

Как научить LLM работать с данными, а не просто писать правдоподобный SQL

Как научить LLM работать с данными, а не просто писать правдоподобный SQL

ПРИНИМАЕМ?



Максим Стаценко

руководитель отдела BigData
в Яндекс Поиске и Рекламе

15+ лет работы с данными

Сейчас — ландшафт данных,
в котором AI приносит пользу бизнесу

В подразделении CDO **больше чем в половине задач** агент
проделывает **большую часть работы**

По оценке исполнителей задач

РЦОИ · Газпром · МОЭК · БРК
ВТБ24 · Яндекс · Mail.ru

ВОЗЬМЕМ ТИПИЧНУЮ ЗАДАЧУ

Вы бы закрыли задачу?

Добавить ежедневную **выручку**
в e-commerce срез поиска

ВОЗЬМЕМ ТИПИЧНУЮ ЗАДАЧУ

Вы бы закрыли задачу?

Добавить ежедневную **выручку**
в e-commerce срез поиска

✓ Сборка проходит

ВОЗЬМЕМ ТИПИЧНУЮ ЗАДАЧУ

Вы бы закрыли задачу?

Добавить ежедневную **выручку**
в e-commerce срез поиска

✓ Сборка проходит ✓ Тесты проходят

ВОЗЬМЕМ ТИПИЧНУЮ ЗАДАЧУ

Вы бы закрыли задачу?

Добавить ежедневную **выручку**
в e-commerce срез поиска

✓ Сборка проходит ✓ Тесты проходят ✓ Агент написал «готово»

ВОЗЬМЕМ ТИПИЧНУЮ ЗАДАЧУ

Вы бы закрыли задачу?

Добавить ежедневную **выручку**
в e-commerce срез поиска

✓ Сборка проходит ✓ Тесты проходят ✓ Агент написал «готово»

Кто бы принял? Поднимите руку.

Первое дело: прочитать новый формат

ВНЕШНЯЯ ВЫГРУЗКА

```
<ValCurs Date="01.04.2025">  
  <Valute>  
    <CharCode>CNY</CharCode>  
    <Nominal>1</Nominal>  
    <Value>12,3456</Value>  
  </Valute>  
</ValCurs>
```

Учебный фрагмент, значение вымышлено

Первое дело: прочитать новый формат

ВНЕШНЯЯ ВЫГРУЗКА

```
<ValCurs Date="01.04.2025">
  <Valute>
    <CharCode>CNY</CharCode>
    <Nominal>1</Nominal>
    <Value>12,3456</Value>
  </Valute>
</ValCurs>
```

ВНУТРЕННЯЯ ЗАПИСЬ

```
{
  "currency": "CNY",
  "nominal": 1,
  "rate": 12.3456
}
```

Учебный фрагмент, значение вымышлено

Тест задаёт условие приёма

01 Зафиксированы вход и ожидаемый результат

Тест задаёт условие приёма

- 01 Зафиксированы вход и ожидаемый результат
- 02 При ошибке тесты падают без вмешательства человека

Тест задаёт условие приёма

- 01 Зафиксированы вход и ожидаемый результат
- 02 При ошибке тесты падают без вмешательства человека
- 03 Агент пишет код

Тест задаёт условие приёма

- 01 Зафиксированы вход и ожидаемый результат
- 02 При ошибке тесты падают без вмешательства человека
- 03 Агент пишет код
- 04 Мы уверены в результате на основе прохождения тестов и крайних случаев.

×1,5

**больше задач
за спринт**

Агенты помогают разбирать
техдолг и бэклог

Собираем консилиум агентов



Собираем консилиум агентов

БЕЗОПАСНОСТЬ

Что может пойти не так?

Собираем консилиум агентов

БЕЗОПАСНОСТЬ

Что может пойти не так?

АРХИТЕКТУРА

Как это будет жить?

Собираем консилиум агентов

БЕЗОПАСНОСТЬ

Что может пойти не так?

