СИСТЕМАТИЗАЦИЯ ИНСТРУМЕНТОВ И МЕТОДОВ ДЛЯ ЭФФЕКТИВНОЙ РЕПЛИКАЦИИ БАЗЫ ДАННЫХ

Т.М. Опарина

старший преподаватель, e-mail: oparina2007@yandex.ru

Омский государственный университет им. Ф.М. Достоевского, Омск, Россия

Аннотация. Рассматриваются основные проблемы, с которыми сталкиваются преподаватели и студенты, изучая тему репликации, а также предлагается возможный путь их решения.

Ключевые слова: преподавание, репликация, безопасность баз данных.

Введение

Изучение основ построения защищённых баз данных является важной частью подготовки специалистов в области безопасности информационных технологий. Одной из ключевых тем в рамках дисциплины «Основы построения защищённых баз данных» является изучение механизмов обеспечения высокой доступности баз данных. В системах, где требуется высокая надёжность, доступность данных и минимизация времени простоя, важна репликация базы данных. Анализ учебных пособий, лабораторных практикумов в области баз данных и безопасности баз данных показал, что в большинстве из них отсутствует практическое освещение данной темы, и лишь в редких случаях она рассматривается с теоретической точки зрения [1–3]. Это создаёт пробел в образовательном процессе, что подчёркивает необходимость более глубокого изучения и внедрения практических аспектов настройки репликации баз данных в учебные программы. Однако обучение студентов и формирование у них практических навыков в этой области сталкивается с рядом проблем.

В данной статье будут рассмотрены основные трудности, с которыми сталкиваются студенты, изучая тему репликации, а также предложены возможные пути их решения.

1. Методологические проблемы при реализации репликации базы данных

Репликация базы данных – это процесс создания и поддержания копий базы данных на нескольких серверах для повышения доступности, надёжности и производительности базы данных. Наиболее часто используется организация репликации, при которой изменения могут выполняться только на одном главном (master) сервере. Изменения, выполненные на этом сервере, переносятся на все копии, которые могут использоваться только для чтения данных, и называются запасными (standby), или репликами (replica) [4].

При реализации репликации в образовательной среде студенты и преподаватели могут столкнуться с рядом проблем:

1. Длительность и сложность настройки при ограниченном времени занятия:

1.1. Многообразие технологий.

Существует множество технологий и подходов к репликации (например, синхронная и асинхронная репликация, мастер-мастер и т. д.). Преподаватели могут испытывать трудности с выбором подходящей технологии и её настройкой на лабораторных занятиях. Разные операционные системы, системы управления базами данных (СУБД) имеют свои уникальные решения и процедуры в области репликации, что может привести порой к сложности реализации и нехватки времени в рамках одного занятия для полноценного разбора.

1.2. Настройка сетевого окружения.

Для успешной реализации репликации необходимо правильно настроить сетевую инфраструктуру. Для настройки репликации требуется как минимум два сервера (мастер и реплика), что в учебных условиях сложно организовать. В этом случае студенты могут столкнуться с проблемами настройки сетевых адресов и маршрутизации, что может затруднить процесс обучения.

2. Проблемы с синхронизацией данных.

При использовании асинхронной репликации могут возникать конфликты данных, когда изменения в одной базе данных не успевают быть отражены в другой. Это может создать ложное представление о целостности данных у студентов и затруднить их понимание процессов, связанных с репликацией.

3. Управление отказами:

3.1. Обнаружение сбоев и восстановление работоспособности системы.

Студенты должны научиться выявлять сбои в системе репликации и восстанавливать нормальную работу СУБД. Это включает в себя диагностику проблем и применение методов восстановления данных.

3.2. Потеря данных.

В случае сбоя системы репликации существует риск потери данных. Необходимо проводить репликацию учебной тестовой базы данных. Обучающиеся должны быть осведомлены о возможных последствиях и о том, как минимизировать риски потери информации.

4. Безопасность данных. Уязвимости при передаче данных в момент репликации базы данных.

Репликация данных между серверами может создавать уязвимости, связанные с безопасностью. Если данные передаются по незащищённым каналам (от мастера к реплике), то злоумышленники могут перехватить информацию, включая учётные данные и личные данные пользователей. Студенты должны осознавать важность шифрования данных при передаче и применять меры

безопасности для защиты информации. Таким образом, тема «Методы шифрования в базе данных» должна стоять перед изучением репликации.

