

Как X5 Tech дает аналитику по данным

Без привлечения аналитиков,
специалистов и иных посредников

Ермаченков Владимир

Руководитель направления
продвинутой аналитики в X5 Tech



Работаю в компании X5 Tech
более 7 лет. Всю свою «сознательную
жизнь»
работал с данными.

Мой путь от аналитика
мастер-данных до руководителя
направления продвинутой аналитики.

Создаю ИИ ассистента на данных для
ускорения принятия решений.



Ермаченков Владимир

Руководитель направления
продвинутой аналитики в X5 Tech

X5

ведущая продуктовая розничная компания.



324

IT-системы
в эксплуатации

400

Активных проектов

426 тыс.

Сотрудников

30+

Цифровых продуктов

10,0 пб

Объем хранения кластера
больших данных

О чем мы сегодня поговорим?

- Проблематика
 - Что мы делаем?
 - Что такое генерация текста в SQL? Зачем?
 - С какими проблемами мы сталкиваемся?
 - Как мы работаем над повышением качества?
 - Как мы отслеживаем качество?
-
- Новая роль каталога данных
 - Синергия инженера данных и разработчика нейронных сетей
 - Что нужно пользователю?
 - Как довериться агенту и все отпустить?
 - Что мы делаем дальше?

Проблематика

Слишком много информации

Разработка – Время

Много источников информации

Отсутствие гибкости

Много пользователей,
и они разные

Низкая скорость принятия
решений



Что мы делаем?

