



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИУ «Информатика и системы управления»

КАФЕДРА ИУ-7 «Программное обеспечение ЭВМ и информационные технологии»

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ
НА ТЕМУ:**

***«Классификация методов достижения семантики
строго однократной доставки в брокерах
сообщений»***

Студент ИУ7-52Б

_____ Княжев А. В.

Руководитель

_____ Кивва К. А.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

УТВЕРЖДАЮ

Заведующий кафедрой ИУ-7
(Индекс)

_____ И. В. Рудаков
(И.О.Фамилия)

«16» сентября 2022 г.

З А Д А Н И Е
на выполнение научно-исследовательской работы

по теме

**«Классификация методов достижения семантики строго однократной доставки в
брокерах сообщений»**

Студент группы ИУ7-52Б

Княжев Алексей Викторович

Направленность НИР

учебная

Источник тематики

НИР кафедры

График выполнения НИР: 25% к 6 нед., 50% к 9 нед., 75% к 12 нед., 100% к 15 нед.

Техническое задание

Провести обзор существующих методов достижения семантики строго однократной доставки в брокерах сообщений. Провести анализ предметной области, сформулировать критерии сравнения методов достижения семантики строго однократной доставки в брокерах сообщений и классифицировать эти методы.

Оформление научно-исследовательской работы:

Расчетно-пояснительная записка на **12-20** листах формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т. п.)

Презентация на **6-10** слайдах.

Дата выдачи задания «16» сентября 2022 г.

Руководитель НИР

_____ **Кивва К. А.**
(Подпись, дата) (И.О.Фамилия)

Студент

_____ **Княжев А. В.**
(Подпись, дата) (И.О.Фамилия)

СОДЕРЖАНИЕ

ОПРЕДЕЛЕНИЯ	4
РЕФЕРАТ	5
ВВЕДЕНИЕ	6
1 Анализ предметной области	7
1.1 Актуальность задачи	7
1.2 Описание принципов работы брокеров сообщений	8
1.3 Необходимые условия для достижения семантики строго однократной доставки	8
1.3.1 Сохранность данных	8
1.3.2 Надежность брокера сообщений	9
1.4 Методы достижения семантики строго однократной доставки . .	9
1.4.1 Методы достижения семантики доставки не менее одного раза	9
1.4.2 Методы достижения семантики доставки не более одного раза	10
1.5 Области применения методов достижения семантики строго однократной доставки в брокерах сообщений	13
1.6 Вывод	14
2 Классификация методов достижения семантики строго однократной доставки в брокерах сообщений	15
2.1 Определение критериев классификации	15
2.1.1 По обеспечиваемой базовой семантике	15
2.1.2 По стороне обеспечения семантики	15
2.1.3 По использованию внешних сервисов	16
2.2 Вывод	16
ЗАКЛЮЧЕНИЕ	17
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	19

ОПРЕДЕЛЕНИЯ

В настоящей расчетно-пояснительной записке применяют следующие термины с соответствующими определениями.

Брокер сообщений — приложение, которое выступает посредником между двумя приложениями — источником и приемником, выполняя преобразование сообщения по протоколу источника в сообщение по протоколу приемника.

Протокол — набор определенных правил или соглашений, определяющий обмен данными между приложениями.

Потребитель — приложение, которое осуществляет чтение сообщений из брокера сообщений, приемник данных.

Производитель — приложение, которое осуществляет запись сообщений в брокера сообщений, источник данных.

Кластер — группа компьютеров, связанных друг с другом, и воспринимаемых как единый ресурс.

Реплика — один из экземпляров приложения. Приложение обычно запускают в нескольких экземплярах с целью обеспечения большей надежности.

Транзакция — единая операция, возможно состоящая из нескольких операций, которая выполняется как единый неделимый блок, либо не выполняется вовсе.

Микросервис — небольшой модуль, отвечающий за небольшой фрагмент бизнес-логики продукта. Минимально зависим от других модулей.

РЕФЕРАТ

Расчетно-пояснительная записка 19 с., 0 рис., 0 табл., 13 источн., 2 прил.

Целью данной научно-исследовательской работы является классификация методов достижения семантики строго однократной доставки в брокерах сообщений.

В результате был проведен обзор существующих методов достижения семантики строго однократной доставки в брокерах сообщений, проведен анализ предметной области, сформулированы критерии сравнения методов достижения семантики строго однократной доставки, классифицированы эти методы и предложены наиболее подходящие области применения данных методов.

