

Magnit Governance изнутри

Открытый стек Governance
для агентских сценариев над Data Lake

Филимонов Михаил

CTO DataPlatform · Магнит · SmartData



Q1 2026: каталог зависит от поставщика

Каталог Informatica / Axon с ежегодным сопровождением

The screenshot displays the Informatica Data Catalog interface. The top navigation bar includes 'DATA CATALOG', 'Enterprise Data Catalog', 'Discovery', and 'Data Asset Analytics'. The left sidebar shows 'Home' and 'V_WHS'. The main content area is titled 'Развитие сети и эксплуатация' with a 'Поделиться' button. Below the title, the breadcrumb path is 'Teradata_TDPED > TERADATA > PRD_VD_DM'. The 'Overview' tab is active, showing a 'Description' section with text about the data domain, a 'Source Description' section with 'Справочник магазинов', and a 'Sample Columns' table. The table has columns: Name, Business Title, Description, and Data Domains. It lists 'WHS_ID' and 'CODE' with their respective descriptions. A 'Reviews' section at the bottom allows users to enter a review. On the right, a 'People' sidebar lists roles like Data Owner, Data Steward, Followers, Metadata Engineer, Metadata Specialist, Subject Matter Experts, and Technical Owner, all marked as 'not assigned'. A 'Custom Attributes' section at the bottom right shows 'Категория информации' as 'Корпоративная информация (К)' and 'Предметная область' as 'Развитие сети и эксплуатация'.

Name	Business Title	Description	Data Domains
WHS_ID		ID места хранения	
		Уникальный корпоративный код/ном...	
		Экспертиза:	
CODE	Корпоративный код торговой точки	Сектор операционного сопровож...	
		Сектор эффективности труда раз...	

- Доработки зависят от провайдера
- Ежегодно платим за сопровождение

- Ограничены в интеграции с агентами
- Команда не может менять каталог сама

Q1 2026: DQ через команду аналитиков

Прежний DQ-движок и каталог проверок



DQ-14: Количество товара на РЦ более 1 000 000 000 шт.
Локальное правило качества данных в PRD_VD_DM | Связанный набор данных : V_REP_COMM_LOG_RC

 Поделиться 

 Действия 

ОБЩАЯ ИНФОРМАЦИЯ

ОТВЕТСТВЕННЫЕ

ОТЧЕТ

СВЯЗИ В ИЕРАХИИ

ОРГАНИЗАЦИОННАЯ МОДЕЛЬ

ИСТОРИЯ ИЗМЕНЕНИЙ

ЗАПРОСЫ

ПОДПИСАТЬСЯ 

ОПРЕДЕЛЕНИЕ

Проверяется количество товара по модулю. Ошибкой считается превышение больше 1 000 000 000 шт.

Контрольный номер:

DQ-14

Тип:

Достоверность

Система измеренная в:

PRD_VD_DM

Способ измерения:

Автоматический Informatica Data Quality

Описание способа измерения:

Измерение качества данных в Informatica Data Quality автоматически запускаются Axon

Предметная область:

Логистика (РЦ, АТП)

КЛАССИФИКАЦИИ

БАЗОВЫЕ КЛАССИФИКАЦИИ

Актуальность описания:

 Активный

Жизненный цикл:

РЕАЛИЗОВАНО

Видимость элемента:

Общедоступный

Критичность:

Low

ДРУГИЕ СВЕДЕНИЯ

Создано:

22-Июн-2023

Автор последнего обновления:

[svc_dg_MD_admin](#)
[svc_dg_MD_admin](#)

- Проверки качества создаются только DQ аналитиками
- T2M проверки > 3 дней

DG по метрике заказов. Проблемы

Задача	Действие	Срок решения
Интегрировать базу в каталог	Тикет на подрядчика	~7 дней/интеграция
Создать проверки на расчет	Тикет на команду DQ аналитиков	~3 дня/проверка
Поставить расчет на мониторинг	Тикет на сопровождение	3 успешных расчета на продакшн
Создание MCP к инструментам	Тикет на подрядчика	Недели/месяцы

