# Databases - Lab

Your lab consists of several parts, explained below. You may work independently on each part. Submit your solutions in the automated judge system: <http://judge.softuni.bg/Contests/70/Databases-Sample-Exam>.

## Part I – SQL Queries

You are given a **MS SQL Server database "Forum"** holding questions, answers, categories, tags, votes and users, available as SQL script. Your task is to write SQL queries for displaying data from the given database.

### All Question Titles

Display all **question titles in descending order**. Name the columns exactly like in the table below.

Steps:

1. Select all question titles from questions table.
2. Add order by descending statement.

|  |
| --- |
| **Title** |
| Can I create a SQLite view UNIONING all databases? |
| Catch multiple Exceptions at once? |
| Cycles in family tree software |
| … |

Submit for evaluation the result grid **with headers**.

1 score

### Answers in Date Range

Find **all answers** created between **15-June-2012** (00:00:00) and **21-Mart-2013** (23:59:59) sorted by date. Name the columns exactly like in the table below.

Steps:

1. Select all answers (only **Content** and **CreatedOn** columns).
2. Add where filter by **CreatedOn** column. You can use between or comparison statements.

|  |  |
| --- | --- |
| **Content** | **CreatedOn** |
| It says here that "... the number-of-calls figures are derived by counting, not sampling. They are completely accurate...". Yet I find my call graph giving me 5345859132+784984078 as call stats to my most called function, where the first number is supposed to be direct calls, and the second recursive calls (which are all from itself). Since this implied I had a bug, I put in long (64bit) counters into the code and did the same run again. My counts : 5345859132 direct, and 78094395406 self-recursive calls. There's a lot of digits there, so I'll point out the recursive calls I measure are 78bn, versus 784m from gprof : a factor of 100 different. Both runs were single threaded and unoptimised code, one compiled -g and the other -pg. | 2012-06-26 10:48:34.000 |
| System databases with ID 5 and 6 will be ReportServer and ReportServerTempDB if you have SQL Server Reporting Services installed. | 2012-06-29 17:57:40.000 |
| … | … |

Submit for evaluation the result grid **with headers**.

2 score

### Users with "1/0" Phones

Display all user **username** and **last names** along with a column named "**Has Phone**" holding "**1**" or "**0**" for all users sorted by their last name. Name the columns exactly like in the table below.

Steps:

1. Select all users (only **Username** and **LastName** columns).
2. Add order by last name statement.
3. Add “**Has Phone**” column. You can use **case** statement in select to check is there a phone and set 1 or 0.

|  |  |  |
| --- | --- | --- |
| **Username** | **LastName** | **Has Phone** |
| Joiner | Abatangelo | 0 |
| Rafael | Afonso | 0 |
| casper | Beyer | 1 |
| … | … | … |

Submit for evaluation the result grid **with headers**.

4 score

### Questions with their Author

Find all **questions along with their user** sorted by Id. Display the question **title** and **author username**. Name the columns exactly like in the table below.

Steps:

1. Select all columns from **Questions** table.
2. Join **Users** table.
3. Add only the required columns in the expression.
4. Add the required aliases.

|  |  |
| --- | --- |
| **Question Title** | **Author** |
| Why is processing a sorted array faster than an unsorted array? | Joiner |
| Why is subtracting these two times (in 1927) giving a strange result? | JonSkeet |
| … | … |

Submit for evaluation the result grid **with headers**.

2 score

### Answers with their Question with their Category and User

Find all **answers along with their questions, along with question category, along with question author** sorted by **Category Name**. Display the answer content, created on, question title, category name and author username. Name the columns exactly like in the table below.