От вопроса до цифры за 20 секунд

T2M
~ 4 дня

Сейчас и там, где это необходимо

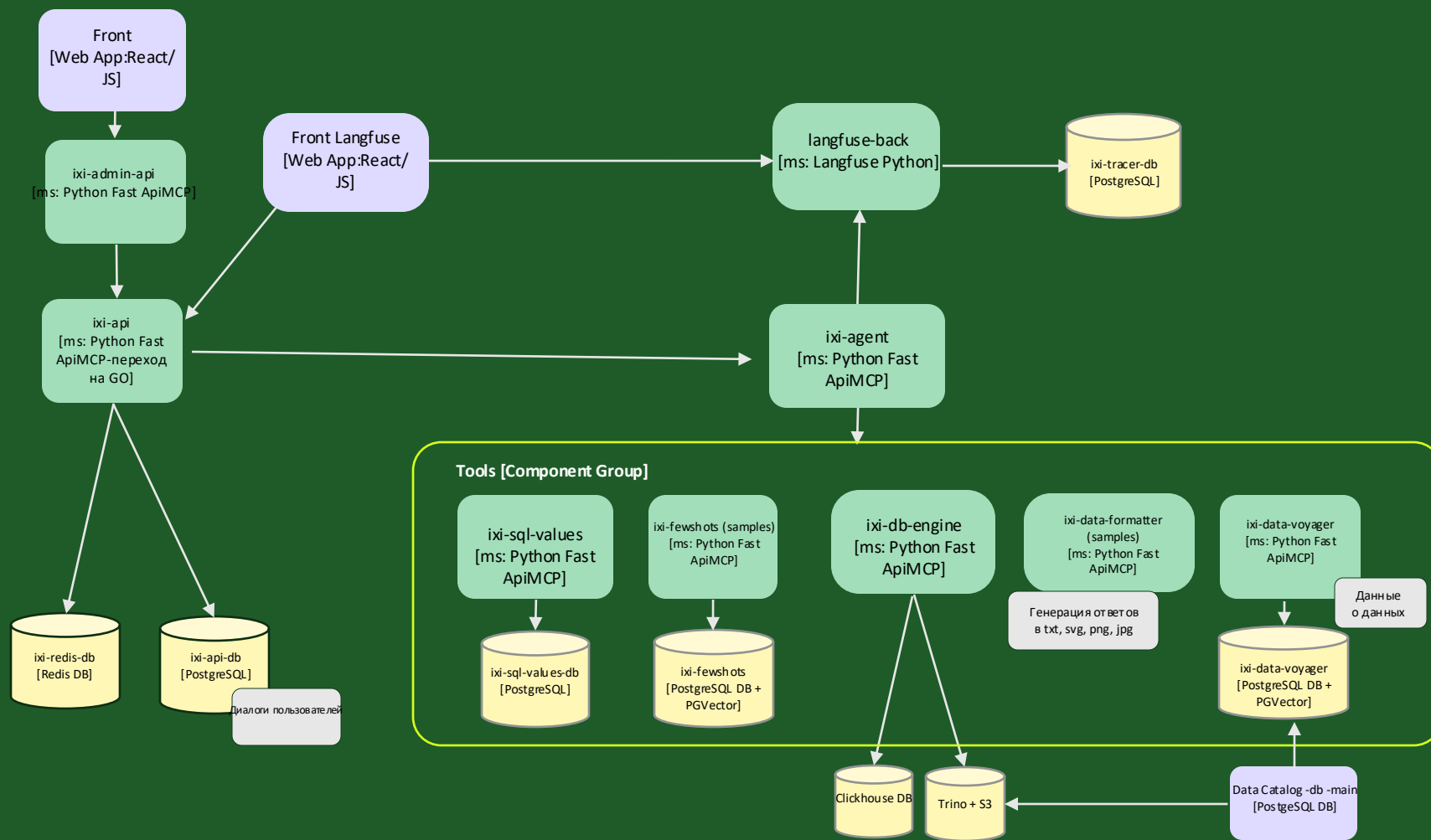
Переводим намерения в факт

Доступ к данным сотрудникам магазинов,
кластеров, дивизионов

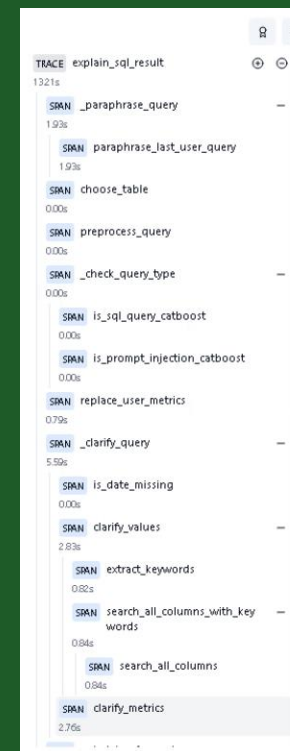
DAU >
300



Архитектура



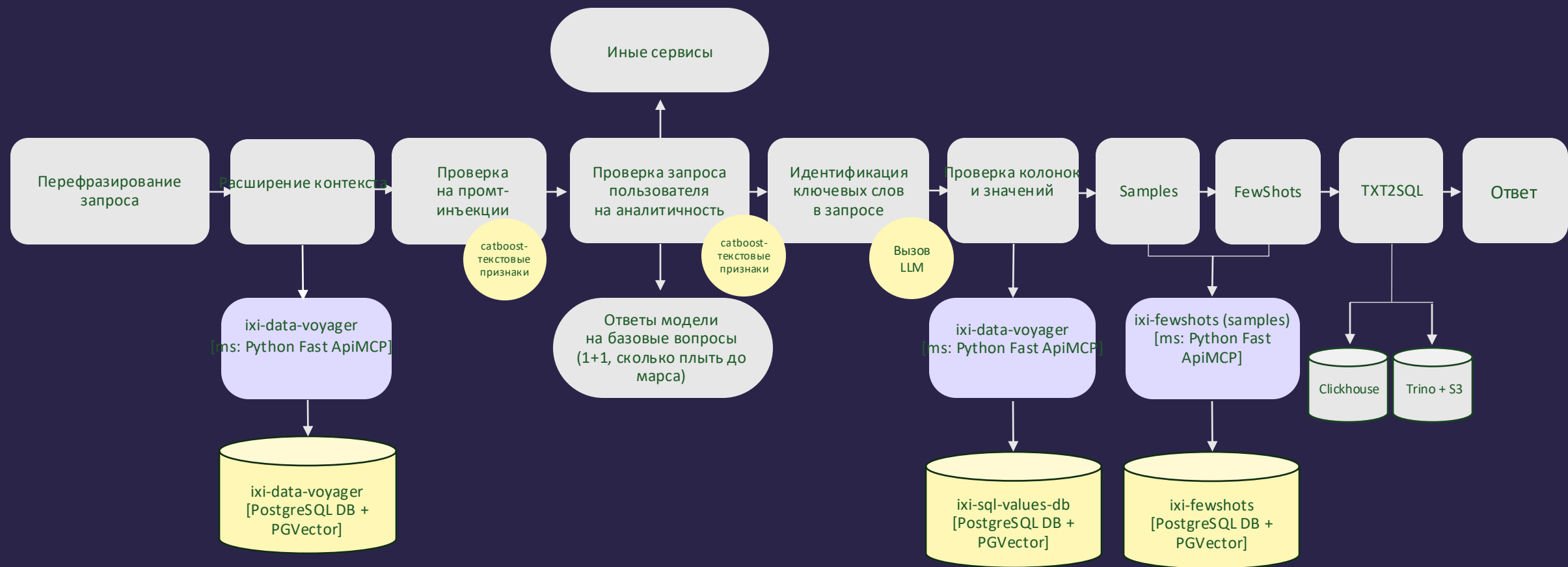
Model GateWay



Qwen3-32B
Qwen2.5-7B
Qwen3Flash Coder- 7b
Think2SQL 7 B
gptoss - 120b



LLM - Pipeline



Что такое TXT2SQL? Зачем?

Топ 3 города по продажам молока.... 15:10 ✓



Проверка запроса пользователя на аналитичность



Перефразирование запроса, очистка



Определение релевантной витрины
под запрос пользователя



Генерация текста в SQL



Выполнение SQL



Интерпретация результата и передача в необходимом пользователю
формате (текст/ таблица/ график)

Что такое TXT2SQL? Зачем?

потери по чижик 12:45

показатели по магазинам 12:45

что по продаже рыбы в пятере 12:45

в каком регионе западают продажи молока 12:45

lfl рто средний чек по переку за прошлую неделю 12:45



Запрос пользователя

DDL схемы

SQL
Generation Hints

SQL Fewshots

SQL Templates

```
Database: database_name
Table: table1

[(day_dt: DATE, Description: Day of the slice. Represents the calendar date for
which the data record is valid, used as the primary time dimension in the dataset.,
Extended description: , Hints: \n\ SPECIAL FUNCTIONS:\n\ day_of_week: toDayOfWeek
(day_dt)\n\ previous_day: day_dt = yesterday()\n\ current_day_check: toDayOfWeek
(day_dt) = toDayOfWeek(today()) - 1\n\ year: toYear(day_dt)\n\ quarter: toQuarter
(day_dt)\n\ month: toMonth(day_dt)\n\ week: toISOWeek(day_dt)\n\ USAGE RULES:\n\
When comparing with dates in functions, always convert string dates to Date type
using toDate()\n\ Use 'day' instead of 'date' or 'DATE' in aliases\n, Synonyms:
['data', 'день', 'дата среза', 'вчера', 'за неделю', 'за месяц'], Anti_aliases: [],
Examples: ['2025-02-02', '2025-02-03', '2025-02-04', '2025-02-05', '2025-02-06',
'2025-02-07', '2025-02-08', '2025-02-09', '2025-02-10', '2025-02-11'],\n
(plan_code: STRING, Description: Plan version. Identifies which planning scenario (e.
g., baseline plan, motivation plan) the budget/forecast figures belong to. Valid
values are strictly limited to the example list., Extended description: , Hints:
\n\ RESTRICTIONS:\n\ DON'T use plan_code in WHERE filter\n\ plan_code is intended
only for CASE WHEN within SUM\n\ USAGE RULES:\n\ Use only in CASE WHEN within SUM
aggregations\n, Synonyms: ['план', 'версия плана', 'код плана', 'тип бюджета'],
Anti_aliases: [], Examples: ['PLAN', 'TPG', 'PLAN03'],\n
(financial_unit_format_nm: STRING, Description: Retail chain. Always choose this
column when you need to filter or group by 'Торговая сеть' ('TC') or 'Бизнес
единица' ('БЕ'). Notice that for this column, only values from the example list are
presented in the table., Extended description: , Hints: \n\ TRIGGER KEYWORDS: trading
network, торговая сеть, TC, БЕ, Бизнес единица\n\ ALIASES: БЕ, Бизнес единица,
торговая сеть, TC\n\ VALUE MAPPING:\n\ trading_networks: 'ТС Пятерочка', 'ТС
Перекресток', 'ТС Чижик', 'Байкал', 'Перекресток Впрок', 'БЕ Х5 Диджитал', 'БЕ Много
Лосося'\n\ FILTER RULES:\n\ Use filter only when user explicitly refers to trading
networks\n\ Do not apply filter for X5 queries - X5 represents entire company\n\ For
filtering by financial_unit_format_nm:\n\ - Use STRICT EQUALITY (=) with FULL COLUMN
VALUE\n\ Example: WHERE financial_unit_format_nm = 'ТС Пятерочка' (complete name)\n\
\ CRITICAL: Use complete column values, not abbreviations! 'TC5' + 'ТС
Пятерочка', Synonyms: ['торговая сеть', 'ТС', 'бизнес единица', 'БЕ', 'сеть',
```



15:36, 09.09.2025 ✓

продажи по молоку в разрезе городов. выведи топ 10 городов по продажам



15:36, 09.09.2025

РТО с НДС, руб по молоку в разрезе городов за последние 7 дней

Фильтры

- "Период" - "02 сентября 2025 по 08 сентября 2025"
- "Товарная категория" - "Молоко"

Результат:

НАСЕЛЕННЫЙ ПУНКТ	РТО с НДС, РУБ
Москва	1 000 000
Чикаго	999 999
Париж	888 888
Пекин	777 777
Нью-Йорк	555 555
Лондон	555 555
Воронеж	444 444
Милан	333 333
Рим	222 222
Лиссабон	111 111

Итог



Интерпретация



Объем контекста



Галлюцинации



Пользователь осознает потребность не сразу

Архитектура данных



Качество

30%
качество
ответов

Что такое TXT2SQL? Зачем?

Выведи топ 10 тт 16:25 ✓



```
store_id:
{ 'type': 'STRING',
'description_en': 'Store identifier. Human-readable short code for a store,
unique within the retail network.',
'description_rus': 'Идентификатор магазина. Краткий
человекочитаемый код магазина, уникальный в пределах торговой
сети.', 'examples': ['HB6P', 'L157', '37FF', '328S', '31FC'], 'translation': None,
'synonyms': ['код магазина', 'ID магазина', 'Сap', 'SAP', 'торговая точка',
'точка', 'тт'],
'anti_aliases': ['store_rk', 'financial_unit_id'], },
```

```
'store_nm':
{ 'type': 'STRING',
'description_en': 'Store name. Concatenation of internal store number and
retail chain brand plus optional work status of the store.',
'description_rus': 'Название магазина. Объединение внутреннего
номера магазина и бренда торговой сети с возможным статусом
работы магазина.',
'examples': [ '18083-Пятерочка'],
'translation': None,
'synonyms': ['магазин', 'название магазина', 'номер', 'маг', 'ТТ'],
'anti_aliases': ['financial_unit_nm'],
```

