

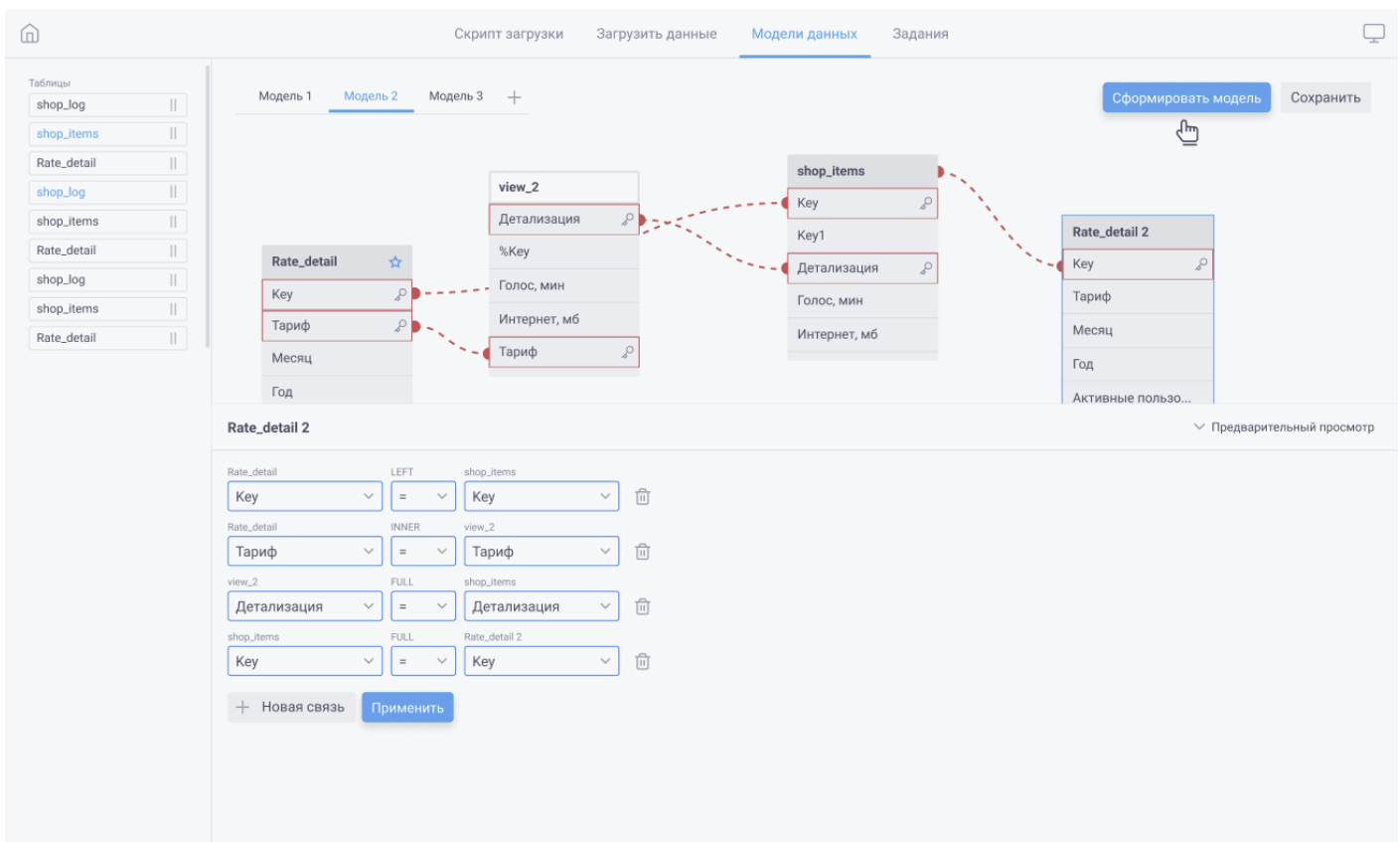
# Модуль подготовки модели данных с помощью ИИ (Фронт)

## Цель

Автоматическое формирование связей между загруженными таблицами в модели данных: определение ключевых полей и типов соединений между ними в зависимости от их данных.