Steps:

1. Select all columns from **Answers** table.
2. Join **Questions** table by QuestionId foreign key.
3. Join **Categories** table by CategoryId foreign key.
4. Join **Users** table by UserId foreign key in Answers table.
5. Add **order by** statement by **category name**.
6. **Select** only the **required columns** with their **aliases**.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Answer Content** | **CreatedOn** | **Answer Author** | **Question Title** | **Category Name** |
| According to the language specification, the language itself must consider string to be exactly the same as the BCL type System.String, nothing else. That is not ambiguous at all. Of course, you can implement your own compiler, using the C# grammar, and use all of the tokens found like that for something arbitrary, unrelated to what is defined in the C# language specification. However, the resulting language would only be a C# lookalike, it could not be considered C# | 2012-01-23 07:37:21.000 | AGreenman | What's the difference between String and string? | C# |
| System.String is THE .net string class - in C# string is an alias for System.String - so in use they are the same. | 2013-09-08 20:35:09.000 | StefanG | What's the difference between String and string? | C# |
| … | … | … | … |  |

Submit for evaluation the result grid **with headers**.

3 score

### Category with Questions

Find all **categories** along with their **questions** **sorted by category name**. Display the category name, question title and created on columns. Name the columns exactly like in the table below.

1. Select all **Questions**.
2. Join **Categories** table (think how to display null values).
3. Add **order by** statement by **category name**.
4. Select only **required columns**.

|  |  |  |
| --- | --- | --- |
| **Name** | **Title** | **CreatedOn** |
| .NET | NULL | NULL |
| Algorithms | NULL | NULL |
| Android | NULL | NULL |
| AngularJS | NULL | NULL |
| Artifficial Intelligence | NULL | NULL |
| C# | What's the difference between String and string? | * + 1. 4:59:51.990 |

Submit for evaluation the result grid **with headers**.

3 score

### \*Users without PhoneNumber and Questions

Find **all users** that have **no phone** and **no questions** sorted by **RegistrationDate**. Show all user data. Name the columns exactly like in the table below.

1. Select all users.
2. Add **where** filter by **null phone number**.
3. Add **order by** statement by **RegistrationDate**.
4. Join **Questions** table and display all users with questions for each user (with questions with null values).
5. Add filter to show only users with no questions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Id** | **Username** | **FirstName** | **PhoneNumber** | **RegistrationDate** | **Email** |
| 31 | quirk | Patrick | NULL | 2010-04-19 | quirk@mail.com |
| 38 | Kaviraj | Kaviraj | NULL | 2011-06-13 | kaviraj@microsoft.com |
| … | … | … | … | … | … |

Submit for evaluation the result grid **with headers**.

5 score

### Earliest and Latest Answer Date

Find the dates of the earliest answer for 2012 year and the latest answer for 2014 year. Name the columns exactly like in the table below.

Steps:

1. Select min and max date from **Answers**.
2. Add **where** filter **by year**.

|  |  |
| --- | --- |
| **MinDate** | **MaxDate** |
| 2012-01-11 00:02:01.000 | 2014-12-28 13:41:33.000 |

Submit for evaluation the result grid **with headers**.

1 score

### Find the latest ten answers

Find the **latest 10 answers** sorted by date of creation in descending order. Display for each ad its **content**, **date** and **author**. Name the columns exactly like in the table below.

|  |  |  |
| --- | --- | --- |
| **Content** | **CreatedOn** | **Author** |
| What's the point with this? It only makes the code slower and harder to read? I can't see a single advantage with it. 1u << n is easier to read for C programmers, and can hopefully be translated into a single clock tick CPU instruction. Your division on the other hand, will be translated to something around 10 ticks, or even as bad as up to 100 ticks, depending on how poorly the specific architecture handles division. As for the bitmap feature, it would make more sense to have a lookup table translating each bit index to a byte index, to optimize for speed. | 2011-01-16 20:42:23.000 | buch |
| We did in fact push back on this change in C# 3 and C# 4. When we designed C# 3 we did realize that the problem (which already existed in C# 2) was going to get worse because there would be so many lambdas (and query comprehensions, which are lambdas in disguise) in foreach loops thanks to LINQ. I regret that we waited for the problem to get sufficiently bad to warrant fixing it so late, rather than fixing it in C# 3. | 2011-01-17 02:09:59.000 | DanielB |
| … | … | … |

Submit for evaluation the result grid **with headers**.

