Если вы видите что-то необычное, просто сообщите мне.

Какую типобезопасную библиотеку базы данных вы должны использовать?

Beam или Squeal: что лучше? Или может быть вы слышали про отличную штуку Selda или Opaleye. Множество мнений, редкие руководства.

Чтобы ответить на вопрос, я взял 7 популярных библиотек для базы данных и реализовал один и тот же проект, на каждой из них.

Участники:

- Beam
- Opaleye
- Squeal
- Persistent + Esqueleto
- Hasql
- Groundhog
- Selda

Почемы мы используем эти библиотеки?

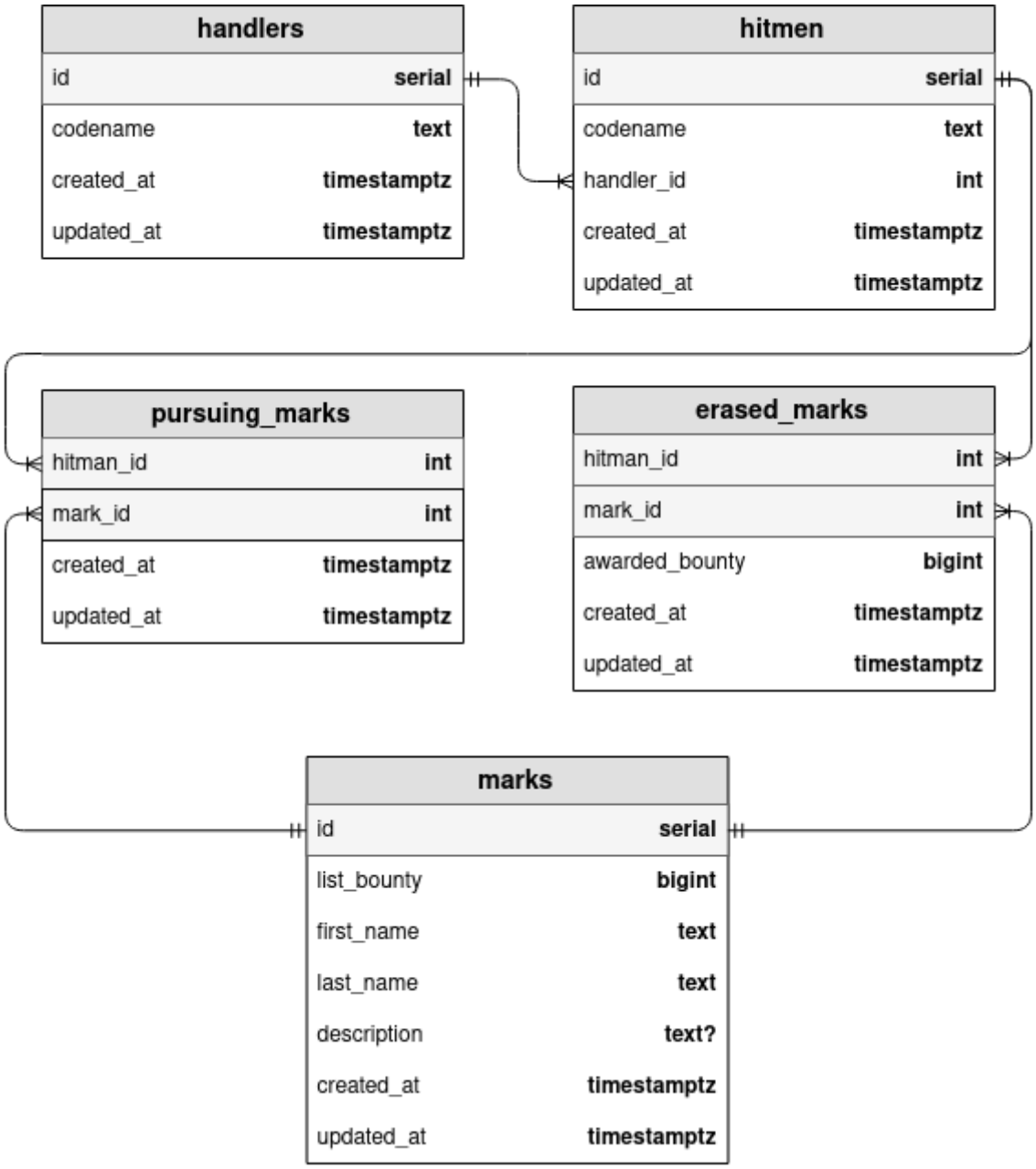
Вполне возможно, что вы как и я сагитировались на преимущества строгой типизации чтобы знать, чтобы писать приложения лучше.(Если нет, то на данный момент будем считать так)