Open-source основа и свои сервисы

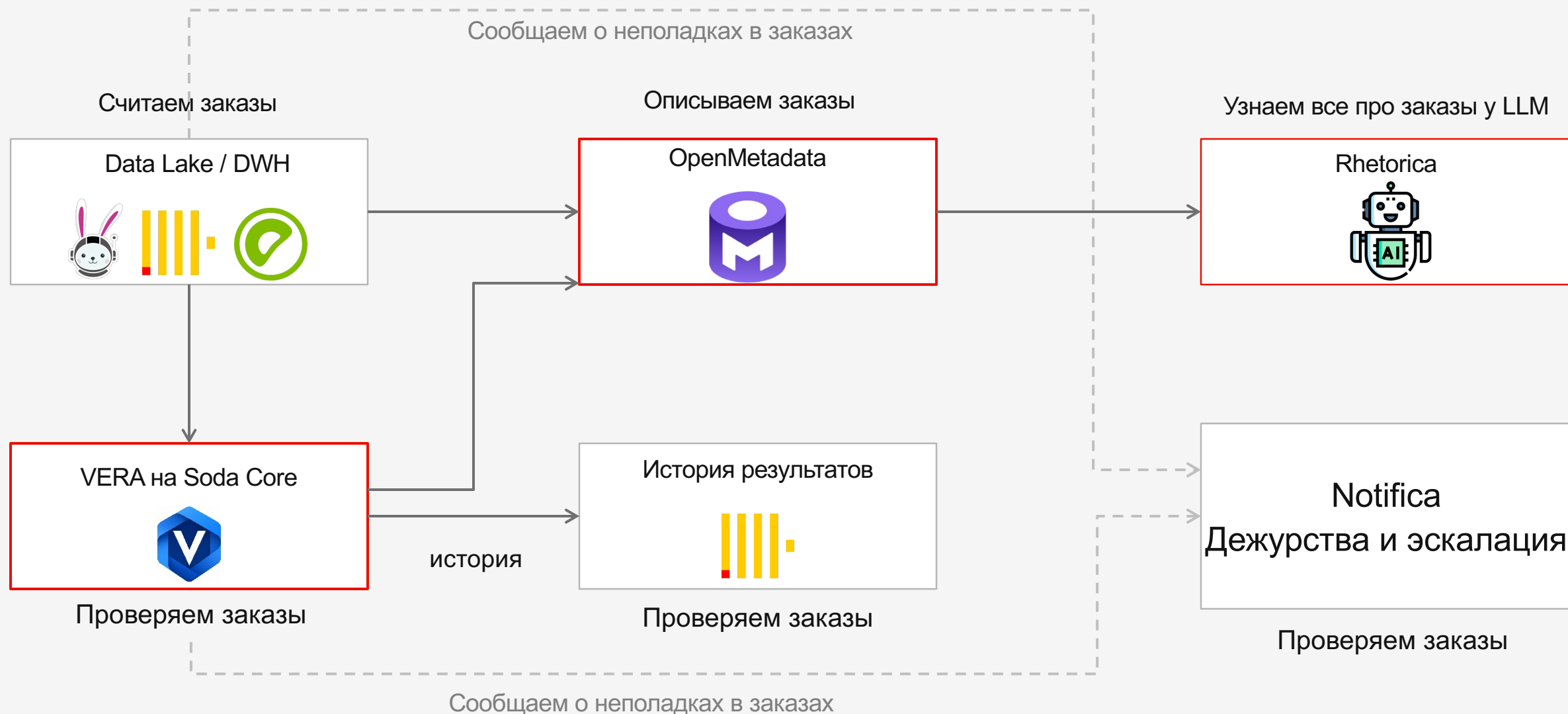
Команда сама развивает инструменты и их интеграции

Задача	Текущее решение	Что развивает команда
Каталог	OpenMetadata	Коннекторы, UI и метаданные
Качество	Soda Core + VERA	Правила, рантаймы и публикацию
Инциденты	Notifica и процесс Incident Management	Дежурства, эскалацию и восстановление
Работа с данными	Rhetorica	Агентов, контекст и каналы работы

OpenMetadata и Soda Core дают открытую основу

VERA, Notifica и Rhetorica развиваем как внутренние решения

Governance заказа: описываем и контролируем



VERA – DQ движок

Vera

Self-service DQ framework Vera для декларативных Soda-проверок, локального запуска и Airflow-интеграции. В документации и пользовательских запросах также может встречаться как Veradq и «Вера».