2 score

### Not Published Ads from the First and Last Month

Find the answers which is not hidden from the first and last month where there are any published answer, from the last year where there are any published answers. Display for each ad its **answer content**, **question** **title** and **category name**. Sort the results by category name. Name the columns exactly like in the table below.

Steps:

1. Select all data from **Answers**.
2. Join **Questions** and **Categories** by foreign keys.
3. Add **order by** statement by **category name**.
4. Select only needed columns.
5. Add where statement. You can use **MONTH** and **YEAR** functions and nested select statements.

|  |  |  |
| --- | --- | --- |
| **Answer Content** | **Question** | **Category** |
| If you ask me, the second example (with the if and is keywords) is both significantly less readable, and simultaneously significantly more error-prone during the maintenance phase of your project. | Catch multiple Exceptions at once? | C# |
| P.S. This can also be done on multi-thread programs if there is a way to collect call-stack samples of the thread pool at a point in time, as there is in Java. | What can I use to profile C++ code in Linux? | C++ |
| … | … | … |

Submit for evaluation the result grid **with headers**.

4 score

### Answers count for Category

Display the **count of answers in each category**. Sort the results by answers count in descending order. Name the columns exactly like in the table below.

Steps:

1. Select all data from **Categories**.
2. Join **Questions** and **Answers** (What type of join do you need?).
3. Order results.
4. **Group** results and add **COUNT** column in select statement.

|  |  |
| --- | --- |
| **Category** | **Answers Count** |
| C# | 85 |
| C++ | 45 |
| … | … |

Submit for evaluation the result grid **with headers**.

2 score

### Answers Count by Category and Username

Display the **count of answers for category and each username**. Sort the results by **Answers count**. Display only non-zero counts. Display only users with phone number. Name the columns exactly like in the table below.

Steps:

1. Select all data from Categories.
2. Join Questions, Answers and Users (What type of joins do you need?).
3. Add needed columns in select statement and group by them.
4. Add order by statement by answers count.

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Username** | **PhoneNumber** | **Answers Count** |
| C# | HHendriks | 0887541254 | 4 |
| C# | StefanG | 0886947513 | 4 |
| C# | Onorio | 0895232112 | 4 |
| … | … | … |  |

Submit for evaluation the result grid **with headers**.

3 score

## Part II – Changes in the Database

You are given a **MS SQL Server database "Forum"** holding questions, answers, categories, tags, votes and users, available as SQL script. Your task is to modify the database schema and data and write SQL queries for displaying data from the database.

### Ads by Country

1. Create a **table Towns(Id, Name)**. Use auto-increment for the primary key. Add a new column **TownId** in the **Users table** to link each user to some town (non-mandatory link). Create a **foreign key** between the Users and Towns tables.

Steps:

* 1. Create table with Id and name. Don’t forget identity and primary key constraint.
  2. Alter table Users and add the new column.
  3. Alter table Users again and add the foreign key constraint to Towns table.

1. Execute the following SQL script (it should pass without any errors):

|  |
| --- |
| INSERT INTO Towns(Name) VALUES ('Sofia'), ('Berlin'), ('Lyon')  UPDATE Users SET TownId = (SELECT Id FROM Towns WHERE Name='Sofia')  INSERT INTO Towns VALUES  ('Munich'), ('Frankfurt'), ('Varna'), ('Hamburg'), ('Paris'), ('Lom'), ('Nantes') |

1. Write and execute a SQL command that **changes the town to "Berlin" for all users with registration date at Wednesday**.

Steps:

1. Write update statement by Users table.
2. Get town id with nested select statement.
3. Add where statement where check the day of the week (You can use **DATEPART** function).
4. Write and execute a SQL command that **changes the question to “Java += operator” of Answers, answered at Monday or Sunday in February**.

Steps:

* 1. Write update statement by Answers table.
  2. Get QuestionId with nested select statement.
  3. Add where filter (You can check the day of week with DATEPART function and the month with MONTH function).

