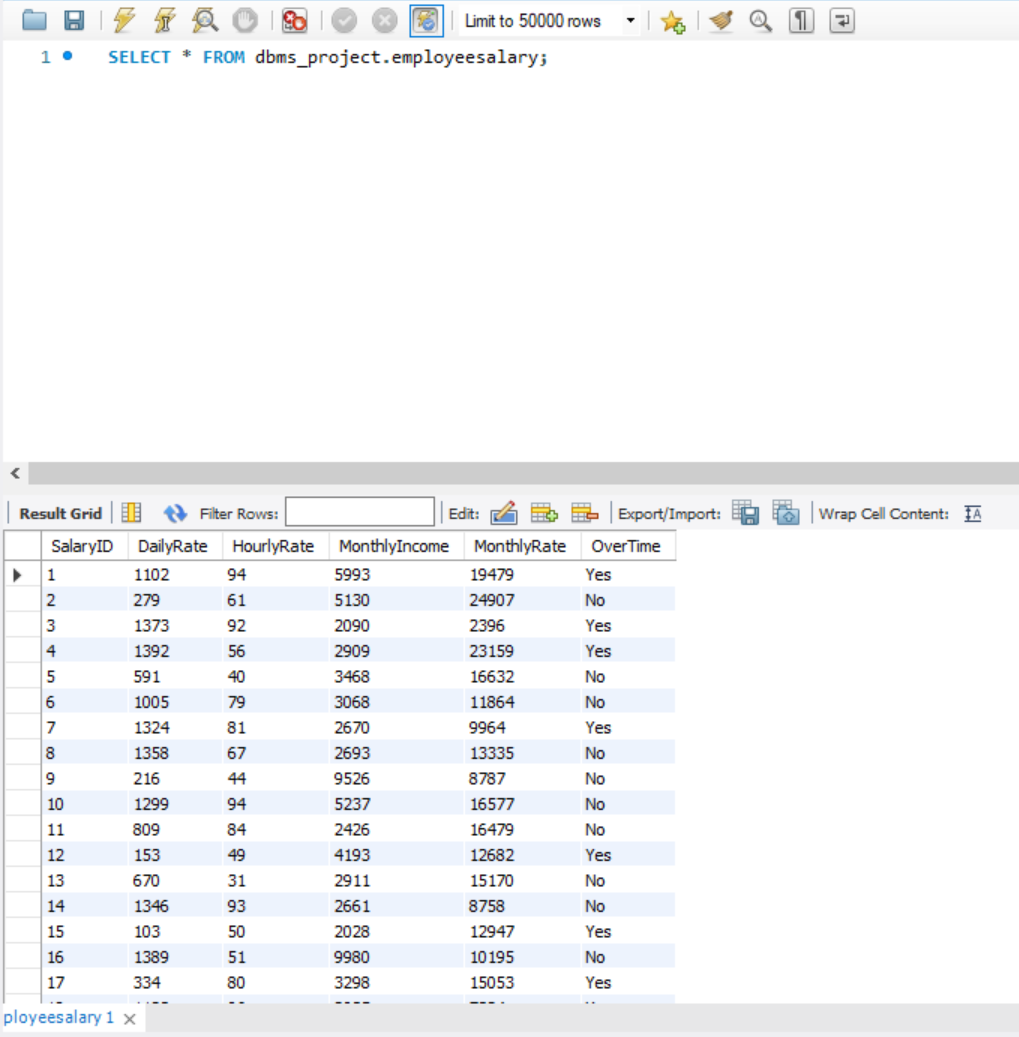
**Table 3: JobDescription**

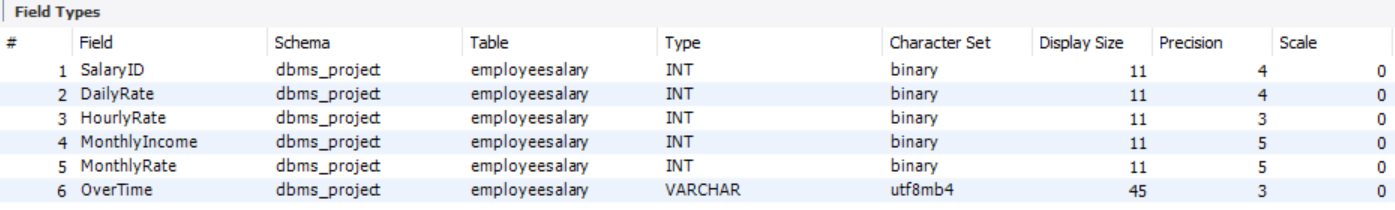


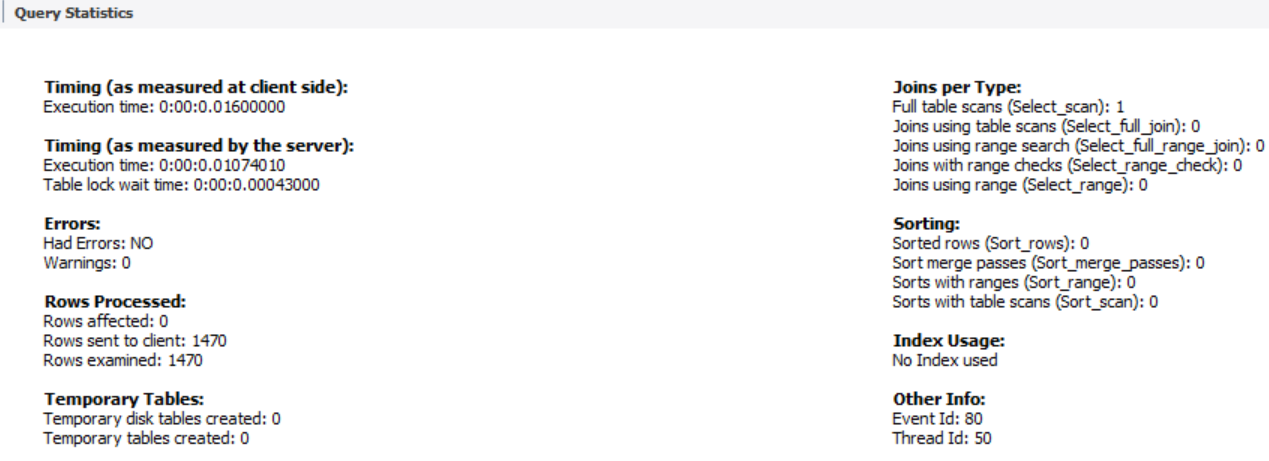




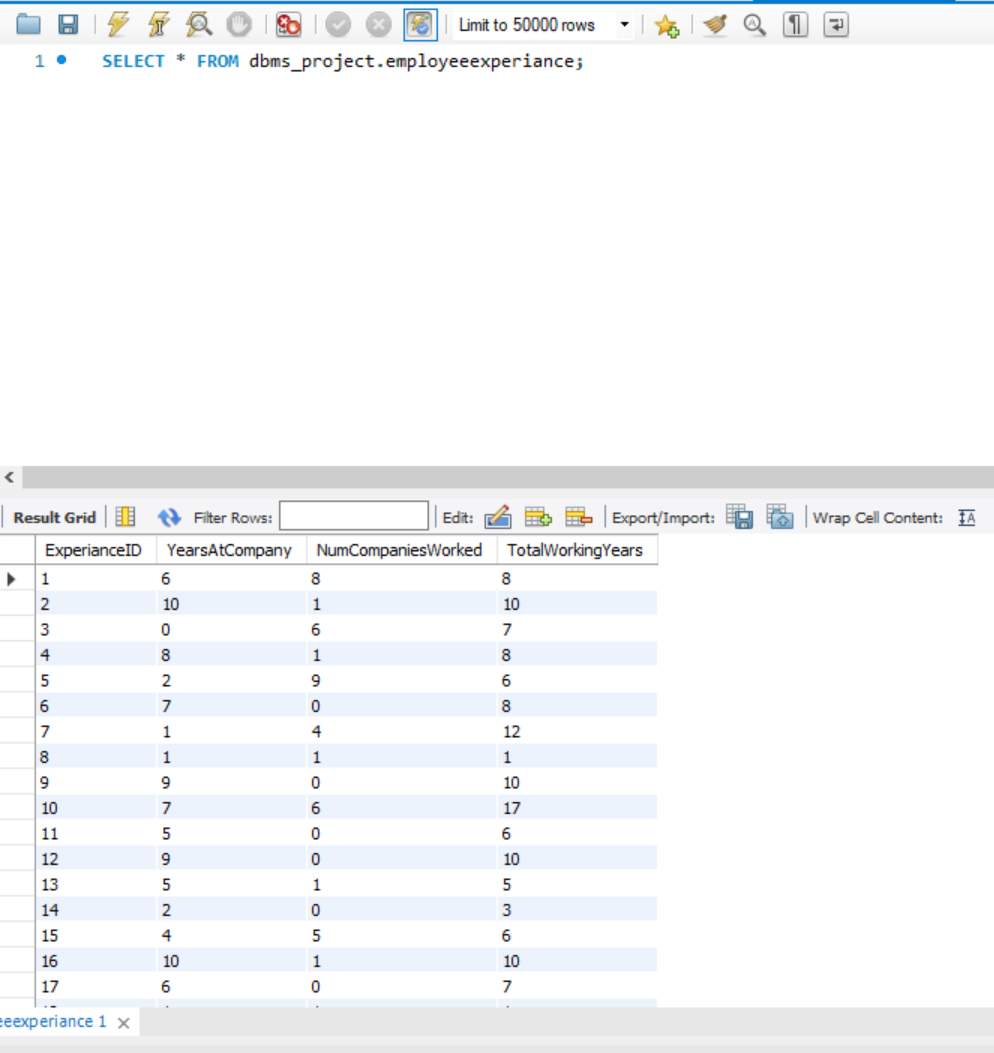
**Table 4:  Salary**

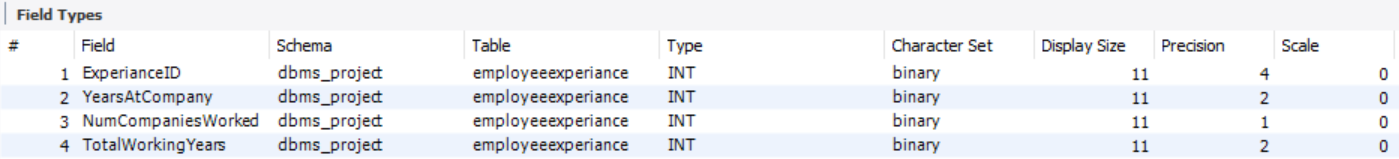






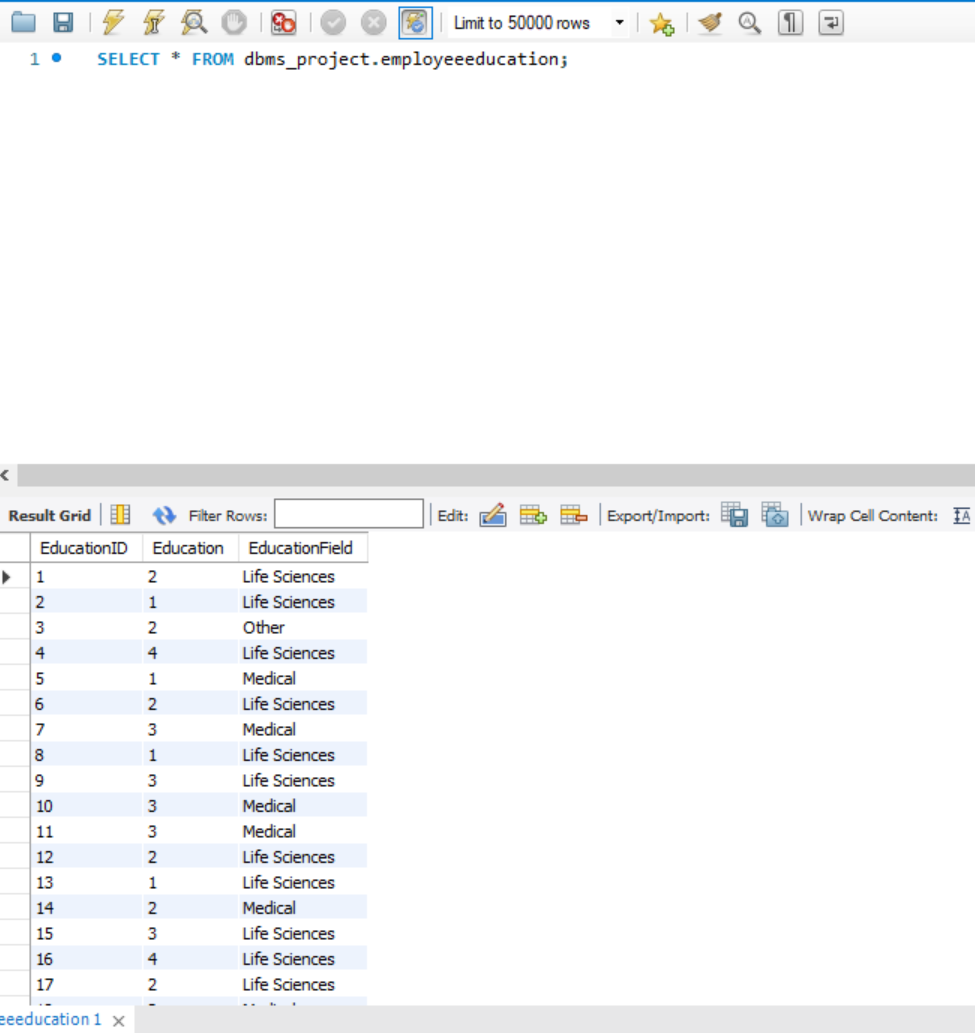
**Table 5: EmployeeExperience**

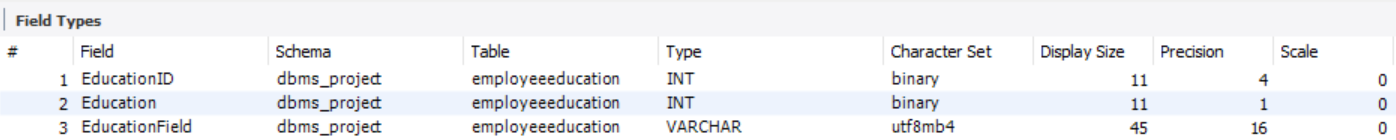


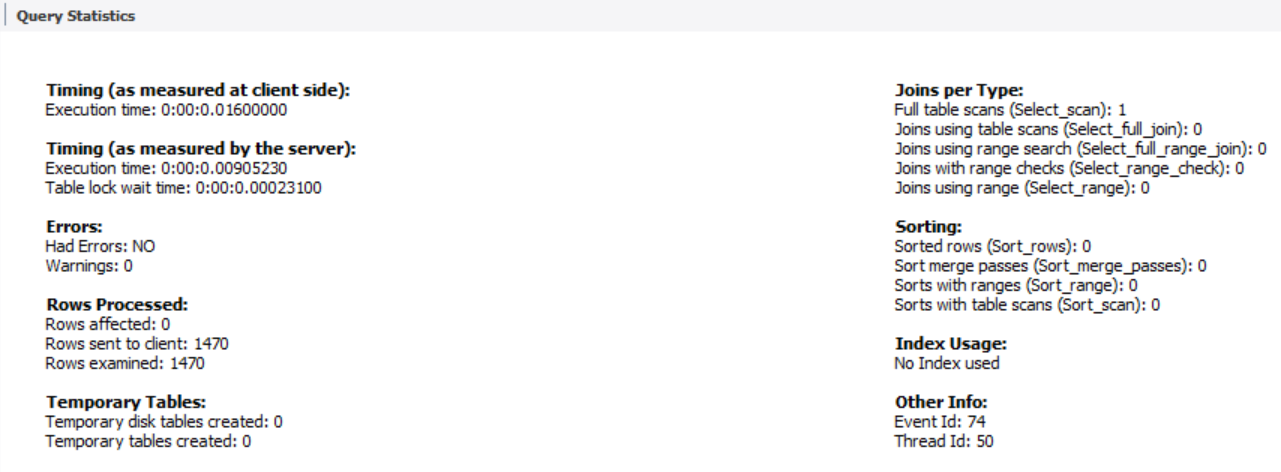




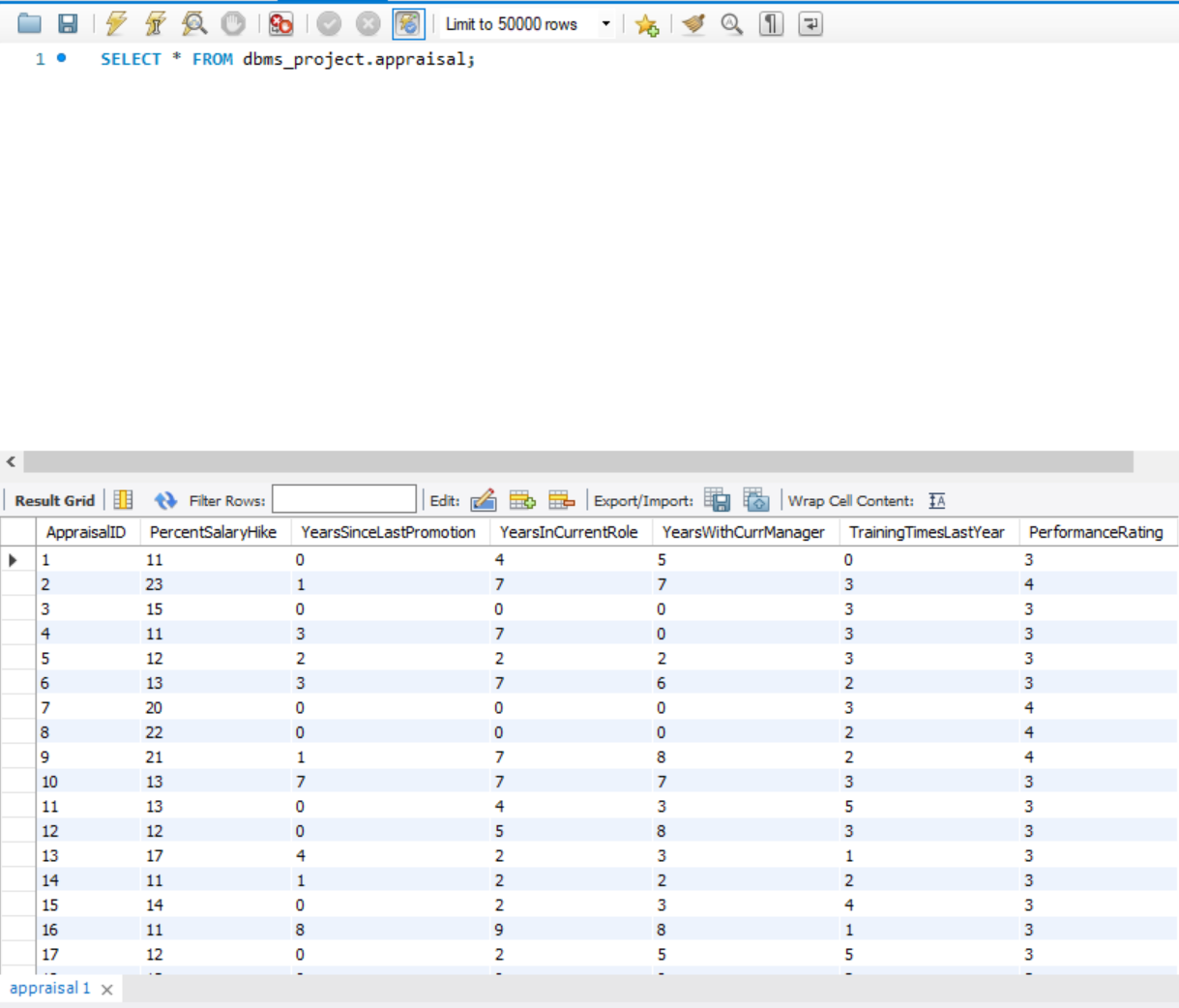
**Table 6:  EmployeeEducation**

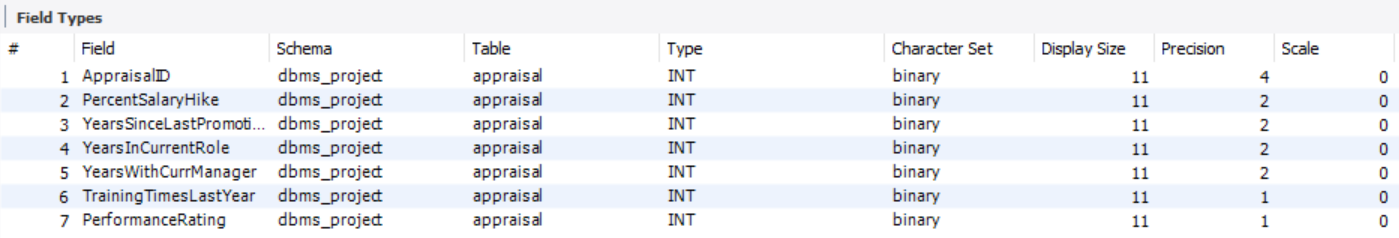


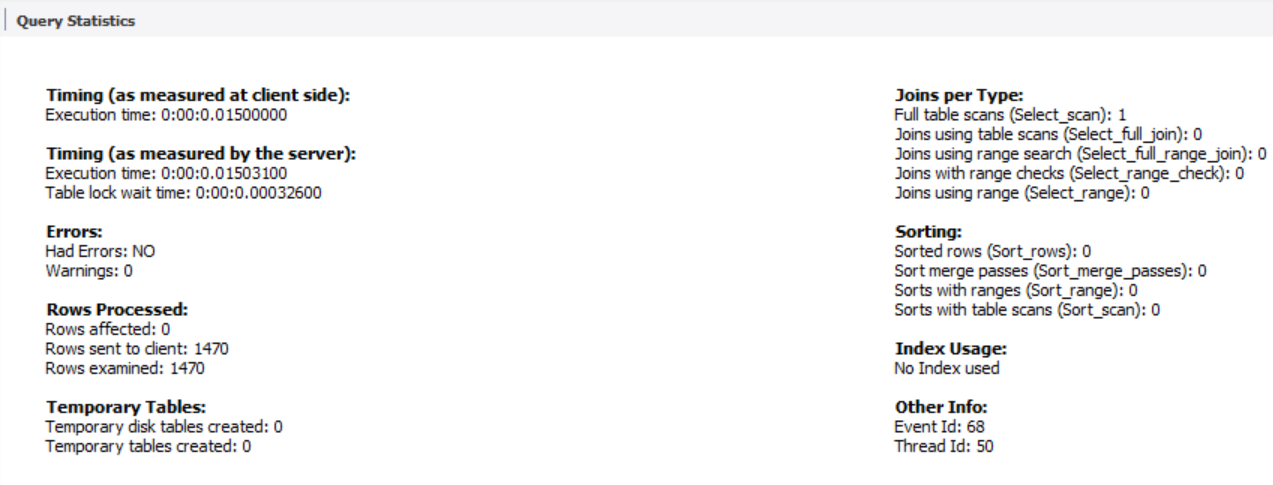




**Table 7:  Appraisal**







**Data in the Database**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table Name** | **Primary Key** | **Foreign Key** |  | **#  of rows** |
| EmployeeDetail | EmployeeNumber | 1. EmployeeEducation\_EducationID 2. JobDescription\_JobDescriptionID 3. SatisfactionLevel\_SatisfactionID 4. EmployeeExperience\_ExperienceID 5. EmployeeSalary\_SalaryID 6. Appraisal\_ApprasialID |  | 1470 |
| SatisfactionLevel | SatisfactionID | None |  | 1470 |
| JobDescription | JobDescriptionID | None |  | 1470 |
| Salary | SalaryID | None |  | 1470 |
| EmployeeExperience | ExperienceID | None |  | 1470 |
| EmployeeEducation | EducationID | None |  | 1470 |
| Appraisal | ApprasialID | None |  | 1470 |

# SQL Queries

## SQL Query 1

### Question

A new research scientist who loves to travel joins the firm and is told by HR that his job role is one of the top two roles in terms of employees that travel frequently. Is HR right in saying so? Why or Why not?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

The HR is correct in saying so because a Research Scientist’s job is second in terms of number of employees who travel frequently.

Returned 9 rows.