Vera - self-service DQ framework для команд, которые хотят описывать проверки качества данных декларативно, запускать их в своем runtime и публиковать результат в стандартные каналы.

В старых материалах и пользовательских запросах `Veradq` и «Вера» означают тот же фреймворк Vera.

Куда идти

роль	РАЗДЕЛ
Пишешь DQ-проверку, YAML, alerts или расписание	Для авторов DQ-проверок
Меняешь код пакета <code>veradq</code> , CLI, Airflow adapter или publishers	Для разработчиков фреймворка
Разбираешься, что это за система	Обзор
Чинишь ошибку запуска или валидации	FAQ / Troubleshooting

VERA: одна проверка в разных рантаймах



При all_success следующий шаг ждёт успешной DQ-задачи

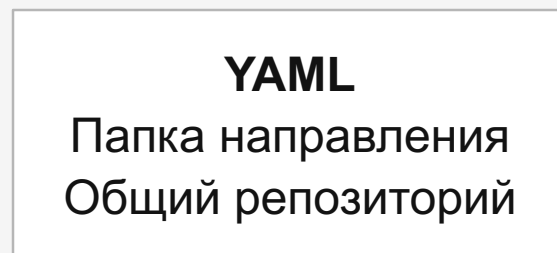
VERA: общий Airflow и свой рантайм

Путь автора до регулярной проверки

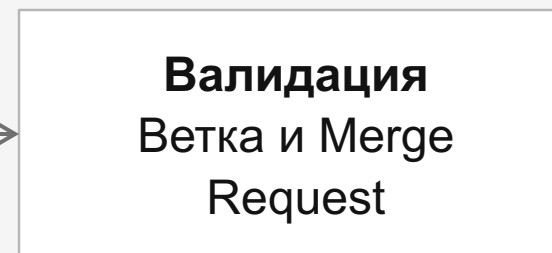
Общий Airflow

Готовая среда
исполнения

1. Подготовить



2. Проверить

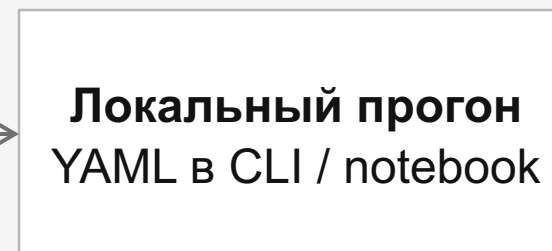
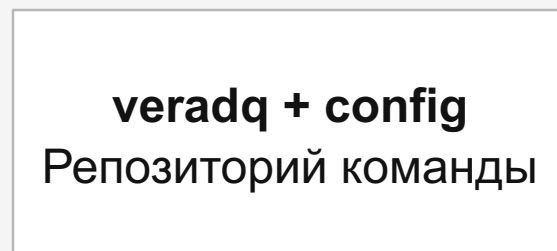


3. Подключить



Свой рантайм

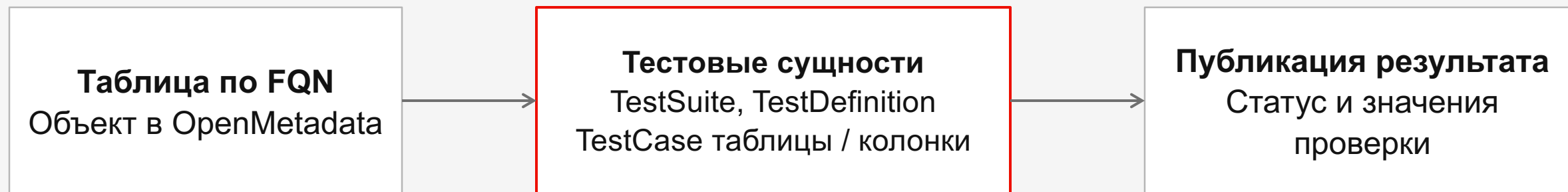
Команда владеет
исполнением



Skills для автора: YAML, локальная валидация и Airflow-интеграция

VERA: проверка и результат сразу в OMD

Результат готов. Recording настроен, таблица уже есть в OMD.



