

Syllabus for the course “SQL and automation of everyday reporting”

Danila Shershukov

Higher School of Economics, ICEF, 2018

1 Course description and scope of application

The aim of this course is to teach students how to use SQL for retrieving data from existing databases and to give them the skills needed for automation of everyday reporting. Knowledge of SQL is a competitive advantage in the job market, while the ability to automate everyday operations increases efficiency and allows to devote more time to issues that need deeper analysis.

The syllabus is designed for teachers who will be conducting the course and students who are willing to attend the course. This syllabus states the requirements for the educational results. It also defines the materials that will be used in class and the types of classes and required reporting forms.

2 Course aims

The course should result in students acquiring the following skills:

- usage of an SQL client with a graphical user interface;
- analyse database structure;
- compose SQL queries for retrieving the required data;
- automate execution of SQL queries;
- automate creation of reports using Excel & VBA or Python.

3 Course requirements

The course is recommended for students who have already completed the “Introduction to Probability Theory and Statistics” course. Moreover, good knowledge of either Excel & VBA or Python is required. It suggests that students have

completed on of the “Information Computer Systems”, “Programming and Data Processing” or “Programming and Databases” course.

4 Topics covered in the course

#	Section title	Teaching hours	Lectures	Seminars
1	Introduction	4	2	2
2	SELECT queries in SQL	12	6	6
3	Applying SQL to real-life problems	6	3	3
4	Automating reporting	6	3	3
5	Exam	4	2	2
Total		32	16	16

5 Forms of knowledge control

Control type	Control form	1	2	3	4	5	6	7	8	9	10	Description
Current	Home assignment	*										HA 1
Current	Home assignment		*									HA 2
Current	Home assignment			*								HA 3
Current	Home assignment				*							HA 4
Current	Home assignment					*						HA 5
Current	Home assignment						*					HA 6
Current	Home assignment							*				HA 7
Current	Home assignment								*			HA 8
Intermediate	Project					*	*	*	*	*		
Final	Exam										*	

Home assignments are to be submitted by e-mail in a week after being published. The topics of home assignments are:

1. Establishing a database connection, executing simple queries to examine database structure.
2. Using conditions in queries, joining two tables in a query.
3. Joining multiple tables.
4. Working with dates and time, aggregating data.

5. Doing more complex aggregations.
6. Using subqueries, solving problems that require a lot of intermediate computations.
7. Using subqueries and window functions, making simple reports.
8. Automating query execution, automating report creation.

As the last home assignment and the project require usage of Excel & VBA or Python, home assignments 1-7 may contain *optional* revision tasks.

The project is submitted by e-mail before the day of the final exam. The project is done individually. While working on the project, the students will:

- selecting metrics that should be reported;
- retrieving data for computing these metrics using SQL;
- visualizing the computed metrics using charts or tables;
- automating creation of the report;
- writing instructions on how to use the automated report.

The final exam is open-book and consists of two equally graded parts. The first part consists of SQL queries only. The second part allows choice between more complex SQL queries and report automation.

6 Topic contents

6.1 Introduction

This part is designed to introduce the core concepts of the SQL language which will serve as a foundation for the rest of the course. After covering the topic, the students must be able to:

- understand the main concepts of how an SQL server works;
- establish a connection to the database;
- use an SQL client with a graphical user interface;
- reproduce the structure of the SQL SELECT query;
- execute simple queries for getting general information about the database.

This part is covered during a single class and consists of a 2-hour lecture and a 2-hour seminar (4 hours total).

Class contents:

- main SQL concepts;
- working with the SQL client;
- practice usage of the SQL client;
- basics of Data Definition and Data Manipulation;
- basics of SELECT queries;
- practice SELECT queries.

Literature:

- Learning SQL [1] chapters 1 and 3;
- Head First SQL[2] chapter 1;
- PostgreSQL manual (9.4) [3] chapters 1 and 4.

6.2 SELECT queries in SQL

This part focuses on the core skills required to make SQL queries: checking conditions, joining tables and aggregating data. Working with dates and time is also included in this part, as it is an important part of reporting. After covering this part, the students must be able to write most of the simple queries they might need and they will be ready to start using subqueries.

This part consists of four topics and is covered in 12 hours. Time distribution between topics is given in table 3.

#	Topic	Class hours	Self-study hours
1	Checking conditions	2	4
2	Table Expressions	4	8
3	Dates and Time	2	4
4	Aggregating data	4	8
Total		12	24

Table 2: time distribution (SELECT queries in SQL)

6.2.1 Checking conditions

Topic contents:

- checking equality and inequality conditions;
- checking for null values;
- checking interval conditions;
- testing strings against patterns;

Literature:

- Learning SQL [1] chapter 4, 11;
- Head First SQL[2] chapter 2;
- PostgreSQL manual (9.4) [3] chapters 4, 9.

