



PostgreSQL Cloud Solutions

Получите навыки работы в различных виртуальных и облачных средах

Длительность курса: 124 академических часа

1 SQL и реляционные СУБД. PostgreSQL в облаках

Цели занятия:

научиться представлять сложность выбора той или иной технологии кластеризации и отказоустойчивости Постгреса лучше представлять, чем мы будем заниматься на курсе

Краткое содержание:

где мы были и где сейчас;
выбор и PostgreSQL;
случай из жизни;
программа курса.

Домашние задания

1 Работа с уровнями изоляции транзакции в PostgreSQL

Цель: - научиться работать с Google Cloud Platform на уровне Google Compute Engine (IaaS) /ЯО

- научиться управлять уровнем изоляции транзакции в PostgreSQL и понимать особенность работы уровней read committed и repeatable read

- создать новый проект в Google Cloud Platform, Яндекс облако или на любых VM, например postgres2022-`<ууууммдд>`, где `ууууммдд` год, месяц и день вашего рождения (имя проекта должно быть уникально на уровне GCP)

- далее создать инстанс виртуальной машины Compute Engine с дефолтными параметрами

- 1-2 ядра, 2-4Гб памяти, любой линукс, на курсе Ubuntu 100%

- добавить свой ssh ключ в GCE metadata

- зайти удаленным ssh (первая сессия), не забывая про ssh-add

- поставить PostgreSQL из пакетов `apt install`

- зайти вторым ssh (вторая сессия)

- запустить везде `psql` из под пользователя

postgres

- выключить auto commit

- сделать в первой сессии новую таблицу и наполнить ее данными

```
create table persons(id serial, first_name text, second_name text);
```

```
insert into persons(first_name, second_name) values('ivan', 'ivanov');
```

```
insert into persons(first_name, second_name) values('petr', 'petrov');
```

```
commit;
```

- посмотреть текущий уровень изоляции: show transaction isolation level

- начать новую транзакцию в обеих сессиях с дефолтным (не меняя) уровнем изоляции

- в первой сессии добавить новую запись

```
insert into persons(first_name, second_name) values('sergey', 'sergeev');
```

- сделать select * from persons во второй сессии

- видите ли вы новую запись и если да то почему?

- завершить первую транзакцию - commit;

- сделать select * from persons во второй сессии

- видите ли вы новую запись и если да то почему?

- завершите транзакцию во второй сессии

- начать новые но уже repeatable read транзакции - set transaction isolation level repeatable read;

- в первой сессии добавить новую запись

```
insert into persons(first_name, second_name) values('sveta', 'svetova');
```

- сделать select * from persons во второй сессии

- видите ли вы новую запись и если да то почему?

- завершить первую транзакцию - commit;

- сделать select * from persons во второй сессии

- видите ли вы новую запись и если да то почему?

- завершить вторую транзакцию

- сделать select * from persons во второй сессии

- видите ли вы новую запись и если да то почему?

ДЗ сдаем в виде миниотчета в markdown в гите

Цели занятия:

разобраться в устройстве Docker
установить Постгрес через docker & docker-compose

Краткое содержание:

устройство docker
установка PostgreSQL в docker
установка PostgreSQL в docker-compose

Домашние задания

- 1 Установка и настройка PostgreSQL в контейнере Docker

Цель: развернуть VM в GCP/ЯО/Аналоги
установить туда докер
установить PostgreSQL в Docker контейнере
настроить контейнер для внешнего подключения

- сделать в GCE/ЯО/Аналоги инстанс с Ubuntu 20.04
- поставить на нем Docker Engine
- сделать каталог /var/lib/postgres
- развернуть контейнер с PostgreSQL 14 смонтировав в него /var/lib/postgres
- развернуть контейнер с клиентом postgres
- подключится из контейнера с клиентом к контейнеру с сервером и сделать таблицу с парой строк
- подключится к контейнеру с сервером с ноутбука/компьютера извне инстансов GCP/ЯО/Аналоги
- удалить контейнер с сервером
- создать его заново
- подключится снова из контейнера с клиентом к контейнеру с сервером
- проверить, что данные остались на месте
- оставляйте в ЛК ДЗ комментарии что и как вы делали и как боролись с проблемами

Цели занятия:

поработать с параметрами конфигурации PostgreSQL; объяснить в чем разница между различными группами параметров; объяснить выбор оптимального значения для параметров.