Недостающие тестовые сущности создаются автоматически

История в ClickHouse

История результатов
С OMD — FQN таблицы и TestCase

Реакция по статусу

Tag / Email / Telegram
Jira: задача или комментарий

dbt run_results.json использует те же publishers. Запуск dbt остаётся в проекте команды.

VERA: Покрытие заказов проверками

checks/orders.yml :

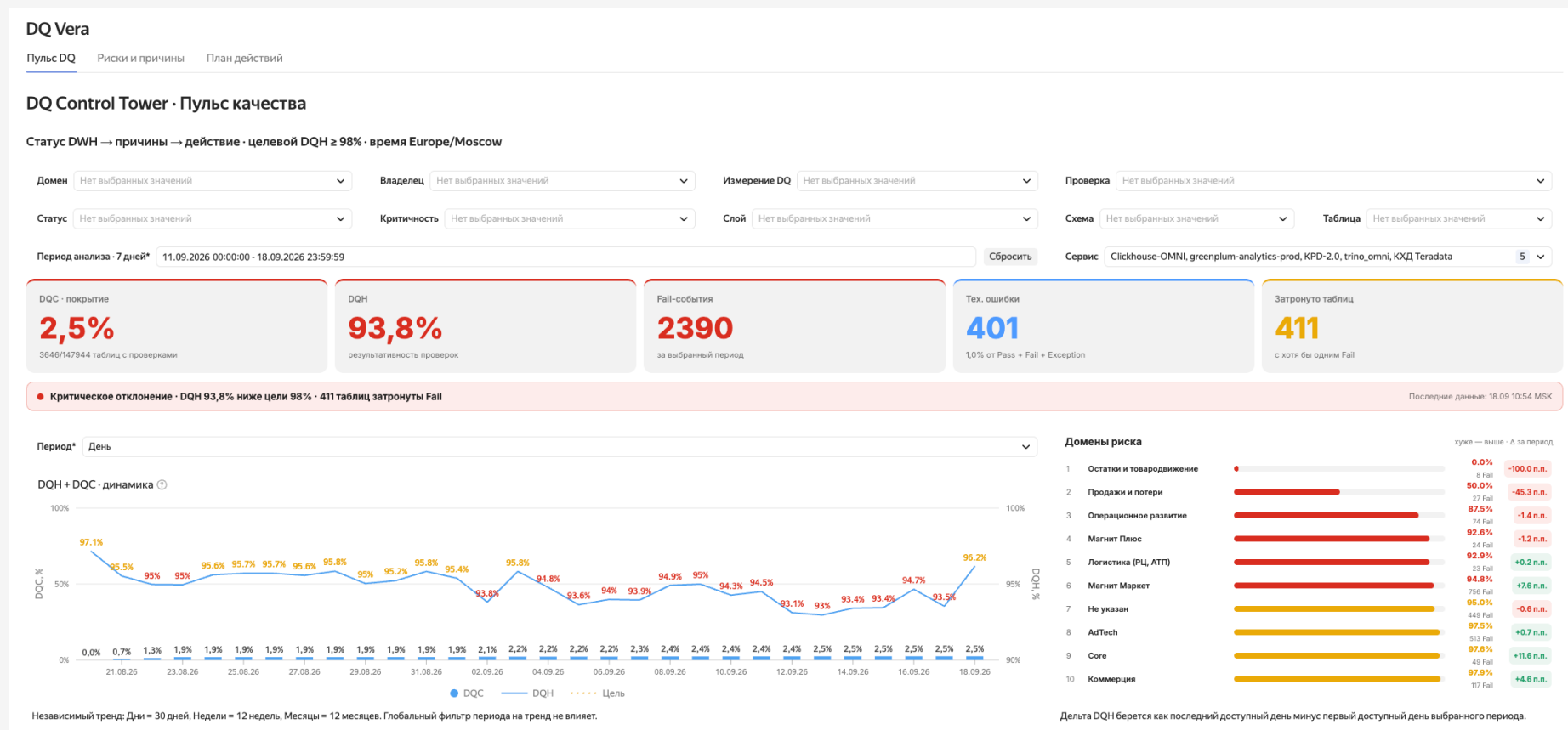
```
models:
  - name: orders
    schema: dm
    tier: 1
    domain: ecommerce
    tags: ["dq", "orders"]
    description: DQ checks for orders mart
    dag:
      schedule: "0 8 * * *"
      catchup: false
      af_pool: dq_pool
    tests:
      soda:
        data_source: trino_prod
        alerting:
          - statuses: ["failed", "warning"]
            alert_msg_block_set: standard
            tag:
              config: dq_ops
              channel: dq-alerts
        recording:
          - statuses: ["success", "failed", "warning"]
            clickhouse: dq_reports
      checks:
        checks for dm.orders:
          - row_count > 0:
              name: orders table must not be empty
          - missing_count(order_id) = 0:
              name: order_id must be filled
          - duplicate_count(order_id) = 0:
              name: order_id must be unique
```

