

# **SQL-Mongo Project – IBM HR Analytics Employee Attrition & Performance**

BUAN 6320

## Contents

Relational Data Model .....	4
Assumptions/Notes About Data Entities and Relationships .....	4
Entity-Relationship Diagram .....	5
Physical MySQL Database .....	6
Assumptions/Notes About Data Set .....	6
Screen shot of Physical Database objects.....	7
Loaded entire dataset into MySQL Workbench.....	7
Screenshots of individual tables .....	8
Data in the Database .....	15
SQL Queries .....	16
SQL Query 2 .....	16
Question .....	16
Notes/Comments About SQL Query and Results (Include # of Rows in Result) .....	16
Translation .....	16
Query .....	16
Explanation .....	16
Screen Shot of SQL Query and Results.....	17
SQL Query 4 .....	18
Question .....	18
Notes/Comments About SQL Query and Results (Include # of Rows in Result) .....	18
Translation .....	18
Query .....	18
Explanation .....	18
Screen Shot of SQL Query and Results.....	19
SQL Query 5 .....	19
Question .....	19
Notes/Comments About SQL Query and Results (Include # of Rows in Result) .....	19
Translation .....	19
Query .....	19
Explanation .....	20
Screen Shot of SQL Query and Results.....	20
SQL Query 6 .....	20

Question .....	20
Notes/Comments About SQL Query and Results (Include # of Rows in Result) .....	20
Translation .....	21
Query .....	21
Explanation .....	21
Screen Shot of SQL Query and Results.....	21
SQL Query 7 .....	21
Question .....	21
Notes/Comments About SQL Query and Results (Include # of Rows in Result) .....	21
Translation .....	22
Query .....	22
Explanation .....	22
Screen Shot of SQL Query and Results.....	22
Data Review for MongoDB .....	23
Assumptions/Notes About Data Collections, Attributes and Relationships between Collections.....	23
Physical Mongo Database.....	24
Assumptions/Notes About Data Set .....	24
Screen shot of Physical Database objects (Database, Collections and Attributes).....	25
Data in the Database .....	26
MongoDB Queries/Code.....	27
Mongo Query 9 .....	27
Question .....	27
Translation .....	27
Query .....	27
Screen Shot of MongoDB Query/Code and Results.....	27
Results.....	28
Mongo Query 10.....	28
Question .....	28
Translation .....	28
Screen Shot of MongoDB Query/Code and Results.....	29
Results.....	29
Mongo Query 11.....	29
Question .....	29

Translation .....	30
Screen Shot of MongoDB Query/Code and Results .....	30
Results.....	30

## Relational Data Model

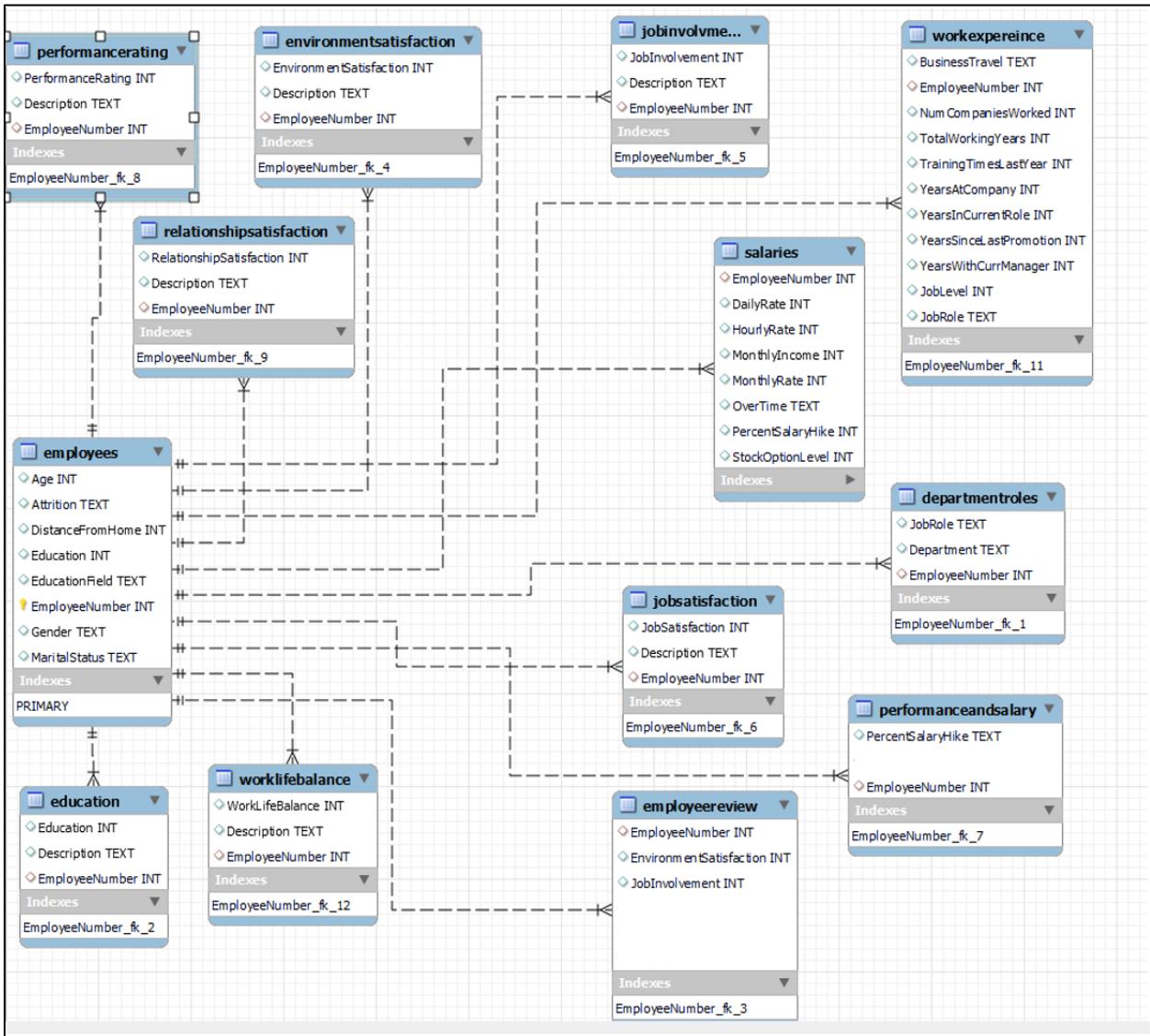
### Assumptions/Notes About Data Entities and Relationships

The database is divided into 13 tables: some directly partitioning data and others describing it.