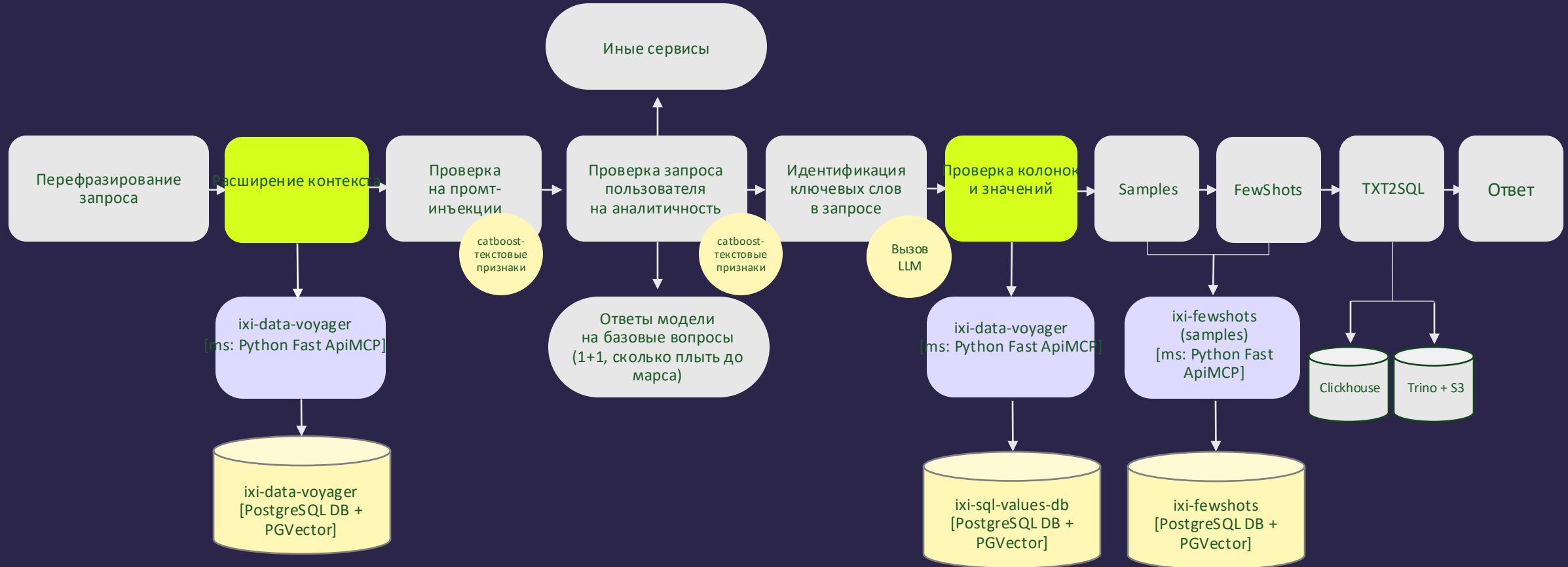
```
select store_id
from table
where ....
limit 10
```

ИЛИ

```
select store_nm
from table
where ....
limit 10
```

Как мы работаем над повышением качества?

Column and Value Retriever



/ multilingual-e5-large (vector search)

/ Полнотекстовый поиск

/ RFF Score To Be

Как мы работаем над повышением качества?

Column and Value Retriever

Сколько автомобилей
было продано в **регионе
Центр**
(за последние
7 дней: с 15 сентября
2025 по 21 сентября
2025)



```
{  
  
  keywords: {  
    filters: [  
      0: "автомобиль"  
      1: "регион Центр"  
    ]  
    metrics: [  
      0: "продажи"  
    ]  
    groupings: [  
    ]  
  }  
}
```



```
macroregion_nm:  
  
  {  
    score: 0.033  
    value: "ЦО Территория Центр"  
    column_name: "macroregion_nm"  
  }  
  
  { score: 0.031  
    value: "Макрорегион Центр Юг"  
    column_name: "macroregion_nm"  
  }
```

Как мы работаем над повышением качества?

Column and Value Retriever

Сколько автомобилей
было продано
в регионе Центр
(за последние
7 дней: с 15 сентября
2025 по 21 сентября
2025)



```
{  
  
  keywords: {  
    filters: [  
      0: "автомобиль"  
      1: "регион Центр"  
    ]  
    metrics: [  
      0: "продажи"  
    ]  
    groupings: [  
    ]  
  }  
}
```



```
{ cached: true results: {} }
```

Как мы работаем над повышением качества?

Column and Value Retriever

```
{  
  answer:  
  
  "### 📊 Результат по запросу продажи автомобилей в регионе Центр за последние 7 дней  
  
  ## 🔍 Фильтры - "День среза" - "с 15 сентября 2025 по 21 сентября 2025" –  
  "Наименование уровня иерархии" - "автомобиль" –  
  "Название макрорегиона" - "центр"  
  
  ## 📋 Результат: Данные не были найдены по заданным фильтрам. Возможно, стоит  
  пересмотреть критерии фильтрации, такие как период, название дивизиона или уровень  
  иерархии.»  
  
  full_explanation: "There is no explanation"  
}
```

- 40%

ниже
галлюцинаций

-50%

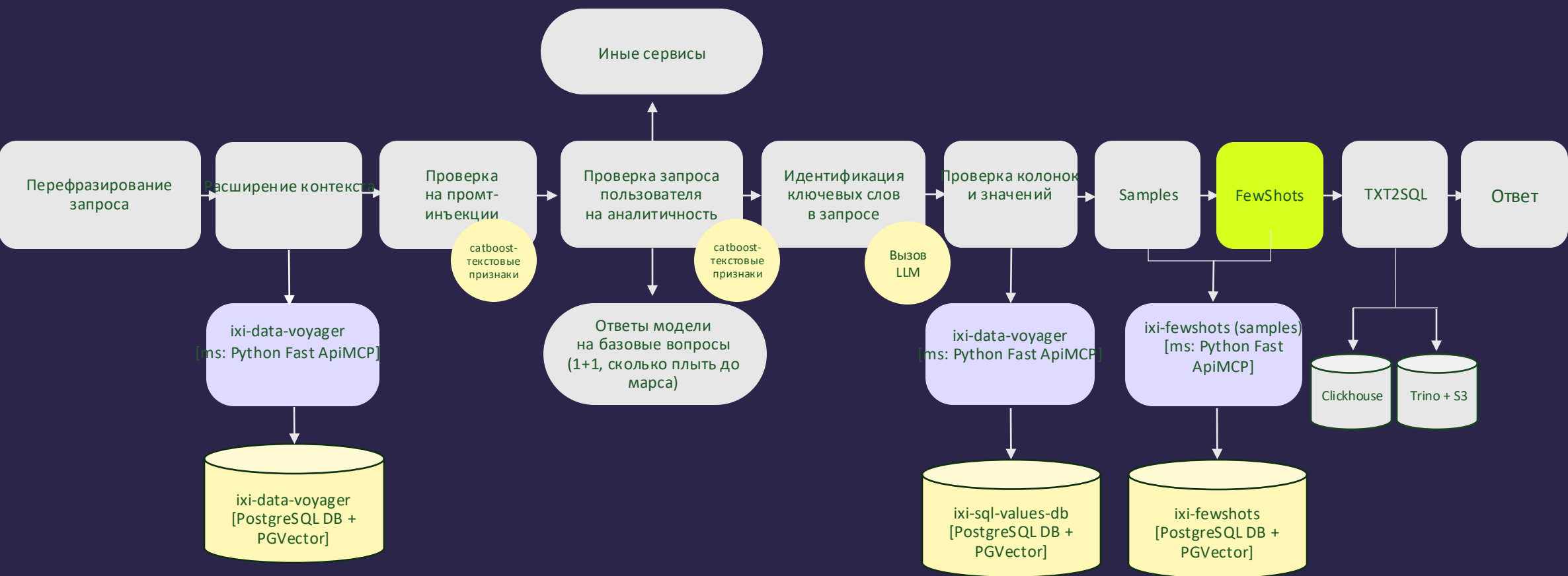
входного
контекста

+ - 10%

качеств
а

Как мы работаем над повышением качества?