Краткое содержание:

настройка сервера;
работа с параметрами;
значения параметров;
практика.

Домашние задания

1 Настройка дисков для Постгреса

Цель: - создавать дополнительный диск для уже существующей виртуальной машины, размечать его и делать на нем файловую систему

- переносить содержимое базы данных PostgreSQL на дополнительный диск
- переносить содержимое БД PostgreSQL между виртуальными машинами

- создайте виртуальную машину с Ubuntu 20.04 LTS (bionic) в GCE типа e2-medium в default VPC в любом регионе и зоне, например us-central1-a или ЯО

- поставьте на нее PostgreSQL 14 через sudo apt

- проверьте что кластер запущен через sudo -u postgres pg_lsclusters

- зайдите из под пользователя postgres в psql и сделайте произвольную таблицу с произвольным содержимым

```
postgres=# create table test(c1 text);
```

```
postgres=# insert into test values('1');
```

```
\q
```

- остановите postgres например через sudo -u postgres pg_ctlcluster 14 main stop

- создайте новый standard persistent диск GKE через Compute Engine -> Disks в том же регионе и зоне что GCE инстанс размером например 10GB - или аналог в другом облаке/ виртуализации

- добавьте свежесозданный диск к виртуальной машине - надо зайти в режим ее редактирования и дальше выбрать пункт `attach existing disk`
- проинициализируйте диск согласно инструкции и подмонтировать файловую систему, только не забывайте менять имя диска на актуальное, в вашем случае это скорее всего будет `/dev/sdb` - <https://www.digitalocean.com/community/tutorials/how-to-partition-and-format-storage-devices-in-linux>
- перезагрузите инстанс и убедитесь, что диск остается примонтированным (если не так смотрим в сторону `fstab`)
- сделайте пользователя `postgres` владельцем `/mnt/data` - `chown -R postgres:postgres /mnt/data/`
- перенесите содержимое `/var/lib/postgres/14` в `/mnt/data` - `mv /var/lib/postgresql/14 /mnt/data`
- попытайтесь запустить кластер - `sudo -u postgres pg_ctlcluster 14 main start`
- напишите получилось или нет и почему
- задание: найти конфигурационный параметр в файлах расположенных в `/etc/postgresql/14/main` который надо поменять и поменяйте его
- напишите что и почему поменяли
- попытайтесь запустить кластер - `sudo -u postgres pg_ctlcluster 14 main start`
- напишите получилось или нет и почему
- зайдите через `psql` и проверьте содержимое ранее созданной таблицы
- задание со звездочкой *: не удаляя существующий GCE инстанс/ЯО сделайте новый, поставьте на его PostgreSQL, удалите файлы с данными из `/var/lib/postgres`, перемонтируйте внешний диск который сделали ранее от первой виртуальной машины ко второй и запустите PostgreSQL на второй машине так чтобы он работал с данными на внешнем диске, расскажите как вы это сделали и что в итоге получилось.

4 **Углубленный анализ
производительности.
Профилирование.
Мониторинг.
Оптимизация**

Цели занятия:

выбирать правильные архитектурные решения;
определить узкие места базы;
оптимизировать параметры;
профилировать запросы

Краткое содержание:

Архитектурные решения;
оптимизация: объекты, подходы;
методики мониторинга и профайлинга;
варианты тюнига;
бест практис.

Домашние задания

1 Тюнинг Постгреса

Цель: Развернуть инстанс Постгреса в VM в GCP
Оптимизировать настройки

Развернуть Постгрес на VM
Протестировать pg_bench
Выставить оптимальные настройки
Проверить насколько выросла
производительность
Настроить кластер на оптимальную
производительность не обращая внимания на
стабильность БД
ДЗ сдаем в виде миниотчета в markdown и
гите

5 **Углубленное
изучение бэкапов и
репликации**

Цели занятия:

уверенно использовать утилиты бэкапа
pg_probackup, WAL-G
уверенно использовать утилиту репликации
pg_rewind

Краткое содержание:

pg_probackup
WAL-G
pg_rewind

Домашние задания

1 Бэкапы Постгреса

Цель: Используем современные решения для бэкапов

Делаю бэкап Постгреса используя WAL-G или pg_probackup и восстанавливаемся на другом кластере