Ваше приложение, допустим, требует возможность хранить данные постоянно. Вы можете использовать `postgresql-simple` для чего угодно, но есть небольшое смущение в том, что придется писать чистые SQL запросы, и надеятся что они работают в языке который хочет делать больше.

Но к счатсью, есть множество возможностей Haskell экосистемы для типобезопасных SQL запросов. Вы можете убедиться, что вы не забыли вылкючить столбцы в ваш вывод, или получить их в неправильном порядке. Вы можете даже переисползовать запросы легко, сочиная их напрямую в Haskell, затем создавая простые запросы для отправки на бэкенд дб. Всё с проверкой типов помогает вам, убеждаясь что вы не создали неправильный запрос и вызвали ошибку выполнения.

Что педставялет из себя проект в примере?

Мы создаем бэкенд для веб-сайта для профессионального киллера. Представим Fiverr или Upwork, только где платят за убийство. У каждого клиллера есть обработчик(один обработчик может обрабатывать несколько клиллеров), и киллер преследует "отметки". Как только работа выполнена, киллер отмечает цели как "удаленные". Мы смоделируем, как это будет в базе данных. Добавленная сущность `erased_marks` вовсе не удаляет `pursuing_marks`.



Для нашего занятия, мы используем Postgres как бэкенд нашей бд. В тоже время библиотеки которые мы рассматриваем(к примеру Beam) доволно агностични относительно самой бд, и может использоватьсья для любой базыданных, другие (как Opaleye и Squeal) работают только с Postgres.

Чтобы обслуживать данные нашего бэкенда, нам нужны запросы. Уточним, это запросы изменяются от простых до запросов которые содеражт объединения, подзапросы, и агрегаторы.

- Получить всех киллеров
- Получить всех киллеров которые предследуют целей(то есть имет неудаленные цели)

- Получить все цели которые были уничтожены за данное время.
- Получить все отметки которые были уничтожены за время определенным киллером.
- Получить общее вознаграждение для всех клиллеров
- Получить общее вознаграждение для опредленного киллера
- Получить для всех киллеров последнее убийство
- Получить последнее убийство для определенного клиллера
- Получить цели, которые имеют только одного преследователя.
- Получить все "возможные отметки"(то есть отметки которые киллер удаляет без дополнительного преследования цели)

Мы так же хотим написать обновления и вставки каждый библиотеки чтобы посмотреть как они обрабатывают их. Это должно напрячь все возможности запросов каждой библиотеки чтобы найти грубые точки.

Ну чтож со всем этим прыгаем в сравнение.

Beam

Beam - это попытка решить проблемы типо-безопасности SQL совместим с абсолютным игнорированием бэкенда. Способ с которым это получается решается добавлением типа параметра в каждый запрос для бэкенда и имеет множество типов классов что определения функциональности. К сожалению, он устарел, очень быстро, особенно когда вам нужно использовать типы классов с именами вроде `HasSqlEqualityCheck backend` - `Int64` and `BeamSqlT071Backend backend`, так как бог запрещает вам использовать `BIGINT`.(А вам нужно обе в одном запросе, между прочим)

Это можно обойти простым игнорированем бекенда целиком и указать определенный бэкенд в каждом запросе, но даже тогда тип вашего запроса будет кучей непостижимых `QExpr`, `QAgg`, и что там еще есть из параметров для запроса определеня объема.