Few-shot стратегия



Как мы работаем над повышением качества?

Few-shot стратегия



Пользователь

Пользователь: Сравни РТО по жвачке между макрорегионами по пятерке за предыдущую неделю



Пользователь

Пользователь: Сравни рто по **<категория товара>** между макрорегионами по **<торговая сеть>** за **<период>**

Как мы работаем над повышением качества?

Few-shot стратегия

/multilingual-e5-large/
/bm25

RFF Score:
 $\text{Score} = \sum (1 / (k + \text{rank}))$

{

Table Description
Metrics
SQL Generation hints
Specific Rules
DDL Statements
Few-shot Block Reference
Examples

}

Few-shot block reference:

The section

Examples: (below the ## DDL statements) contains few-shot examples – pairs(“Question: ...”, “SQL: ...”).

They illustrate style, aliasing, and safe patterns; treat them as reference material only.

Rules for using few-shot ## Examples:: 1. Enrichment, not replacement: Borrow patterns (aliasing, ILIKE, NULLIF, date math, etc.) but never copy concrete filters, columns, or time ranges unless the user explicitly requests the same value.

2. Preserve user intent: The final SQL must answer exactly what the user asked; do not silently add dimensions, metrics, or filters absent from the request.

3. Structure discipline: Keep the canonical order: SELECT ... FROM ... WHERE ... GROUP BY ... HAVING ... ORDER BY ... Use lowercase English aliases and quote identifiers only when required.

4. Composable decomposition first, then compression: • If a complex question can be expressed more clearly through CTEs, do so.....

....

+15%

```

SELECT store_id,
store_nm,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * sum(cy_d_lfl_turnover_vat_rub_amt) /
nullIf(sum(py_d_lfl_turnover_vat_rub_amt, 0.0) - 1)
    ELSE 100 * (1.0 * sum(lfl_turnover_vat_rub_amt) / nullIf(sum(py_wd_lfl_turnover_vat_rub_amt, 0.0) - 1)
    END, 2) AS lfl_rto_percent,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * sum(rtl_bxn_lfl_d_cnt) /
nullIf(sum(py_dd_lfl_rtl_bxn_cnt, 0.0) - 1)
    ELSE 100 * (1.0 * sum(rtl_bxn_lfl_cnt) / nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0) - 1)
    END, 2) AS lfl_traffic_percent,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * ((1.0 * sum(cy_d_lfl_turnover_vat_rub_amt) /
nullIf(sum(rtl_bxn_lfl_d_cnt, 0.0)) / nullIf((1.0 * sum(py_d_lfl_turnover_vat_rub_amt) /
nullIf(sum(py_wd_lfl_turnover_vat_rub_amt, 0.0)) / nullIf((1.0 *
sum(py_wd_lfl_turnover_vat_rub_amt) / nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0)) / nullIf((1.0 *
sum(py_wd_lfl_turnover_vat_rub_amt) / nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0)) / nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0)) - 1)
    END, 2) AS lfl_avg_check_percent,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1

```

...

```

OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * sum(cy_d_lfl_turnover_base_uom_amt) /
nullIf(sum(py_d_lfl_turnover_base_uom_amt, 0.0) - 1)
    ELSE 100 * (1.0 * sum(lfl_turnover_base_uom_amt) / nullIf(sum(py_wd_lfl_turnover_base_uom_amt, 0.0) - 1)
    END, 2) AS lfl_rto_pcs,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * (sum(cy_d_lfl_turnover_base_uom_amt) /
nullIf(sum(rtl_bxn_lfl_d_cnt, 0.0)) / (sum(py_d_lfl_turnover_base_uom_amt) / nullIf(sum(py_dd_lfl_rtl_bxn_cnt, 0.0)) - 1)
    ELSE 100 * (1.0 * (sum(lfl_turnover_base_uom_amt) / nullIf(sum(rtl_bxn_lfl_cnt, 0.0)) /
(sum(py_wd_lfl_turnover_base_uom_amt) / nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0)) - 1)
    END, 2) AS lfl_number_pcs_per_receipt,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * (sum(plu_unique_lfl_d_cnt) /
nullIf(sum(rtl_bxn_lfl_d_cnt, 0.0)) / (sum(py_dd_lfl_plu_unique_cnt) / nullIf(sum(py_dd_lfl_rtl_bxn_cnt, 0.0)) - 1)
    ELSE 100 * (1.0 * (sum(plu_unique_lfl_cnt) / nullIf(sum(rtl_bxn_lfl_cnt, 0.0)) / (sum(py_wd_lfl_plu_unique_cnt) /
nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0)) - 1)
    END, 2) AS lfl_number_of_plu_per_receipt,
ROUND(SUM(turnover_vat_rub_amt, 1) AS rto_with_vat_rub,
ROUND(100.0*sum(turnover_vat_rub_amt)/nullIf(sum(CASE
    WHEN plan_code='PLAN' THEN budget_sales_rub_amt
    ELSE 0
    END), 0.0), 1) AS rto_plan_fulfillment_percent,
ROUND(SUM(CASE
    WHEN plan_code='PLAN' THEN budget_sales_rub_amt
    ELSE 0
    END), 1) AS plan_rto_with_vat_rub,
ROUND(100.0* SUM(turnover_vat_rub_amt)/NULLIF(SUM(budget_motiv_plan_tmrv_rub_amt, 0.0), 1) AS
motiv_plan_fulfillment_percent,

```

...