## Концепт для интерфейса

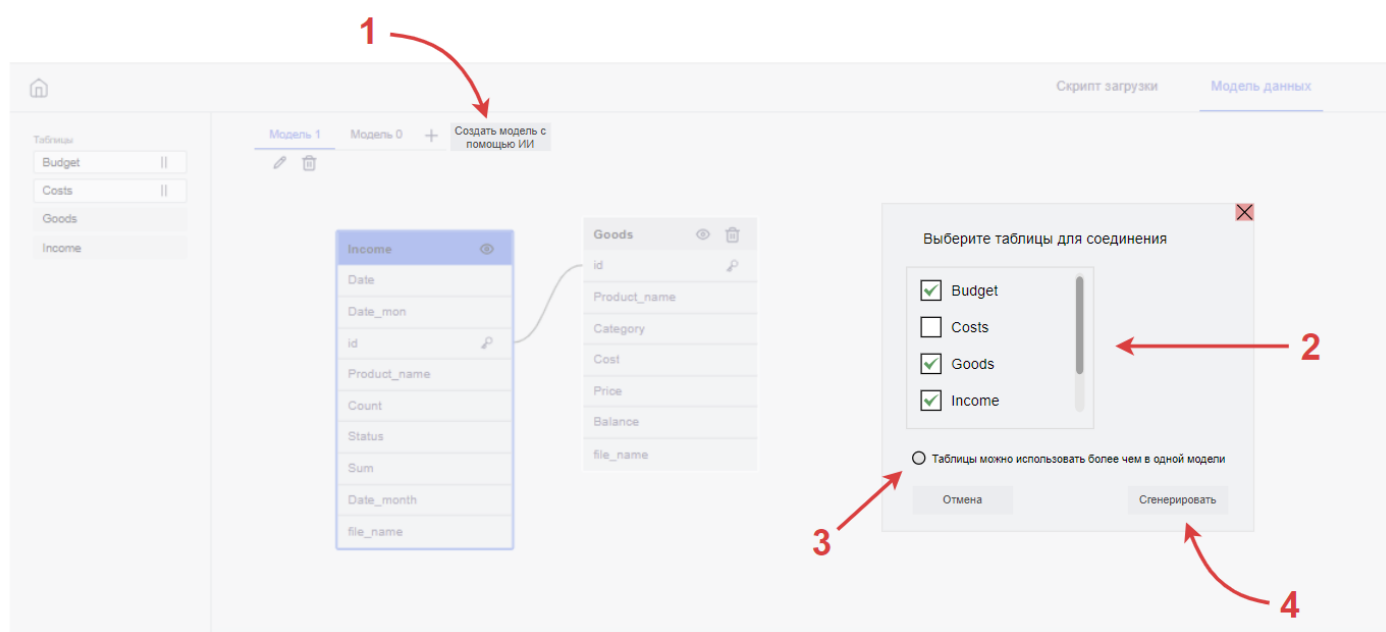
Рисунок первичного концепта, текущий см. в разделе **"Задание на разработку для фронта"**