Ключевые слова: брокер сообщений, однократная доставка, микросервисы, распределенные системы, событийная архитектура, классификация.

ВВЕДЕНИЕ

С ростом сложности программных продуктов возникает необходимость задумываться о сложности управления приложением. Важным аспектом в решении данной задачи является архитектура приложения. В современном мире все более и более актуальным решением становится микросервисная архитектура, в которой бизнес-логика продукта разбита на отдельные части со своей кодовой базой — микросервисы [1].

Однако, при использовании микросервисной архитектуры, возникает проблемы выбора технологии обмена данными между сервисами. Одной из распространенных технологий такого рода являются брокеры сообщений [2].

Большинство брокеров сообщений гарантируют доставку сообщений «хотя бы один раз», то есть не исключают вероятность появления дубликатов. Такое поведение допустимо для большого количества продуктов, но, например, для банковской сферы является недопустимым [3]. Для такого рода систем необходимо обеспечивать семантику строго однократной доставки — гарантию того, что сообщение дойдет до получателя строго один раз.

Целью данной научно-исследовательской работы является классификация методов достижения семантики строго однократной доставки в брокерах сообщений. Для достижения поставленной цели необходимо решить следующие задачи:

- провести анализ предметной области;
- провести обзор существующих методов достижения семантики строго однократной доставки в брокерах сообщений;
- сформулировать критерии сравнения методов достижения семантики строго однократной доставки в брокерах сообщений;
- классифицировать методы достижения семантики строго однократной доставки.

1 Анализ предметной области

В данной части рассматривается актуальность задачи и существующие методы достижения семантики строго однократной доставки в различных брокерах сообщений.

1.1 Актуальность задачи

В современных брокерах сообщений используется три основных семантики передачи сообщений [4]:

- «хотя бы один раз» — сообщение будет передано хотя бы один раз. Гарантируется отправка сообщения, но не гарантируется отсутствие дубликатов. Вероятность появления дубликата в таком случае мала, но она есть. Например, если сообщение было сохранено брокером, но при этом до поставщика данных не дошла информация о принятии сообщения, он попытается отправить его еще один раз, из-за этого появится дубликат;
- «не более одного раза» — поставщик данных ни в каком случае не попытается повторно записывать данные в брокер. Такой подход не гарантирует доставку сообщения в случае ошибки;
- «строго однократная доставка» — сообщение гарантированно будет передано, причем строго один раз.

Несмотря на то, что последняя семантика более применима, так как гарантирует как доставку сообщения, так и отсутствие дубликатов, ее техническая реализация является наиболее сложной, в связи с такими нерешенными проблемами, как задача о двух генералах [5].

Строго однократная доставка сообщений необходима в банковской сфере из-за недопустимости повторных финансовых операций, при синхронизации критично важных данных с использованием брокера сообщений, при сборе и отправке аналитических данных.

1.2 Описание принципов работы брокеров сообщений

В большинстве брокеров сообщений хранение организовано в формате очереди. То есть когда поставщик данных отправляет сообщение брокеру, оно помещается в очередь. После принятия сообщения, брокер отправляет подтверждение принятия сообщения поставщику.

Соответственно, от подтверждения зависит дальнейшее поведение поставщика, например, в случае отсутствия подтверждения может произойти повторная отправка сообщения. Брокер может предоставлять несколько стратегий подтверждения принятия сообщений [6]:

1. отсутствие ожидания ответа от брокера. В таком случае, отправка сообщения сразу будет считаться успешной. При данной стратегии, при возникновении ошибок записи, поставщик данных все равно будет считать отправку сообщения успешной, соответственно, никаких действий принято не будет, и сообщение потеряется;
2. ожидание ответа от ведущего узла. Если брокер сообщений состоит из нескольких узлов, то для подтверждения успешности операции достаточно получить подтверждение успешной записи только с одного узла;
3. ожидание ответа от всех узлов. Для подтверждения успешной записи необходимо подтверждение от всех узлов.

1.3 Необходимые условия для достижения семантики строго однократной доставки

1.3.1 Сохранность данных

Для гарантии доставки сообщения необходимо обеспечивать сохранность данных. Это означает, что сообщения должны сохраняться на диске в течение некоторого времени, в течение которого эти данные могут быть нужны.