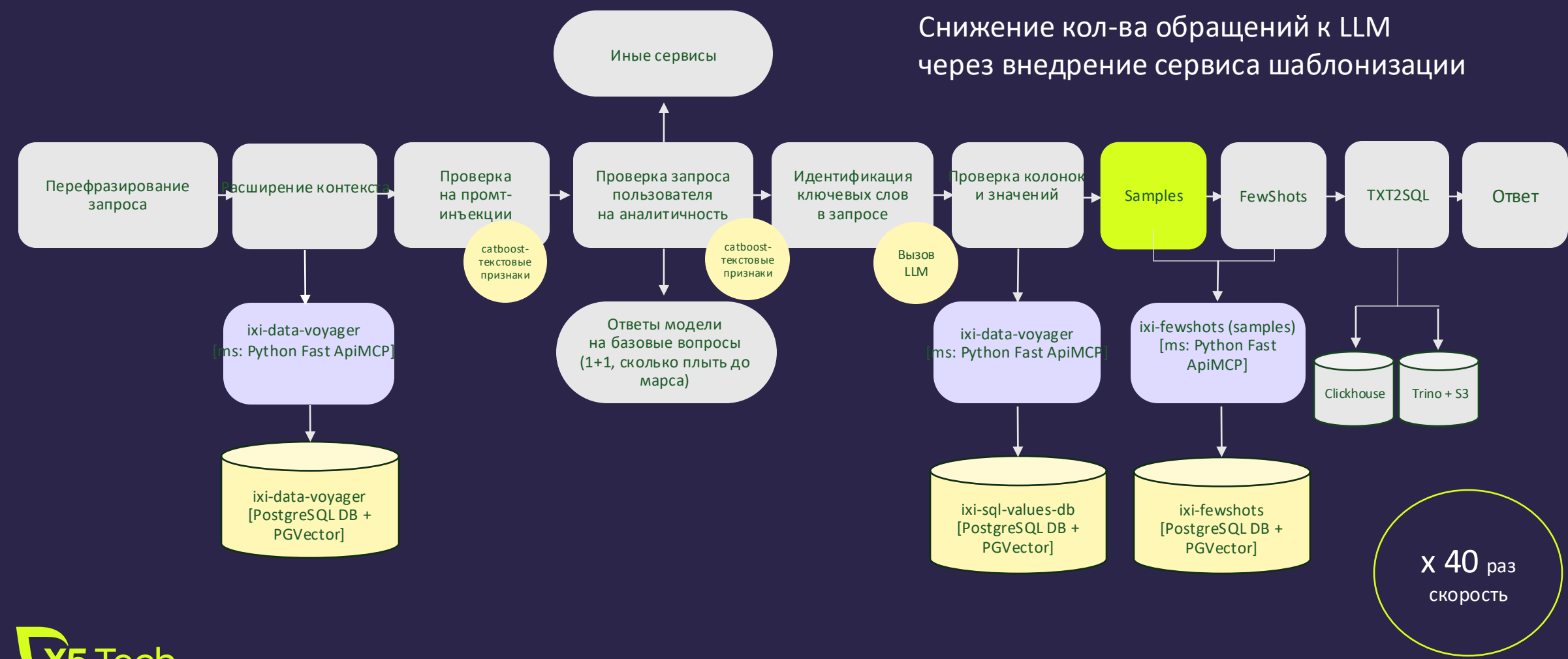
```

ROUND(SUM(turnover_vat_rub_amt) / NULLIF(SUM(rtl_bxn_cnt, 0), 1) AS avg_check_rub,
ROUND(100.0 * (sum(turnover_bw_no_vat_rub_amt) - sum(prime_cost_no_vat_rub_amt)) /
nullIf(sum(turnover_bw_no_vat_rub_amt, 0.0), 1) AS front_margin_percent,
ROUND(100.0*sum(bread_turnover_vat_rub_amt)/nullIf(sum(turnover_vat_rub_amt, 0.0), 1) AS
sp_bread_rto_with_vat_per,
ROUND(SUM(bread_turnover_base_uom_amt, 0) AS bread_turnover_base_uom_amt,
ROUND(100.0 * sum(lty_turnover_vat_rub_amt) / nullIf(sum(turnover_vat_rub_amt, 0.0), 1) AS
loyalty_card_rto_per,
ROUND(100.0 * sum(fpl_rtl_txn_cnt) / nullIf(sum(rtl_bxn_cnt, 0.0), 1) AS loyalty_card_txn_per,
ROUND(SUM(turnover_base_uom_amt, 1) AS turnover_base_uom_amt,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN sum(py_d_turnover_vat_rub_amt)
    ELSE sum(py_wd_turnover_vat_rub_amt)
    END, 1) AS py_rto_with_vat_rub,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * ((SUM(turnover_vat_rub_amt) -
SUM(py_d_turnover_vat_rub_amt) / nullIf(SUM(py_d_turnover_vat_rub_amt, 0.0)))
    ELSE 100 * (1.0 * ((SUM(turnover_vat_rub_amt) - SUM(py_wd_turnover_vat_rub_amt) /
nullIf(SUM(py_wd_turnover_vat_rub_amt, 0.0)))
    END, 1) AS yoy_rto_with_vat_per
FROM hldm_dtl_edw.kbi_operational_metric
WHERE day_dt = yesterday()
    AND management_cluster_nm ILIKE 'Князь Погром 78'
    AND financial_unit_format_nm ILIKE 'ТС Питание'
    AND store_id <> ''
    AND store_nm <> ''
GROUP BY store_id,
store_nm
ORDER BY store_id

```

Как мы работаем над повышением качества?

Шаблонизация SQL



```
SELECT store_id,
store_nm,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * sum(cy_d_lfl_turnover_vat_rub_amt) /
nullIf(sum(py_d_lfl_turnover_vat_rub_amt, 0.0) - 1)
    ELSE 100 * (1.0 * sum(lfl_turnover_vat_rub_amt) / nullIf(sum(py_wd_lfl_turnover_vat_rub_amt, 0.0) - 1)
    END, 2) AS lfl_rto_percent,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * sum(rtl_bxn_lfl_d_cnt) /
nullIf(sum(py_dd_lfl_rtl_bxn_cnt, 0.0) - 1)
    ELSE 100 * (1.0 * sum(rtl_bxn_lfl_cnt) / nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0) - 1)
    END, 2) AS lfl_traffic_percent,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * ((1.0 * sum(cy_d_lfl_turnover_vat_rub_amt) /
nullIf(sum(rtl_bxn_lfl_d_cnt, 0.0)) / nullIf((1.0 * sum(py_d_lfl_turnover_vat_rub_amt) /
nullIf(sum(py_dd_lfl_rtl_bxn_cnt, 0.0)), 0.0) - 1)
        ELSE 100 * ((1.0 * sum(lfl_turnover_vat_rub_amt) / nullIf(sum(rtl_bxn_lfl_cnt, 0.0)) / nullIf((1.0 *
sum(py_wd_lfl_turnover_vat_rub_amt) / nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0)), 0.0) - 1)
    END, 2) AS lfl_avg_check_percent,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
```

...

```
OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * sum(cy_d_lfl_turnover_base_uom_amt) /
nullIf(sum(py_d_lfl_turnover_base_uom_amt, 0.0) - 1)
    ELSE 100 * (1.0 * sum(lfl_turnover_base_uom_amt) / nullIf(sum(py_wd_lfl_turnover_base_uom_amt, 0.0) - 1)
    END, 2) AS lfl_rto_pcs,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * (sum(cy_d_lfl_turnover_base_uom_amt) /
nullIf(sum(rtl_bxn_lfl_d_cnt, 0.0)) / (sum(py_d_lfl_turnover_base_uom_amt) / nullIf(sum(py_dd_lfl_rtl_bxn_cnt, 0.0)) - 1)
        ELSE 100 * (1.0 * (sum(lfl_turnover_base_uom_amt) / nullIf(sum(rtl_bxn_lfl_cnt, 0.0)) /
(sum(py_wd_lfl_turnover_base_uom_amt) / nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0)) - 1)
    END, 2) AS lfl_number_pcs_per_receipt,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * (sum(plu_unique_lfl_d_cnt) /
nullIf(sum(rtl_bxn_lfl_d_cnt, 0.0)) / (sum(py_dd_lfl_plu_unique_cnt) / nullIf(sum(py_dd_lfl_rtl_bxn_cnt, 0.0)) - 1)
        ELSE 100 * (1.0 * (sum(plu_unique_lfl_cnt) / nullIf(sum(rtl_bxn_lfl_cnt, 0.0)) / (sum(py_wd_lfl_plu_unique_cnt) /
nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0)) - 1)
    END, 2) AS lfl_number_of_plu_per_receipt,
ROUND(SUM(turnover_vat_rub_amt, 1) AS rto_with_vat_rub,
ROUND(100.0*sum(turnover_vat_rub_amt)/nullIf(sum(CASE
    WHEN plan_code='PLAN' THEN budget_sales_rub_amt
    ELSE 0
    END), 0.0), 1) AS rto_plan_fulfillment_percent,
ROUND(SUM(CASE
    WHEN plan_code='PLAN' THEN budget_sales_rub_amt
    ELSE 0
    END), 1) AS plan_rto_with_vat_rub,
ROUND(100.0* SUM(turnover_vat_rub_amt)/NULLIF(SUM(budget_motiv_plan_tmrv_rub_amt, 0.0), 1) AS
motiv_plan_fulfillment_percent,
```

...