Когда вы прошли это всё, собрка и создание запросов в `Beam` довольно приятны, подзапросы могут быть переиспользованны достаточно легко используя сущности `Beam` монад для этих запросов, объединений в одну строку. Легко определить запросы которые возвращают к этому еще и кортоежи, без надобности определения новых типов. Это потому что типы `Beam`

противны. Вам нужно обхдить все эти псевдонимы и семьи умных типов но в тоже время... к чему Intepeживания, когда другие библиотеки делают всё тоже самое без костылей?

Еще одни красный флаг - нет простого путит получить определенных строк, поэтому нужно явно собирать и группировать все толбцы, что вам нужны. `Beam` очень похож на не правиольный тип `SUM`, по крайней мере в Postgres, он не правильно меняет типа колонок. Для примера, в Postgres обход по `BIGINIT` колонке выдаст резульат `NUMERIC`, но в стране `Beam` если у вашего типа данных атрибут `Int64`, то он всё равно постарается выдать вам `Int64`, что становится ошибкой выполнения.

В любом случае я не могу рекомендовать `Beam`, он слишком привередлив.

```
latestHits :: Q Postgres HitmanDB s
  ( HitmanT (QExpr Postgres s)
  , MarkT (QExpr Postgres s)
  )
latestHits = do
  hitman <- allHitmen
  mark <- allMarks

  (minHID, minCreated, minMarkID) <- minID
  (latestHID, latestCreated) <- latest

  guard_ (minHID ==. latestHID)
  guard_ (just_ minCreated ==. latestCreated)
  guard_ (minHID ==. HitmanID (_hitmanID hitman))
  guard_ (minMarkID ==. just_ (_markID mark))

  pure (hitman, mark)

  where minID = aggregate_ (\em -> ( group_ (_erasedMarkHitman em)
                                   , group_ (_erasedMarkCreatedAt em)
                                   , min_ (getMarkID $ _erasedMarkMark em)
                                   ))
                           allErasedMarks

        latest = aggregate_ (\em -> ( group_ (_erasedMarkHitman em)
                                    , max_ (_erasedMarkCreatedAt em)
                                    ))
```

```
                          allErasedMarks


        getMarkID (MarkID id) = id
```

Opaleye

Opaleye это SQL DSL разработанная для Postgres. Из коробки, это решение «просто работает» без всяких настроек или полного игнорирования частей ядра библиотеки.

Написание запросов работает. Создание запросов работает. Типы (относительно) простые и с ними легко работать.

Главное отличие между Opaleye и другими библиотеками это способ определения схемы таблицы. Все определения схем выполняются на уровне определений, таким образом, чаще всего не зависят от самого типа домена. Это связано со способом настройки(используется product-profunctros), вы можете легко абстрагировать общие столбцы. На пример: я использовал это чтобы абстрагировать определения `created_at` и `updated_at` временные отметки. Дальше, Opaleye, показывает отличия между временем записи и временем чтения данных, ну что ж, это просто, скажем, для определенных колонок нельзя писать с помощь `insert/updates`(как было сказано выше с отметками времени).

Так как прошлые версии Opaleye имеет проблемы с правильным типом агрегатора к примеру `sums`, что касается Opaleye версии 0.6.7006.1, библиотека имеет улучшеную функцию для обработки. Вдобавок, теперь возможно использовать библиотеку целиком с помощь интерфейса монады вместо ссылок, избегая перегрузки познания, который в прошлом был необходим. Одно из препятствий которое необходимо будет изучить это `product-profunctors` используются везде. Однако можно легко обойтись без глубокого знания этой темы. Нужно просто добавить `p2` и `p3` везде где вам говорит документация.

