ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ»

Факультет Программной Инженерии и Компьютерной Техники

Дисциплина:
«*Распределённые системы хранения данных*»

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №3

Выполнили:
Студенты гр. P33151
Соловьев Артемий Александрович
Понамарев Степан Андреевич

Проверил:
Перцев Тимофей Сергеевич

Санкт-Петербург
2024г.

Задание

Внимание! У разных вариантов разный текст задания!

Цель работы - настроить процедуру периодического резервного копирования базы данных, сконфигурированной в ходе выполнения лабораторной работы №2, а также разработать и отладить сценарии восстановления в случае сбоев.

Узел из предыдущей лабораторной работы используется в качестве основного. Новый узел используется в качестве резервного. Учётные данные для подключения к новому узлу выдаёт преподаватель. В сценариях восстановления необходимо использовать копию данных, полученную на первом этапе данной лабораторной работы.

Требования к отчёту

Отчет должен быть самостоятельным документом (без ссылок на внешние ресурсы), содержать всю последовательность команд и исходный код скриптов по каждому пункту задания. Для демонстрации результатов приводить команду вместе с выводом (самой наглядной частью вывода, при необходимости).

Этап 1. Резервное копирование

- Настроить резервное копирование с основного узла на резервный следующим образом:
  Периодические обособленные (standalone) полные копии.
  Полное резервное копирование (pg_basebackup) по расписанию (cron) два раза в сутки. Необходимые файлы WAL должны быть в составе полной копии, отдельно их не архивировать. Срок хранения копий на основной системе - 1 неделя, на резервной - 1 месяц. По истечении срока хранения, старые архивы должны автоматически уничтожаться.
- Подсчитать, каков будет объем резервных копий спустя месяц работы системы, исходя из следующих условий:
  - Средний объем новых данных в БД за сутки: 550МБ.
  - Средний объем измененных данных за сутки: 750МБ.
- Проанализировать результаты.

Этап 2. Потеря основного узла

Этот сценарий подразумевает полную недоступность основного узла. Необходимо восстановить работу СУБД на РЕЗЕРВНОМ узле, продемонстрировать успешный запуск СУБД и доступность данных.

Этап 3. Повреждение файлов БД

Этот сценарий подразумевает потерю данных (например, в результате сбоя диска или файловой системы) при сохранении доступности основного узла. Необходимо выполнить полное восстановление данных из резервной копии и перезапустить СУБД на ОСНОВНОМ узле.

Ход работы:

- Симулировать сбой:
  - удалить с диска директорию конфигурационных файлов СУБД со всем содержимым.

- Проверить работу СУБД, доступность данных, перезапустить СУБД, проанализировать результаты.
- Выполнить восстановление данных из резервной копии, учитывая следующее условие:
    - исходное расположение директории PGDATA недоступно - разместить данные в другой директории и скорректировать конфигурацию.
- Запустить СУБД, проверить работу и доступность данных, проанализировать результаты.

Этап 4. Логическое повреждение данных

Этот сценарий подразумевает частичную потерю данных (в результате нежелательной или ошибочной операции) при сохранении доступности основного узла. Необходимо выполнить восстановление данных на ОСНОВНОМ узле следующим способом:

- Генерация файла на резервном узле с помощью pg_dump и последующее применение файла на основном узле.

Ход работы:

- В каждую таблицу базы добавить 2–3 новые строки, зафиксировать результат.
- Зафиксировать время и симулировать ошибку:
    - в любой таблице с внешними ключами подменить значения ключей на случайные (INSERT, UPDATE)
- Продемонстрировать результат.
- Выполнить восстановление данных указанным способом.
- Продемонстрировать и проанализировать результат.

Для подключения:

1) Для подключения к helios:
   ssh s334645@se.ifmo.ru
2) Для подключения к основному узлу:
   ssh postgres1@pg156
3) Для подключения к резервному узлу:
   ssh postgres0@pg191

Этап 1. Резервное копирование

Настройка резервного копирования

Создаем пользователя с привилегией REPLICATION

```
CREATE USER backup_user WITH REPLICATION;
```

Добавляем в postgesql.conf

```
archive_mode = on
archive_command = 'scp %p lab3'
```

Создание первоначальной копии

```
pg_basebackup --progress -p 9143 -U backup_user --format=t --wal-method=fetch -D $HOME/backup
```

Формат – tar файлы, включаем необходимые журналы транзакций в нашу резервную копию, включаем отчет о прогрессе.

```
[postgres1@pg156 ~/lab2]$ ./create_backup.sh
51926/51926 КБ (100%), табличное пространство 3/3
```

Перенос копии на резервный узел

```
rsync -avzP $HOME/backup/* postgres0@pg191:~/backup/
```

```
sending incremental file list
16384.tar
         10.752 100%    9,59MB/s    0:00:00 (xfr#1, to-chk=3/4)
16385.tar
         10.752 100%   10,25MB/s    0:00:00 (xfr#2, to-chk=2/4)
backup_manifest
        181.836 100%  173,41MB/s    0:00:00 (xfr#3, to-chk=1/4)
base.tar
     53.154.304 100% 1013,84MB/s    0:00:00 (xfr#4, to-chk=0/4)

sent 26.207 bytes  received 60.385 bytes  57.728,00 bytes/sec
total size is 53.357.644  speedup is 616,20
```

Подсчет размера копий

Размер начального бэкапа: 10 Мб

```
[postgres0@pg191 ~]$ du -sh backup/
 10M    backup/
```

При полном резервном копировании общий объем всех копий будет равен:

$$\frac{2 \cdot 10 + (30 - 1) \cdot 550}{2} \cdot 30 = 239{,}550 \text{ Гб}$$

При инкрементном копировании через месяц объем всех копий будет равен:

$$10 + 750 \cdot 30 = 22{,}510 \text{ Гб}$$

Очевидно, инкрементальные бекапы намного эффективнее.

Этап 2. Потеря основного узла