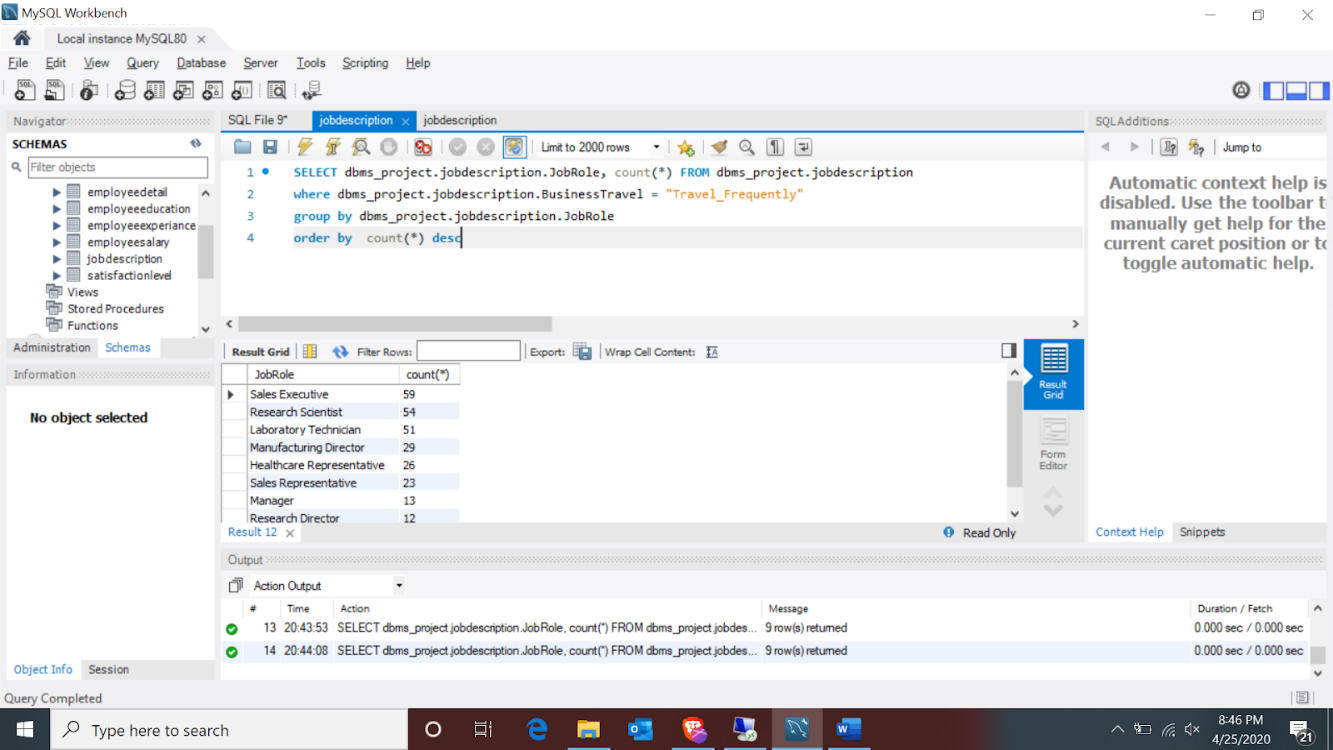
### Translation

Select the jobRole and count jobRole from the job description table where business travel is equal to Travel Frequently grouping by their job role and order the count in ascending order.

Cleanup

Select jobrole, count(\*) from jobdescription where businesstravel = Travel\_Frequently group by jobrole order by count ascending.

### Screen Shot of SQL Query and Results



## SQL Query 2

### Question

The company has been paying gas expenses for miles traveled by employees between their home and work. If they want to increase the per mile compensation, which department's employees will gain the least?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

Employees from the Human Resources department will gain the least if the company wants to increase the per mile compensation for miles traveled by employees between their home and work.

Returned 3 rows

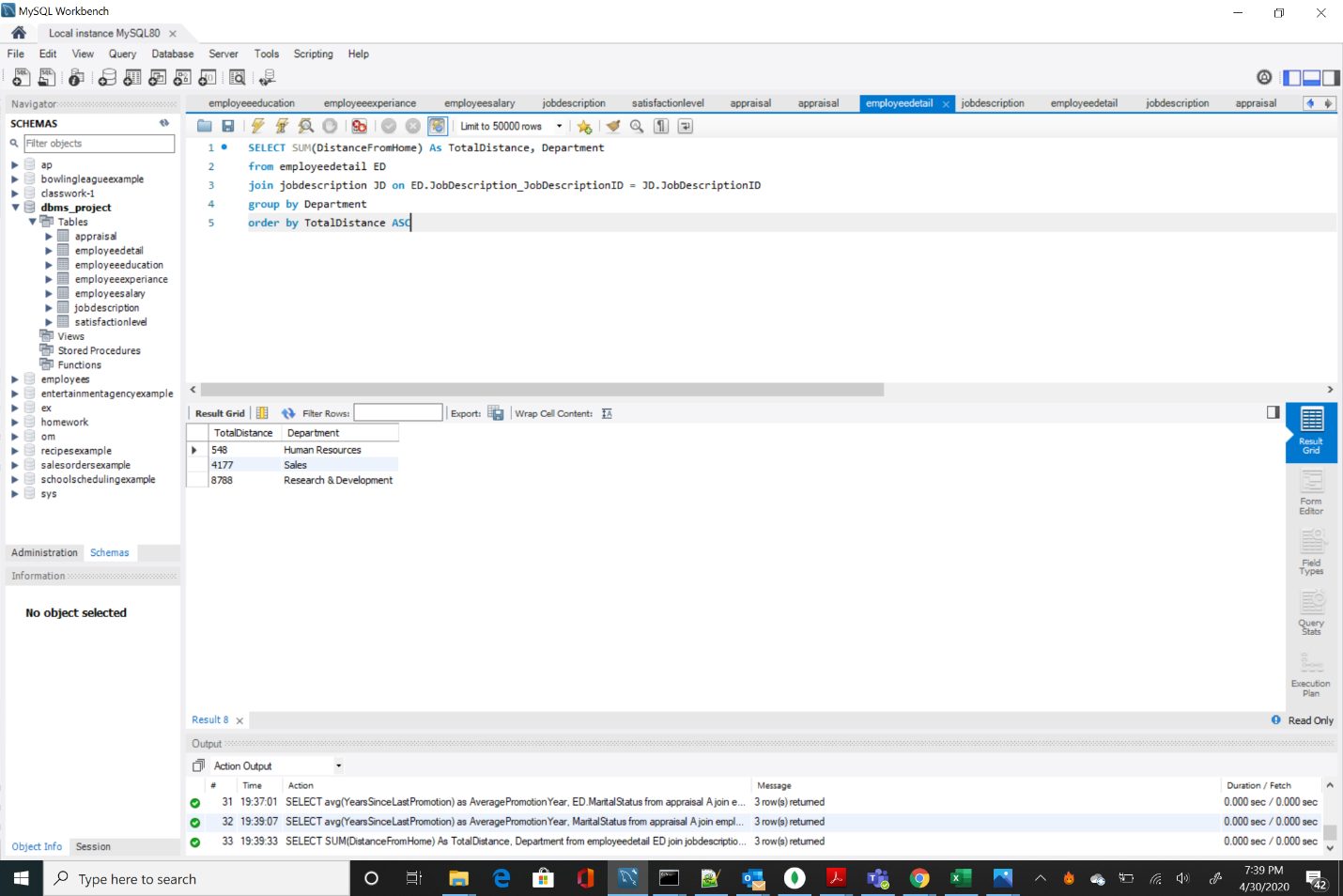
### Translation

Select the sum of distancefromhome as TotalDistance and the Department from the employeedetail table joined with jobdescription table on jobdescriptionID grouped by Department and order by TotalDistance in ascending order.

### Cleanup

Select department, sum(DistanceFromHome) as TotalDistance from employeedetail join jobdescription on employeedetail.JobDescriptionID = JobDescription.JobDescriptionID group by Department order by TotalDistance ascending.

### Screen Shot of SQL Query and Results



## SQL Query 3

### Question

A new employee with a Marketing degree wants to work in HR. Do you believe the company might be able to give him a chance to work in HR? Why or Why not?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

It is unlikely that the company will give a chance to a new employee with a Marketing degree to work in HR because as per the current employees education field in HR, there is no employee with a Marketing degree.

Returned 1 rows

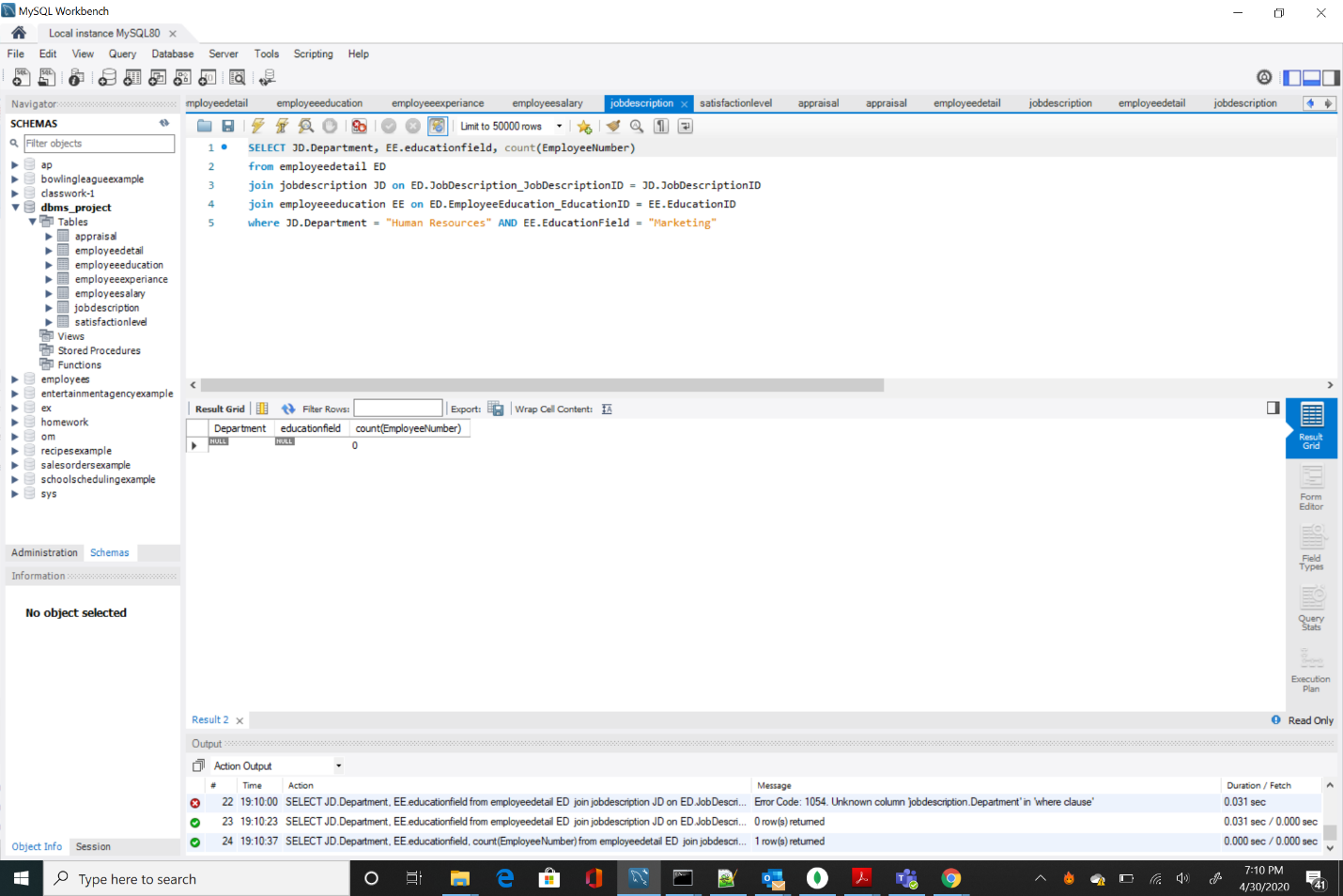
### Translation

Select the JobDescription.Department, employeedetail.educationfield and count employeenumber from the employeedetail table then join jobdescription table on JobDescriptionID then join employeeeducation table with employeedetail table on EducationID where department is equal to Human Resources and EducationField is equal to marketing.