Задание повышенной сложности*
под нагрузкой*
бэкап снимаем с реплики**

6 Кластер Patroni on-premise 1

Цели занятия:

разворачиваем отказоустойчивый DCS на базе ETCD

разворачиваем HA Patroni на пустых данных

разворачиваем HA Patroni на существующих данных

настраиваем отказоустойчивое решение с использованием PGbouncer + HAproxy + keepalived

Краткое содержание:

HA Etcd

HA Patroni

PGbouncer

HAproxy

Keepalived

7 Кластер Patroni on-premise 2

Цели занятия:

разворачиваем отказоустойчивый DCS на базе ETCD

разворачиваем HA Patroni на пустых данных

разворачиваем HA Patroni на существующих данных

настраиваем отказоустойчивое решение с использованием PGbouncer + HAпроху + keepalived

Краткое содержание:

HA Etcd

HA Patroni

PGbouncer

HAпроху

Keepalived

Домашние задания

1 Кластер Patroni

Цель: Развернуть HA кластер

Создаем 3 VM для etcd + 3 VM для Patroni +1 HA проху (при проблемах можно на 1 хосте развернуть)

Инициализируем кластер

Проверяем отказоустойчивость

*настраиваем бэкапы через wal-g или pg_probackuper

8 Введение в Kubernetes. Часть 1

Цели занятия:

научиться работать в Kubernetes

Краткое содержание:

архитектура и базовые сущности Kubernetes: Pod, Deployment, ReplicaSet, Service, Ingress.

Цели занятия:

научиться работать в Kubernetes

Краткое содержание:

архитектура и базовые сущности Kubernetes:
StatefulSet, ConfigMap, Persistent Volume,
Persistent Volume Claim.

Домашние задания

1 Постгрес в minikube

Цель: Развернуть Постгрес в миникубе

Устанавливаем minikube
Разворачиваем PostgreSQL 14 через
манифест

Задание повышенной сложности*
Разворачиваем PostgreSQL 14 с помощью
helm

1 Работа с большим объемом реальных данных

Цели занятия:

объяснить природу больших данных и знать места их обитания;
загрузить большие данные в PostgreSQL;
объяснить как работать с большими данными.

Краткое содержание:

что такое большие данные;
особенности работы с большими данными;
как загрузить большие данные в PostgreSQL;
практика.

Домашние задания

1 Разворачиваем и настраиваем БД с большими данными

Цель: - знать различные механизмы загрузки данных
- уметь пользоваться различными механизмами загрузки данных

Необходимо провести сравнение скорости работы запросов на различных СУБД

- 1) Выбрать одну из СУБД
- 2) Загрузить в неё данные (от 10 до 100 Гб)
- 3) Сравнить скорость выполнения запросов на PostgreSQL и выбранной СУБД
- 4) Описать что и как делали и с какими проблемами столкнулись

2 Варианты кластеров высокой доступности для PostgreSQL

Цели занятия:

выбрать оптимальный вариант высокой доступности для PostgreSQL;
проанализировать варианты классических HA кластеров для PostgreSQL;
настроить классический PostgreSQL HA кластер.

Краткое содержание:

высокая доступность;
кластер высокой доступности;
PostgreSQL HA кластеры;
практика.

3 Работа с кластером высокой доступности

Цели занятия:

выбрать оптимальный вариант высокой доступности для PostgreSQL;
проанализировать варианты классических HA кластеров для PostgreSQL;
настроить классический PostgreSQL HA кластер.

Краткое содержание:

высокая доступность;
кластер высокой доступности;
PostgreSQL HA кластеры;
практика.