АРХИТЕКТУРА

Как это будет жить?

ПРОДУКТ

Ту ли задачу решили?

Польза появилась не сразу

ПОЧТИ ГОД

0

выигрыша по времени:
ревью отвлекало разработчиков

Польза появилась не сразу

ПОЧТИ ГОД

0

выигрыша по времени:
ревью отвлекало разработчиков

СЕЙЧАС

до 10%

экономии времени
разработчика

Конфиг написан. Кто отвечает за запуск?

```
# Учебный фрагмент
schedule: "0 3 * * *"
retries: 3
catchup: true
resources:
  cpu: 16
  memory: 64Gi
```

Популярный инструмент

→ **агент знает шаблон**

Внутренний инструмент

→ **нужны примеры и документация**

Конфиг написан. Кто отвечает за запуск?

```
# Учебный фрагмент
schedule: "0 3 * * *"
retries: 3
catchup: true
resources:
  cpu: 16
  memory: 64Gi
```

Популярный инструмент

→ **агент знает шаблон**

Внутренний инструмент

→ **нужны примеры и документация**

Значения и последствия

запуска проверяет человек

Первого апреля график сломался



Первого апреля график сломался



Быстрее.
При том же результате.

SQL / Python / Java / Scala

Быстрее. При том же результате.

SQL / Python / Java / Scala

Эталон результата + критерий приёмки

Быстрее. При том же результате.

SQL / Python / Java / Scala

Эталон результата + критерий приёмки

Несколько тысяч ядер сэкономлено

Опыт спикера и его команд

Кажется, он устроил нагрузочное тестирование

```

101 SELECT s.gordi, w.gordi, d.gordi, 'gordi' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
102 SELECT s.sevastianov, w.sevastianov, d.sevastianov, 'sevastianov' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
103 SELECT s.karmiryan, w.karmiryan, d.karmiryan, 'karmiryan' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
104 SELECT s.revonenkod, w.revonenkod, d.revonenkod, 'revonenkod' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
105 SELECT s.ravasiliev, w.ravasiliev, d.ravasiliev, 'ravasiliev' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
106 SELECT s.maxstatsenko, w.maxstatsenko, d.maxstatsenko, 'maxstatsenko' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
107 SELECT s.khomyuk, w.khomyuk, d.khomyuk, 'khomyuk' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
108 SELECT s.annmelnik, w.annmelnik, d.annmelnik, 'annmelnik' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
109 SELECT s.evg_i, w.evg_i, d.evg_i, 'evg-i' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
110 SELECT s.saltykov, w.saltykov, d.saltykov, 'saltykov' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
111 SELECT s.kochurovvs, w.kochurovvs, d.kochurovvs, 'kochurovvs' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d;

```

Код агента из коробки тоже приходится оптимизировать

Кажется, он устроил нагрузочное тестирование

```

101 SELECT s.gordi, w.gordi, d.gordi, 'gordi' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
102 SELECT s.sevastianov, w.sevastianov, d.sevastianov, 'sevastianov' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
103 SELECT s.karmiryan, w.karmiryan, d.karmiryan, 'karmiryan' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
104 SELECT s.revonenkod, w.revonenkod, d.revonenkod, 'revonenkod' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
105 SELECT s.ravasiliev, w.ravasiliev, d.ravasiliev, 'ravasiliev' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
106 SELECT s.maxstatsenko, w.maxstatsenko, d.maxstatsenko, 'maxstatsenko' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
107 SELECT s.khomyuk, w.khomyuk, d.khomyuk, 'khomyuk' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
108 SELECT s.annmelnik, w.annmelnik, d.annmelnik, 'annmelnik' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
109 SELECT s.evg_i, w.evg_i, d.evg_i, 'evg-i' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
110 SELECT s.saltykov, w.saltykov, d.saltykov, 'saltykov' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d UNION ALL
111 SELECT s.kochurovvs, w.kochurovvs, d.kochurovvs, 'kochurovvs' FROM $staff_counts AS s CROSS JOIN $wau_counts AS w CROSS JOIN $dau_counts AS d;

```

CROSS JOIN × CROSS JOIN

Код агента из коробки тоже приходится оптимизировать

ВОЗВРАЩАЕМСЯ К ИСХОДНОЙ ЗАДАЧЕ

ОСТАЛОСЬ ПОСЧИТАТЬ ВЫРУЧКУ.