The employees table is the main table and is related to other tables using the employee number as the foreign key. All other tables have employee number as foreign key so they are connected in the schema and the schema follows 3NF because we have initially screened the schema for all the rules from 1NF to 3NF.

The column Over18 has been removed because whether an employee is aged over 18 or not can be determined directly using the Age column. Also, the column EmployeeCount has been omitted from the database because the row-wise count of employee will be 1 (the entry in all rows of this column).

## Entity-Relationship Diagram



## Physical MySQL Database

### Assumptions/Notes About Data Set

Include any assumptions made about data such as empty fields, sparse data, bad data, etc.

1. Education and education field are independent of each other. For every employee, you cannot find the education field just by education because they are independent of each other. We need employee number since it is the primary key. The education field cannot be null.
2. The job role and job level do not depend on each other. We need the employee number to find the job role and job level.
3. Over 18 will be determined by the age in the employees table.
4. Hourly rate is the hourly salary paired to the employee. Standard hours is 80 per week. Anything above 80 is considered over time.
5. Daily rate is the amount of money paid per day.
6. All employees are paid a monthly income.
7. The employee number is the primary key which is used to determine employee information.
8. Employee\_number field is foreign key in almost every table connecting the schema with other tables ensuring the schema is in 3NF normalized form.
9. It is assumed that every employee works in 1 department only and that multiple departments cant employ the same employee.
10. Below are the functional dependencies identified for the normalization –
  - a. Education  $\rightarrow$  hourlyrate, joblevel, jobrole, totalworkingyears, department
  - b. Worklifebalance  $\rightarrow$  jobsatisfaction, totalworkingyears, relationshipsatisfaction
  - c. Performancerating  $\rightarrow$  yearssincepromotion, yearsincurrentrole
  - d. salaries  $\rightarrow$  dailyrate, hourlyrate, monthlyincome, monthlyrate, overtime, percentsalaryhike, stockoptionlevel
  - e. workexperience  $\rightarrow$  numcompaniesworked, totalworkingyears, trainingtimeslastyear, yearincompany, yearincurrentrole, yearssincepromotion, yearswithcurrentmanager, joblevel, jobrole
  - f. employeereview  $\rightarrow$  employeesatisfaction, jobinvolvement
  - g. performanceandsalary  $\rightarrow$  percentsalaryhike, performancerating

## Screen shot of Physical Database objects

Loaded entire dataset into MySQL Workbench

The screenshot shows the MySQL Workbench interface with a query editor and a results grid.

Query Editor:

```
employeeattrition
1 •  SELECT * FROM ibmhranalyticsdb.employeeattrition;
```

Result Grid:

ID	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfied
1	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	2
2	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	3
3	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	4
4	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	4
5	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	1
6	32	No	Travel_Frequently	1005	Research & Development	2	2	Life Sciences	1	8	4
7	59	No	Travel_Rarely	1324	Research & Development	3	3	Medical	1	10	3

Output:

#	Time	Action	Message	Duration / Fetch
190	20:50:38	SELECT * FROM ibmhranalyticsdb.worklifebalance LIMIT 0, 1000	4 row(s) returned	0.000 sec / 0.000 sec
191	20:51:38	SELECT * FROM ibmhranalyticsdb.employeeattrition LIMIT 0, 1000	1000 row(s) returned	0.000 sec / 0.000 sec

## Screenshots of individual tables

### Department Roles

The screenshot shows a database interface with a query window titled "departmentroles". The query is:

```
1 • SELECT * FROM ibmhranalyticsdb.departmentroles;
```

The result grid displays the following data:

JobRole	Department	EmployeeNumber
Sales Executive	Sales	1
Research Scientist	Research & Development	2
Laboratory Technician	Research & Development	4
Research Scientist	Research & Development	5
Laboratory Technician	Research & Development	7
Laboratory Technician	Research & Development	8
Laboratory Technician	Research & Development	10
Laboratory Technician	Research & Development	11

Output details:

#	Time	Action
1	22:01:24	SELECT * FROM ibmhranalyticsdb.departmentroles LIMIT 0, 2000

Message: 1470 row(s) returned

Duration / Fetch: 0.000 sec / 0.000 sec

### Education:

The screenshot shows a database interface with a query window titled "education". The query is:

```
1 • SELECT * FROM ibmhranalyticsdb.education;
```

The result grid displays the following data:

Education	Description	EmployeeNumber
5	Doctor	1
5	Doctor	2
5	Doctor	4
5	Doctor	5
5	Doctor	7
5	Doctor	8
5	Doctor	10
5	Doctor	11