VERA: использование и свежесть результатов

3646 активных проверок

5000+ таблиц под контролем

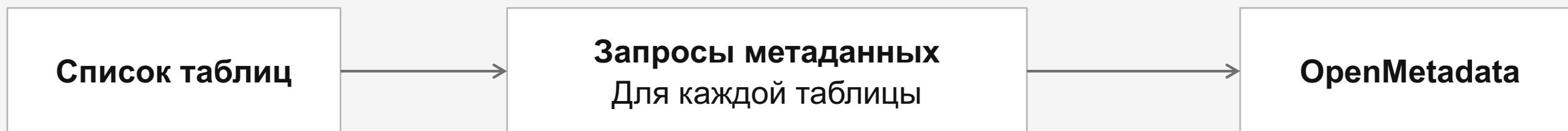
18 команд-авторов



Результаты по доменам, нарушения, ошибки исполнения и свежесть проверок

Greenplum и Teradata: пакетный сбор по схеме

Было - обход по таблицам



Стало - пакетирование конкретного вида метаданных



Меньше обращений к источнику. Устранение нагрузки и ускорение опроса метаданных

Greenplum и Teradata: Ускорение опроса

Database Services > greenplum-analytics-prod

greenplum-analytics-prod

Запустить Автопилот

Домены

Владельцы

Критичность

--

stf_artemov_so

--

Статистика

Базы данных 1

Агенты 3

Соединение

Развёртывание завершено

38 557 849

3 из 3 агентов завершено · всё в порядке

объектов загружено

ошибки

Агенты метаданных

Извлекайте метаданные, использование, происхождение и профилирование из источника.