Загрузчик готов. История загружена.

event_id	segment	amount_rub	tax
s1	ecom	120	20
s1	ecom	120	20
r1	ecom	-24	-4
x1	other	50	10

Какую формулу выберете вы?

event_id	segment	amount_rub	tax
s1	ecom	120	20
s1	ecom	120	20
r1	ecom	-24	-4
x1	other	50	10

A amount_rub

B amount_rub - tax

Сначала нужно выбрать определение

event_id	segment	amount_rub	tax
s1	ecom	120	20
s1	ecom	120	20
r1	ecom	-24	-4
x1	other	50	10

Уверенный ответ не заменяет согласованное правило

Сначала нужно выбрать определение

event_id	segment	amount_rub	tax
s1	ecom	120	20
s1	ecom	120	20
r1	ecom	-24	-4
x1	other	50	10

Налог входит в выручку?

Уверенный ответ не заменяет согласованное правило

Сначала нужно выбрать определение

event_id	segment	amount_rub	tax
s1	ecom	120	20
s1	ecom	120	20
r1	ecom	-24	-4
x1	other	50	10

Налог входит в выручку?

Повтор s1 — дубль?

Уверенный ответ не заменяет согласованное правило

Сначала нужно выбрать определение

event_id	segment	amount_rub	tax
s1	ecom	120	20
s1	ecom	120	20
r1	ecom	-24	-4
x1	other	50	10

Налог входит в выручку?

Повтор s1 — дубль?

Как учитываем возврат?

Уверенный ответ не заменяет согласованное правило

Сначала нужно выбрать определение

event_id	segment	amount_rub	tax
s1	ecom	120	20
s1	ecom	120	20
r1	ecom	-24	-4
x1	other	50	10

- Налог входит в выручку?
- Повтор s1 — дубль?
- Как учитываем возврат?
- Как задаём e-commerce срез?

Уверенный ответ не заменяет согласованное правило

Одно слово. Восемь метрик.

revenue_rub_with_tax

...

Как выбрать, если попросили просто «выручку»?

Одно слово. Восемь метрик.

revenue_rub_with_tax

revenue_rub_wo_tax

...

Как выбрать, если попросили просто «выручку»?

Одно слово. Восемь метрик.

revenue_rub_with_tax

revenue_rub_wo_tax

revenue_cur_with_tax

...

Как выбрать, если попросили просто «выручку»?

А теперь примите этот SQL

事件	分组	金额	税额
s1	电商	120	20
s1	电商	120	20
r1	电商	-24	-4
x1	其他	50	10

```
SELECT SUM(金额 - 税额)
FROM 交易
WHERE 分组 = '电商';
```

А теперь примите этот SQL

事件	分组	金额	税额
s1	电商	120	20
s1	电商	120	20
r1	电商	-24	-4
x1	其他	50	10

```
SELECT SUM(金额 - 税额)
FROM 交易
WHERE 分组 = '电商';
```

Запрос выглядит правдоподобно.
Что он должен был посчитать?

ВОТ ГДЕ МЕНЯЕТСЯ ЗАДАЧА

SQL ВЫПОЛНИЛСЯ.

Исполнение проверили. Смысл — ещё нет.

ВОТ ГДЕ МЕНЯЕТСЯ ЗАДАЧА

SQL ВЫПОЛНИЛСЯ. ОТВЕТ ВЕРНЫЙ?

Исполнение проверили. Смысл — ещё нет.