5. Ограниченные ресурсы.

В образовательном учреждении часто отсутствуют необходимые ресурсы (память, процессор, дисковое пространство) для полноценного обучения репликации баз данных. Студенты могут не иметь доступа к реальным техническим инструментам или программам, что ограничивает их практический опыт. Вместо работы с реальными данными студенты могут использовать симуляционные модели, которые требуют меньше ресурсов. Это позволит им понять основные принципы репликации данных без необходимости в больших объёмах информации.

6. Теоретические знания без практики.

Обучение может быть сосредоточено на теоретических аспектах без достаточного количества практических занятий. Это может привести к недостаточному пониманию студентами материала, связанного с репликацией баз данных.

2. Формирование компетенций в области репликации баз данных

Целью занятия по настройке репликации является изучение процессов и методов, позволяющих создавать копии данных с одного сервера на другой, обеспечивая тем самым высокую доступность, отказоустойчивость и балансировку нагрузки в СУБД. В ходе занятия студенты должны ознакомиться с различными типами синхронной репликации, а также получить практические навыки по настройке репликации в конкретной СУБД.

В настоящее время для изучения дисциплины «Основы построения защищённых баз данных» на факультете цифровых технологий и кибербезопасности Омского государственного университета им. Ф.М. Достоевского используется PostgreSQL.

Практическая часть занятия по репликации данных предполагает решение двух задач в PostgreSQL с использованием двух виртуальных машин:

1. Настройка синхронной репликации в PostgreSQL, когда одна виртуальная машина является мастером, а вторая – репликой, состоит из следующих этапов:
 - Создание двух виртуальных машин VMware (VM1 и VM2) с установленной операционной системой (например, Ubuntu Linux). Для реализации репликации используются две виртуальные машины под Ubuntu. Одна является мастером, а вторая – репликой.
 - Установка PostgreSQL на обе виртуальные машины.
 - Настройка сети виртуальной машины VM1 с IP-адресом и виртуальной машины VM2 с IP-адресом.
 - Настройка основной (мастер) машины.
 - Настройка резервной (реплика) машины.

- Создание серверов базы данных и проверка репликации.
2. Настройка синхронной репликации в PostgreSQL, когда одна виртуальная машина является мастером, а вторая – репликой, но при этом если сбой происходит на машине, которая является мастером, то вторая машина становится мастером на время сбоя первой. Работа с базой данных осуществляется с той машины, которая является мастером в кластере (под кластером здесь понимается серверный процесс, а также набор баз данных под его управлением). Только мастер может вносить изменения в базу данных, реплика же находится в режиме чтения, и с неё вносить изменения в базу данных нельзя.

Решение этой задачи связано с дополнительными техническими сложностями при настройке отказоустойчивого кластера PostgreSQL, с использованием Patroni и Etcd, но является более актуальной. Patroni – Framework, разработанный для построения кластера PostgreSQL высокой доступности на основе потоковой репликации [5]. Если мастер-узел становится недоступным, Patroni автоматически выбирает новую реплику и переводит её в режим мастера. Etcd – это распределённая система хранения ключей-значений, используемая для хранения конфигурации и метаданных. Patroni использует Etcd для синхронизации состояния узлов кластера, чтобы обеспечить отказоустойчивость. Etcd обеспечивает непрерывность работы кластера в случае сбоя отдельных узлов.

Это современное решение подходит для сред, где требуется минимизировать время простоя и обеспечить бесперебойную работу баз данных.

Настройка отказоустойчивого кластера PostgreSQL с использованием Patroni и Etcd предполагает:

- Создание двух виртуальных машин VMware (VM1 и VM2) с установленной операционной системой (например, Ubuntu Linux). Для реализации репликации используются две виртуальные машины под Ubuntu. Одна является мастером, а вторая – репликой.
- Установку PostgreSQL на обе виртуальные машины.
- Настройку сети виртуальной машины VM1 с IP-адресом и виртуальной машины VM2 с IP-адресом.
- Установку Etcd.
- Настройку Etcd (создание конфигурационного файла для каждой виртуальной машины). Студентам предлагается шаблон конфигурационного файла (см. рис. 1) с описанием каждого параметра.
- Настройку автоматизации запуска Etcd как сервиса в системе.
- Проверку корректной работы Etcd.
- Установку и настройку Patroni (создание конфигурационного файла Patroni для каждой виртуальной машины). Студентам также предлагается шаблон конфигурационного файла (см. рис. 2);
- Проверку корректной работы Patroni.
- Создание серверов базы данных и тестирование работы репликации.

```

1 name: "node1"
2 data-dir: "node1.etcd"
3 initial-advertise-peer-urls: http://192.168.1.36:2380
4 listen-peer-urls: http://192.168.1.36:2380
5 listen-client-urls: http://192.168.1.36:2379,http://127.0.0.1:2379
6 advertise-client-urls: http://192.168.1.36:2379
7 initial-cluster: "node1=http://192.168.1.36:2380,node2=http://192.168.1.37:2380"
8 initial-cluster-state: "new"
9 initial-cluster-token: "etcd-cluster"