+ Добавить агента

Metadata Agent

Метаданные

At 10:00 AM, Only on monday, wednesday, and friday

Успешно

38 557 объектов · завершено 1 день назад

Недавние выполнения

Показать историю выполнений

2 ошибки

92 предупреждения

Логи

Запустить

greenplum-analytics-prod_dbt_VTH2rEHO

Формирование

At 01:00 AM, Every day

Успешно

18 524 модели · завершено 2 часа назад

Недавние выполнения

Показать историю выполнений

847 ошибки

7783 предупреждения

Логи

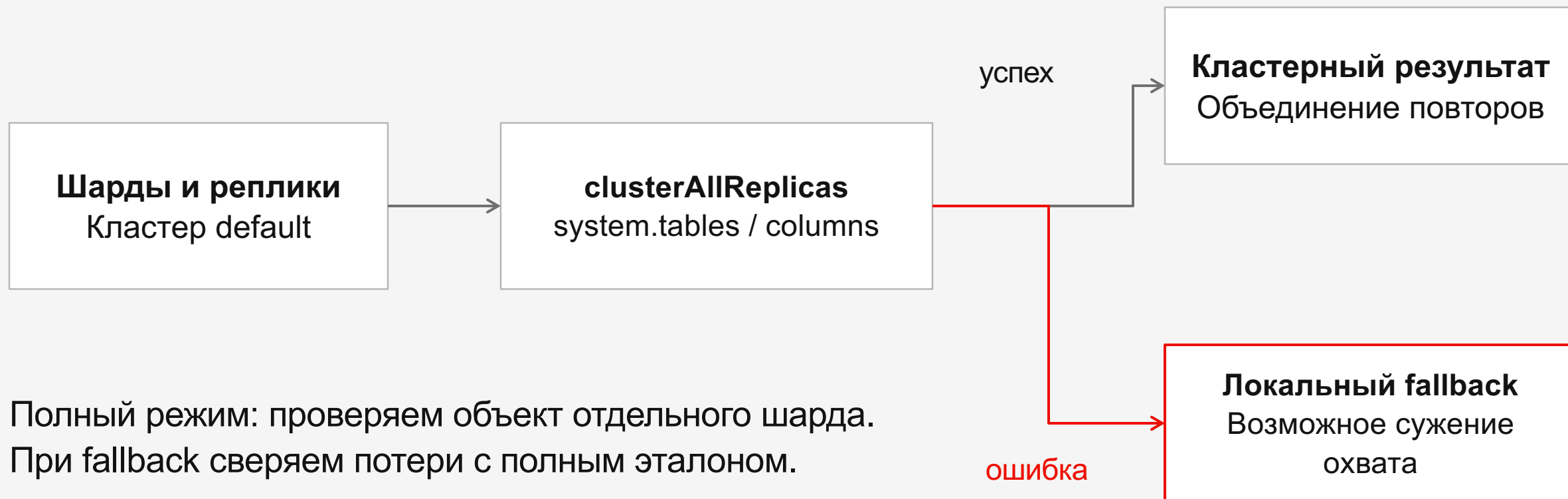
Запустить

X 10

10+ Часов → 1 Час

Ежедневный опрос GreenPlum на 39.000 объектов

ClickHouse: охват метаданных и резервный сбор



Полный режим: проверяем объект отдельного шарда.
При fallback сверяем потери с полным эталоном.

Дедупликация метаданных не проверяет согласованность схем между шардами

DataLens: Происхождение заказа от источника до отчета



Datalens Tander prod > Заказы

Заказы

nbbprqu0vud448

Домены

Владельцы

Критичность

Сертификация

Типы объектов «Дашборд»

Проект

--

--

--

--

Dashboard

Сектор нестандартных обращений / Поиск

Использование

0-й процентиль

Детали

Лента активности и задачи 0

Происхождение

Контракт

Дополнительные свойства

Найти объект «Происхождение»

Происхождение

Анализ воздействия

Всё время

oms_service_cart

Clickhouse-OMNI > default > dm_nrt

Table

12 Столбцы

oms_service_oms_order

Clickhouse-OMNI > default > dm_nrt

Table

22 Столбцы

1985646021966827109...