Генерация
3
секунды

...

```
ROUND(SUM(turnover_vat_rub_amt) / NULLIF(SUM(rtl_bxn_cnt, 0), 1) AS avg_check_rub,
ROUND(100.0 * (sum(turnover_bw_no_vat_rub_amt) - sum(prime_cost_no_vat_rub_amt)) /
nullIf(sum(turnover_bw_no_vat_rub_amt, 0.0), 1) AS front_margin_percent,
ROUND(100.0*sum(bread_turnover_vat_rub_amt)/nullIf(sum(turnover_vat_rub_amt, 0.0), 1) AS
sp_bread_rto_with_vat_per,
ROUND(SUM(bread_turnover_base_uom_amt, 0) AS bread_turnover_base_uom_amt,
ROUND(100.0 * sum(lty_turnover_vat_rub_amt) / nullIf(sum(turnover_vat_rub_amt, 0.0), 1) AS
loyalty_card_rto_per,
ROUND(100.0 * sum(fpl_rtl_txn_cnt) / nullIf(sum(rtl_bxn_cnt, 0.0), 1) AS loyalty_card_txn_per,
ROUND(SUM(turnover_base_uom_amt, 1) AS turnover_base_uom_amt,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN sum(py_d_turnover_vat_rub_amt)
    ELSE sum(py_wd_turnover_vat_rub_amt)
    END, 1) AS py_rto_with_vat_rub,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * ((SUM(turnover_vat_rub_amt) -
SUM(py_d_turnover_vat_rub_amt) / nullIf(SUM(py_d_turnover_vat_rub_amt, 0.0)))
        ELSE 100 * (1.0 * ((SUM(turnover_vat_rub_amt) - SUM(py_wd_turnover_vat_rub_amt) /
nullIf(SUM(py_wd_turnover_vat_rub_amt, 0.0)))
    END, 1) AS yoy_rto_with_vat_per
FROM hldm_dtl_e dw.ki_operational_metric
WHERE day_dt = yesterday()
    AND management_cluster_nm ILIKE 'Князь Полюс 78'
    AND financial_unit_format_nm ILIKE 'ТС.Пятигорск'
    AND store_id <> ''
    AND store_nm <> ''
GROUP BY store_id,
store_nm
ORDER BY store_id
```

Как мы измеряем?



Claude Sonnet 4
as LLM as a Judge

Тестирование
на 250+ запросах

Как мы измеряем?

Элемент	Что это
Вопрос	Размеченный запрос пользователя
Эталонный SQL	Подготовленный SQL запрос инженером
Сгенерированный SQL	SQL запрос, сгенерированный LLM
Эталонный DF	Результат отработки подготовленного SQL запроса
Сгенерированный DF	Результат генерации SQL LLM
Semantic Score	1/0 - вывод судьи на основании анализа эталонного и сгенерированного SQL запросов
DF Score	1/0 - вывод судьи на основании анализа эталонного и сгенерированного DF

Вердикт судьи

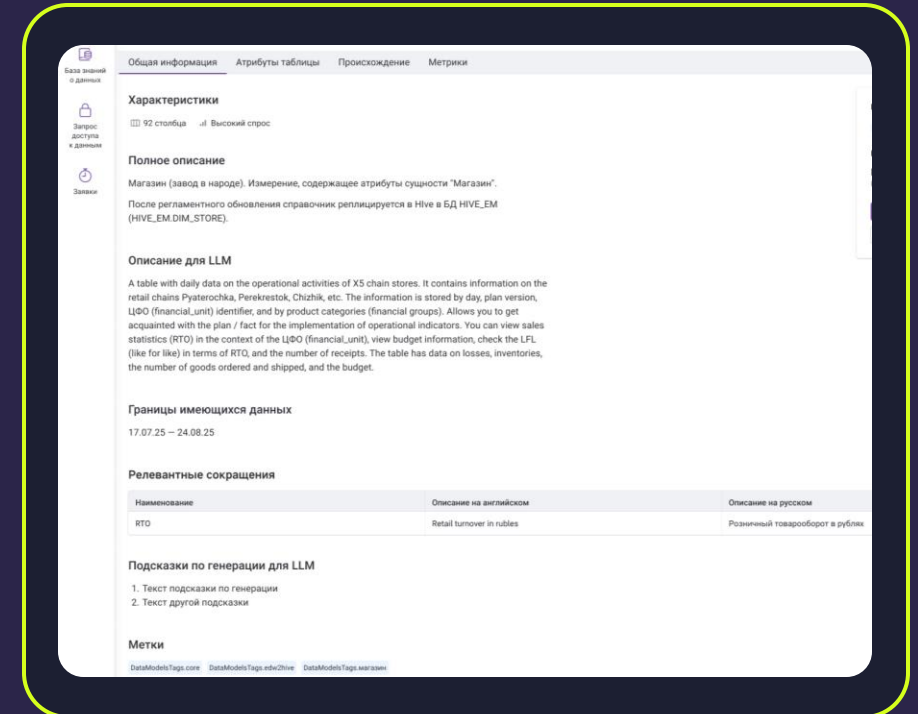
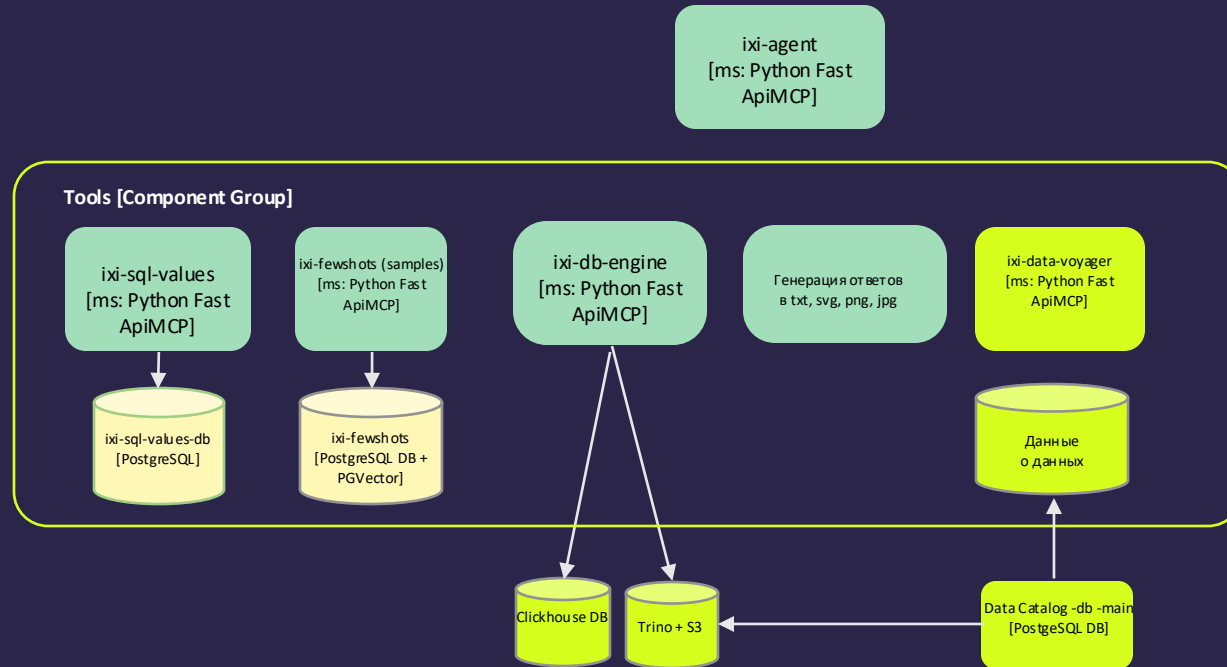
Написанный SQL-запрос некорректно отвечает на пользовательский запрос.

В пользовательском запросе требуется информация о площади магазинов конкретного кластера 'Казань 17' за текущий месяц, вывод данных должен включать идентификатор магазина, название, общую площадь и торговую площадь для каждого магазина, принадлежащего кластеру.

Как мы измеряем?