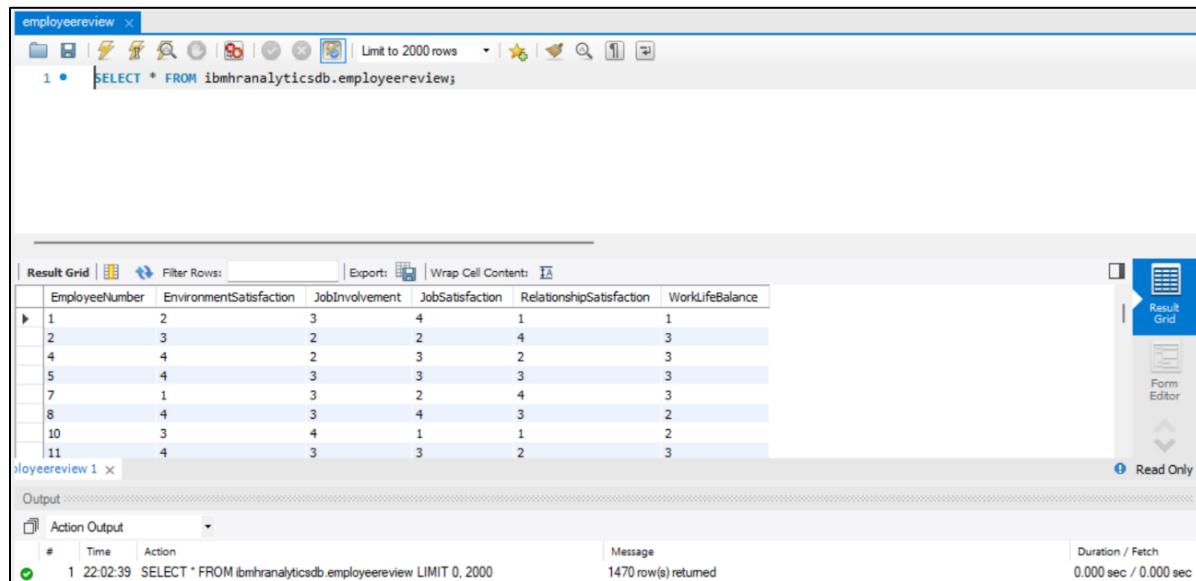
Output details:

#	Time	Action
1	22:02:21	SELECT * FROM ibmhranalyticsdb.education LIMIT 0, 2000

Message: 1470 row(s) returned

Duration / Fetch: 0.000 sec / 0.000 sec

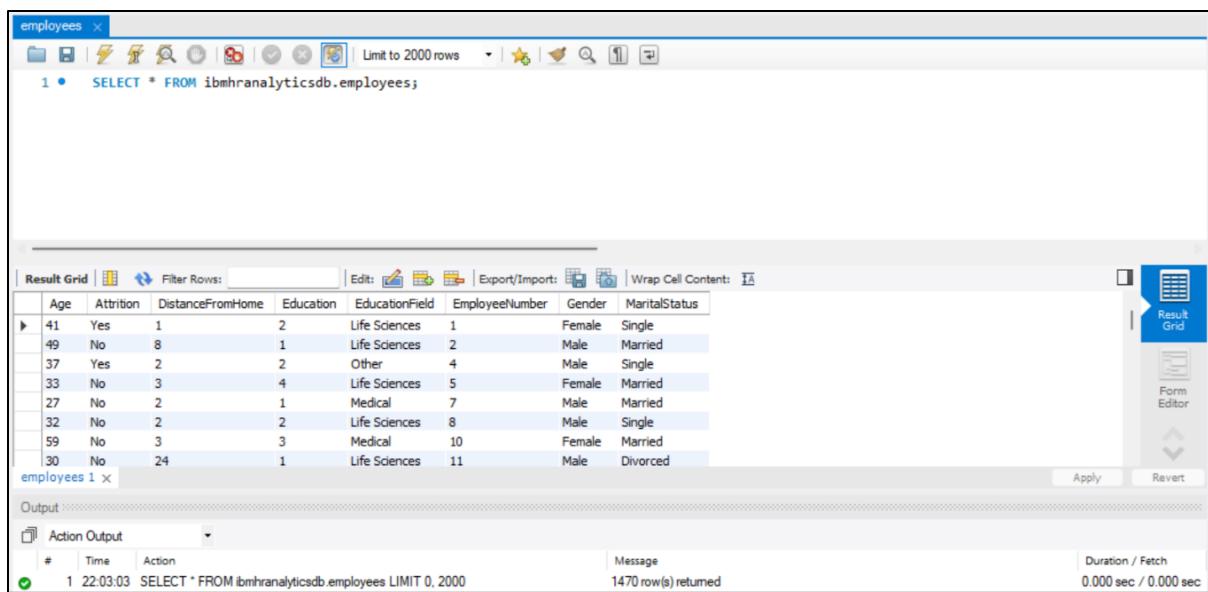
## Employee Review:



The screenshot shows a database query results window titled "employee review". The query executed is "SELECT \* FROM ibmhranalyticsdb.employeereview;". The results are displayed in a "Result Grid" table with the following columns: EmployeeNumber, EnvironmentSatisfaction, JobInvolvement, JobSatisfaction, RelationshipSatisfaction, and WorkLifeBalance. The data consists of 11 rows of employee reviews. The "Output" section at the bottom shows the query log with one entry: "1 22:02:39 SELECT \* FROM ibmhranalyticsdb.employeereview LIMIT 0, 2000" and a message indicating "1470 row(s) returned".

EmployeeNumber	EnvironmentSatisfaction	JobInvolvement	JobSatisfaction	RelationshipSatisfaction	WorkLifeBalance
1	2	3	4	1	1
2	3	2	2	4	3
4	4	2	3	2	3
5	4	3	3	3	3
7	1	3	2	4	3
8	4	3	4	3	2
10	3	4	1	1	2
11	4	3	3	2	3

## Employees:



The screenshot shows a database query results window titled "employees". The query executed is "SELECT \* FROM ibmhranalyticsdb.employees;". The results are displayed in a "Result Grid" table with the following columns: Age, Attrition, DistanceFromHome, Education, EducationField, EmployeeNumber, Gender, and MaritalStatus. The data consists of 11 rows of employee information. The "Output" section at the bottom shows the query log with one entry: "1 22:03:03 SELECT \* FROM ibmhranalyticsdb.employees LIMIT 0, 2000" and a message indicating "1470 row(s) returned".