Datalens Tander prod

Dashboarddatamodel

24 Столбцы

Заказы

Datalens Tander prod

Заказы

Datalens Tander prod

Dashboard

1 Диаграммы

OMD: подключения, покрытие и использование

30+ prod-подключений

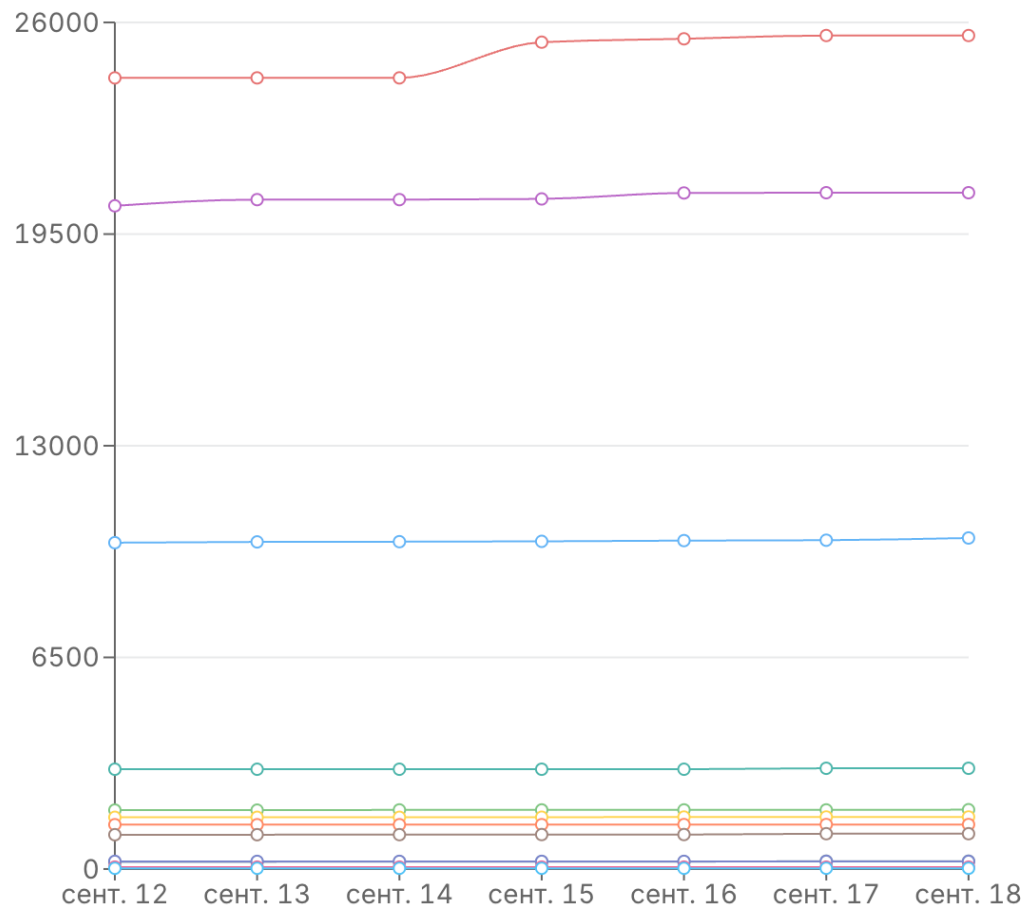
69550 таблиц

35% покрытия описанием

140+ DAU каталога

Все объекты данных

Количество объектов данных разных типов.



Все Объекты

+3.06%

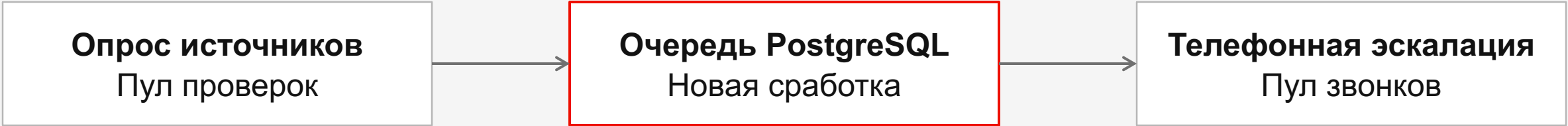
69550 изменено 7 дней назад

Q Search...

Tables	25598	
Charts	20771	
Pipelines	10167	
Stored Procedures	309	
Topics	1822	
Dashboard Data Models	160	
Dashboards	1368	
Database Schemas	108	
Metrics	242	
Containers	235	
Data Products	58	

Notifica: сигналы и правила эскалации

Источник	Что запускает эскалацию	Как управляем повторами
TAG	Новые сообщения по фильтру	Дедупликация по post ID
Grafana	Нарушение порога панели	Новый цикл после нормы
Airflow	Нет success / запуска к сроку	Повторы в пределах дня
OpenMetadata	Новый незакрытый DQ-инцидент	Одна цепочка на проверку



Дежурный определяется по календарю в момент звонка

Notifica: настройка силами команды

1. Команда

Календарь и роли

2. Подключение

Доступ к источнику

3. Мониторинг

Условие и эскалация

4. Ручная проверка

Без звонка

MAGNIT TECH
NOTIFICA

Обзор

Мониторинги

Команды

Сотрудники

Подключения

Журнал сработок

Статус сервиса

SWAT > Дежурства

КОМАНДНЫЙ КАЛЕНДАРЬ

Дежурства · SWAT

Один дежурный на смену. Все даты и время — MCK.

Настроить цепочку

Разовое назначение

ОСНОВНОЙ РЕЖИМ

Цепочка дежурных

С 18.09.2026 17:00 MCK

1. Михаил Филимонов · 48 ч → 2. Илья Кулиев · 48 ч → 3. Антон Ребров · 48 ч → 4. Никита Царёв · 48 ч → повтор

Разовые назначения необязательны: они заменяют дежурного на указанное время, не сдвигая цепочку. Остановка цепочки сохраняет разовые назначения и историю.

