

HW #11: Cassandra + Spark (bonus track)

Deadline: 02.09.2019 08:00

1. Описание задания.	1
2. Критерии оценивания	2
3. Задача	2
4. Дорешка.	3

1. Описание задания.

В этом задании вам необходимо разработать дизайн приложения, выполняющего обработку и хранение данных. Назовем это приложение машиной состояний пользователей (далее - МСП).

Источником данных является распределенное веб приложение, на которое заходят пользователи и выполняют разные действия - от просмотра полезной информации до совершения покупок. Мы хотим адаптировать поведение приложения под различных пользователей. Например, если пользователь недавно совершил покупку и опять зашел на сайт, то мы хотим сделать акцент на рекомендации для этого пользователя.

Для другого пользователя, который заходил на сайт несколько раз, но так ничего не приобрел, мы бы хотели показывать всплывающее окошко с помощью и возможностью связаться с нашими консультантами.

Таким образом, у каждого пользователя нашего приложения есть статус (или состояние), который может меняться. Состояния зависят от поведения пользователя на сайте и могут быть следующими:



- NEW - состояние “по-умолчанию” для всех новых пользователей
- BOUNCED - пользователь зашел на сайт, ничего не купил и больше не проявлял активности (мы считаем, что пользователь переходит в состояние BOUNCED спустя 24 часа после того, как зашел на сайт и не проявил активности)
- ACTIVE - пользователь совершал хотя бы одно действие (не обязательно покупку) за последнюю неделю
- LAPSED - пользователь, который имел статус ACTIVE, но не проявлял никакой активности в течение последней недели
- RECENT_PURCHASE - пользователь совершил покупку за последние 48 часов

У каждого пользователя есть свой уникальный неизменный `user_id`, который остается у него даже если он длительное время не посещал наше приложение.

2. Критерии оценивания

Балл за задачу складывается из:

- 20% - рабочая схема
- 10% - присутствует обоснование отказоустойчивости и высокой доступности
- 10% - присутствует обоснование консистентности данных, исключения состояний гонки
- 60% - 0-1 сек задержки между обновлением состояния и его фактическим изменением, 30% - 1-60 сек задержки, 0% - 60+ сек задержки

Штрафы:

- 10% - за несоответствие правилам оформления задания
- 30% - за просрочку дедлайн

3. Задача

Вам необходимо разработать дизайн МСП. Представьте, что вы являетесь техническим лидером проекта и хотите презентовать дизайн коллегам на митинге и обосновать выбор тех или иных компонентов. Результат этой задачи должен быть оформлен в виде небольшой презентации в Google Slides.

В этой задаче вам не нужно писать код. От вас требуется дизайн решения с указанием компонент и обоснованием их выбора. При этом не забывайте, что в любой



распределенной системе присутствует вероятность сбоев, состояний гонки и т. п. - ваш дизайн должен все это учитывать.

Требования к МСП:

- Сервис должен предоставлять API, которое позволит получать текущий статус пользователя по его user_id
- Изменившееся состояния пользователя должно отражаться в МСП как можно быстрее. Это не означает, что система должна работать в режиме реального времени, но и обновление состояния через 24 часа после его фактического изменения не допускается. Идеальное решение должно обеспечивать обновление состояния спустя не более чем 1 секунду с момента его фактического изменения
- Приложение МСП должно масштабироваться горизонтально так, чтобы в случае роста нагрузки нам не пришлось изменять его дизайн
- Действия пользователей могут направляться вам любым удобным образом (отправляться в топик(и) Kafka, прилетать через REST API и т. д)

Требования к составу компонент:

- Вы можете использовать ЛЮБЫЕ системы (RDBMS, NoSQL, Hadoop, Spark, Kafka и т. д.)
- Основное требование - выбор должен быть обоснованным и должен учитывать особенности выбранных решений
- Например, если вы будете использовать Cassandra для решения этой задачи, то вам необходимо детально описать схему всех таблиц, структуру композитных ключей и указать, как будет происходить чтение и запись в них.

Deadline: 02.09.2019 08:00

Оформление задания:

- Код задания (Short name): **HW11:NoSQL**
- Решения задач должны содержаться в одной презентации Google Slides.
- Присылайте выполненное задание на почту bigdata_mf2019q2@bigdatateam.org с темой письма "Short name. ФИО". Например: "**HW11:NoSQL**. Иванов Иван Иванович". В теле письма укажите ссылку на презентацию.
- Перед отправкой письма, оставьте, пожалуйста, отзыв о домашнем задании по ссылке: http://rebrand.ly/mf2019q2_feedback_11_cassandraspark. Это позволит нам скорректировать учебную нагрузку по следующим заданиям (в зависимости от того, сколько часов уходит на решение ДЗ), а также ответить на интересующие вопросы.



Любые вопросы / комментарии / предложения можно писать:

- в телеграм-канале: http://rebrand.ly/mf2019q2_telegram_join
- На почту: bigdata_mf2019q2@bigdatateam.org

4. Дорешка.

Решения после получения фидбека на решение ДЗ можно улучшить. Разрешается одна досылка в течение 1й недели после окончания дедлайна за ДЗ. Соответственно, фидбек за дорешенные ДЗ вы получите в течение 24 часов после окончания deadline следующего ДЗ.

Дорешивать неработающие задания - нельзя. Это позволит исправить дисбаланс между
присланными **работающими** заданиями **после** deadline

VS

присланными **НЕработающими** заданиями **до** deadline

Всем удачи!