Age	Attrition	DistanceFromHome	Education	EducationField	EmployeeNumber	Gender	MaritalStatus
41	Yes	1	2	Life Sciences	1	Female	Single
49	No	8	1	Life Sciences	2	Male	Married
37	Yes	2	2	Other	4	Male	Single
33	No	3	4	Life Sciences	5	Female	Married
27	No	2	1	Medical	7	Male	Married
32	No	2	2	Life Sciences	8	Male	Single
59	No	3	3	Medical	10	Female	Married
30	No	24	1	Life Sciences	11	Male	Divorced

### Environment Satisfaction:

The screenshot shows a database interface with a query window titled "environmentsatisfaction x". The query is:

```
1 • | SELECT * FROM ibmhranalyticsdb.environmentsatisfaction;
```

The results are displayed in a "Result Grid" table:

	EnvironmentSatisfaction	Description	EmployeeNumber
▶	2	Medium	1
	3	High	2
	4	Very High	4
	4	Very High	5
	1	Low	7
	4	Very High	8
	3	High	10
	4	Very High	11

The interface includes a sidebar with "Result Grid", "Form Editor", and "Read Only" buttons. Below the grid is an "Output" section with an "Action Output" table:

#	Time	Action	Message	Duration / Fetch
1	22-03-27	SELECT * FROM ibmhranalyticsdb.environmentsatisfaction LIMIT 0, 2000	1470 row(s) returned	0.016 sec / 0.000 sec

### Job Involvement:

The screenshot shows a database interface with a query window titled "jobinvolvement x". The query is:

```
1 • | SELECT * FROM ibmhranalyticsdb.jobinvolvement;
```

The results are displayed in a "Result Grid" table:

	JobInvolvement	Description	EmployeeNumber
▶	3	High	1
	2	Medium	2
	2	Medium	4
	3	High	5
	3	High	7
	3	High	8
	4	Very High	10
	3	High	11

The interface includes a sidebar with "Result Grid", "Form Editor", and "Read Only" buttons. Below the grid is an "Output" section with an "Action Output" table:

#	Time	Action	Message	Duration / Fetch
1	22-03-44	SELECT * FROM ibmhranalyticsdb.jobinvolvement LIMIT 0, 2000	1470 row(s) returned	0.000 sec / 0.000 sec

## Job Satisfaction:

The screenshot shows a database interface with a query window and a results grid.

**Query Window:**

```
jobsatisfaction
1 • | SELECT * FROM ibmhranalyticsdb.jobsatisfaction;
```

**Results Grid:**

	JobSatisfaction	Description	EmployeeNumber
▶	4	Very High	1
2	Medium	2	
3	High	4	
3	High	5	
2	Medium	7	
4	Very High	8	
1	Low	10	
3	High	11	

**Action Output:**

#	Time	Action	Message	Duration / Fetch
1	22:04:01	SELECT * FROM ibmhranalyticsdb.jobsatisfaction LIMIT 0, 2000	1470 row(s) returned	0.015 sec / 0.000 sec

## Performance Rating:

The screenshot shows a database interface with a query window and a results grid.

**Query Window:**

```
performancerating
1 • | SELECT * FROM ibmhranalyticsdb.performancerating;
```

**Results Grid:**

	PerformanceRating	Description	EmployeeNumber
▶	3	High	1
4	Very High	2	
3	High	4	
3	High	5	
3	High	7	
3	High	8	
4	Very High	10	
4	Very High	11	

**Action Output:**

#	Time	Action	Message	Duration / Fetch
1	22:04:45	SELECT * FROM ibmhranalyticsdb.performancerating LIMIT 0, 2000	1470 row(s) returned	0.015 sec / 0.000 sec

## Performance and Salary:

The screenshot shows a database interface with a query window titled "performanceandsalary". The query is:

```
1 • | SELECT * FROM ibmhranalyticsdb.performanceandsalary;
```

The results grid displays the following data:

PercentSalaryHike	PerformanceRating	EmployeeNumber
<20	3	1
>=20	4	2
<20	3	4
<20	3	5
<20	3	7
<20	3	8
>=20	4	10
>=20	4	11

The output pane shows the following log entry:

#	Time	Action	Message	Duration / Fetch
1	22:04:23	SELECT * FROM ibmhranalyticsdb performanceandsalary LIMIT 0, 2000	1470 row(s) returned	0.000 sec / 0.000 sec

## Relationship Satisfaction:

The screenshot shows a database interface with a query window titled "relationshipsatisfaction". The query is:

```
1 • | SELECT * FROM ibmhranalyticsdb.relationshipsatisfaction;
```

The results grid displays the following data:

RelationshipSatisfaction	Description	EmployeeNumber
1	Low	1
4	Very High	2
2	Medium	4
3	High	5
4	Very High	7
3	High	8
1	Low	10
2	Medium	11

The output pane shows the following log entry:

#	Time	Action	Message	Duration / Fetch
1	22:05:06	SELECT * FROM ibmhranalyticsdb relationshipsatisfaction LIMIT 0, 2000	1470 row(s) returned	0.000 sec / 0.000 sec

## Salaries:

The screenshot shows the DB2 Workbench interface with a query window titled 'salaries'. The query executed is: `SELECT * FROM ibmhranalyticsdb.salaries;`. The result grid displays 1470 rows of data from the 'salaries' table. The columns are EmployeeNumber, DailyRate, HourlyRate, MonthlyIncome, MonthlyRate, Overtime, PercentSalaryHike, and StockOptionLevel. The data includes various employee details like rate per hour or day, monthly income, overtime status, and stock options level. The interface also shows an action output log at the bottom.

