

SQL и автоматизация текущей отчётности

Данила Шершуков

29 января 2019 г.

План презентации

- 1 Что такое SQL и зачем он нужен?
- 2 Чему можно научиться на этом факультативе?
- 3 Что изменилось по сравнению с прошлым годом?
- 4 План факультатива
- 5 График занятий и оценивание
- 6 Технические детали
- 7 Отзывы студентов
- 8 Мои контакты

Что такое SQL?

- SQL (Structured Query Language) — это язык команд для систем управления базами данных (СУБД/DBMS).
- SQL состоит из нескольких частей:
 - Data Query Language (SELECT: fetches data)
 - Data Manipulation Language (INSERT, UPDATE, DELETE: modifies data)
 - Data Control Language (manages permissions)
 - Data Definition Language (defines database schema)

SQL — это не единый язык

- Существуют разные диалекты SQL (у каждого производителя баз данных свой):
 - PostgreSQL, MySQL — диалекты от крупных проектов с открытым исходным кодом (и возможностью их бесплатного использования). PostgreSQL — одно из популярных решений в высоконагруженных проектах в IT-компаниях. Открытый исходный код позволяет доработать СУБД под свои нужды, если в этом есть необходимость.
 - Oracle, MSSQL — диалекты от Oracle Corporation и Microsoft для их серверов. Популярны корпоративные решения.
 - SQLite — диалект, используемый в одноимённой библиотеке (бэкенд для большого количества программ).

SQL — это не единый язык

- Существует стандарт SQL. Он платный (его нельзя прочитать просто так), и не все диалекты ему полностью соответствуют.
- Диалекты могут различаться типами данных и некоторыми расширениями и функциями.
- Несмотря на различия между диалектами, основа одна и та же.
- Мы будем работать с PostgreSQL.

Зачем нужен SQL?

- Для аналитика:
 - извлечение данных из базы
 - базовый анализ данных
 - позволяет автоматизировать составление отчётов (не надо извлекать данные вручную)
- Для бизнеса:
 - надёжное хранение данных
- Для студента:
 - конкурентное преимущество при поиске работы (стажировки)
 - инструмент для работы с данными, которые не помещаются в памяти (может пригодиться при написании диплома или курсовой)

Чему можно научиться на этом факультативе?

- Основное:
 - свободно извлекать данные из базы (наболее важная часть SQL для аналитика)
 - автоматизировать извлечение данных и составление отчётов на базе этих данных
- Дополнительное (необязательные задания):
 - генерировать отчёты в html на основе шаблонов, в том числе с картинками (SQL + pandas + matplotlib + Jinja2)
 - разработать собственную базу данных

Сдача и проверка домашних заданий

- Домашние задания будут сдаваться через сайт и проверяться автоматически.
- По результатам проверки можно будет получить детальные комментарии (по запросу).
- Проект также будет сдаваться через сайт (поэтапно), но проверяться будет вручную.

Содержание

- Больше не будет отдельных занятий по Python.
- Для студентов, заинтересованных в Python, будут проводиться короткие tutorials в конце каждого занятия.
- Примерные темы tutorials:
 - введение в Python
 - работа с pandas
 - работа с matplotlib & seaborn
 - HTML и работа с Jinja2
 - генерация отчётов на Python

Софт

- Мы больше не будем пользоваться pgAdmin4, вернёмся на pgAdmin3.

Темы

- Вступление: учимся основам SQL
- SELECT-запросы в SQL: "набиваем руку"
 - проверка условий
 - табличные выражения
 - даты и время
 - агрегация данных
- Применение SQL для решения практических задач: то, что тяжело учить, но очень нужно в жизни
 - подзапросы
 - window functions
- Автоматизация составления отчётов: пробуем, каково это на практике

Расписание

- факультатив будет идти в течение 10 недель с перерывом на время мартовской сессии
- полторы-две пары в неделю, пожелания насчёт дня недели принимаются