В конце концов `Opaleye` просто работает, и это мой личный совет. В ней есть немного абстрактная кривая изучения но то дает вам общие возможности и комбинирование частей вашей запросов, это моё личный фаворит среди DB библиотек которые мы рассматриваем.

```
latestHits :: Select (HitmanF, MarkF)
latestHits = do
```

```
  (hID, created, mID) <- byDate
  (maxHID, maxCreated) <- maxDates
  h <- selectTable hitmenNoMeta
  m <- selectTable marksNoMeta

  viaLateral restrict $ hID .== maxHID .&& created .== maxCreated
  viaLateral restrict $ hID .== hitmanID h
  viaLateral restrict $ mID .== markID m

  pure (h, m)

  where byDate = aggregate (p3 (groupBy, groupBy, min)) $ do
          ( ErasedMark { erasedMarkHitmanID = hitmanID
                       , erasedMarkMarkID = markID
                       }
            , (createdAt, _)
            ) <- selectTable erasedMarkTable
          pure (hitmanID, createdAt, markID)
        maxDates = aggregate (p2 (groupBy, max)) $ do
          ( ErasedMark { erasedMarkHitmanID = hitmanID }
            , (createdAt, _)
            ) <- selectTable erasedMarkTable
          pure (hitmanID, createdAt)
```

Squeal

Squeal странный ребенок, менее удобный DSL, он предполагает глубокое встраивание SQL в самом Haskell. Поэтому он гораздо ближе к написанию реального SQL запроса, и не пытается обастрагироваться от этого, чтобы ваши SQL ключевые слова были в нуном месте в вашем запросе.

Эта жестокость делает использование Squeal болезненным, так как используется соблюдение типов, например. вы делаете `WHERE` после таблицы которую вы объединяете из принесенного в рамках. Так как Squeal использует чистую комбинаторноние применение, вместо стрелочной или монадной у других библиотек, использование становится упражнением жонглирования вложенных скобок и постоянного прыгания между

различными уровнями вложнеий. Честно говоря, ощущается как куча.

Squeal так же использует `OverLoadedLabels` для выбора колонок и таблиц, и идет даже дальше чем всё что есть в этом списке, он просит не только ввести вашу колонку, но так же просит отслеживать какое название вы использовали для каждой колонки. Какая, восхитительная, но так же очень раздражает когда создание подзапроса в другом и в этом случае необходимо явно перевыбрать результаты подзапроса используя тоже самое имя.

Эта настройчивость в названии столбцов ведет к множеству проблем в том числе. Способ которым вы возвращаете объекты ваших типов доменов это явное имя столбца запроса такого же как свойство вашего типа данных при использовании SQL запроса `AS`. Нет возможности просто определить ответственность один раз и забыть, что значит, что даже если вы просто выбираете все сущности из одной таблицы, вам нужно явно переназвать все колонки. Красота! Хотите получить для конкретного случая кортеж данных из быстрого запроса? Извините, нельзя так, кортеж не имеет называных полей, как вы можете назвать вашу колонку правильно? На деле, каждый раз когда вы хотите вернуть данные из новой формы, вам нужно определить полностью новый тип данных для этогог и перенаследовать специальные типы классов Squeal.

Пока работал с Squeal, чувствовал себя будто я не останавливался споткаться. Тип запросов Squeal имеет тип параметров для обоих запросов входящих\исходящих, но они не похожи на возможность передать их как параметры позапроса? Поэтому вам нужно закончить просто копировать кодов запросов. Иногда использовать подзапросы просто... которые вызывают ошибку выполнения непонятно почему, даже несмотря на проверку типов. Я надеюсь вы не когда не ошибетесь в названии колонки, или Squeal кинет вам в море непостижимых ошибок типов.

Ктоме всего, Squeal успешен в качестве инструмента встраивания SQL в Haskell, но проиграл в возможности описать общие SQL шаблоны без серьезных последствий. Не рекомендую.

```
latestHits :: Query_ Schema () HitInfo
latestHits = select_
  (#minid ! #hitman_id `as` #hiHitmanID :* #minid ! #mark_id `as` #hiMarkID)
  ( from ((subquery ((select_
           ( #em ! #hitman_id
          :* #em ! #created_at
```

```
        :* (fromNull (literal @Int32 (-1)) (min_ (All (#em ! #mark_id)))) `as` #mark_id
         )
         ( from (table (#erased_marks `as` #em))
           & groupBy (#em ! #hitman_id :* #em ! #created_at ))) `as` #minid ))
    & innerJoin (subquery ((select_
                 ( #em ! #hitman_id
                :* (max_ (All (#em ! #created_at))) `as` #created_at
                 )
                 ( from (table (#erased_marks `as` #em))
                   & groupBy (#em ! #hitman_id) )) `as` #latest))
        (#minid ! #created_at .== #latest ! #created_at)) )
```