Домашние задания

1 Развернуть HA кластер

Цель: - развернуть высокодоступный кластер PostgreSQL собственными силами
- развернуть высокодоступный сервис на базе PostgreSQL на базе одного из 3-ки ведущих облачных провайдеров - AWS, GCP и Azure

Выбрать один из вариантов и развернуть кластер. Описать что и как делали и с какими проблемами столкнулись

Вариант 1

- How to Deploy PostgreSQL for High Availability

Вариант 2

- Introducing pg_auto_failover: Open source extension for automated failover and high-availability in PostgreSQL

Для гурманов

- Настройка Active/Passive PostgreSQL Cluster с использованием Pacemaker, Corosync, и DRBD (CentOS 5,5)
-
-
-

Задание повышенной сложности*

Создать два кластера GKE в разных регионах

Установить на первом Patroni HA кластер

Установить на втором Patroni Standby кластер

Настроить TCP LB между регионами

Сделать в каждом регионе по клиентской VM

Проверить как ходит трафик с клиентской VM

Описать что и как делали и с какими проблемами столкнулись

4 Способы горизонтального масштабирования PostgreSQL

Цели занятия:

проанализировать горизонтальное масштабирование PostgreSQL;

объяснить, когда стоит использовать горизонтальное масштабирование, а когда нет;

настроить параллельный кластер.

Краткое содержание:

параллельные кластеры БД;

способы горизонтального масштабирования PostgreSQL;

практика.

5 Работа с горизонтально масштабируемым кластером 1

Цели занятия:

разворачивать мультимастер кластера;
настроить мультимастер кластера;
настроить геокластер.

Краткое содержание:

мультимастер кластера;
географически распределенные кластера.

Домашние задания

1 Multi master

Цель: - развернуть multi master кластер PostgreSQL своими руками
- развернуть PostgreSQL like географически распределенный сервис от одного из 3-х крупнейших облачных провайдеров - AWS, GCP и Azure

1 вариант:

Развернуть CockroachDB в GKE или GCE
Потесировать dataset с чикагскими такси
Или залить 10Гб данных и протестировать скорость запросов в сравнении с 1 экземпляром PostgreSQL
Описать что и как делали и с какими проблемами столкнулись

2 вариант:

Переносим тестовую БД 10 Гб в географически распределенный PostgreSQL like сервис
Описать что и как делали и с какими проблемами столкнулись

6 **Работа с горизонтально масштабируемым кластером 2**

Цели занятия:

развернуть параллельные кластера

Краткое содержание:

Greenplum

Yogabyte

Домашние задания

1 Parallel cluster

Цель: Развернуть один из вариантов параллельного кластера

Развернуть Yogabyte или Greenplum в GKE или GCE

Потесировать dataset с чикагскими такси

Или залить 10Гб данных и протестировать скорость запросов в сравнении с 1 инстансом PostgreSQL

Описать что и как делали и с какими проблемами столкнулись

Задание повышенной сложности*

Развернуть оба варианта и протестировать производительность

7 PostgreSQL и Google Kubernetes Engine

Цели занятия:

развернуть HA и мультимастер PostgreSQL кластера в k8s.

Краткое содержание:

работа с k8s
multi-master
high availability

Домашние задания

1 Работа с PostgreSQL в Kubernetes

Цель: - запустить HA и multi master PostgreSQL кластер в Kubernetes

Развернуть CitusDB в GKE, залить 10 Гб чикагского такси. Шардировать. Оценить производительность. Описать проблемы, с которыми столкнулись

Задание повышенной сложности*
залить все чикагское такси, оценить производительность

8 Pentaho

Цели занятия:

научиться использовать кубы данных

Краткое содержание:

pentaho bi suite, PDI,
трансформации данных,
pentaho schema workbench

9 **Самый простой
back-end на
NodeJS**

Цели занятия:

научиться разрабатывать самый простой back-end;
писать простейшие приложения на JS;
пользоваться платформой NodeJS;
научиться разрабатывать простой back-end для Web-приложений на NodeJS.

Краткое содержание:

JavaScript;
типы данных JS;
переменные и функции в JS;
“Серверный JS” - NodeJS;
ExpressJS - фреймворк для построения API на NodeJS;
API и статический веб-контент.

10 **Веб-приложение с
UI и БД**

Цели занятия:

научиться писать простейший UI;
подключать приложение к БД.

Краткое содержание:

ExpressJS - фреймворк для построения API на NodeJS;
API и статический веб-контент;
HTML, jquery, AJAX;
подключение к БД PostgreSQL.

1 PostgreSQL и Google Cloud Platform

Цели занятия:

поработать в GCP;
воспользоваться PostgreSQL based сервисами в GCP;
воспользоваться PostgreSQL like сервисами в GCP.

Краткое содержание:

работа в GCP;
PostgreSQL based сервисы в GCP;
PostgreSQL like сервисы в GCP.