Домашние работы, проект, экзамен, оценивание

- всего будет 8 домашних работ, каждая на одну неделю (30% от итоговой оценки)
- проект начнётся в середине февраля и закончится в середине апреля (30% от итоговой оценки)
- экзамен в середине апреля, open book, with Internet access (40% от итоговой оценки)

Необходимый для работы софт

- pgAdmin3 — SQL-клиент с графическим интерфейсом (<https://www.openscg.com/bigsql/pgadmin3/>, проект, который поддерживает pgAdmin3 после выхода pgAdmin4)
- psql — консольный SQL-клиент, будет использоваться для автоматизации
- MS Excel или Python3 + PyCharm (использование Anaconda не рекомендуется) — для проекта

Отзывы студентов: как получены

- В мае 2018 проводился опрос студентов, сдававших итоговый экзамен на этом факультативе (30 человек).
- Были получены ответы от 10 студентов.
- На следующих слайдах привожу ответы на вопрос "Пожалуйста, расскажи будущим студентам об этом курсе и о своих впечатлениях о нём в двух-трёх предложениях."
- Надо понимать, что приводимые ответы сильно смещены из-за того, что опрос прошло лишь десять студентов.

Отзывы студентов (1-3)

- Курс очень интересный и прикладной, к тому же помогает понять, чем именно хочется заниматься в будущем. Я бы всем его советовала.
- Невероятно полезный прикладной курс для всех, кто так или иначе собирается работать с большими объемами данных. Это не обязательно Data Science, это понадобится даже в околораудиторской аналитике. При этом курс может освоить любой студент, даже не знакомый с программированием вообще.
- Курс достаточно интересный и полезен тем, кто хочет заниматься или работать в сфере анализа больших объемов информации. Более того, это даёт некоторое преимущество над другими, ведь почти у каждого есть питон в резюме, но далеко не у каждого есть sql.

Отзывы студентов (4-5)

- Впечатления от курса самые лучшие. Курс даёт возможность с минимальными затратами времени (а его всегда мало) получить большое количество практических навыков. Курс пригодится практически любому, кто планирует связать свою деятельность или её часть с работой и вычислениями из больших массивов данных. Если знания курса и не пригодятся вам напрямую, то уж точно будут весомым преимуществом для претендента на позицию в финансовой сфере.
- Курс "Базы данных и автоматизация отчетности" даёт базовые знания для дальнейшего развития в работе с данными и знакомит с инструментами, необходимыми для экономии времени на задачах, требующих рутинных действий. Также неплохо смотрится в резюме, если идти на соответствующие позиции.

Отзывы студентов (6-8)

- Курс очень полезный и применимый в практике. Позволяет не только овладеть инструментом, но и понять организацию баз данных. Незаменим в работе аналитика, даже если придется писать на других языках или (простите) экселе.
- Впечатления только хорошие. Владение SQL очень ценится HRами и безусловно для общего развития это очень нужный скилл. Советую сразу приходить с ПК на винде, потому что на собственном опыте Мак не очень компетентен в вопросах с SQL.
- Для будущих студентов - курс очень крутой, хорошая атмосфера на занятиях, если хотите научиться полезному скиллу без лишнего давления, очень рекомендую. Не стоит бояться, что будет очень сложно, материал, конечно, непростой, но Данила всегда убеждает, что все всё успевают, плюс очень круто сделана практическая часть.

Отзывы студентов (9-10)

- Курс Данилы Шершукова по автоматизации с SQL дает весьма солидные навыки для работы с базами данных. Курс больше рассчитан на студентов, у которых опыт работы с SQL небольшой, и которым хотелось бы получить практические знания. Как опытный программист-любитель могу заверить, что глубоких познаний в программировании не понадобится, и могу порекомендовать курс всем желающим научиться уверенно работать с базами данных!
- Хороший курс, который дает представление об SQL в целом. Помимо это дает базовые знания питона, что поможет в дальнейшем углубить свои знания в этом аспекте.

Мои контакты

VK — https://vk.com/danila_shershukov

Telegram — @server_pomnit

Email — danila.shershukov@yandex.ru

Мои контакты

Спасибо за внимание!