EmployeeNumber	DailyRate	HourlyRate	MonthlyIncome	MonthlyRate	Overtime	PercentSalaryHike	StockOptionLevel
1	1102	94	5993	19479	Yes	11	0
2	279	61	5130	24907	No	23	1
4	1373	92	2090	2396	Yes	15	0
5	1392	56	2909	23159	Yes	11	0
7	591	40	3468	16632	No	12	1
8	1005	79	3068	11864	No	13	0
10	1324	81	2670	9964	Yes	20	3
11	1358	67	2693	13335	No	22	1

## Work Experience:

The screenshot shows the DB2 Workbench interface with a query window titled 'workexperience'. The query executed is: `SELECT * FROM ibmhranalyticsdb.workexperience;`. The result grid displays 1470 rows of data from the 'workexperience' table. The columns are BusinessTravel, EmployeeNumber, NumCompaniesWorked, TotalWorkingYears, TrainingTimesLastYear, YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, and YearsV. The data includes travel frequency, number of companies worked, total working years, training times last year, and years at current role. The interface also shows an action output log at the bottom.

BusinessTravel	EmployeeNumber	NumCompaniesWorked	TotalWorkingYears	TrainingTimesLastYear	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsV
Travel_Rarely	1	8	8	0	6	4	0	5
Travel_Frequently	2	1	10	3	10	7	1	7
Travel_Rarely	4	6	7	3	0	0	0	0
Travel_Frequently	5	1	8	3	8	7	3	0
Travel_Rarely	7	9	6	3	2	2	2	2
Travel_Frequently	8	0	8	2	7	7	3	6
Travel_Rarely	10	4	12	3	1	0	0	0

## Work-Life Balance:

The screenshot shows a database interface with the following details:

- Title Bar:** worklifebalance
- Toolbar:** Includes icons for file operations, search, and navigation.
- Query Editor:** Displays the SQL command: `SELECT * FROM ibmhranalyticsdb.worklifebalance;`
- Result Grid:** Shows the data from the query:

	WorkLifeBalance	Description	EmployeeNumber
1	Low	1	
3	High	2	
3	High	4	
3	High	5	
3	High	7	
2	Medium	8	
2	Medium	10	
3	High	11	
- Output Panel:** Shows the execution log:

#	Time	Action	Message	Duration / Fetch
1	22:06:03	SELECT * FROM ibmhranalyticsdb.worklifebalance LIMIT 0, 2000	1470 row(s) returned	0.000 sec / 0.000 sec
- Status:** Read Only

## Data in the Database

<b>Table Name</b>	<b>Primary Key</b>	<b>Foreign Key</b>	<b># of Rows in Table</b>
Department roles		EmployeeNumber	1470
Education		EmployeeNumber	1470
Employee review		EmployeeNumber	1470
Employees	EmployeeNumber		1470
Environmentsatisfaction		EmployeeNumber	1470
Jobinvolvement		EmployeeNumber	1470
Jobsatisfaction		EmployeeNumber	1470
performanceandsalary		EmployeeNumber	1470
Performance rating		EmployeeNumber	1470
Relationshipsatisfaction		EmployeeNumber	1470
Salaries		EmployeeNumber	1470
Work experience		EmployeeNumber	1470
worklifebalance		EmployeeNumber	1470

## SQL Queries

### SQL Query 2

#### Question

Which department's employee is the most likely to have the shortest commute between home and work?

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

1. This query filters the department name based on the employee with the least distance from home value.
2. The result set contains 1 row

#### Translation

Select the department from departmentroles table where employeeNumber in the departmentroles table equals the selection of employeeNumber from the selection of employeeNumber and the minimum distanceFromHome from the employees table.

#### Query

```
select department from departmentroles  
where employeeNumber = (select a.employeeNumber from (select employeeNumber,  
min(distanceFromHome) from employees) a);
```

#### Explanation

From the query, an employee from Sales has the shortest commute between home and work.

## Screen Shot of SQL Query and Results

```
2  /*  
3   2. Which department's employee is the most likely to have the shortest commute between home and work?  
4 */  
5  
6 •  select department from departmentroles  
7  ⚡ where employeeNumber = (select a.employeeNumber  
8    from (select employeeNumber, min(distanceFromHome)  
9         from employees) a)  
10 ;  
11  
12 -- From the query, an employee from Sales has the shortest commute between home and work.  
13
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

department
Sales

departmentroles 1 ×

Output

Action Output	
# Time Action Message	
1 16:10:47 USE 'ibmhranalyticsdb'	0 row(s) affected
2 16:10:48 select department from departmentroles where employeeNumber = (select a.employeeNumber from (select employeeNumber, min(distanceFromHom... 1 row(s) returned	

## SQL Query 4

### Question

The HR department feels they have the highest job satisfaction while Research & Development department feels their department has the highest environment satisfaction. Who is right?

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

1. This query filters the department name and the employees' average job satisfaction and average environment satisfaction.
2. The result set contains 3 rows

### Translation

Select the department, average JobSatisfaction, average EnvironmentSatisfaction from jobsatisfaction table, departmentroles table, and environmentsatisfaction table where the employeenumber of jobsatisfaction table equals the employeenumber of departmentroles table, and the employeenumber of departmentroles table equals the employeenumber of environmentsatisfaction table, and then group by department.

### Query

```
select department, avg(JobSatisfaction) as avg_job_satisfaction, avg(EnvironmentSatisfaction) as avg_env_satisfaction  
from jobsatisfaction j, departmentroles d, environmentsatisfaction e  
where j.employeenumber = d.employeenumber  
and d.employeenumber = e.employeenumber  
group by department;
```

### Explanation

From the above query it is clear that Research & Development have a higher average environment satisfaction while the highest job satisfaction is in the sales department and not the HR department.

## Screen Shot of SQL Query and Results

The screenshot shows a SQL query editor interface. At the top, there is a multi-line text area containing a SQL script. Below this is a 'Result Grid' table showing the output of the query. At the bottom, there is a 'Result 3' section with an 'Output' tab displaying a log of the query execution.