Модель данных	Кол-во колонок	Кол-во строк	LLM score
Клиентские метрики	150	> 1 млрд	0,65
Показатели по магазинам	171	> 1 млрд	0,52
Доступность товаров на полке	112	18 млн	0,76
NPS показатели	26	>300 тысяч	0,78
HR показатели по магазинам (текучка, производительность, скорость)	45	11 млн	0,49

Новая роль каталога данных



Новая роль каталога данных

Синергия инженера данных и LLM инженера

1 Описываем показатель

2 Формируем ака

3 Добавляем, не путать с чем

4 Указываем, где искать

5 Даем подсказку

Снижаем риск амбивалентности
Роутинг запросов на основании источника

```
{
  "id": "rto_lfl_vat_rub",

  /* === Бизнес-описание === */
  "name": "PTO LFL с НДС, руб",
  "unit": "руб",
  "aliases": ["PTO LFL с НДС", "LFL turnover VAT incl"],
  "anti_aliases": ["≠ rto.total.vat.rub", "≠ rto.lfl.novat.rub"],

  /* === Матрица «источник × бизнес-функция» === */
  "usages": [
    {
      "business_function": "operational_sales",
      "description": "Операционный контроль динамики продаж в LFL-магазинах",
      "sources": [
        {
          "system": "clickhouse",
          "mcp": "mcp-clickhouse",
          "table": "fct_daily_sales",
          "column": "revenue_lfl_vat_rub",
          "extra_filters": {"channel": ["offline", "online"], "status": "closed_shift"}
        }
      ]
    }
  ],
  /* === LLM-подсказка === */
  "llm_hint": "When user mentions 'sales performance' or 'operations', use usage.operational_sales. When user mentions 'loyalty' or 'customer segment', use usage.customer_analytics."
}
```

+ - 10%

Что нужно пользователю?

1 Скорость

2 Точность

3 То, что он хочет. А что он хочет?



15:43, 17.09.2025 ✓

показатели по тс5 за вчера



15:44, 17.09.2025

Вчера по торговой сети «ТС Пятерочка» были получены следующие данные по ключевым метрикам розничного товарооборота и выполнению плана:

Фильтры

- Торговая сеть - "ТС Пятерочка"
- Дата - "2025-09-16"

Результат:

- LFL PTO % (с онлайн продажами)* - хх%
- LFL трафик (с онлайн продажами) % - хх%
- LFL средний чек (с онлайн продажами) % - хх%
- LFL PTO в кг/шт % (с онлайн продажами)* - хх%
- LFL количество товаров в чеке (с онлайн продажами) % - хх%
- LFL количество PLU в чеке (с онлайн продажами) % - хх%
- PTO с НДС, руб - хх%
- Выполнение плана PTO/% - хх%
- План PTO (PLAN), руб - хх%
- Выполнение мотивационного плана/% - хх%
- Средний чек, руб - хх%
- Фронт маржа/% - хх%
- Доля PTO группы «Хлеб СП» (%) - хх%
- Сумма продаж по хлебу СП, штук - хх%
- Доля PTO с картами лояльности/% - хх%
- Доля чеков с картами лояльности/% - хх%
- PTO в базовых единицах, кг/шт - хх%
- PTO за прошлый год по дням недели, руб - хх%
- PTO год-год/% (день недели в день недели) - хх%



```
SELECT store_id,
store_nm,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * sum(cy_d_lfl_turnover_vat_rub_amt) /
nullIf(sum(py_d_lfl_turnover_vat_rub_amt, 0.0) - 1)
    ELSE 100 * (1.0 * sum(lfl_turnover_vat_rub_amt) / nullIf(sum(py_wd_lfl_turnover_vat_rub_amt, 0.0) - 1)
    END, 2) AS lfl_rto_percent,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * sum(rtl_bxn_lfl_d_cnt) /
nullIf(sum(py_dd_lfl_rtl_bxn_cnt, 0.0) - 1)
    ELSE 100 * (1.0 * sum(rtl_bxn_lfl_cnt) / nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0) - 1)
    END, 2) AS lfl_traffic_percent,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * ((1.0 * sum(cy_d_lfl_turnover_vat_rub_amt) /
nullIf(sum(rtl_bxn_lfl_d_cnt, 0.0)) / nullIf((1.0 * sum(py_d_lfl_turnover_vat_rub_amt) /
nullIf(sum(py_wd_lfl_turnover_vat_rub_amt, 0.0)) / nullIf((1.0 *
sum(py_wd_lfl_turnover_vat_rub_amt) / nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0)) / nullIf((1.0 *
sum(py_wd_lfl_turnover_vat_rub_amt, 0.0)) - 1)
    END, 2) AS lfl_avg_check_percent,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
```

...

```
OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * sum(cy_d_lfl_turnover_base_uom_amt) /
nullIf(sum(py_d_lfl_turnover_base_uom_amt, 0.0) - 1)
    ELSE 100 * (1.0 * sum(lfl_turnover_base_uom_amt) / nullIf(sum(py_wd_lfl_turnover_base_uom_amt, 0.0) - 1)
    END, 2) AS lfl_rto_pcs,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * (sum(cy_d_lfl_turnover_base_uom_amt) /
nullIf(sum(rtl_bxn_lfl_d_cnt, 0.0)) / (sum(py_d_lfl_turnover_base_uom_amt) / nullIf(sum(py_dd_lfl_rtl_bxn_cnt, 0.0)) - 1)
    ELSE 100 * (1.0 * (sum(lfl_turnover_base_uom_amt) / nullIf(sum(rtl_bxn_lfl_cnt, 0.0)) /
(sum(py_wd_lfl_turnover_base_uom_amt) / nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0)) - 1)
    END, 2) AS lfl_number_pcs_per_receipt,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * (sum(plu_unique_lfl_d_cnt) /
nullIf(sum(rtl_bxn_lfl_d_cnt, 0.0)) / (sum(py_dd_lfl_plu_unique_cnt) / nullIf(sum(py_dd_lfl_rtl_bxn_cnt, 0.0)) - 1)
    ELSE 100 * (1.0 * (sum(plu_unique_lfl_cnt) / nullIf(sum(rtl_bxn_lfl_cnt, 0.0)) / (sum(py_wd_lfl_plu_unique_cnt) /
nullIf(sum(py_wd_lfl_rtl_txn_cnt, 0.0)) - 1)
    END, 2) AS lfl_number_of_plu_per_receipt,
ROUND(SUM(turnover_vat_rub_amt, 1) AS rto_with_vat_rub,
ROUND(100.0*sum(turnover_vat_rub_amt)/nullIf(sum(CASE
    WHEN plan_code='PLAN' THEN budget_sales_rub_amt
    ELSE 0
    END), 0.0), 1) AS rto_plan_fulfillment_percent,
ROUND(SUM(CASE
    WHEN plan_code='PLAN' THEN budget_sales_rub_amt
    ELSE 0
    END), 1) AS plan_rto_with_vat_rub,
ROUND(100.0* SUM(turnover_vat_rub_amt)/NULLIF(SUM(budget_motiv_plan_tmrv_rub_amt, 0.0), 1) AS
motiv_plan_fulfillment_percent,
```