1.3.2 Надежность брокера сообщений

Брокер должен состоять как минимум из нескольких реплик, чтобы при выходе узлов из строя, данные не были утеряны. Если условиться, что в некотором кластере брокера сообщений будет n узлов, то допустим выход из строя до $n - 1$ узлов при сохранении сообщений в системе [4].

1.4 Методы достижения семантики строго однократной доставки

Не существует единого метода достижения семантики строго однократной доставки. Однако существуют методы, композиция которых позволяет сделать достижимой строго однократную доставку. Условно эти методы можно разделить на две группы: методы, обеспечивающие доставку не менее одного раза, и методы, обеспечивающие доставку не более одного раза. Совместное применение методов доставки не менее и не более одного раза дает строго однократную доставку. Кроме того, необходимо совмещать указанные методы, работающие как на стороне связки производитель-брокер, так и брокер-потребитель.

1.4.1 Методы достижения семантики доставки не менее одного раза

Получение подтверждения от реплик брокера

При отправке сообщения в брокер, производитель данных ожидает подтверждение получения сообщения от всех реплик брокера [6]. Только после подтверждения всеми экземплярами, сообщение считается успешно отправленным. Таким образом, метод позволяет гарантировать доставку сообщения в брокер.

Метод двойного опроса

В качестве дополнительного подтверждения получения брокером сообщения может быть использован метод двойного опроса [7]: производитель ожидает два подтверждения получения сообщения. Если первое подтверждение оказалось

ошибочным, например из-за слишком долгого времени ответа, то производитель ожидает второй ответ. Если он оказывается успешным, то повторных попыток отправить сообщение не производится.

Хранение информации о прочитанных данных на стороне потребителя

Для выдачи нужного сообщения потребителям необходимо хранить индекс последнего прочитанного сообщения. Варианта расположения хранимого индекса два: можно хранить его на стороне сервера (брокера сообщений) [8], можно на стороне клиента (потребителя) [9]. Возможность возникновения ошибок сети заметно усложняет реализацию семантики строго однократной доставки в брокерах, хранящих позицию прочитанного сообщения на стороне сервера, так как в таком случае позиция может быть неверно посчитана из-за сетевых ошибок при отправке потребителем подтверждения прочтения сообщения. Поэтому для обеспечения семантики строго однократной доставки необходимо хранить информацию о текущем сдвиге в очереди на стороне потребителя данных.

1.4.2 Методы достижения семантики доставки не более одного раза

Идемпотентность сохранения сообщения на стороне брокера

Идемпотентной операцией называют операцию, выполнение которой несколько раз влечет тот же результат, что и однократное выполнение [10]. С помощью идемпотентности можно исключить дублирование записей в брокере сообщений. При повторной попытке записи того же сообщения, повторного сохранения не производится. Так как в соответствии с 1.3 гарантируется доставка сообщения хотя бы один раз, то метод позволяет частично обеспечить семантику строго однократной доставки.

Реализуется такой подход следующим образом [11]: на стороне поставщика данных сообщению присваивается уникальный идентификатор. Брокер при по-

лучении сообщения сохраняет его идентификатор, и при последующем принятии данных, проверит наличие в хранилище этого идентификатора. Если он найден, значит это сообщение уже было принято и сохранено брокером, и необходимости в повторном сохранении этого сообщения нет. Иначе происходит сохранение нового сообщения с записью его идентификатора в хранилище.

Обработка состояния прочтения на стороне потребителя

Логику по обеспечению однократной доставки можно частично делегировать потребителю. При действительной логике «хотя бы одной доставки» со стороны сервера, на стороне потребителя можно контролировать отсутствие дубликатов следующим образом [2]: каждому сообщению присваивается уникальный идентификатор. При получении очередного сообщения, потребитель проверяет, сохранен ли идентификатор сообщения в базе данных (в качестве базы данных в данном случае можно использовать, например, Redis [9]). Если сохранен, то сообщение игнорируется как дубликат, если нет, то оно обрабатывается, а его идентификатор помещается в базу данных.

Транзакции

Транзакции позволяют работать с несколькими сообщениями атомарно, то есть, при попытке отправить несколько сообщений, они либо будут отправлены все, либо не будет отправлено ни одно. Поведение потребителей сообщений может быть настроено двумя способами:

1. принимать сообщения в том формате, в котором они помещаются в брокер, в том числе принимать незавершенные транзакции;
2. принимать завершенные транзакции.