Cleanup

SELECT JD.Department, EE.educationfield, count(EmployeeNumber) from employeedetail ED join jobdescription JD on ED.JobDescription\_JobDescriptionID = JD.JobDescriptionID join employeeeducation EE on ED.EmployeeEducation\_EducationID = EE.EducationID where JD.Department = Human Resources AND EE.EducationField = Marketing

### Screen Shot of SQL Query and Results



## SQL Query 4

### Question

Sales feels that their environment satisfaction score is higher than HR, but HR job satisfaction score is lower than Research & Development. Are they right?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

They are only partially correct. Sales department (2.6784) does not have an environment satisfaction score higher than HR (2.6825). HR (2.6032) job satisfaction score is lower than Research and Development (2.7263).

Returned 3 rows.

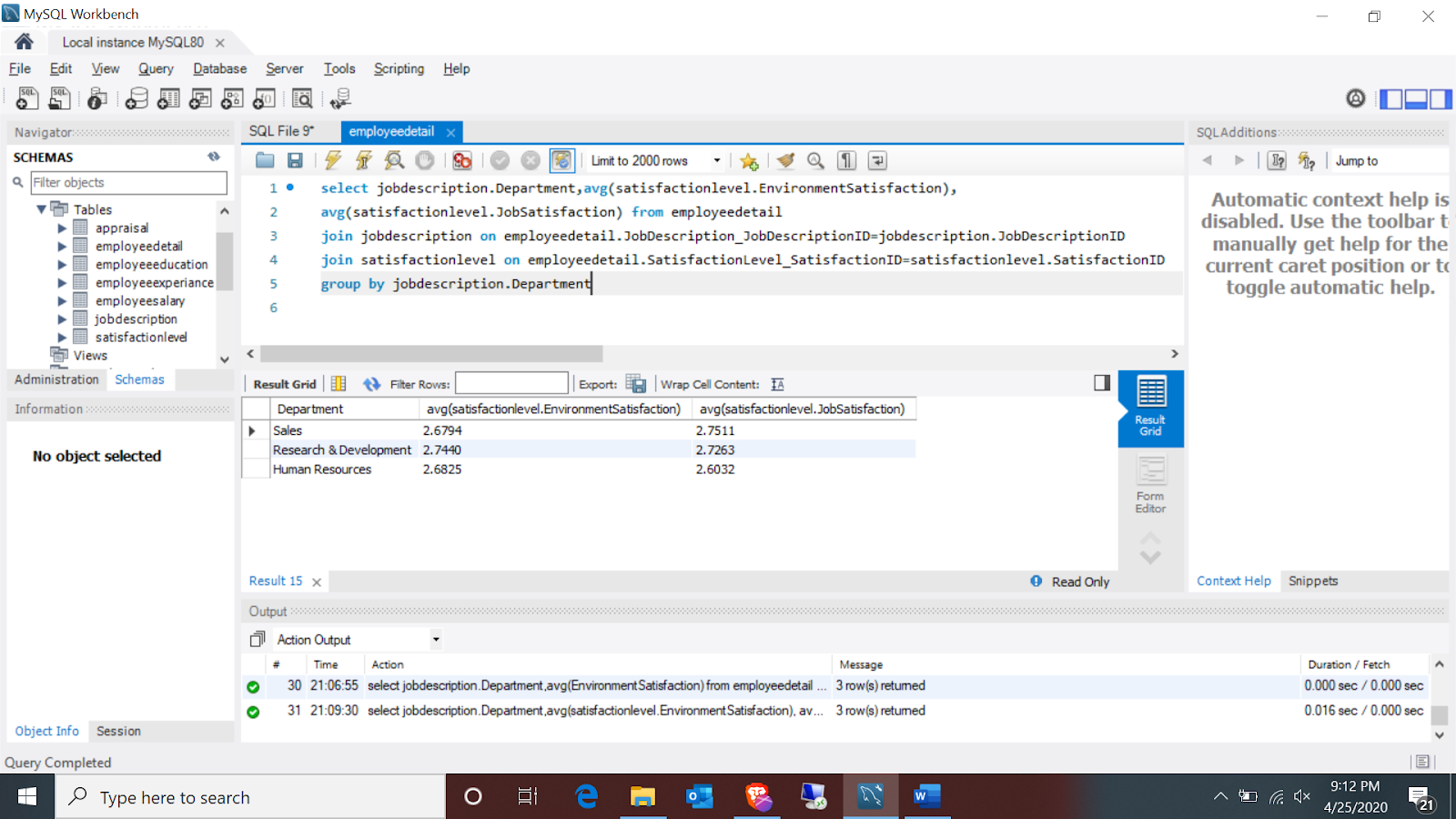
### Translation

Select the department, the average of environmentalsatisfaction and the jobsatisfaction from the employeedetail table then join with the jobdescription table on jobdescriptionID then Join satisfactionlevel table with the employeedetail table on SatisfactionID Grouped by Department in the JobDescription Table

### Cleanup

Select department, environmentalsatisfaction and jobsatisfaction from employeedetail, join jobdescription on employeedetail.jobdescription\_jobdescriptionID = JobDescription.JobDescriptionID, Join satisfactionlevel on employeedetail.satisfactionLevel\_SatisfactionID = SatisfactionLevel.SatisfactionID Group by JobDescription.Department

### Screen Shot of SQL Query and Results



## SQL Query 5

### Question

An employee from Life Sciences education field working in Research & Development department has complained to HR that employees with her educational background are paid more in Sales department than in Research & Development. What insight can you provide to prove or disprove that statement?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

The employee is correct. Average monthly income of employees with a Life Sciences education background is higher in Sales ($7246) than in Research and Development ($6179).

Returned 3 rows

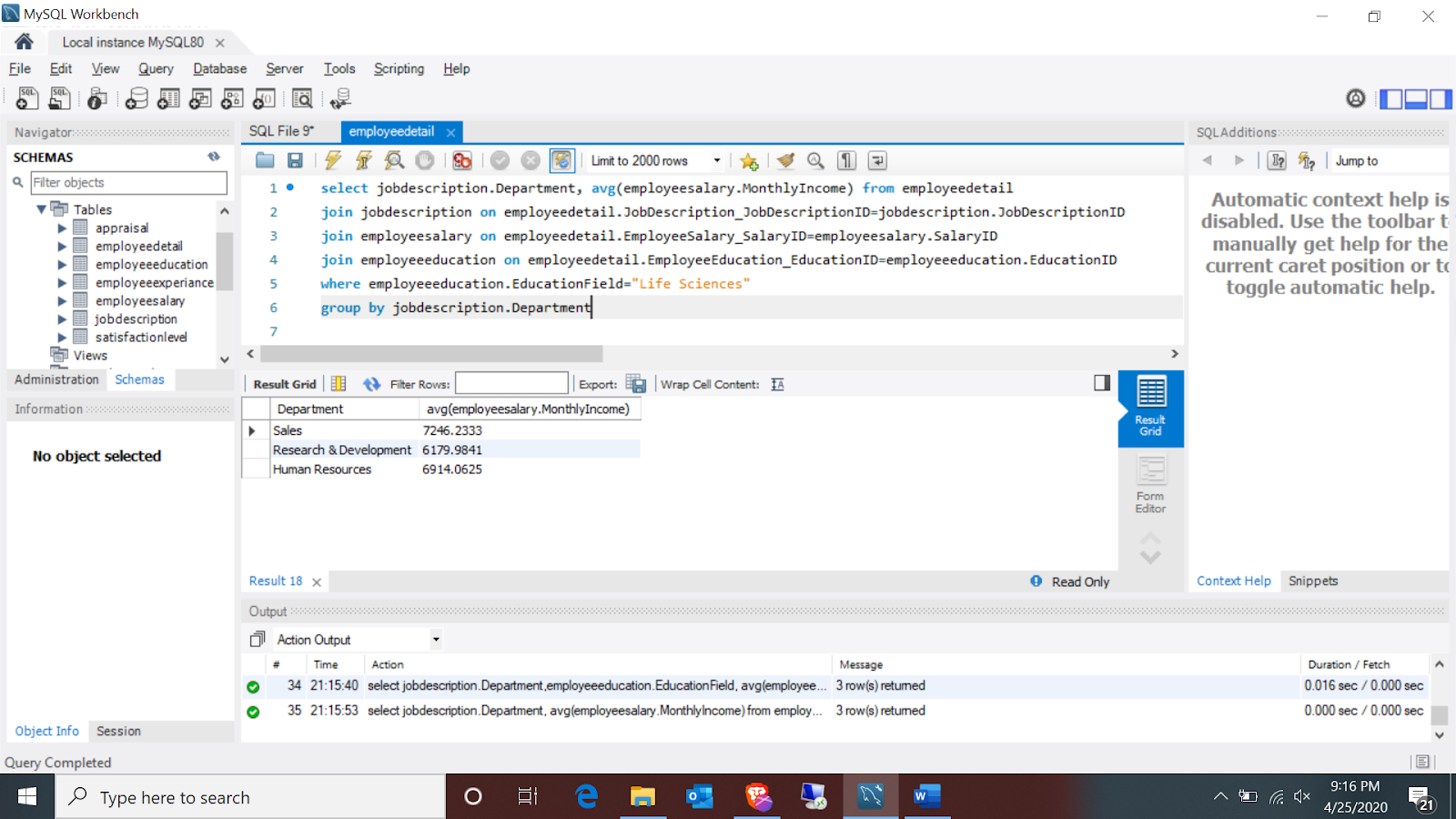
### Translation

Select the department and the average monthly income from the employeedetail table then join with the jobdescription table on jobdescriptionID, then join employeesalary table with employeedetail table on salaryID then join employeeeducation table with the employeedetail table on EducationID where the employeeeducation field in the employeeeducation table is equal to Life Sciences and grouped by department in the jobdescription table.

### Cleanup

Select department, monthlysalary from employeedetail , join jobdescription on employeedetail.jobdescription\_jobdescriptionID=jobdescription.jobdescriptionID, join employeesalary on employeedetail.employeesalary\_salaryID=employeesalar.salaryID where employeeeducation.educationfield = “Life Sciences” , group by jobdescription.department

### Screen Shot of SQL Query and Results



## SQL Query 6

### Question

A press article in a business magazine has said that at this company, Marital status of women in Research & Development has severely affected their promotion rates. What initial finding can you obtain from the data to help articulate the company's response in this regard?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

Comparing Marital status of women data across the Research & Development and the company.

|  |  |  |
| --- | --- | --- |
| Research & Development  (Average Promotion Years) | Across Company  (Average Promotion Years) | Marital Status |
| 2.2946 | 2.0955 | Single |
| 2.1198 | 2.3199 | Married |
| 2.6265 | 2.5726 | Divorced |

Looking at the above data if Marital status of women is single then promotion in Research & Development would delay approx. 0.2 years than rest of the company. If status is Married, she will be promoted early than rest of the company. If status is divorced, then there is a delay which can be considered very minute. So, the Magazine claims are false.

Returned 3 rows each.