2 PostgreSQL и AWS

Цели занятия:

поработать в AWS;
воспользоваться PostgreSQL based сервисами в AWS;
воспользоваться PostgreSQL like сервисами в AWS.

Краткое содержание:

работа в AWS;
PostgreSQL based сервисы в AWS;
PostgreSQL like сервисы в AWS.

3 PostgreSQL и Azure

Цели занятия:

поработать в Azure;
воспользоваться PostgreSQL based сервисами в Azure;
воспользоваться PostgreSQL like сервисами в Azure.

Краткое содержание:

работа в Azure;
PostgreSQL based сервисы в Azure;
PostgreSQL like сервисы в Azure.

Домашние задания

1 PostgreSQL и Azure, GCP, AWS, ЯО

Цель: - научиться пользоваться PostgreSQL based и like сервисами в Azure, GCP, AWS, ЯО

1 вариант:

Воспользоваться PostgreSQL based и like сервисами в AWS. Описать что и как делали и с какими проблемами столкнулись

2 вариант:

Воспользоваться PostgreSQL based и like сервисами в GCP. Описать что и как делали и с какими проблемами столкнулись

3 вариант:

Воспользоваться PostgreSQL based и like сервисами в Azure. Описать что и как делали и с какими проблемами столкнулись

4 вариант:

Воспользоваться PostgreSQL based и like сервисами в ЯО. Описать что и как делали и с какими проблемами столкнулись

4 PostgreSQL и Яндекс Облако

Цели занятия:

поработать в ЯОблако;
воспользоваться PostgreSQL based сервисами в ЯО;
воспользоваться PostgreSQL like сервисами в ЯО.

Краткое содержание:

почему Yandex.Cloud - кейсы когда Postgres нужен именно там;
Managed Services vs Compute Cloud vs Kubernetes;
знакомство с консолью и визуальным интерфейсом Yandex.Cloud;
знакомство с возможностями и фичами Managed Services for PostgreSQL;
типы развертывания PostgreSQL для разных задач;
оптимизация работы PostgreSQL в Yandex.Cloud;
масштабирование PostgreSQL в Yandex.Cloud;
мониторинг PostgreSQL в Yandex.Cloud.

5 PostgreSQL и SberCloud

Цели занятия:

поработать в SberCloud;
воспользоваться PostgreSQL based сервисами в SberCloud;
воспользоваться PostgreSQL like сервисами в SberCloud.

Краткое содержание:

почему SberCloud - кейсы когда Postgres нужен именно там;
знакомство с консолью и визуальным интерфейсом SberCloud;
знакомство с возможностями и фичами Managed Services for PostgreSQL;
типы развертывания PostgreSQL для разных задач;
масштабирование PostgreSQL в SberCloud;
мониторинг PostgreSQL в SberCloud.

1 Выбор темы и организация проектной работы

Цели занятия:

выбрать и обсудить тему проектной работы;
спланировать работу над проектом;
ознакомиться с регламентом работы над проектом.

Краткое содержание:

правила работы над проектом и специфика проведения итоговой защиты;
требования к результату проекта и итоговой документации.

Домашние задания

1 Проект

Цель: Создать свой проект

1. Выбрать тему проектной работ
 2. Согласовать ее с руководителем курса
 3. Реализовать свой проект на 60%+
 4. Сделать презентацию своего проекта и защитить его на защите
-

2 Консультация по дипломному проекту

Цели занятия:

получить ответы на вопросы по проекту, ДЗ и по курсу.

Краткое содержание:

вопросы по улучшению и оптимизации работы над проектом;
затруднения при выполнении ДЗ;
вопросы по программе.

**3 Защита
проектных
работ**

Цели занятия:

защитить проект и получить рекомендации экспертов.

Краткое содержание:

презентация проектов перед комиссией;
вопросы и комментарии по проектам.

4 Итоги курса

Цели занятия:

узнать, как получить сертификат об окончании курса, как взаимодействовать после окончания курса с OTUS и преподавателями, какие вакансии и позиции есть для выпускников (опционально - в России и за рубежом) и на какие компании стоит обратить внимание.

Краткое содержание:

организационные вопросы;
рынок вакансий по направлению;
статистика курса и вопросы по курсу.