1. Delete all answers with negative sum of votes.

Steps:

* 1. Write delete statement by Answers table
  2. Add where statement. You can sum the values of the votes with nested select statement and check if the sum is negative.

Hint: Think how to delete votes with answers.

1. Add a **new question** holding the following information: Title="Fetch NULL values in PDO query", Content="When I run the snippet, NULL values are converted to empty strings. How can fetch NULL values?", CreatedOn={current date and time}, Owner="darkcat", Category="Databases".

Hint: You can use **GETDATE** function for current datetime and nested select statements for user id and category id.

1. Find the **count of the answers for the users from town**. Display the town name, username and **answers count**. Sort the results by answers count in descending order, then by username. Name the columns exactly like in the table below.

Steps:

* 1. Create temporary table [#AnswerIds] with one column AnswerId (int)
  2. Insert into [#AnswerIds] table all answer ids where sum of answer votes are negative number.
  3. Delete votes where sum of answer votes are negative number.
  4. Delete answers which ids are in table [#AnswerIds]
  5. Drop temporary table [#AnswerIds]

|  |  |  |
| --- | --- | --- |
| Town | Username | AnswersCount |
| Sofia | Chad | 7 |
| Sofia | DanielB | 7 |
| Sofia | HHendriks | 7 |
| Sofia | micori | 7 |
| … | … | … |

Submit for evaluation the result grid **with headers**.

20 score

## Part III – Stored Procedures

You are given a **MS SQL Server database "Forum"** holding questions, answers, categories, tags, votes and users, available as SQL script. Your task is to write some stored procedures, views and other server-side database objects and write some SQL queries for displaying data from the database.

**Important:** start with a clean copy of the "**Forum**" database.

### Create a View and a Stored Function

1. Create a **view "AllQuestions"** in the database that holds information about questions and users (use RIGHT OUTER JOIN): If you execute the following SQL query:

|  |
| --- |
| SELECT \* FROM AllAnswers |

The result should be the table below:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **UId** | **Username** | **FirstName** | **LastName** | **Email** | **PhoneNumber** | **RegistrationDate** | **QId** | **Title** | **Content** | **CategoryId** | **UserId** | **CreatedOn** |
| 1 | Joiner | Adriano | Abatangelo | Joiner@gmail.com | NULL | 2014-07-06 | 2 | Why is processing a sorted array faster than an unsorted array? | Here is a piece of C++ code that seems very peculiar. For some strange reason, sorting the data miraculously makes the code almost six times faster. Initially, I thought this might be just a language or compiler anomaly. So I tried it in Java. With a somewhat similar, but less extreme result. My first thought was that sorting brings the data into the cache, but my next thought was how silly that is, because the array was just generated. | 1 | 1 | 2011-10-12 15:34:27.123 |
| … | … | … | … | … | … | … | … | … | … | … | … | … |

1. Using the view above, create a stored function "**fn\_ListUsersQuestions**" that **returns a table**, holding **all users** in descending order as first column, along with **all titles of their questions** (in ascending order), separated by ", " as second column.

If your function is correct and you execute the following SQL query:

|  |
| --- |
| SELECT \* FROM fn\_ListUsersAds() |

The result should be the table below:

|  |  |
| --- | --- |
| **UserName** | **Questions** |
| AGreenman | What are the differences between a pointer variable and a reference variable in C++?, Should 'using' statements be inside or outside the namespace? |
| Ankit | Deep cloning objects |
| ben | NULL |
| … | … |

Name the columns exactly like in the table below. Submit for evaluation the result grid with headers.

15 score

## Exam Information

To avoid locale-specific problems, use the "**English / United States**" as your locale. The decimal point is ".", the CSV column separator is ",", the month names are in English, etc.

You are allowed to use any resources you have like Internet, software, existing code.

You are not allowed to get help from other people: Skype, ICQ, FB, email, talks, phone calls, etc. are forbidden.

Exam time: **5 hours**.