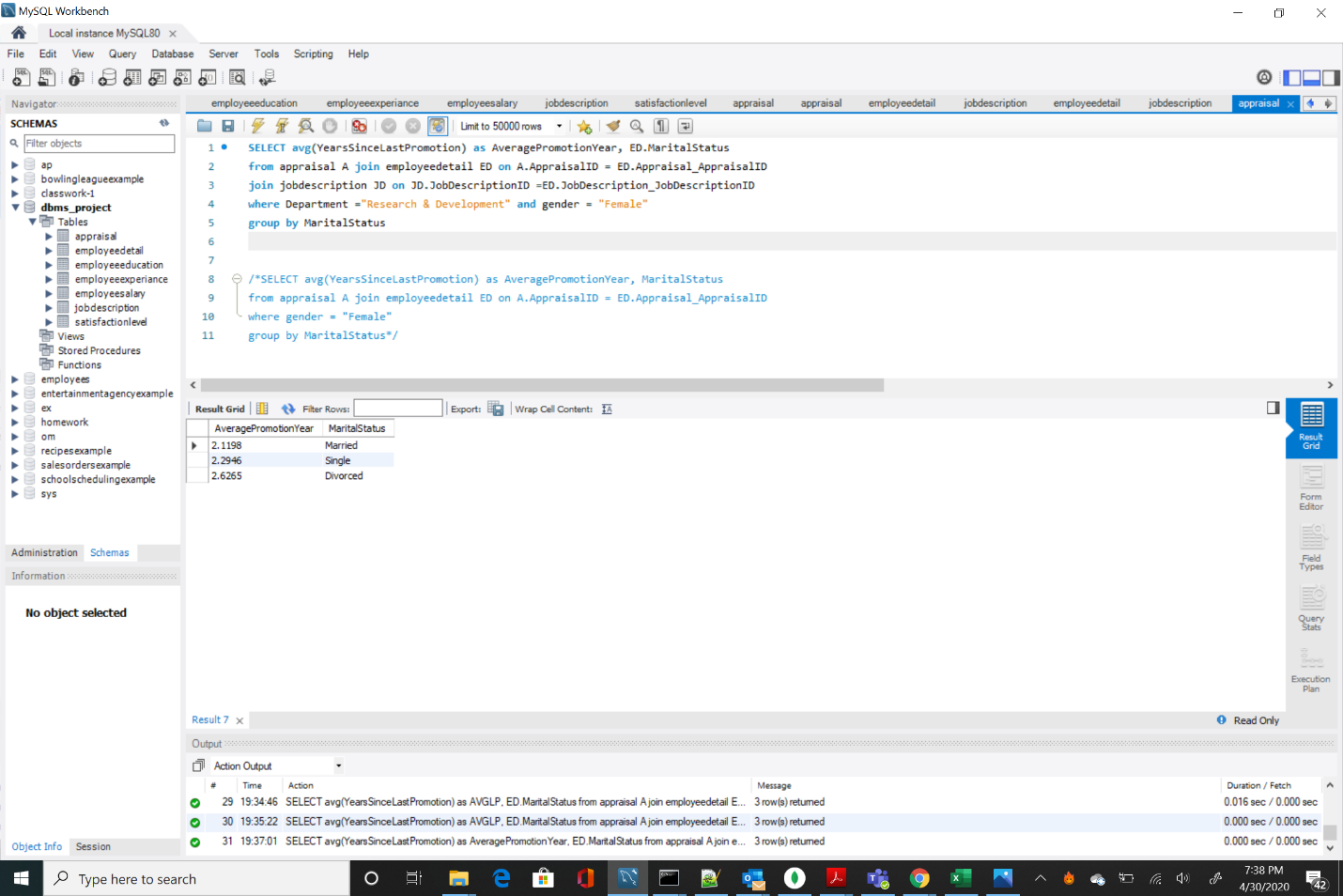
### Translation

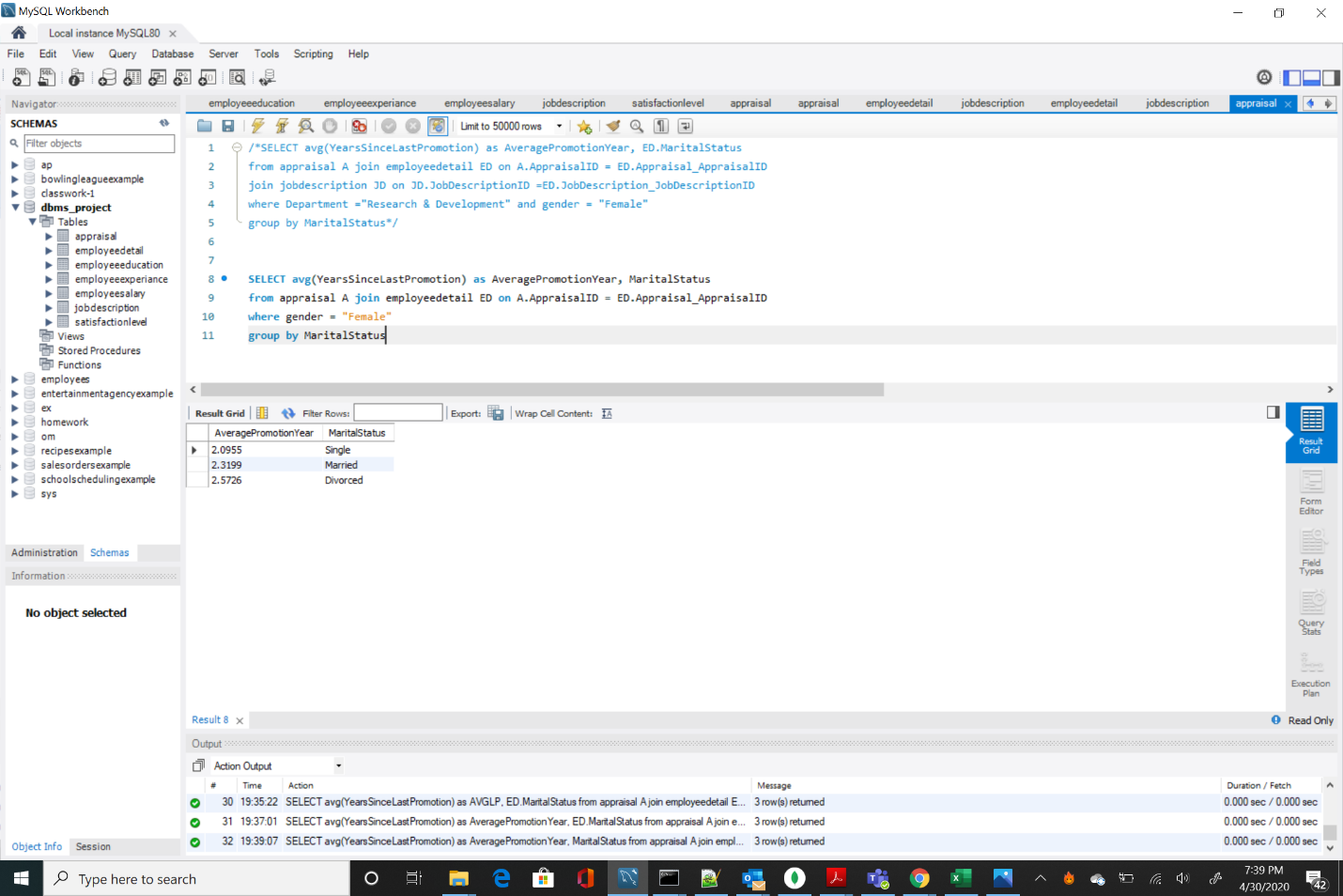
1. Select the average yearssincelastpromotion as AveragePromotionYear and ED.maritalStatus from the appraisal table and join EmployeeDetail ED on A.AppraisalID = ED.Appraisal\_AppraisalID join Jobdescription JD on JD.JobDescription\_jobDescriptionID = ED.JobDescription\_JobDescriptionID Where Department = "Research & Development“ and the gender = "female“ grouped by the MaritalStatus
2. Select the average YearsSinceLastPromotion as AveragePromotion and ED.maritalStatus from the appraisal table and join EmployeeDetail ED on A.AppraisalID = ED.Appraisal\_AppraisalID where the gender = "female“ grouped by the MaritalStatus

### Cleanup

1. Select avg YearsSinceLastPromotion as AveragePromotion and ED.maritalStatus from appraisal join EmployeeDetail ED on A.AppraisalID = ED.Appraisal\_AppraisalID join Jobdescription JD on JD.JobDescription\_jobDescriptionID = ED.JobDescription\_JobDescriptionID Where Department = "Research & Development“ , gender = "female“ group by MaritalStatus
2. Select avg YearsSinceLastPromotion as AveragePromotion and ED.maritalStatus appraisal and join EmployeeDetail ED on A.AppraisalID = ED.Appraisal\_AppraisalID where gender = "female“ group by MaritalStatus

### Screen Shot of SQL Query and Results





# Physical Mongo Database

## Assumptions/Notes About Data Set

1. Employee's work experience history would be summed up in the experience table.
2. Satisfaction table has 4 sub-sections, and each can have a rating from 1 to 4.
3. Stock option level is the relative level of stocks the employee owns. Higher the level of stock option higher the number of stocks the employee owns.
4. We are assuming each employee has its own daily and monthly rate which is used to bill client for providing their services to them.
5. Monthly income is the main salary of the employee and hourly rate will be used as an addon if employee works overtime means more than standard hours to calculate his month end payout.
6. We have assumed each employee age is over 18 and that is not going to change in future.
7. There are 6 different education fields in our database and each field has 5 levels of education that an employee can have. Higher the level of education higher the education employee has.
8. There are 9 job roles and 5 job levels. Healthcare Representatives, Sales Executives and Manufacturing Directors are at levels 2-4. Laboratory Technician, HR and Research Scientists are at levels 1-3. Managers and Research Directors are at level 3-5. Sales Representatives are at level 1-2. These levels represent the hierarchy of the chain of command in the specific roles.
9. We are assuming the there are no multiple entries of each employee.
10. We are assuming employee didn’t take the multiple surveys.
11. We have assumed employee count for each employee will be 1 and that is not going to change in futures.
12. All rates are in dollar units.