```
14  /*  
15   4. The HR department feels they have the highest job satisfaction while  
16   Research & Development department feels their department has the highest environment satisfaction. Who is right?  
17 */  
18  
19 • select department, avg(JobSatisfaction) as avg_job_satisfaction, avg(EnvironmentSatisfaction) as avg_env_satisfaction  
20   from jobsatisfaction j, departmentroles d, environmentsatisfaction e  
21   where j.employeenumber = d.employeenumber  
22   and d.employeenumber = e.employeenumber  
23   group by department  
24   j  
25
```

department	avg_job_satisfaction	avg_env_satisfaction
Sales	2.7511	2.6794
Research & Development	2.7263	2.7440
Human Resources	2.6032	2.6825

Result 3 x

Output

Action Output	#	Time	Action	Message
1 16:11:48 select department, avg(JobSatisfaction) as avg_job_satisfaction, avg(EnvironmentSatisfaction) as avg_env_satisfaction from jobsatisfaction j, depart...	3	row(s) returned		

## SQL Query 5

### Question

An employee in Sales department has complained to HR saying that females are paid less than males in the company, in all departments.

What insight can you provide to prove or disprove that statement?

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

1. This query filters the department name, gender and the average monthly income grouped by department and gender.
2. The result set contains 6 rows

### Translation

Select the department, gender, average monthlyincome from the salaries table, departmentroles table, and employees table where the employeenumber of salaries table equals the employeenumber of departmentroles table and the employeenumber of employees table equals the employeenumber of departmentroles table, and then group by the department of departmentroles table and the gender of employees table, and order by the department of departmentroles table and the gender of employees table.

### Query

```
select d.department, e.gender, avg(s.monthlyincome) as avegrage_salary_by_dept_gender  
from salaries s, departmentroles d, employees e
```

```

where s.employeenumber = d.employeenumber
and e.employeenumber = d.employeenumber
group by d.department, e.gender
order by d.department, e.gender;

```

### Explanation

From the above query, it is clear that females have a higher monthly income as compared to males in all departments.

### Screen Shot of SQL Query and Results

The screenshot shows a SQL query editor interface. At the top, there is a question:

```

29  /*
30  5. An employee in Sales department has complained to HR saying that females are paid less than males in the company, in all departments.
31  What insight can you provide to prove or disprove that statement?
32  */
33
34 • select d.department, e.gender, avg(s.monthlyincome) as avegrage_salary_by_dept_gender
35   from salaries s, departmentroles d, employees e
36   where s.employeenumber = d.employeenumber
37   and e.employeenumber = d.employeenumber
38   group by d.department, e.gender
39   order by d.department, e.gender
40 ;

```

Below the question is the query itself. The result grid shows the following data:

department	gender	avegrage_salary_by_dept_gender
Human Resources	Female	7264.0000
Human Resources	Male	6371.0233
Research & Development	Female	6513.6913
Research & Development	Male	6129.8883
Sales	Female	6972.1270
Sales	Male	6949.6459

The output pane at the bottom shows the command run and the message "6 row(s) returned".

### SQL Query 6

#### Question

A press article in a business magazine has said that at this company, married men have higher performance ratings than divorced or single men.

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)

1. This query filters the data based on the marital status and average performance rating of employees grouped by the marital statuses.
2. The result set contains 3 rows

## Translation

Select the maritalstatus, average performancerating from the performancerating table, and the employees table where the employeenumber of employees table equals the employeenumber of performancerating table and the gender of employees table equals male, and then group by the maritalstatus of employees table.

## Query

```
select e.maritalstatus, avg(p.performancerating) as avg_performance_rating from performancerating p,
employees e
where e.employeenumber = p.employeenumber
and e.gender = 'Male'
group by e.maritalstatus;
```

## Explanation

From the above query it is evident that single men have the highest average performance rating followed by married men and then by divorced men.

## Screen Shot of SQL Query and Results

The screenshot shows a SQL query editor interface. The query is:

```
/*
6. A press article in a business magazine has said that at this company, married men have higher performance ratings than divorced or single men.
What initial finding can you obtain from the data to help articulate the company's response in this regard?
*/
49 • select e.maritalstatus, avg(p.performancerating) as avg_performance_rating from performancerating p, employees e
50 where e.employeenumber = p.employeenumber
51 and e.gender = 'Male'
52 group by e.maritalstatus
53 ;
54
55 -- From the above query it is evident that single men have the highest average performance rating followed by married men and then by divorced men.
```

The results table shows the average performance rating by marital status:

maritalstatus	avg_performance_rating
Married	3.1471
Single	3.1624
Divorced	3.1381

The message bar at the bottom indicates the query was run at 16:12:41 and returned 3 rows.

## SQL Query 7

### Question

If the company wants to cut travel costs, which department should the company focus on?

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

1. This query filters the data based on the department name, business travel occurrence and count of employees travelling per department.

## 2. The result set contains 9 rows

### Translation

Select the department, businesstravel, and the count of businesstravel from the workexperience table and the departmentroles table where the employee number of workexperience table equals the employee number of departmentroles table, and then group by the department of departmentroles table and the businesstravel of workexperience table, and order by the count of businesstravel from the workexperience table by descending.

### Query

```
select d.department, w.businesstravel, count(w.businesstravel) as count from workexperience w,  
departmentroles d  
where w.employeeNumber = d.employeeNumber  
group by d.department, w.businesstravel  
order by count(w.businesstravel) desc;
```

### Explanation

From the above query, Research and development department has the people who travel. Hence Research & Development should be the first department to focus on for cutting travel costs.

### Screen Shot of SQL Query and Results

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