6.2.2 Table Expressions

Topic contents:

- types of JOIN statements;
- using complex conditions in JOIN statements;

- caveats of joining many tables;

Literature:

- Learning SQL [1] chapters 3, 5, 10;
- Head First SQL[2] chapter 10;
- PostgreSQL manual (9.4) [3] chapters 4, 7.

6.2.3 Dates and Time

Even though dates and time are a rather narrow subject, they require extra attention. Mistakes in working with dates and time easily lead to inconsistency of reports.

Topic contents:

- date and time types, timezones;
- functions for working with dates and time;
- common errors while working with time and dates and how to avoid them;

Literature:

- PostgreSQL manual (9.4) [3] chapter 9.

6.2.4 Aggregating Data

Topic contents:

- GROUP BY clause;
- aggregate functions;
- common errors and how to avoid them;

Literature:

- Learning SQL [1] chapter 8;
- Head First SQL[2] chapter 6;
- PostgreSQL manual (9.4) [3] chapter 9.

6.3 Applying SQL to real-life problems

This part is designed to cover more complex concepts like subqueries and window functions. By this time, students must have accumulated enough knowledge to be fluent with the basic concepts, so most of the tasks in class and in assignments are now given in terms of real-life problems. After covering this part, the students must be able to:

- compose queries that are complex enough to cover their needs;
- identify data that has to be queried to solve real-life problems.

This part consists of two topics and is covered in 6 hours. Time distribution between topics is given in table 3.

#	Topic	Class hours	Self-study hours
1	Subqueries	4	8
2	Window functions	2	4
	Total	12	24

Table 3: time distribution (SELECT queries in SQL)

6.3.1 Subqueries

Topic contents:

- using subqueries as table expressions;
- using subqueries in the WITH clause;
- practicing on real-life problems to learn to structure the query effectively.

Literature:

- Learning SQL [1] chapter 9;
- Head First SQL[2] chapter 9;
- PostgreSQL manual (9.4) [3] chapter 7.

6.3.2 Window Functions

Topic contents:

- applying common aggregation functions to windows;
- window functions that require sorting;
- practicing on real-life problems to learn to structure the query effectively.

Literature:

- PostgreSQL manual (9.4) [3] chapter 9.

6.4 Automating reporting

In this part students train to automate their reports using a programming language (VBA in Excel or Python¹) and automate query execution using scripts.

This part is covered in 6 hours.

Topic contents:

- using an SQL client with a command line interface;
- automating query execution using scripts and OS scheduling mechanisms;
- using Excel & VBA or Python to automatically prepare detailed reports.

¹As it is hard to show how to solve a problem in two languages at once, demo solutions will be demonstrated during the lectures and at seminars the students will be free to choose the language they work with. Helpful notes will be provided both for students who use VBA and students who use Python.

Literature:

- Excel documentation [4];
- Python library reference [5] and other Python documentation;
- PostgreSQL manual (9.4) [3] References, part II, psql.

7 Grading

The course grade consists of the grades for home assignments (M_{HA}), the grade for the project ($M_{project}$) and the grade for the exam (M_{exam}) and is calculated as follows:

$$M_{total} = 0.3M_{HA} + 0.3M_{project} + 0.4M_{exam}$$

The grade for home assignments is the average of all grades for individual home assignments:

$$M_{HA} = \frac{1}{n} \sum_{i=1}^n M_{HA_i}$$

The grade for the project is a weighted sum of the grades for each of the project parts.

Project part	Weight
Selecting metrics that should be reported	0.05
Retrieving data for computing these metrics using SQL	0.40
Visualizing the computed metrics using charts or tables	0.25
Automating creation of the report	0.10
Writing instructions on how to use the automated report	0.20

Both parts of the exam have equal weight. All assignments and tasks are graded according to a 1 to 10 scale.

8 Literature

- [1] Alan Beaulieu. *Learning SQL*. O'REILLY, 2009.
- [2] Lynn Beighley. *Head First SQL*. O'REILLY, 2007.
- [3] The PostgreSQL Global Development Group. *PostgreSQL 9.4 official manual*.
- [4] Microsoft. Excel help. <https://support.office.com/en-US/Excel>.

- [5] Guido van Rossum and the Python development team. *The Python Library Reference*.

9 Technical resources

As all classes within the course imply usage of computers, a computer class with a projector is required.

The software that will be used during the course is:

- SQL client with a graphical user interface, e.g. pgAdmin;
- SQL client with a command line interface, e.g. psql;
- MS Excel or Python 3
- (optional) IDE for Python 3, e.g. PyCharm Community Edition, is recommended if the student is going to use Python for automatization.

MS Excel is usually installed in the computer classes by default. The remaining software is distributed freely.