При использовании метода транзакций приложению конечным пользователем назначается идентификатор. Благодаря использованию идентификатора брокер может гарантировать [12]:

- идемпотентность поставки данных — при появлении в сети поставщика с существующим идентификатором, запущенные поставщики с тем же идентификатором прекращают свою работу. Таким образом, если существует несколько реплик приложения поставщика, в один момент времени может работать только одна;
- транзакционное восстановление между сессиями приложения-поставщика. Если экземпляр приложения завершается (например, завершается с ошибкой), то для следующего экземпляра приложения существует гарантия того, что все незавершенные транзакции будут либо выполнены успешно, либо прерваны.

Гарантии, получаемые при использовании транзакций, более характерны для поставщиков данных, для потребителей данных они не настолько сильны, так как нет гарантии того, что потребитель будет использовать все сообщения одной транзакции. Данное поведение возникает по следующим причинам:

- при сжатии в целях экономии памяти часть сообщений могут заменяться более новыми версиями [12];
- потребители могут обращаться к конкретным сообщениям внутри транзакции;
- при записи нескольких сообщений в рамках одной транзакции сообщения могут быть записаны в разные очереди. Но потребитель может не иметь доступ ко всем очередям, поэтому не сможет прочитать все сообщения в рамках одной транзакции.

Транзакции полезны, когда данные о текущем прочитанном элементе передаются в брокер сообщений. При такой схеме, фактическое чтение сообщения и отправку брокеру пометки о его прочтении можно совершать атомарно, в рамках одной транзакции. Таким образом, исключается ситуация, когда сообщение было прочитано, при этом оно не было помечено как прочитанное.

Хранение информации о прочитанных данных на стороне потребителя

Метод, описанный в 1.4.1, также позволяет обеспечивать доставку не более одного раза на стороне потребителя данных, так как при успешном прочтении сообщение обязательно будет помечено как прочитанное, так как это действие делается на стороне потребителя. Таким образом исключается ситуация, когда сообщение было прочитано, при этом оно не было помечено как прочитанное.

1.5 Области применения методов достижения семантики строго однократной доставки в брокерах сообщений

В данной части рассмотрены наиболее подходящие области применения методов достижения семантики строго однократной доставки в брокерах сообщений.

Семантика строго одновременной доставки в брокерах сообщений является нужным инструментом во многих сферах. Ниже приведены некоторые из них.

- **Банковская сфера.** Брокеры сообщений могут быть использованы для потоковой передачи данных о банковских операциях. Очевидно, что в такой системе недопустимы как потери данных, так и их дубликация, так как в первом случае это грозит потерей банковского перевода, а во втором — двойным списанием средств.
- **Аналитические данные.** Некоторые аналитические данные могут быть довольно важными и не терпеть потери части данных или появление «мусорных» данных, которые и возникают в ходе появления дубликатов.
- **Синхронизация данных.** Брокеры сообщений можно использовать для синхронизации данных между различными хранилищами и базами данных. Одной из таких систем является Kafka Connect [13], работающий на базе Kafka. Для такой системы нежелательны потери данных, но при этом дубликация данных тоже нежелательна, так как она может привести к

искажению данных в целевой базе данных.

- **Отправка уведомлений.** При использовании брокера сообщений в качестве шины уведомлений, например, для социальной сети, нежелательны потери или дубликации данных, так как в первом случае пользователю не придет уведомление о чем-то важном, а во втором случае ему придет одно и то же уведомление несколько раз.

1.6 Вывод

В данной части рассмотрены актуальность задачи и существующие методы достижения семантики строго однократной доставки в различных брокерах сообщений. Обеспечение этой семантики представляет собой комбинацию методов достижения семантики доставки не менее и не более одного раза. Кроме того, применяемые методы могут работать как со стороны брокера или производителя, так и со стороны потребителя. Например, в брокере сообщений может использоваться одновременно идемпотентность, получение подтверждения от всех реплик брокера и хранение информации о прочитанных сообщениях на стороне потребителя. Кроме того, были рассмотрены наиболее подходящие области применения методов достижения семантики строго однократной доставки в брокерах сообщений.

2 Классификация методов достижения семантики строго однократной доставки в брокерах сообщений

В данной части рассматриваются критерии классификации методов достижения семантики строго однократной доставки в брокерах сообщений.

2.1 Определение критериев классификации

2.1.1 По обеспечиваемой базовой семантике

По обеспечиваемой семантике можно выделить методы, которые позволяют обеспечить доставку сообщений не более одного раза, и методы, которые позволяют обеспечить доставку сообщений хотя бы один раз.