```

Рис. 1. Пример конфигурационного файла Etcd для виртуальной машины VM1

```

1 scope: postgres
2 namespace: /db/
3 name: fatal1
4
5 restapi:
6 listen: 192.168.1.36:8008
7 connect_address: 192.168.1.36:8008
8
9 etcd3:
10 hosts: 192.168.1.36:2379,192.168.1.37:2379
11
12 bootstrap:
13 dcs:
14 ttl: 30
15 loop_wait: 10
16 retry_timeout: 10
17 maximum_lag_on_failover: 1048576
18 postgresql:
19 use_pg_rewind: true
20 initdb:
21 - encoding: UTF8
22 - data-checksums
23 pg_hba:
24 - host replication replicator 127.0.0.1/32 md5
25 - host replication replicator 192.168.1.36/0 md5
26 - host replication replicator 192.168.1.37/0 md5
27 - host all all 0.0.0.0/0 md5
28 users:
29 admin:
30 password: admin
31 options:
32 - createrole
33 - createdb
34
35 postgresql:
36 listen: 192.168.1.36:5432
37 connect_address: 192.168.1.36:5432
38 data_dir: /mnt/patroni
39 bin_dir: /usr/lib/postgresql/16/bin
40 pgpass: /tmp/pgpass
41 authentication:
42 replication:
43 username: replicator
44 password: '1234'
45 superuser:
46 username: postgres
47 password: '1234'
48 parameters:
49 unix_socket_directories: ''
50 wal_level: replica
51 max_wal_senders: 10
52 max_replication_slots: 10
53 synchronous_commit: local
54
55 tags:
56 nofailover: false
57 noloadbalance: false
58 clonefrom: false
59 nosync: false

```

Рис. 2. Пример конфигурационного файла Patroni для виртуальной машины VM1

Заключение

Формирование практических навыков студентов при изучении темы «Настройка репликации базы данных» является важной задачей в рамках дисциплины «Основы построения защищённых баз данных». Несмотря на существующие проблемы, их решение возможно через использование современных инструментов, таких как виртуальные машины, системы автоматизации процесса репликации, упрощение настройки окружения, разработка подробных практических заданий. Это позволит студентам не только освоить теорию, но и приобрести необходимые навыки для успешной профессиональной деятельности в области информационных технологий.

Литература

1. Агафонов А.А., Юмаганов А.С. Безопасность систем баз данных: учеб. пособие. Самара: Изд-во Сам. ун-та, 2023. 272 с.
2. Дудник Е.А. Информационная безопасность баз данных: метод. указ. для самостоят. работы студентам направления подготовки «Информатика и вычислительная техника». Рубцовск: РИИ, 2021. 32 с.
3. Пригонюк Н.Д., Петров В.И. Основы построения защищенных баз данных: учеб. пособие. Воронеж: ООО «МИР», 2019. 76 с.
4. Новиков Б.А. Основы технологий баз данных: учеб. пособие. М.: ДМК Пресс, 2020. 582 с.
5. Отказоустойчивый кластер СУБД Postgres Pro на основе Patroni. URL: <https://postgrespro.ru/clusters/patroni> (дата обращения: 24.11.2024).

SYSTEMATIZATION OF TOOLS AND METHODS FOR EFFECTIVE DATABASE REPLICATION

Т.М. Опарина

Assistant Professor, e-mail: oparina2007@yandex.ru

Dostoevsky Omsk State University, Omsk, Russia

Abstract. This article examines the main problems that teachers and students encounter why studying the topic of replication, and also suggests a possible way to solve them encounters.

Keywords: teaching, replication, database security.

Дата поступления в редакцию: 04.02.2025