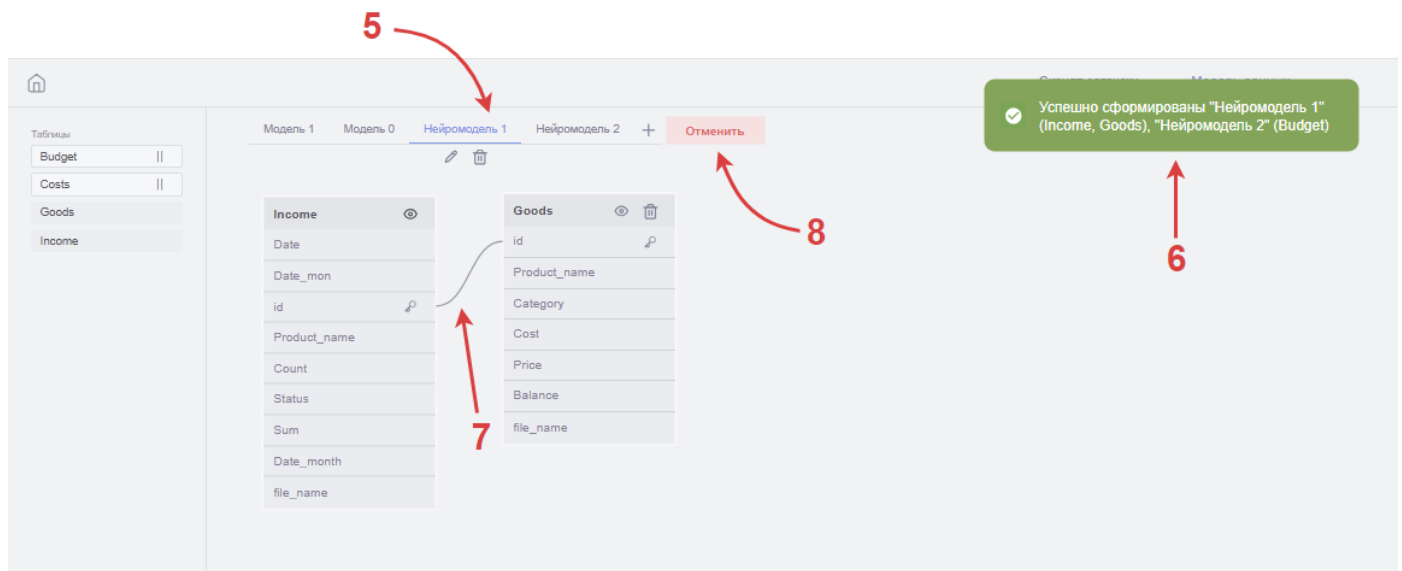


В диспетчере данных в разделе Модель данных необходимо добавить управляющую кнопку "Создать модель данных с помощью ИИ". При нажатии на кнопку должно открываться модальное окно со списком таблиц для соединения и кнопкой, разрешающей использовать таблицы более чем в одной модели. Выбор таблиц можно отменить или нажать на кнопку "Сгенерировать", после чего в модели данных появятся группы, созданные ИИ. Об успешной работе нейросети должно сообщаться системным сообщением, в котором необходимо перечислить созданные модели.

После создания таких моделей кнопка "Создать модель данных с помощью ИИ" трансформируется в кнопку "Отмена", которая позволяет одним нажатием вернуть модель данных в исходное состояние. Чтобы сохранить результат работы нейросети необходимо так же как и в случае с ручным проектированием нажать на кнопку "Сохранить" в правом верхнем углу окна. После нажатия кнопки "Отмена" или "Сохранить" возвращается возможность создать ещё одну модель с помощью ИИ.

## Задание на разработку для фронта





Кнопка "**Создать модель с помощью ИИ**" **(1)** – расположена рядом с кнопкой добавления модели данных, перемещается вместе с ней при добавлении/удалении моделей. При нажатии на кнопку открывается модальное окно с выбором таблиц для соединения.

## Окно выбора таблиц

В данном окне должен быть список с названиями всех таблиц **(2)**, полученных из скрипта загрузки. Рядом с каждым названием необходимо установить флажок для возможности множественного выбора. Если названия таблиц не помещаются в видимое поле списка, то по правой его границе должен размещаться ползунок для скrolла по списку.

Под списком должна быть кнопка-переключатель "Таблицы можно использовать более чем в одной модели" **(3)**. Если кнопка неактивна, то каждая таблица сможет появиться только в одной модели – в наиболее подходящей по логике применения. Если кнопка активна, то число повторов таблицы в разных моделях неограниченно, но в рамках одной модели таблица может появиться лишь однократно (нейросеть не создаёт копии таблиц). *По умолчанию* кнопка неактивна.