Что нужно знать для приёмки

01 Определения

Что нужно знать для приёмки

01 Определения

02 Lineage и доверие к источникам

Что нужно знать для приёмки

01 Определения

02 Lineage и доверие к источникам

03 Ньюансы данных

Что нужно знать для приёмки

- 01 Определения
- 02 Lineage и доверие к источникам
- 03 Нюансы данных
- 04 Skills для типовых задач

Что нужно знать для приёмки

- 01 Определения
- 02 Lineage и доверие к источникам
- 03 Нюансы данных
- 04 Skills для типовых задач
- 05 Evals

У метрики должен быть паспорт

ОПРЕДЕЛЕНИЕ

Что именно называем выручкой

У метрики должен быть паспорт

ОПРЕДЕЛЕНИЕ

Что именно называем выручкой

ФОРМУЛА

Налоги, возвраты, валюта, дубли

У метрики должен быть паспорт

ОПРЕДЕЛЕНИЕ

Что именно называем выручкой

ФОРМУЛА

Налоги, возвраты, валюта, дубли

ИСТОЧНИК И СРЕЗЫ

Где считать и как группировать

У метрики должен быть паспорт

ОПРЕДЕЛЕНИЕ

Что именно называем выручкой

ФОРМУЛА

Налоги, возвраты, валюта, дубли

ИСТОЧНИК И СРЕЗЫ

Где считать и как группировать

ВЛАДЕЛЕЦ

Кто подтверждает правило

Начать можно с одного Markdown-файла

Метрики e-commerce

Выручка

Определение: [согласовать с владельцем]

Формула: [явно описать правило]

Источник: [каноническая таблица]

Срезы: [допустимые измерения]

Владелец: [ответственный]

Если определение не выбрано — уточни его.

Один домен. Понятные правила. Доступ в контексте агента.

Как меняется качество с семантикой

Эксперимент	Без семантики	С семантикой
dbt · сильная модель	62,5%	87,2%
dbt · слабая модель	62,5%	100%
Yandex · слабая модель	10%	50%

Разные условия и модели. Это не универсальный benchmark.

100 000

ТАБЛИЦ.
КАКАЯ ПОДХОДИТ?

100 000

ТАБЛИЦ. КАКАЯ ПОДХОДИТ?

Похожее название — слабое доказательство

Популярный дашборд. Правдоподобные числа.

Популярный дашборд.
Правдоподобные числа.

3 МЕСЯЦА

таблица не обновлялась

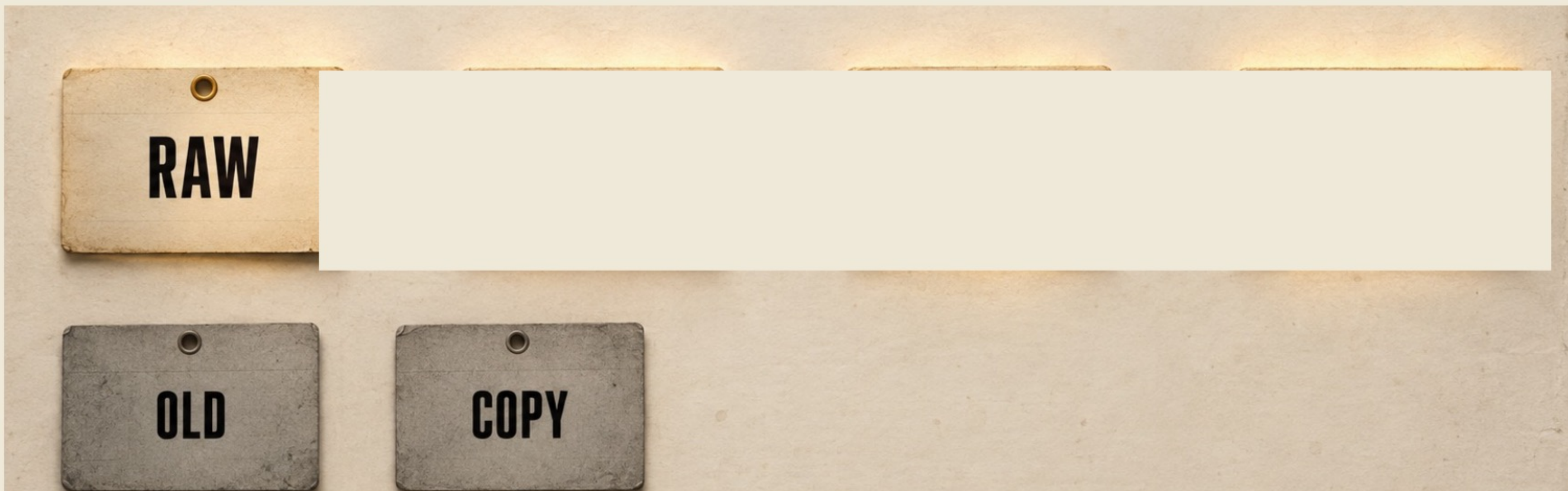
Популярный дашборд.
Правдоподобные числа.