...

```
ROUND(SUM(turnover_vat_rub_amt) / NULLIF(SUM(rtl_bxn_cnt, 0), 1) AS avg_check_rub,
ROUND(100.0 * (sum(turnover_bw_no_vat_rub_amt) - sum(prime_cost_no_vat_rub_amt)) /
nullIf(sum(turnover_bw_no_vat_rub_amt, 0.0), 1) AS front_margin_percent,
ROUND(100.0*sum(bread_turnover_vat_rub_amt)/nullIf(sum(turnover_vat_rub_amt, 0.0), 1) AS
sp_bread_rto_with_vat_per,
ROUND(SUM(bread_turnover_base_uom_amt, 0) AS bread_turnover_base_uom_amt,
ROUND(100.0 * sum(lty_turnover_vat_rub_amt) / nullIf(sum(turnover_vat_rub_amt, 0.0), 1) AS
loyalty_card_rto_per,
ROUND(100.0 * sum(fpl_rtl_txn_cnt) / nullIf(sum(rtl_bxn_cnt, 0.0), 1) AS loyalty_card_txn_per,
ROUND(SUM(turnover_base_uom_amt, 1) AS turnover_base_uom_amt,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN sum(py_d_turnover_vat_rub_amt)
    ELSE sum(py_wd_turnover_vat_rub_amt)
    END, 1) AS py_rto_with_vat_rub,
ROUND(CASE
    WHEN toDayOfMonth(min(day_dt)) =1
        AND toDayOfMonth(dateAdd(DAY, 1, max(day_dt))) =1
        OR toDayOfMonth(min(day_dt)) =1
        AND toMonth(min(day_dt)) =toMonth(today())
        AND dateAdd(DAY, 1, max(day_dt))=today() THEN 100 * (1.0 * ((SUM(turnover_vat_rub_amt) -
SUM(py_d_turnover_vat_rub_amt) / nullIf(SUM(py_d_turnover_vat_rub_amt, 0.0)))
    ELSE 100 * (1.0 * ((SUM(turnover_vat_rub_amt) - SUM(py_wd_turnover_vat_rub_amt) /
nullIf(SUM(py_wd_turnover_vat_rub_amt, 0.0)))
    END, 1) AS yoy_rto_with_vat_per
FROM hldm_dtl_e dw.ki_operational_metric
WHERE day_dt = yesterday()
    AND management_cluster_nm ILIKE 'Князь Полюс 78'
    AND financial_unit_format_nm ILIKE 'ТС.Платежия'
    AND store_id <> ''
    AND store_nm <> ''
GROUP BY store_id,
store_nm
ORDER BY store_id
```

Генерация
3
секунды

А что с агентом?

1 Гибкость

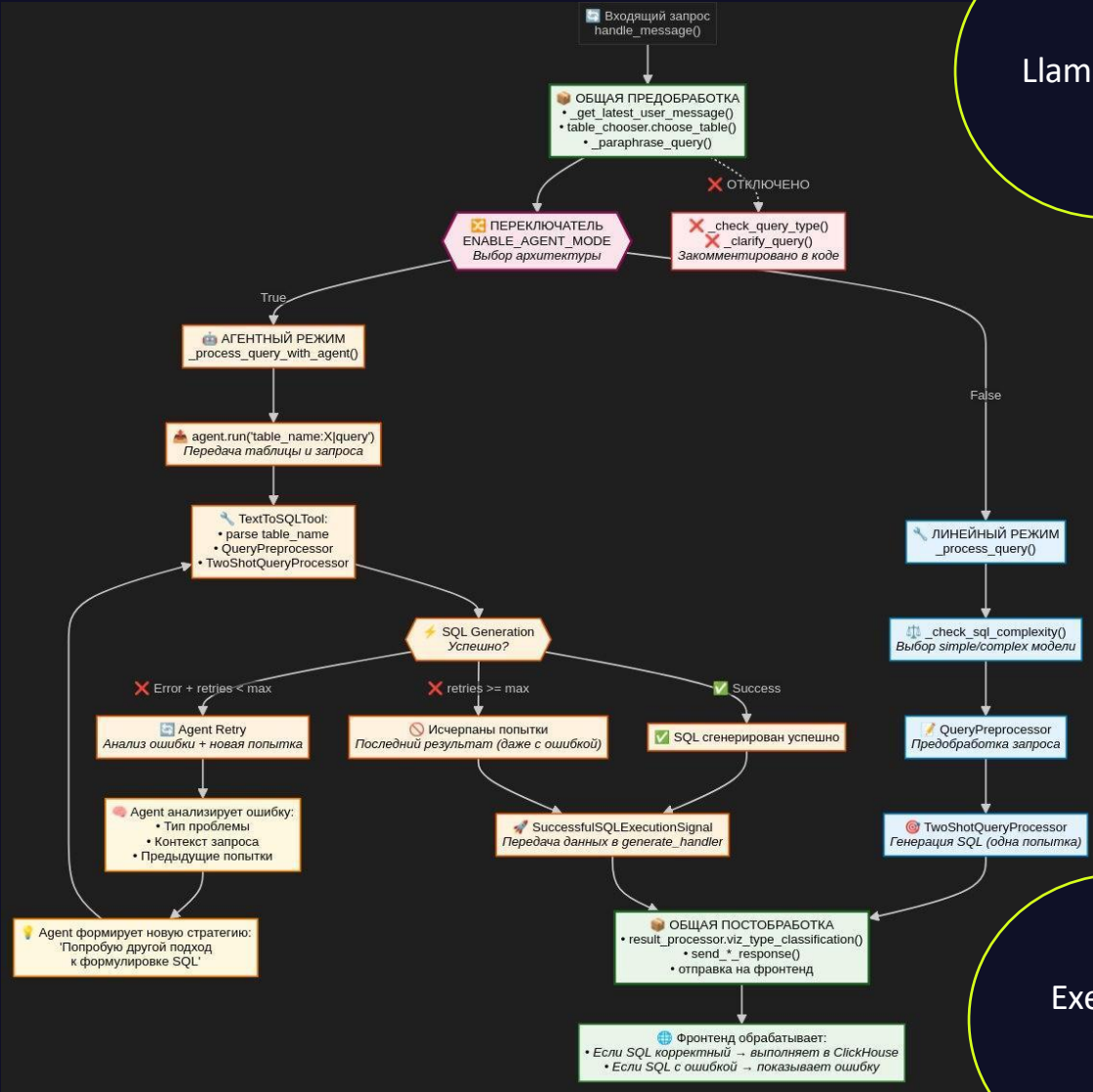
tool
call

2 Автономия

Retry

3 Скейлинг

yaml



Llama Index

Execution
T x2



1 минуту назад
Я добавил параметр seed = 42

1 минуту назад
Давайте квантизацию поменяем?

1 минуту назад
Давайте ролевку прикрутим?

1 минуту назад
Ну что модель когда обучим?

1 минуту назад
Включай beam search

1 минуту назад
Да тут классификация типовая

1 минуту назад
MP подготовил, заливайте

1 минуту назад
Вот новый релиз Qwen

1 минуту назад
Мы тут фьюшоты обновили

1 минуту назад
Мы тут немного описание скорректировали

1 минуту назад
Сделаем саммари в перефразере?

1 минуту назад
Давайте structured output сделаем?

1 минуту назад
Давайте structured output сделаем?

1 минуту назад
Расширяем контекст через LLM

1 минуту назад
Давайте в контекст инфу из ddl statements сразу добавим?

1 минуту назад
Обновил разметку, залейте пжл

1 минуту назад
А чо по ризонингу в sql?

1 минуту назад
Расширяем контекст через LLM

TEST

Дальнейшие планы

- Баланс по уточнениям
 - Динамическое обновление базы фьюшотов
 - Роутинг запросов и подключение новых СУБД
 - Ускорение агентского подхода
-
- Обработка текстовых документов с применением уже имеющихся данных
 - Обучение модели
 - Персонализация/ Personal Memory



@DEATHCOREISLOVE



Ермаченков Владимир

Руководитель направления
продвинутой аналитики в X5 Tech

Сравнение с другими LLM

Таблица	LLM	Exec Accuracy
Таблица по магазинам	Qwen3-32b (текущая)	0.65
	Qwen-235b	0.68
	gptoss-120b	0.64
	Think2sql-7b	0.61
	Qwen3-Coder-Flash	0.65

Что такое TXT2SQL? Зачем?

Данные по
магазинам

Клиентские
метрики

Доступность
товаров

NPS
метрики

HR показатели

etc

N – Таблиц

> 1,5 млрд записей в таблице