Persistent + Esqueleto

Persistent - это легкий уровень для произведения простых CRUD операций. Esqueleto - SQL DSL над Persistent, добавлюящий возможность делать объединения и более сложные запросы.

Много говорить про Persistent не будем, так как библиотека просто предоставляет слой, давайте говорить о Esqueleto. Esqueleto значит быть очень легким языком запроса, в тоже время предоставлять достаточно мощности выполнять ожидаемые вещи от всяких хорошо написанных Haskell библиотек, как некоторые типы безопасности и немного композиционности.

Для меня, однако, я нашел что этот фокус на простых интерфейсах запросов сделал библиотеку такой, что она требует столько же ментальной энергии, сколько написание простого sql заапроса, если не больше. Выглядит это как пол пути применения, где есть некоторая проверка запросов, что вы пишите которая имеет смысл. но библиотеки все еще требует ответственность писать синтаксически верно и правильно сформированные запросы.

На пример: ваш запрос будет щастливо компилироваться без предупреждений но генерировать синтаксически неверные запросы SQL во вермя исполнения если вы забыли `ON` в вашем объединении. Или счастливо падать во время выполнения когда вы пытаетесь выбрать смесь колонок агрегатных и не агрегатных колонок. Запрос DSL сам по себе проктически 1:1 транслируется в чистый SQL включая множество возможных путей

злоупотребения ими. Поэтому я надеюсь что вы уже знакомы с SQL и его кварками, так как Esqueleto делает всего пару попыток скрыть SQL бородавок от вас.

На вершине этого, Esqueleto далеко позади по возможстям поддержки типичной RDBMS функциональности. Заметное отсутствие это можетсов возможностей объединять подзапросы, и вам необходимо писать все ваши запросы используя только один `SELECT` и улучшать условия объединения на существующих таблицах. У меня получилось реализовать все запросы по киллерам, но это потребовало серьезное управление запросами, чтобы все они заработали.

Вывод: даже не смотря на то что у меня получилось реализовать проект на Esqueleto, я чувствовал как будто я не получаю достаточно от простого написания SQL запросов. В других способах выглядело слишком ограниченно, из-за библиотеки каким-то образом открывает базовый набор функций. Вы можете даже познать простоту и гибкость написания SQL, или строгую безопасность типа и сложную компонуемость как в Opaleye. Esqueleto ощущается как неполучившаяся попытка быть маленьким со всех сторон.

```
latestHits :: MonadIO m => SqlPersistT m [(Entity Hitman, Maybe (Entity Mark))]
latestHits = select $
  from $ \(hitman `LeftOuterJoin` emark1
                  `LeftOuterJoin` emark2
                  `LeftOuterJoin` mark) -> do
    on (emark1 ?. ErasedMarkHitmanId ==. emark2 ?. ErasedMarkHitmanId &&.
        emark1 ?. ErasedMarkCreatedAt <. emark2 ?. ErasedMarkCreatedAt &&.
        emark1 ?. ErasedMarkMarkId >. emark2 ?. ErasedMarkMarkId)
    on (emark1 ?. ErasedMarkHitmanId ==. just (hitman ^. HitmanId))
    on (emark1 ?. ErasedMarkMarkId ==. mark ?. MarkId)
    where_ (isNothing $ emark2 ?. ErasedMarkCreatedAt)
    where_ (isNothing $ emark2 ?. ErasedMarkMarkId)
    pure (hitman, mark)
```

Revision #5
Created 2022-01-27 14:10:50 UTC by gasick
Updated 2023-04-16 19:38:12 UTC by gasick