```
55 -- From the above query it is evident that single men have the highest average performance rating followed by married men and then by divorced men.  
56  
57 /*  
58 7. If the company wants to cut travel costs, which department should the company focus on?  
59 */  
60  
61 • select d.department, w.businesstravel, count(w.businesstravel) as count from workexperience w, departmentroles d  
62 where w.employeeNumber = d.employeeNumber  
63 group by d.department, w.businesstravel  
64 order by count(w.businesstravel) desc  
65 ;  
66
```

Result Grid

department	businesstravel	count
Research & Development	Travel_Rarely	682
Sales	Travel_Rarely	315
Research & Development	Travel_Frequently	182
Research & Development	Non-Travel	97
Sales	Travel_Frequently	84
Sales	Non-Travel	47

Output

Action Output
# Time Action
1 16:13:06 select d.department, w.businesstravel, count(w.businesstravel) as count from workexperience w, departmentroles d where w.employeeNumber = d.e... 9 row(s) returned

## Data Review for MongoDB

### Assumptions/Notes About Data Collections, Attributes and Relationships between Collections

- The raw dataset is loaded into Mongo Server because it is capable of handling unstructured data too. This is now also used to compare result sets fetched from Mongo with SQL result sets
- Data is loaded into Mongo using Compass and visualized using Compass and Mongo Shell
- The data base name is HR and the collection name is hra.
- The data consists of 1470 rows/documents. Each document pertains to a different employee.
- The database has only one collection of all the documents. So, we don't need to join or perform any \$lookup operation on the documents.
- Easier access is available because there is only one collection.
- The attributes are numeric and string in nature with each field being of the respective data type.

## Physical Mongo Database

### Assumptions/Notes About Data Set

- We have imported raw csv file given into MongoDB server through Mongo compass and there are no duplicate values and each record is identified by unique id '\_id' column.
- Attrition implies that the particular person is no more employed.
- The data base name is HR and the collection name is hra.
- The data consists of 1470 rows/documents. Each document pertains to a different employee.
- The database has only one collection of all the documents. So, we don't need to join or perform any \$lookup operation on the documents.
- Easier access is available because there is only one collection.
- Any person rejoining the company would have a different EmployeeNumber and his/his details would be listed as any new employee's would, i.e., YearsAtCompany and YearsWithCurrManager will be calculated with respect to this position and this EmployeeNumber. NumCompaniesWorked will also include this company.
- TrainingTimesLastYear are the number of trainings the employee has undertaken with this company in the last year
- StockOptionLevel, which is part of salary for some employees
- All employees are paid monthly salaries. The hourly rate, daily rate and monthly rate are the rates that were paid to contractors that are not on the company payroll but recruited on the same day. The contract employees are not listed here.
- Every department has specific roles in it, which do not overlap with any other department
- For employees with rating = 3 salary hike is < 20% while those with rating = 4 get a hike of  $\geq 20\%$

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

The screenshot shows the MongoDB Compass interface. The top bar displays the title "BUAN 6320\_SQL-MongoProject-Phase 1 and Phase 2\_Group2" and the connection details "MongoDB Compass - cluster0.v9yvrcb.mongodb.net/HR.hra". The status bar indicates "71%" battery, "Sat Dec 10 11:52 AM", and a search bar.

The left sidebar shows the database structure with "cluster0.v9yvrcb" selected. Under "Databases", the "HR" database is expanded, showing the "hra" collection, which is currently selected. Other collections listed are "Test", "admin", "config", and "local".

The main pane is titled "HR.hra" and displays the following information:

- Count: 1.5k DOCUMENTS, 1 INDEXES
- Documents tab (selected): Shows a list of document fields:
  - `_id: ObjectId('638e8513eb1f367fd6e1de57')`
  - `Age: 41`
  - `Attrition: "Yes"`
  - `BusinessTravel: "Travel_Rarely"`
  - `DailyRate: 1102`
  - `Department: "Sales"`
  - `DistanceFromHome: 1`
  - `Education: 2`
  - `EducationField: "Life Sciences"`
  - `EmployeeCount: 1`
  - `EmployeeNumber: 1`
  - `EnvironmentSatisfaction: 2`
  - `Gender: "Female"`
  - `HourlyRate: 94`
  - `JobInvolvement: 3`
  - `JobLevel: 2`
  - `JobRole: "Sales Executive"`
  - `JobSatisfaction: 4`
  - `MaritalStatus: "Single"`
  - `MonthlyIncome: 5993`
  - `MonthlyRate: 19479`
  - `NumCompaniesWorked: 8`
  - `Over18: "Yn"`
  - `OverTime: "Yes"`
  - `PercentSalaryHike: 11`
- Aggregations, Schema, Explain Plan, Indexes, Validation tabs
- Filter, Find, More Options buttons
- Add Data, Export Collection buttons
- Document list: 1 - 20 of 1470

## Data in the Database

<b>Collection Name</b>	<b>Relationships With Other Collections (if any)</b>	<b># of Documents in Collection</b>
hra	NA	1470

# MongoDB Queries/Code

## Mongo Query 9

### Question

A new employee from a Life Sciences education field wants to work in Sales. Do you believe the company might be able to give him a chance to work in Sales? Why or Why not?

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

1. The aggregate function has grouped by Department and Education field to find out the degrees that have been employed by respective department using \$group. Then the \$match statement is filtering out the result only for Sales department because we are only concerned about Sales department.

Number of documents - 5

### Translation

Group documents by Department and EducationField and count the number of people working in each department where Department='Sales'

### Query

```
db.hra.aggregate([{$match:{'Department':{$eq:"Sales"}},{$group:{_id:{Department:"$Department",EducationField:"$EducationField"},count:{$sum:1}}}]})
```

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

The screenshot shows the MongoDB Compass interface. The left sidebar lists databases (cluster0.v9yvrcb.mongodb.net) and collections (HR, hra). The main area displays the HR.hra collection with 1.5k documents and 1 index. A query results table shows the following data:

_id	Department	EducationField	count
{ _id: { Department: 'Sales', EducationField: 'Marketing' }, count: 159 }	Sales	Marketing	159
{ _id: { Department: 'Sales', EducationField: 'Technical Degree' }, count: 34 }	Sales	Technical Degree	34
{ _id: { Department: 'Sales', EducationField: 'Medical' }, count: 88 }	Sales	Medical	88
{ _id: { Department: 'Sales', EducationField: 'Life Sciences' }, count: 150 }	Sales	Life Sciences	150
{ _id: { Department: 'Sales', EducationField: 'Other' }, count: 15 }	Sales	Other	15

The bottom of the screen shows the mongo shell command entered:

