



ДИСЦИПЛИНА

«Базы данных (PostgreSQL)»

Куратор: АИС Город

Нередко DBA и backend-разработчикам, занимающимся работой с базами данных, необходимо постигать новые предметные области, где эти самые базы данных уже составлены. Не всегда они содержат консистентные данные, да и не всегда грамотно спроектированы с точки зрения решаемой задачи.

Одной из таких задач является работа с адресной системой России, которая состоит из Государственного Адресного Реестра (ГАР) и Федеральной Информационной Адресной Системы (ФИАС). Не вдаваясь в подробности, ГАР — это формат представления всевозможных адресов России, а ФИАС — это система, которая предоставляет выгрузки ГАР в формате, например, XML. Ознакомиться с ФИАС и ГАР можно по ссылке <https://fias.nalog.ru/>.

В рамках полуфинала Олимпиады вам предоставлен дамп базы данных для PostgreSQL 15 со схемой «gar», в котором по таблицам разложены данные из адресного реестра по некоторым регионам России. Вам необходимо решить несколько задач:

0. Развернуть дамп базы данных у себя локально.
1. Согласно предметной области определиться с основными уникальными и неуникальными индексами.
2. Написать запросы из списка ниже.
3. Найти «испорченные» данные в дампе согласно логике предметной области.

Разворачивание дампа

Дамп базы данных представлен по ссылке:
<https://cloud.aisgorod.ru/s/0a9cYIG2IAA7nmT>

Определение недостающих ключей и индексов

В рамках этого задания необходимо описать и обосновать дополнительные индексы, которые по вашему мнению необходимо добавить для нормального функционирования представленной базы данных. В качестве примера использования БД можно взять за основу запросы из следующих заданий.

Не допускается создание индексов без обоснования. Также недопустимо создание индексов на абсолютно все поля таблиц.

Написание запросов: общие правила

На текущем этапе Олимпиады проверяется ваше умение разбираться в предметной области и писать запросы на языке SQL для выборки данных согласно заданию.

Ниже представлены запросы двух уровней: базового и продвинутого. Они отличаются сложностью, выставляемыми баллами и погружением в предметную область. Рекомендуется начать с «базовых» запросов, и только потом переходить к «продвинутым».

Для каждого задания запрос должен быть один. CTE и JOIN использовать, конечно же, не запрещается, как и не запрещается использовать подзапросы. В рамках проверки написанного вами запроса может случиться такое, что база данных будет отвечать медленно. Очень медленно. Значит вы нашли потенциальный индекс, который можно указать в задании №1.

В базе данных реализована историчность данных. Например, данные об одном и тот же доме могут повторяться десятки раз, если с этим домом что-то происходило.

Поэтому возникает общее правило: если в задании на написание запросов не указано иное, то записи необходимо учитывать только актуальные. Если просят вывести адрес и «пройтись» по иерархии от дома к улице, от улицы к городу, от города к субъекту РФ, то стоит брать по умолчанию административную иерархию. В ГАР есть два вида иерархии: административная и муниципальная. Так исторически сложилось.

Написание запросов: общие правила

- Б1. Выведите в алфавитном порядке все актуальные улицы в городе Барыш Ульяновской области, которые начинаются на букву «Л».
- Б2. Сколько актуальных СНТ находится в Ульяновской области?
- Б3. Выведите улицу в городе Ульяновск, на которой больше всего актуальных домов. Также выведите количество домов на этой улице.
- Б4. Сколько актуальных парковочных мест в доме «г. Ульяновск, ул. Воробьева, д. 77»?
- Б5. Выведите топ-10 актуальных на текущий момент домов с самой большой историей изменений. В первом столбце - object_id, во втором - количество изменений, в третьем - дата начала действия последнего изменения.
- Б6. Выведите топ-10 актуальных домов с наибольшим количеством квартир. В первом столбце - object_id, во втором - количество квартир, в третьем - адрес дома по административной иерархии без указания типа адресного объекта, например, «Ульяновская, Ульяновск, Федерации, 140».
- Б7. Выведите топ-10 помещений в Ульяновской области с наибольшим количеством комнат внутри? Выведите их object_id и количество комнат внутри помещений.

Б8.

Найдите в Ульяновской области процент актуальных помещений, имеющих кадастровый номер, к общему числу актуальных помещений. Пример кадастрового номера: «73:24:040404:123».

Б9.

Есть ли в Ульяновской области тракты? Например, есть проспекты, улицы, переулки, а тракты? Выведите все их названия.

Б10.

Вывести топ-5 лет, начиная с 2000, в которые в городе Ульяновск Ульяновской области было создано наибольшее количество актуальных на данный момент улиц. В первом столбце - год, во втором - количество созданных улиц (в обратном порядке).

Написание запросов: продвинутый уровень

П1.

Выведите название улицы в городе Ульяновск Ульяновской области, на которой больше всего домов, имеющих разные почтовые индексы. Вторым столбцом выведите количество, а третьим столбцом сами индексы почтовых отделений через «,», к которым относятся дома этой улицы.

П2.

Выведите актуальный полный адрес в муниципальном делении для дома, который когда-то был номером 7 по улице Карла Маркса в городе Ульяновск Ульяновской области.

П3.

Иногда в рамках одного района субъекта РФ (административное деление) существует несколько деревень или сёл с одним названием. Необходимо вывести все такие названия по полной форме административного деления, например, «Ульяновская обл., Майнский р-н, с. Абрамовка». Вторым столбцом вывести количество таких повторов.

П4.

Выведите полные номера актуальных домов и их ОКТМО по улице «Верхняя площадка» в городе Ульяновске Ульяновской области по административной иерархии.

П5.

Выведите актуальные названия улиц в г. Димитровград Ульяновской области по административному делению, поменявших название в период с 2015 года. Вторым столбцом укажите год последнего изменения названия улицы.

П6.

Выведите топ-10 улиц по количеству повторений названия в Ульяновской области.

П7.

Выведите дома, которые были разделены более чем на две части. В первом столбце укажите полный адрес исходного дома по административной иерархии. Во втором столбце - массив полных названий новых частей также по административной иерархии.

П8.

Выведите топ-5 нормативных документов, которые внесли наибольшее количество изменений в названия адресных объектов и домов. Вывести название, номер, дату и количество изменений.

Написание запросов: продвинутый уровень

Это задание является дополнительным и позволит продемонстрировать, что вы хорошо разобрались в предметной области, что можете заметить, что «что-то тут не так».

Проблем можно найти более 10 штук. Можно найти сильно больше, но можете ориентироваться на это значение. Например, возможно где-то нарушена ссылочная целостность (в схеме нет внешних ключей, но по логике предметной области они есть), или где-то в данных размещено такое, чего там быть точно не может.

Опишите как можно больше случаев, где лично вы видите проблемы с данными. Если вы заметили массовую проблему, не надо перечислять все строки, где она присутствует. Достаточно описать диапазон или просто название таблицы с описанием проблемы, конечно же.



Написание запросов: продвинутый уровень

По ссылке <https://cloud.aisgorod.ru/s/Ec3E4yiVontheq2> представлен документ «Шаблон решения.docx». Необходимо взять за пример указанный документ, заполнить своими ответами и прислать на проверку.

Допускается использовать формат документа, отличный от docx, однако должна быть обеспечена возможность копирования текста запросов и команд к PostgreSQL для их корректной проверки.