## Screen shot of Physical Database objects (Database, Collections and Attributes)

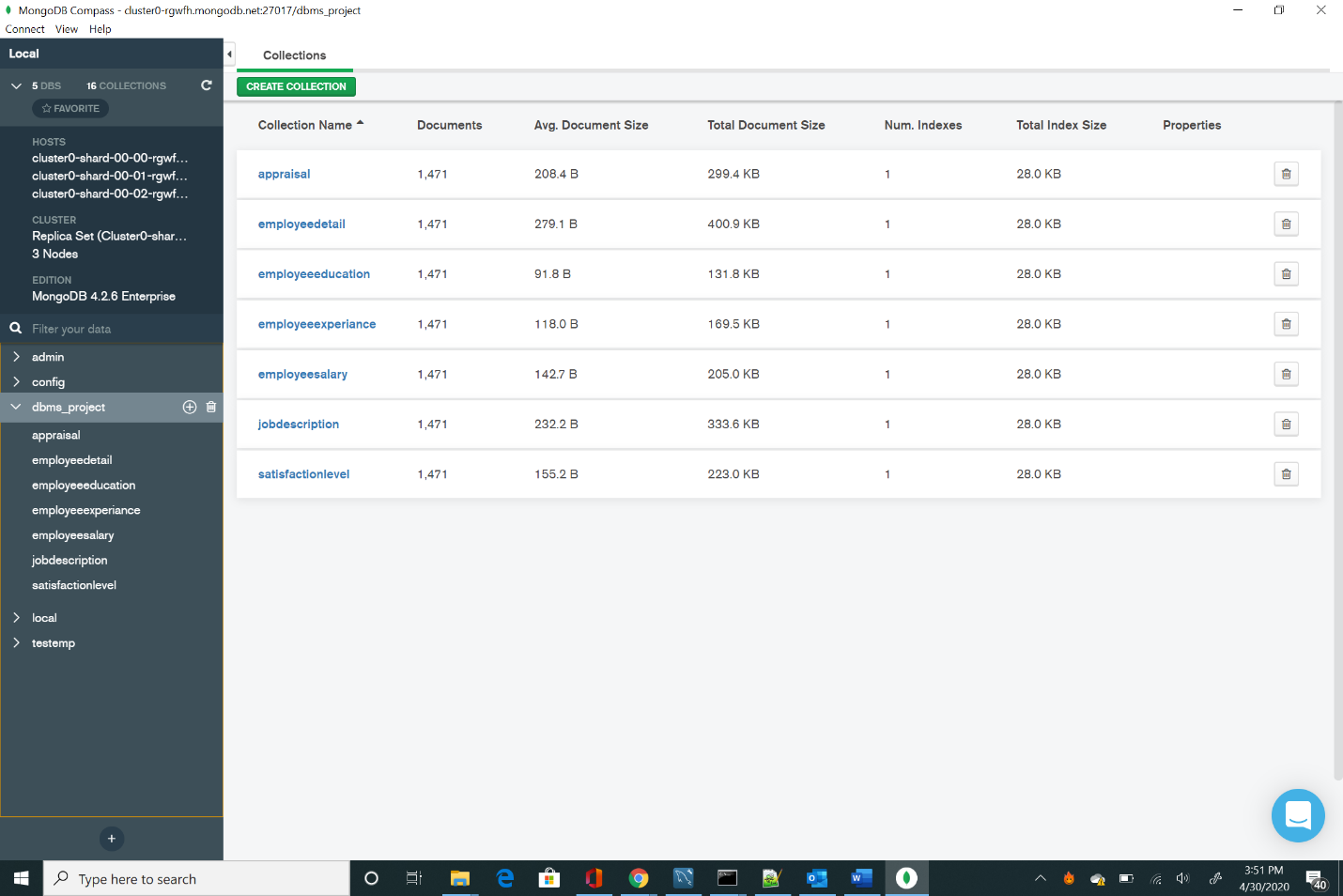


Table 1 Appraisal

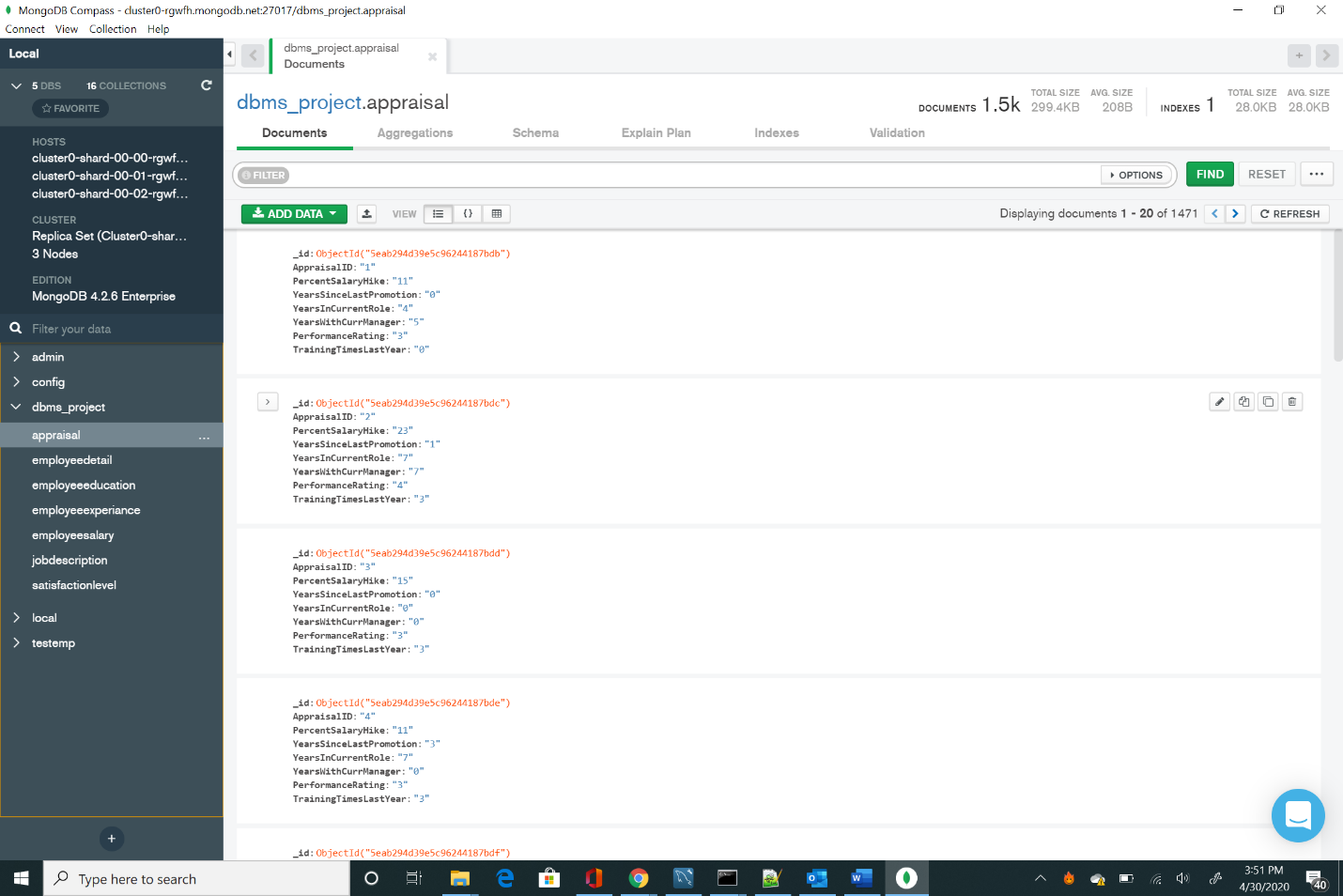


Table2 employeedetail

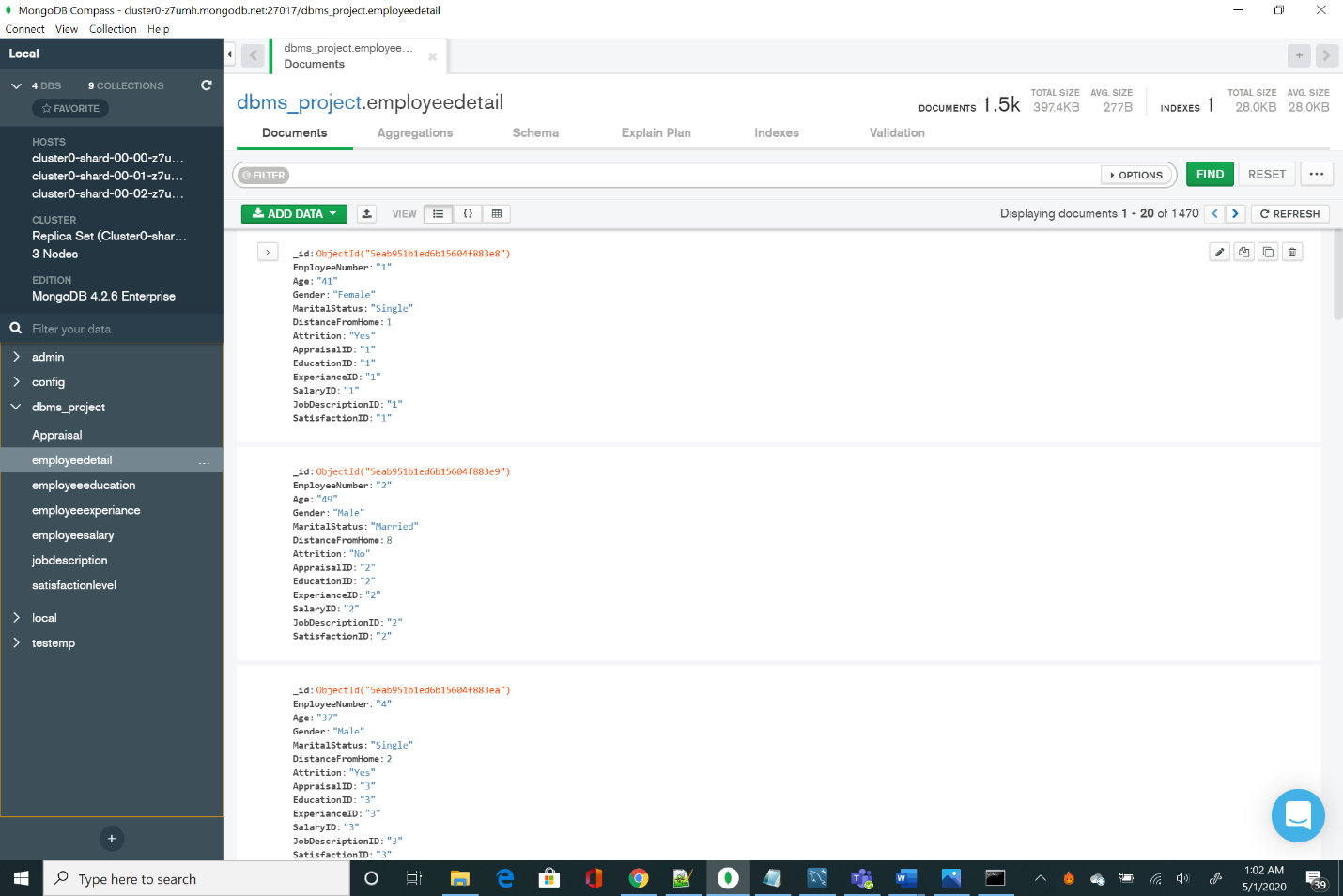


Table3 employeeeducation