3 МЕСЯЦА

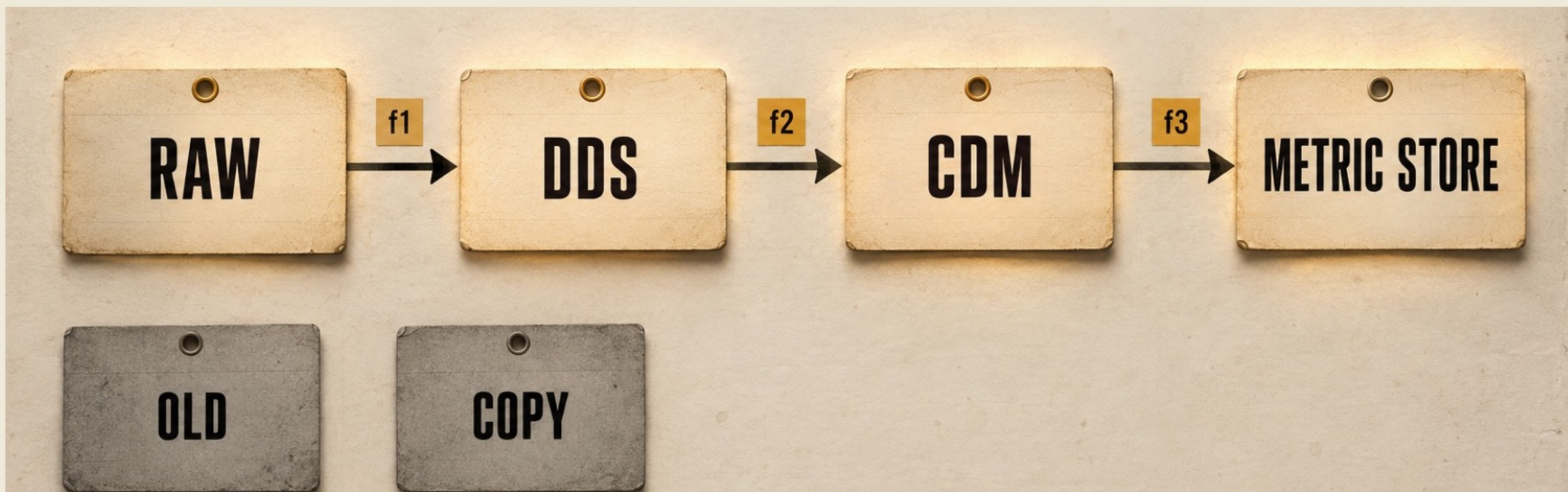
таблица не обновлялась

Нас спасло знание CDO об инциденте

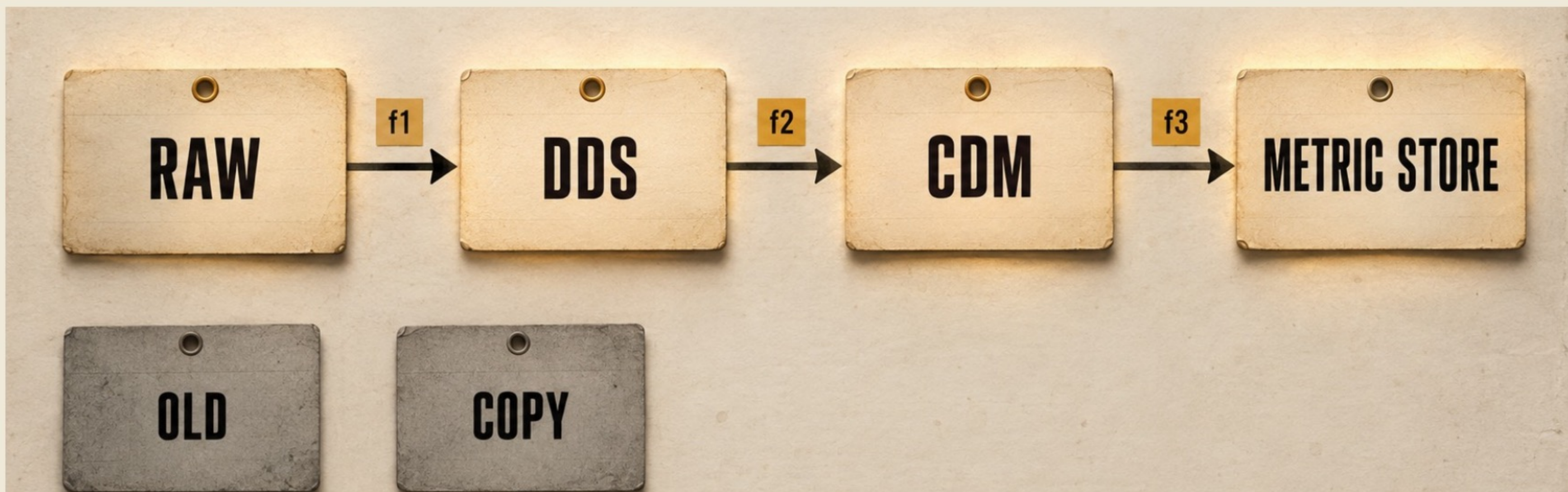
Lineage показывает путь к эталону



Lineage показывает путь к эталону



Lineage показывает путь к эталону



Происхождение → преобразования → проверка совпадения

Агент оставил проверяющему следы

```
SELECT SUM(金额)  
FROM 汇总;
```

Golden layer

1 шаг от Metric Store

Обновилась вчера

```
SELECT SUM(金额)  
FROM 原始;
```

RAW

Нет связи с Metric Store

Не обновлялась год

Слой · свежесть · связь с эталоном

Агент оставил проверяющему следы

```
SELECT SUM(金额)  
FROM 汇总;
```

Golden layer

1 шаг от Metric Store

Обновилась вчера

Оснований больше

```
SELECT SUM(金额)  
FROM 原始;
```

RAW

Нет связи с Metric Store

Не обновлялась год

Слой · свежесть · связь с эталоном

Агент оставил проверяющему следы

```
SELECT SUM(金额)  
FROM 汇总;
```

Golden layer

1 шаг от Metric Store

Обновилась вчера

Оснований больше

```
SELECT SUM(金额)  
FROM 原始;
```

RAW

Нет связи с Metric Store

Не обновлялась год

Нужна проверка источника

Слой · свежесть · связь с эталоном

Нюансы, которых нет в схеме таблицы

ФРОД

Исключаем по умолчанию

Нюансы, которых нет в схеме таблицы

ФРОД

Исключаем по умолчанию

14 ФЕВРАЛЯ

Логи неполные после инцидента

Нюансы, которых нет в схеме таблицы

ФРОД

Исключаем по умолчанию

14 ФЕВРАЛЯ

Логи неполные после инцидента

HTML5-АНИМАЦИЯ

Для продаж — картинка

Как решать типовую задачу

Уточни определение

Найди канонический источник

Примени нюансы

Проверь и объясни результат

ВНУТРЕННИЕ EVALS ANTHROPIC

≤21% без skills

Результат их evals, не прогноз для вашей системы

Как решать типовую задачу

Уточни определение

Найди канонический источник

Примени нюансы

Проверь и объясни результат

ВНУТРЕННИЕ EVALS ANTHROPIC

≤21% без skills

>95% со skills

Результат их evals, не прогноз для вашей системы

METADATA IS THE NEW CODE.