```
> db.hra.aggregate([{$match:{'Department':{$eq:"Sales"}},{$group:{_id:{Department:"$Department",EducationField:"$EducationField"},count:{$sum:1}}}}])
```

## Results

Yes, we believe the company might be able to give him a chance because in the Sales department there are 150 people with education field in Life sciences which is the second highest after Marketing. Since he also holds an education field in life sciences there are high chances he might be employed by the Sales department.

## Mongo Query 10

### Question

HR feels that their environment satisfaction score is higher than Sales 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)

1. The 2 queries are grouping documents by department and finding the aggregate sum of the respective job satisfaction scores and the environmentsatisfaction scores. Then one can easily figure out the lowest and the highest scores for each department and answer the query.
2. Number of documents – 3 & 3

### Translation

Group documents by Department and sum the jobsatisfaction scores to find overall job satisfaction per department

Group documents by Department and sum the EnvironmentSatisfaction scores to find overall EnvironmentSatisfaction per department

### Query

```
db.hra.aggregate([{$group:{_id:"$Department",JobSatisfaction:{$sum:"$JobSatisfaction"}}}])
```

```
db.hra.aggregate([{$group:{_id:"$Department",EnvironmentSatisfaction:{$sum:"$Environment Satisfaction"}}}])
```

## Screen Shot of MongoDB Query/Code and Results

The screenshot shows the MongoDB Compass application running on a Mac OS X desktop. The left sidebar lists databases: HR (selected), Test, admin, config, and local. The main pane displays the 'HR.hra' collection with 1.5k documents and 1 index. A search bar at the top right contains the query '{ field: 'value' }'. Below it are tabs for Documents, Aggregations, Schema, Explain Plan, Indexes, and Validation. A 'Filter' dropdown is open. The 'Documents' tab shows a list of fields: EmployeeCount: 1, EmployeeNumber: 4, EnvironmentSatisfaction: 4, Gender: "Male", HourlyRate: 92, JobInvolvement: 2, JobLevel: 1, JobRole: "Laboratory Technician", JobSatisfaction: 3, MaritalStatus: "Single", MonthlyIncome: 2090, and MonthlyRate: 2396. At the bottom, a terminal window shows MongoDB shell code for aggregation:

```
> db.hra.aggregate([{$group:{_id:"$Department",JobSatisfaction:{$sum:"$JobSatisfaction"}}}] )
< {_id: 'Sales', JobSatisfaction: 1227 }
{ _id: 'Research & Development', JobSatisfaction: 2620 }
{ _id: 'Human Resources', JobSatisfaction: 164 }
> db.hra.aggregate([{$group:{_id:"$Department",EnvironmentSatisfaction:{$sum:"$EnvironmentSatisfaction"}}}])
< {_id: 'Sales', EnvironmentSatisfaction: 1195 }
{ _id: 'Research & Development', EnvironmentSatisfaction: 2637 }
{ _id: 'Human Resources', EnvironmentSatisfaction: 169 }
Atlas atlas-ky9qo1-shard-0 [primary] HR>
```

## Results

No the environment satisfaction score for HR is not higher than Sales, it is in fact the lowest amongst all the departments. Yes, they right when they say the Job satisfaction of HR department is lower than that of Research and Development. HR department has the lowest score for both job satisfaction and environment satisfaction.

## Mongo Query 11

### Question

An employee from Medical education field working in Sales department has spread a rumor saying that employees with his educational background are paid more in Research & Development than in Sales. What insight can you provide to prove or disprove that statement?

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

1. Group documents by Department and Education Field and find the average of the monthly income for each to find out the which department has on average a higher salary. Out of all the documents returned \$match is used to filter out the documents where EducationField is Medical since we are only concerned for Medical EducationField.
2. Number of documents – 3

## Translation

Group documents by Department and Education Field, calculate average salary using monthlyincome, match where EducationField is equal to Medical.

## Query

```
db.hra.aggregate([{$match:{EducationField:{$eq:"Medical"}},{$group:{_id:{Department:"$Department",EducationField:"$EducationField"},AverageMonthlySalary:{$avg:"$MonthlyIncome"}}}]
```

## Screen Shot of MongoDB Query/Code and Results

The screenshot shows the MongoDB Compass interface. On the left, the sidebar displays databases (cluster0.v9yvrcb.mongodb.net/HR.hra) and collections (hra). The main area is titled 'HR.hra' and shows the 'Documents' tab. A search bar at the top right contains the query: 'Type a query: { field: 'value' }'. Below it, there are buttons for 'Filter' (with a dropdown), 'Reset', 'Find', and a copy icon. The results table shows one document with the following fields and values:

Field	Value
JobInvolvement	3
JobLevel	1
JobRole	"Research Scientist"
JobSatisfaction	3
MaritalStatus	"Married"
MonthlyIncome	2989
MonthlyRate	23159
NumCompaniesWorked	1
Over18	"y"
OverTime	"Yes"
PercentSalaryHike	11

At the bottom of the interface, the mongo shell command is shown:

```
> db.hra.aggregate([{$match:{EducationField:{$eq:"Medical"}},{$group:{_id:{Department:"$Department",EducationField:"$EducationField"},AverageMonthlySalary:{$avg:"$MonthlyIncome"}}}])
```

And the results from the shell:

```
< _id: { Department: 'Research & Development', EducationField: 'Medical' }, AverageMonthlySalary: 6539.2231404958675 } { _id: { Department: 'Sales', EducationField: 'Medical' }, AverageMonthlySalary: 6377.227272727273 } { _id: { Department: 'Human Resources', EducationField: 'Medical' }, AverageMonthlySalary: 6594.076923076923 }
```

## Results

As we can see from the results average monthly salary for Medical Education field employees is highest for HR department and lowest for Sales department. Yes the claims of this person are true as the average salary for Medical Educationfield employees is higher in Research and Development than Sales department.