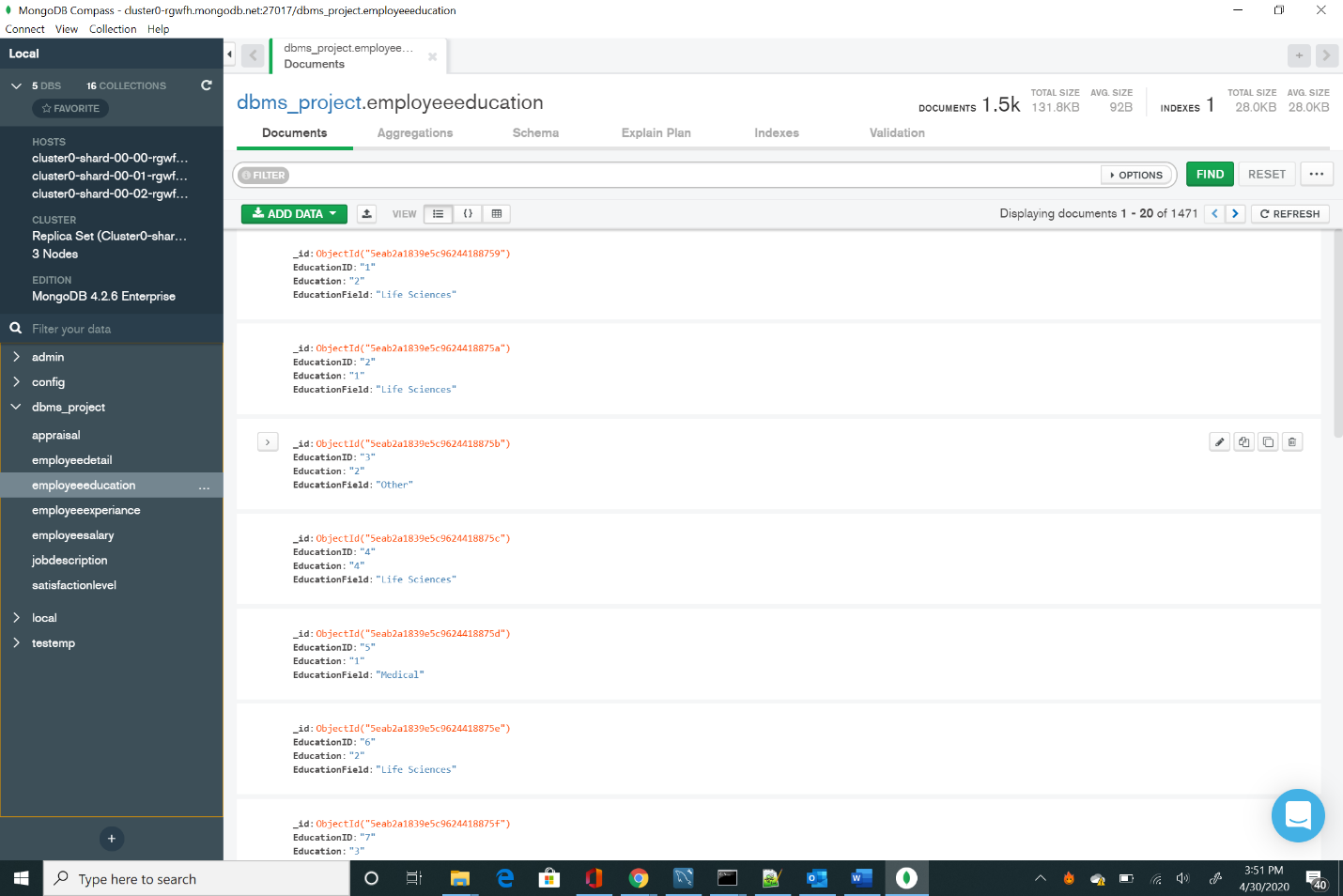


Table4 employeeexperiance

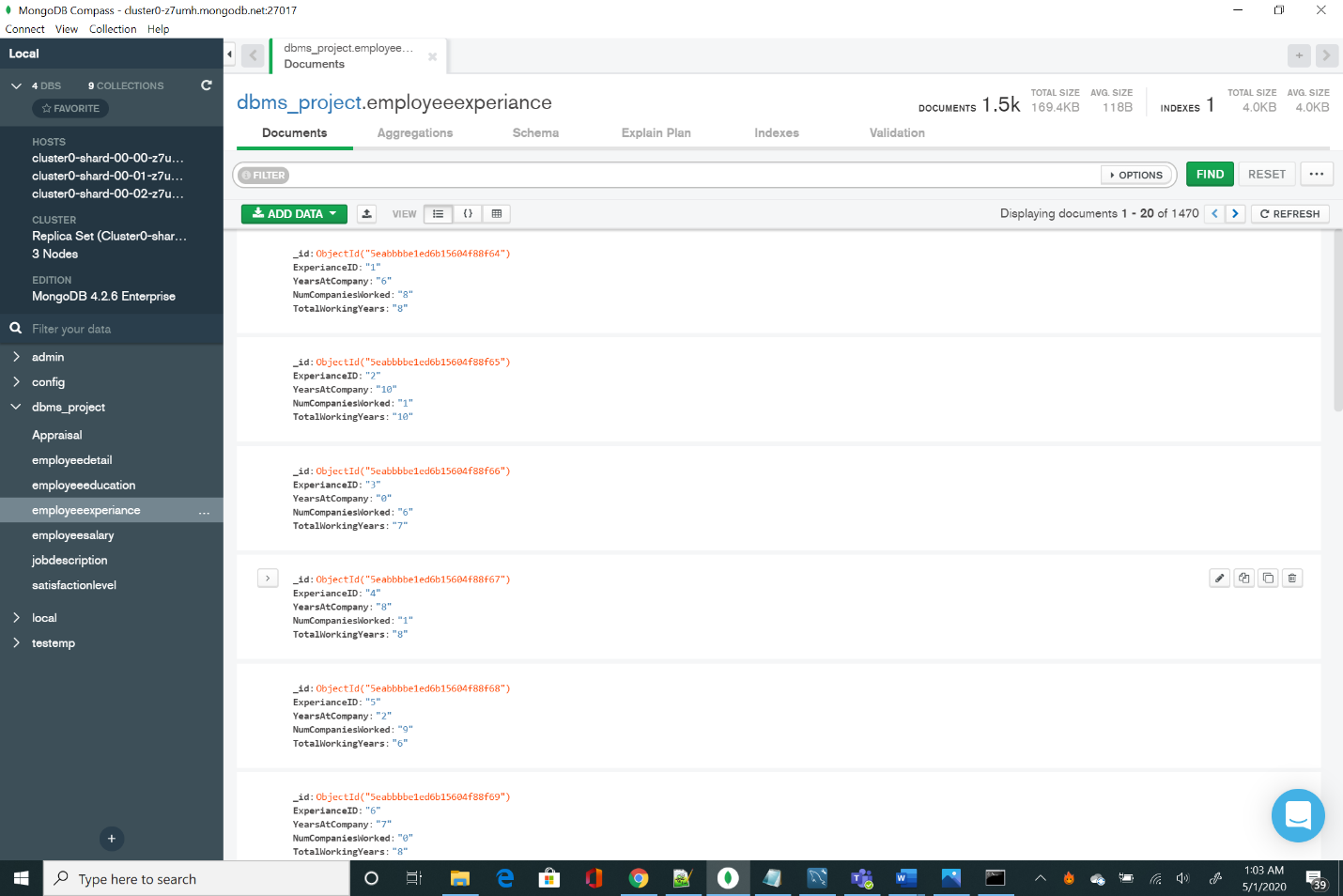


Table5 employeesalary

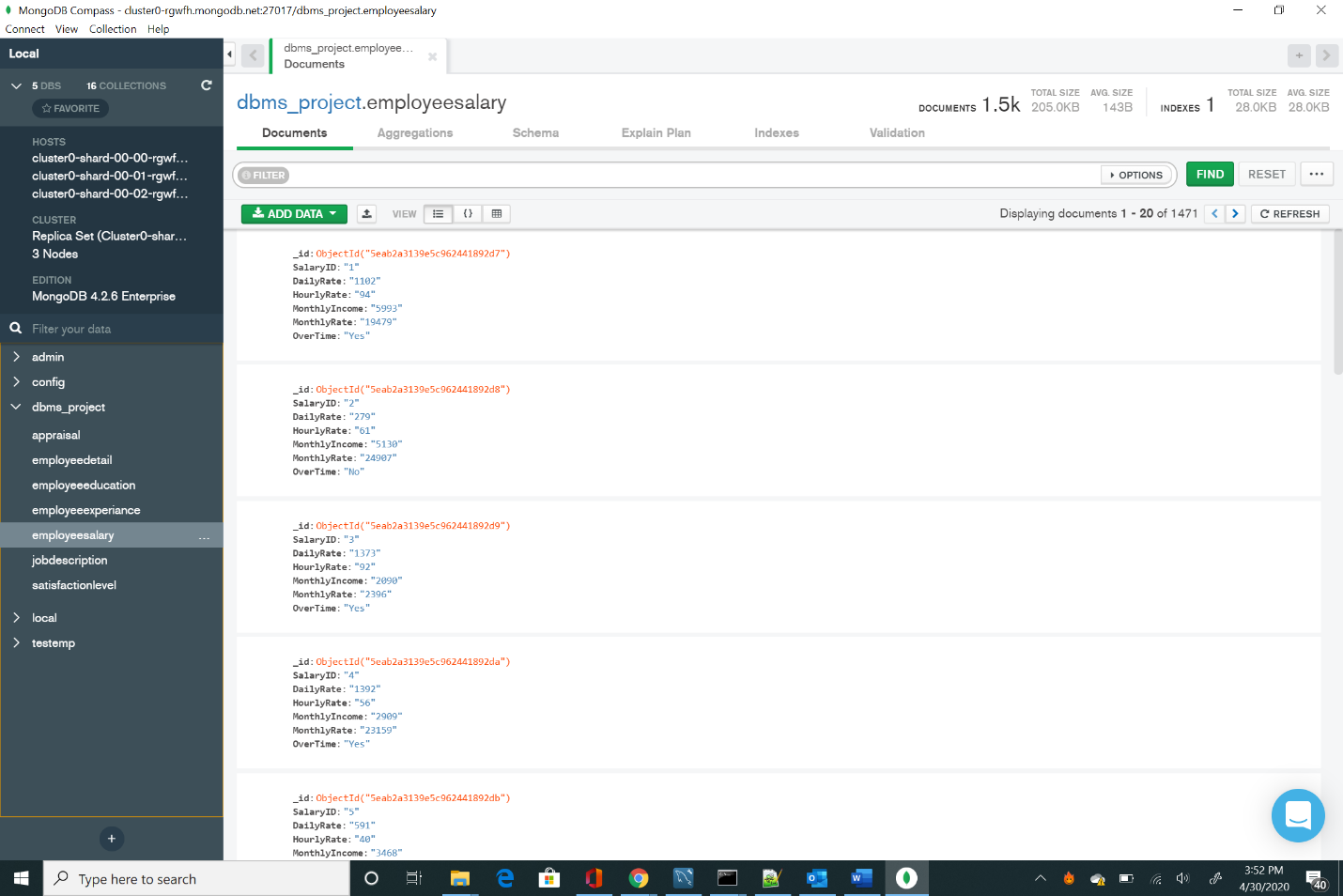


Table6 jobdescription

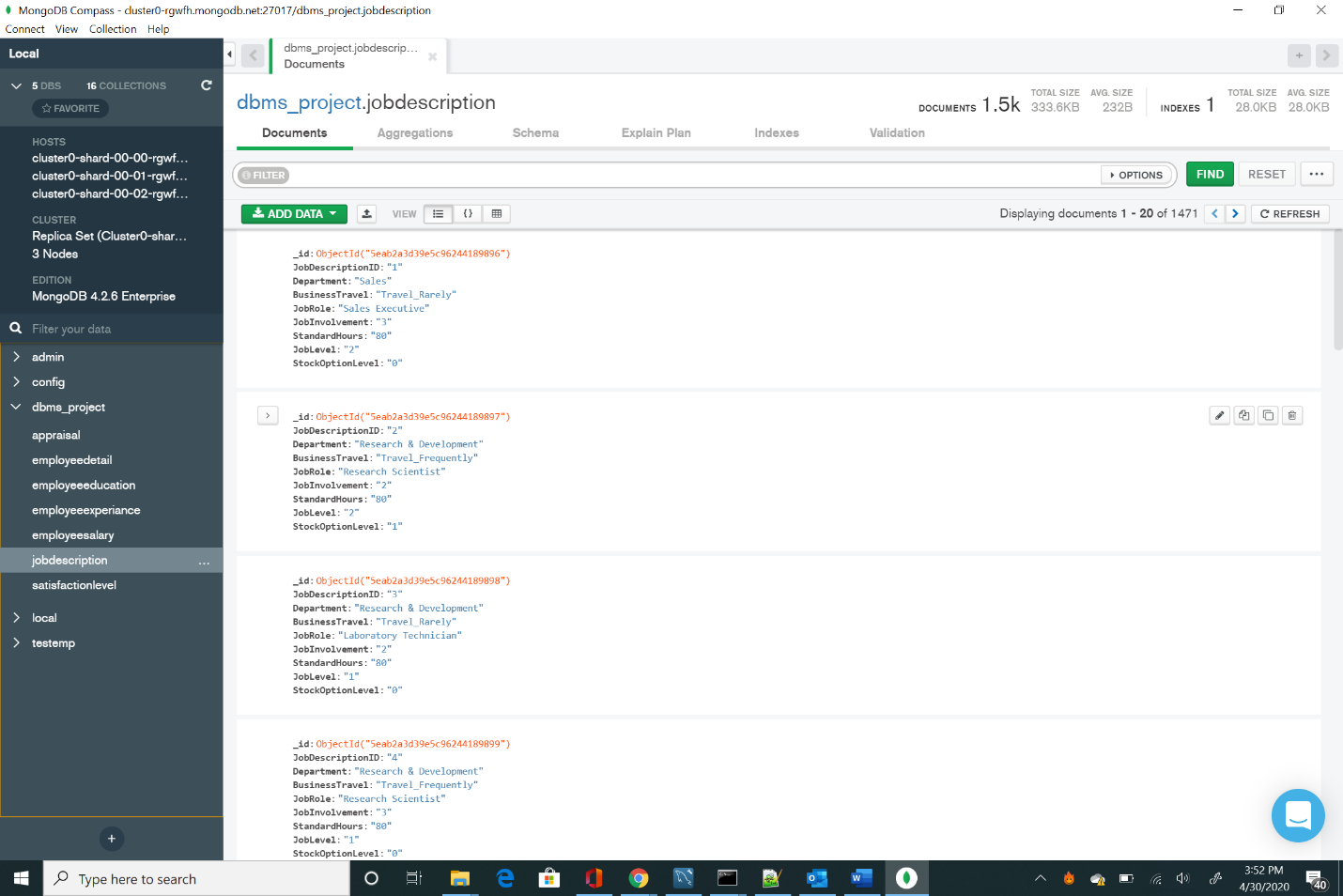
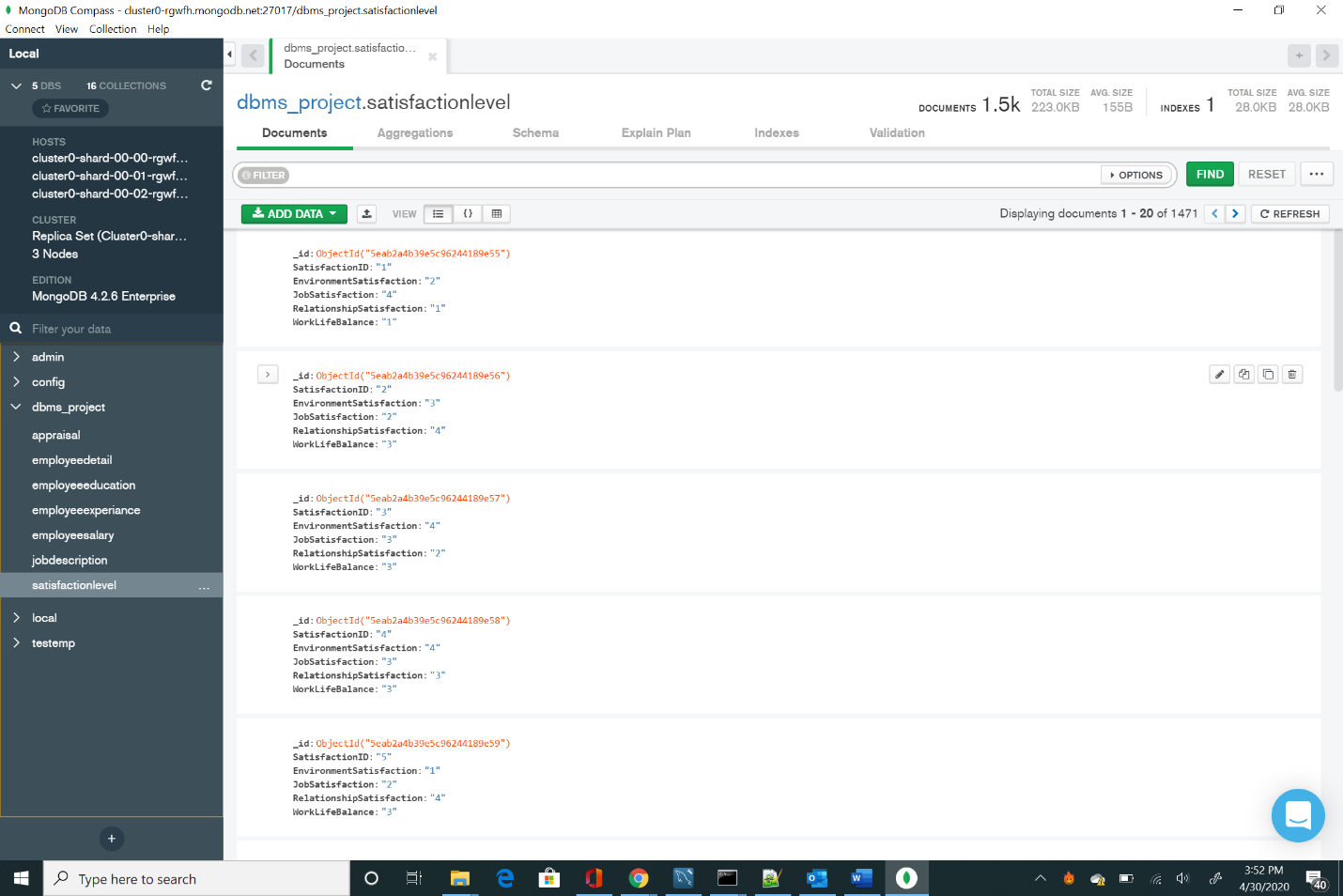