Не более одного раза: идемпотентность, обработка состояния прочтения на стороне потребителя, транзакции, хранение информации о прочитанных данных на стороне потребителя.

Хотя бы один раз: получение подтверждения от реплик брокера, метод двойного опроса, хранение информации о прочитанных данных на стороне потребителя.

2.1.2 По стороне обеспечения семантики

По стороне обеспечения семантики строго однократной доставки можно выделить методы, которые обеспечиваются на стороне производителя данных, на стороне потребителя данных, и на стороне брокера сообщений.

На стороне производителя: получение подтверждения от реплик брокера, метод двойного опроса.

На стороне потребителя: транзакции, обработка состояния прочтения на стороне потребителя, хранение информации о прочитанных данных на стороне потребителя.

На стороне брокера сообщений: получение подтверждения от реплик брокера, идемпотентность, транзакции.

2.1.3 По использованию внешних сервисов

По использованию внешних сервисов можно выделить методы, которые можно реализовать исключительно с использованием средств брокера сообщений, и методы, для реализации которых потребуется использование внешних сервисов.

Без использованием внешних сервисов: получение подтверждения от реплик брокера, идемпотентность, транзакции, метод двойного опроса, хранение информации о прочитанных данных на стороне потребителя.

С использованием внешних сервисов: обработка состояния прочтения на стороне потребителя.

2.2 Вывод

В данной части были рассмотрены критерии классификации методов достижения семантики строго однократной доставки в брокерах сообщений и выполнена классификация этих методов.

ЗАКЛЮЧЕНИЕ

В данной работе были рассмотрены методы достижения семантики строго однократной доставки в брокерах сообщений.

Произведен анализ предметной области: рассмотрена актуальность задачи и произведено разбиение задачи осуществления семантики строго однократной доставки на методы, обеспечивающие доставку не более одного и не менее одного раза, которые должны применяться в совокупности. Описан принцип работы брокеров сообщений, а также необходимые условия для достижения семантики строго однократной доставки: сохранность данных, надежность брокера сообщений и хранение информации о прочитанных данных на стороне потребителя.

Определены критерии классификации методов достижения семантики строго однократной доставки в брокерах сообщений, а также произведена классификация этих методов на основе определенных критериев.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Шитько А. М.* Проектирование микросервисной архитектуры программного обеспечения // Труды БГТУ. Серия 3: Физико-математические науки и информатика. — 2017. — 9 (200). — С. 122—125.
2. *Кондратьев С. В., Хайрисламов Д. А.* Сравнительный анализ методов достижения exactly once семантики в распределённом брокере сообщений Apache Kafka // Материалы конференций ГНИИ «НАЦРАЗВИТИЕ». — 2021. — С. 37—39.
3. EOS: Exactly-Once E-Service Middleware / G. Shegalov [и др.] // VLDB'02: Proceedings of the 28th International Conference on Very Large Databases. — Elsevier. 2002. — С. 1043—1046.
4. Exactly-Once Semantics Are Possible: Here's How Kafka Does It. — 2017. — URL: <https://www.confluent.io/blog/exactly-once-semantics-are-possible-heres-how-apache-kafka-does-it/>.
5. *Байгулов Р., Сковиков А.* Финансовые технологии-блокчейн // Экономика и управление народным хозяйством: генезис, современное состояние и перспективы развития. — 2017. — С. 152—159.
6. *Narkhede N., Shapira G., Palino T.* Kafka: the definitive guide: real-time data and stream processing at scale. — "O'Reilly Media, Inc.", 2017.
7. JetStream NATS Docs. — 2022. — URL: <https://clck.ru/33RcEP>.
8. RabbitMQ documentation. — 2022. — URL: <https://www.rabbitmq.com/documentation.html>.
9. Kafka documentation. — 2022. — URL: <https://kafka.apache.org/documentation/>.
10. *Дудников П. И., Самборский С. Н.* Эндоморфизмы полумодулей над полукольцами с идемпотентной операцией // Известия Российской академии наук. Серия математическая. — 1991. — Т. 55, № 1. — С. 93—109.
11. Idempotent Producer. — 2014. — URL: <https://clck.ru/33RcF2>.

12. Exactly Once Delivery and Transactional Messaging in Kafka . The definitive design. — 2017. — URL: <https://clck.ru/33RcFX>.
13. Kafka Connect. — 2022. — URL: <https://clck.ru/33RcDv>.