При нажатии на кнопку "Отмена" внизу окна выбора таблиц окно должно закрываться, все введённые изменения отменяются, подтверждать отмену не требуется. При нажатии на кнопку "Сгенерировать" **(4)** на бэкенд отправляются следующие данные из скрипта загрузки:

- Названия таблиц вида Table
- Названия столбцов с принадлежностью таблице вида Table.Column
- Типы данных для каждого из столбцов вида Int32, String, Decimal (2)
- Сами значения из столбца, если объём невелик, или подвыборка из столбца (н-р, как результат сэмплирования), если данных достаточно много. Если строк в таблице больше 1 000 000, то необходимо ограничить выборку до этого значения с помощью Limit или Sample (второй вариант предпочтительнее ввиду необходимости передавать статистические данные, которые могут быть неравномерно распределены по таблице).

- Информация о полях в сортировке и первичных ключах вида Order By (список полей со всех таблиц) и Primary Key (список полей со всех таблиц)
- Разрешено ли использовать одну таблицу более чем в одной модели (булево) – результат нажатия на кнопку (3)

После отправки данных окно выбора таблиц закрывается, а в модели данных происходит ряд изменений.

## Нейромоделели

Нейросеть возвращает набор связанных таблиц, разбитых по группам. Каждая такая группа станет отдельной нейромоделью **(5)**, которая добавится в список существующих моделей.

Текущее рабочее название – "Нейромодель n", где n – порядковый номер модели, созданной нейросетью, без учёта других моделей (созданных вручную).

(Дополнительно) Возможное рабочее название – должно предлагаться нейросетью в зависимости от контекста использования данной группы таблиц.

Результат успешного создания нейромоделей должен сопровождаться системным сообщением **(6)** – "Успешно сформированы {Список моделей} ({Список таблиц в каждой модели})"

Внутри каждой группы для каждой пары таблиц нейросеть передаёт ключевые поля, по которым строится связь, и тип JOIN между ними. Эта информация должна использоваться для формирования связей между таблицами на фронте и отрисовки соединений по ключевым полям **(7)**. Поскольку модели формируются автоматически, необходимо определять положение таблиц в модели данных. Таблицы должны добавляться в том порядке, в котором они возвращаются из бэкенда. Первая таблица возникает в том месте, где она возникла бы, если бы была добавлена вручную пользователем нажатием на её название в списке таблиц слева от модели данных. Последующие таблицы должны возникать правее предыдущих на расстоянии  $1,5 \cdot x \cdot (n-1)$  от левой границы первой таблицы, где x – ширина таблицы,

n – порядковый номер таблицы.

Созданные нейросетью модели ведут себя аналогично обычным: в них можно добавлять и удалять таблицы и связи. Сохраняется результат стандартной кнопкой "Сохранить" в правом верхнем углу. Чтобы удалить все созданные нейросетью модели можно воспользоваться кнопкой "Отменить" – при нажатии на неё должны удаляться все нейромоделели, даже если они были любым образом изменены и переименованы. Не требует подтверждения.

## Работа на кластере (на будущее)

Поскольку в КХ соединение JOIN для больших объёмов данных работает плохо, необходимо минимизировать их применение, соединив наибольшее возможное число таблиц с фактами в одну (например, с помощью UNION ALL).

Порядок действий:

- Нейросеть должна определить, какие из имеющихся таблиц можно отнести к справочникам (к которым будут подтягиваться данные из одной общей таблицы). Справочник можно идентифицировать по наличию пар "ключ-значение", согласно которым возможно формирование связей с таблицей фактов. Остальные таблицы относятся к фактическим.
- Фактические таблицы объединяются скриптом в одну таблицу:  
`SELECT Столбцы FROM Таблица1 UNION ALL Столбцы FROM Таблица2 UNION ALL Столбцы FROM Таблица3...`  
Скрипт не запускается, а используется для создания таблицы предварительного просмотра, демонстрируемой в отдельном модальном окне.
- Если пользователя устраивает структура, он нажимает кнопку "Продолжить". Иначе – "перегенерировать" (модель запускает ещё одну попытку классификации таблиц на справочники и факты).
- После подтверждения срабатывает сгенерированный скрипт как в обычной ситуации.

---

Revision #20

Created 12 November 2024 07:57:37 by Артём

Updated 21 January 2025 05:45:06 by Артём