Table7 satisfactionlevel



## Data in the Database

|  |  |  |
| --- | --- | --- |
| **Collection Name** | **Relationships with Other Collections (if any)** | **# of Documents in Collection** |
| employeedetail |  | 1470 |
| appraisal |  | 1470 |
| employeeeducation |  | 1470 |
| employeeexperiance |  | 1470 |
| employeesalary |  | 1470 |
| jobdescription |  | 1470 |
| satisfactionlevel |  | 1470 |

# 

# MongoDB Queries/Code

## Mongo Query 1

### Question

A new research scientist who loves to travel joins the firm and is told by HR that his job role is one of the top two roles in terms of employees that travel frequently. Is HR right in saying so? Why or Why not?

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

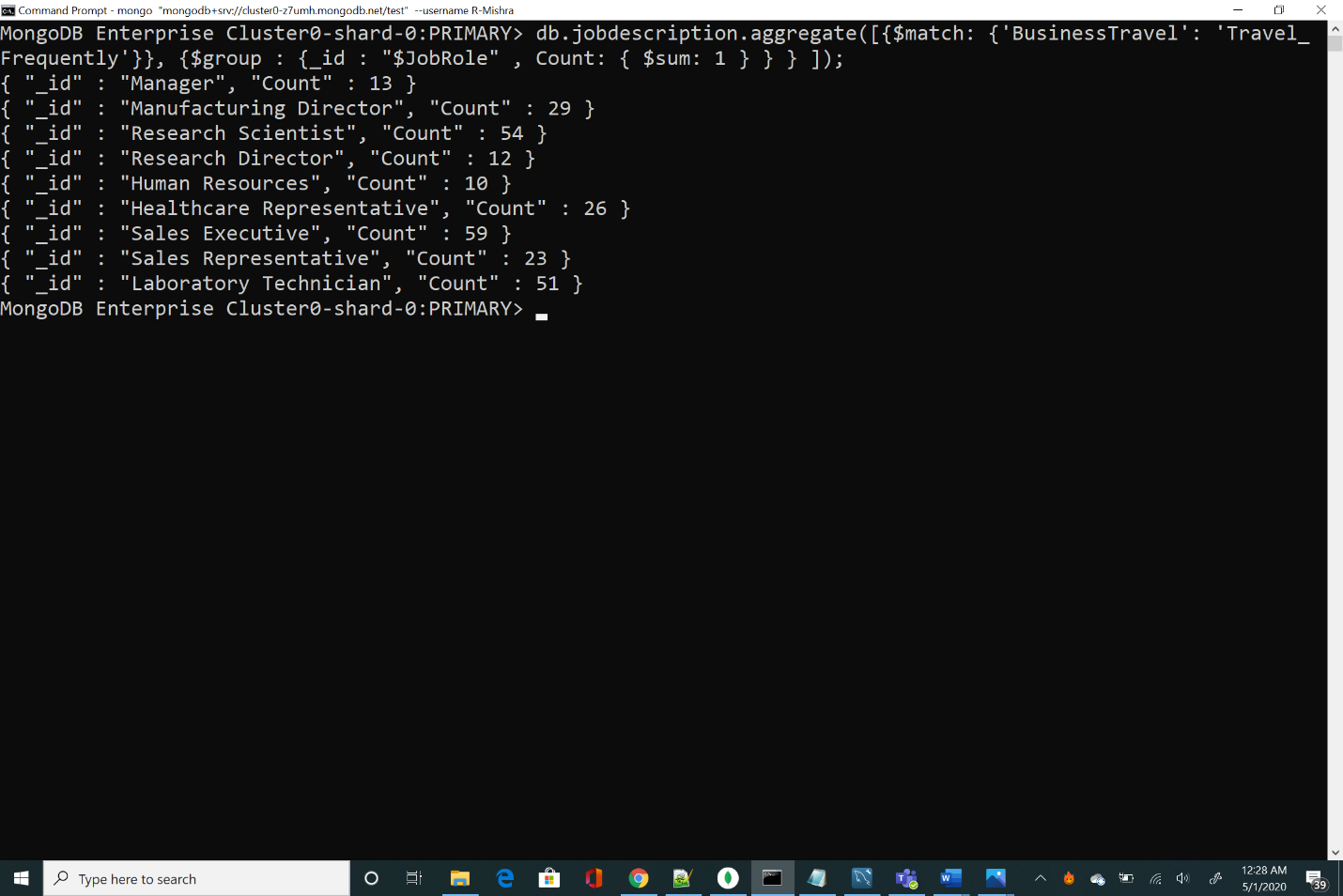
The HR is correct in saying so because a Research Scientist’s job is second in terms of number of employees who travel frequently.

Returned 9 document.

### Translation

On dbms\_project db aggregate jobdescription, match BusinessTravel is equal to Travel\_Frequently and group them by JobRole and count it.

### Screen Shot of MongoDB Query/Code and Results



## Mongo Query 2

### Question

The company has been paying gas expenses for miles traveled by employees between their home and work. If they want to increase the per mile compensation, which department's employees will gain the least?

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

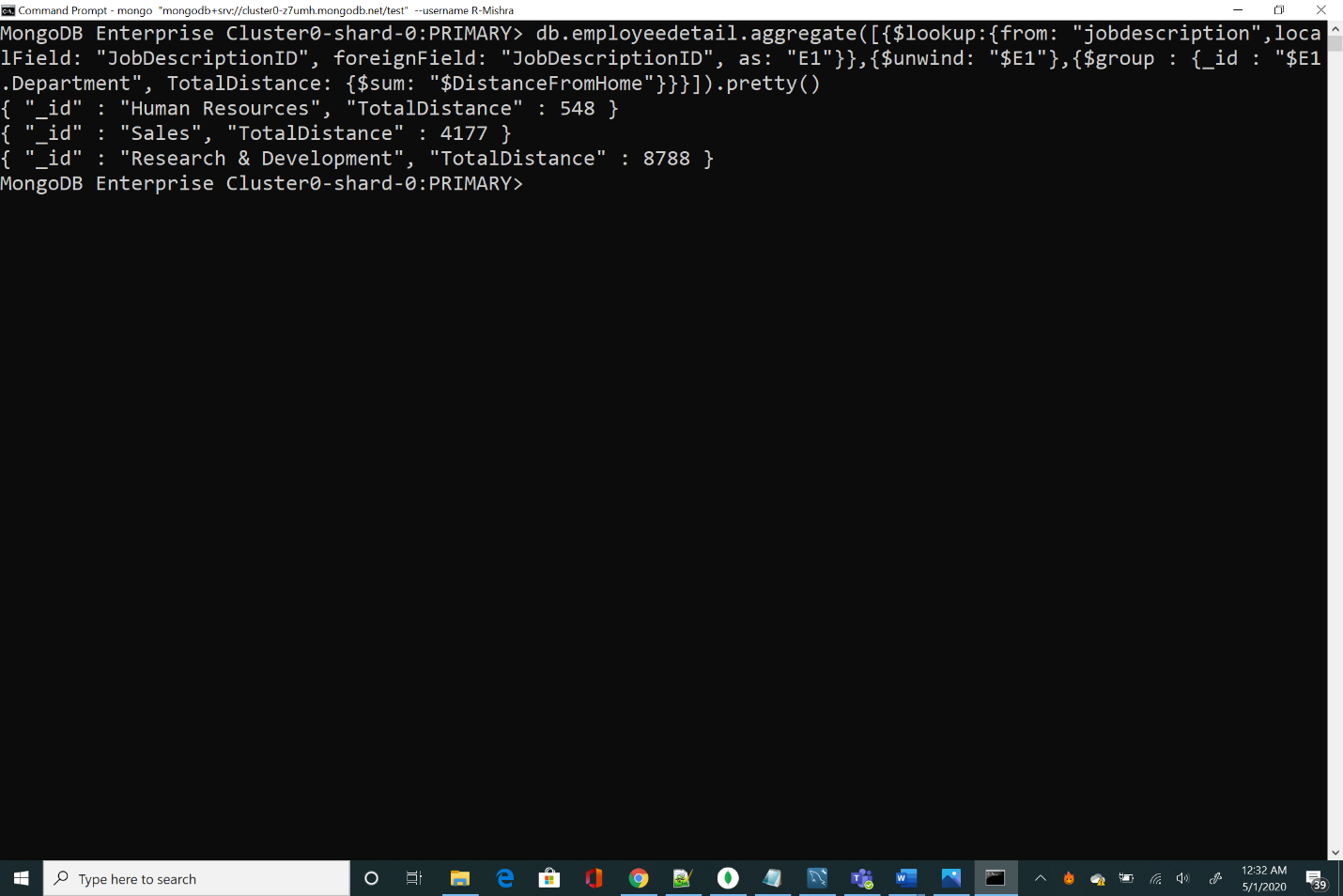
Employees from the Human Resources department will gain the least if the company wants to increase the per mile compensation for miles traveled by employees between their home and work.

Returned 3 document.

### Translation

On dbms\_project db aggregate employeedetail and lookup jobdescription where localField and ForeignField is JobDescriptionID as E1. Unwind E1, group E1 department and sum DistanceFromHome as TotalDistance.

### Screen Shot of MongoDB Query/Code and Results



## Mongo Query 3

### Question

Sales feels that their environment satisfaction score is higher than HR, but HR job satisfaction score is lower than Research & Development. Are they right?

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

They are only partly correct. Sales department (2.6784) does not have an environment satisfaction score higher than HR (2.6825). HR (2.6032) job satisfaction score is lower than Research and Development (2.7263).

Returned 3 document.

### Translation

On dbms\_project db aggregate employeedetail and lookup jobdescription where localField and ForeignField is JodDescriptionID As E1 now lookup satisfactionlevel where localField and ForeignField is SatisfactionID as E2 group E1.Department and calculate average E2.JobSatisfaction as avgJobSatisfaction and average E2.EnvironmentSatisfaction as average EnvSatisfaction.

### Screen Shot of MongoDB Query/Code and Results