Изменили описание —
изменили поведение агента

METADATA IS THE NEW CODE.

Изменили описание —
изменили поведение агента

Проверяем на задачах:

верный ответ · источник · правило
и обоснованный отказ

Изменили ETL — обновите знания

01 PR с изменением ETL

Знания меняются вместе с системой

Изменили ETL — обновите знания

01 PR с изменением ETL

02 Документация витрины и её потомков

Знания меняются вместе с системой

Изменили ETL — обновите знания

- 01 PR с изменением ETL
- 02 Документация витрины и её потомков
- 03 Прогон evals

Знания меняются вместе с системой

Изменили ETL — обновите знания

- 01 PR с изменением ETL
- 02 Документация витрины и её потомков
- 03 Прогон evals
- 04 Подтверждение владельца

Знания меняются вместе с системой

ТЕПЕРЬ ЕСТЬ, ЧТО ПРОВЕРЯТЬ.

✓ Код и тесты

ТЕПЕРЬ ЕСТЬ, ЧТО ПРОВЕРЯТЬ.

- ✓ Код и тесты
- ✓ Определение метрики

ТЕПЕРЬ ЕСТЬ, ЧТО ПРОВЕРЯТЬ.

- ✓ Код и тесты
- ✓ Определение метрики
- ✓ Источник и его свежесть

ТЕПЕРЬ ЕСТЬ, ЧТО ПРОВЕРЯТЬ.

- ✓ Код и тесты
- ✓ Определение метрики
- ✓ Источник и его свежесть
- ✓ Нюансы и результаты проверок

Где агент уже помогает

Задача	Помощь агента	Приёмка
Типовая разработка	Большая часть автономно	ТЗ и тесты
Конфиги и bash скрипты	Подготовка изменений	Ревью последствий
Оптимизация	Большая часть автономно	Эталон + ресурсы
Инциденты	Помощь дежурному	
ETL со смысловой задачей	При доступной семантике	Семантика + ревью

Начните с небольшого домена

01 Сфокусируйтесь на цели сделать copilot, а не end2end

Начните с небольшого домена

- 01 Сфокусируйтесь на цели сделать copilot, а не end2end
- 02 Опишите семантику и соберите eval

Начните с небольшого домена

- 01 Сфокусируйтесь на цели сделать copilot, а не end2end
- 02 Опишите семантику и соберите eval
- 03 Растите адопшен, ищите энтузиастов и реальные примеры успеха

Начните с небольшого домена

- 01 Сфокусируйтесь на цели сделать copilot, а не end2end
- 02 Опишите семантику и соберите eval
- 03 Растите адопшен, ищите энтузиастов и реальные примеры успеха
- 04 Расширяйтесь по мере получения успехов

A steam locomotive is shown in a dark, industrial setting, possibly a factory or a tunnel. The locomotive is moving along tracks, and steam is visible coming from the engine. The lighting is dramatic, with strong highlights on the locomotive's wheels and boiler, and deep shadows in the surrounding environment. The overall tone is gritty and industrial.

Надёжная AI-аналитика начинается не с модели, а с **порядка в данных.**

Модели уже умеют многое.

Подготовим данные, с которыми они работают.

ВОПРОСЫ?

t.me/apxumar 

maxstatsenko@yandex-team.ru

maxstatsenko@gmail.com

«Если искусственный интеллект — это наше будущее,
то большие данные — уголь для паровоза,
на котором мы в него въедем»