Октябрь 2026

Пн	Вт	Ср	Чт	Пт	Сб	Вс
28 Михаил Филимонов 28.09 00:00 — 28.09 17:00 По цепочке Илья Кулиев 28.09 17:00 — 30.09 17:00 По цепочке	29 Илья Кулиев 28.09 17:00 — 30.09 17:00 По цепочке	30 Илья Кулиев 28.09 17:00 — 30.09 17:00 По цепочке Антон Ребров 30.09 17:00 — 02.10 17:00 По цепочке	1 Антон Ребров 30.09 17:00 — 02.10 17:00 По цепочке	2 Антон Ребров 30.09 17:00 — 02.10 17:00 По цепочке Никита Царёв 02.10 17:00 — 04.10 17:00 По цепочке	3 Никита Царёв 02.10 17:00 — 04.10 17:00 По цепочке	4 Никита Царёв 02.10 17:00 — 04.10 17:00 По цепочке Михаил Филимонов 04.10 17:00 — 06.10 17:00 По цепочке
5 Михаил Филимонов	6 Михаил Филимонов	7 Илья Кулиев	8 Илья Кулиев	9 Антон Ребров	10 Антон Ребров	11 Никита Царёв

MAGNIT TECH
NOTIFICA

Обзор

Мониторинги

Команды

Сотрудники

Подключения

Журнал сработок

Статус сервиса

Не рассчитан dds cheque_ea

Описание мониторинга не добавлено.

Приостановить

Настроить

ИСТОЧНИК СИГНАЛА

Airflow

Норма

Периодичность

Каждые 5 минут

Следующая проверка

19.09.2026 04:20:00 MCK

Подключение

Airflow DWH OMNI

Airflow API

https://airflow-dwh.prod.analytics.corp/api/v1 ↗

Дедлайн

04:20 MCK, ежедневно

На что звонить

Отсутствие Success

Повтор без подтверждения

Через 15 минут

DAG ID

dds_cheque_ea

DAG в Airflow

Открыть DAG ↗

Проверка

DAG

Последняя проверка

18.09.2026 04:20:00

Последнее значение

dag:success

Проверить сейчас без звонка

ДОСТУП

Команды

1

DW DWH M+

2 участников

Настройка календаря дежурств в командах

FQN автоматически извлекается из URL

Notifica: подтверждение и хронология

Без подтверждения сервис продолжает карту эскалации

MAGNIT TECH
NOTIFICA

Обзор

Мониторинги

Команды

Сотрудники

Подключения

Журнал сработок

Статус сервиса

Моя карточка

М

Михаил Фили...

filimonov_mia@magnit...

ПРОХОЖДЕНИЕ МОНИТОРИНГА

Хронология сработки

Обновить

Время фиксации этапов — МСК (UTC+3). Ответ и нажатая кнопка фиксируются при получении статуса от шлюза телефонии.

12:57:24

16.09.2026

Запущен мониторинг

Начато чтение канала TAG.
Реакция в TAG: Отправлено · 12:57:29 МСК

12:57:24

16.09.2026

Выявлено нарушение

Канал: <https://tag.magnit.ru/data-management/channels/alerts-dwh-critical> · Сообщение содержит: KubePodNotReady · Новых сообщений: 1

12:57:24

16.09.2026

Карта эскалации поставлена в очередь

Ожидает свободного обработчика звонков.

12:57:29

16.09.2026

Запущена карта эскалации

Звонок по карте эскалации. Порядок звонков: Никита Царёв

12:57:29

16.09.2026

Звонок сотруднику

Никита Царёв
Сотрудник №1 · круг 1 · попытка 1
Ответ в TAG: Отправлено · 12:57:30 МСК

12:58:12

16.09.2026

Звонок принят

Никита Царёв. Шлюз сообщил ANSWERED.
Сотрудник №1 · круг 1 · попытка 1
ID звонка: 31656
Ответ в TAG: Отправлено · 12:58:14 МСК

12:58:12

16.09.2026

Подтверждение: нажата 1

Никита Царёв. Нажата 1. Обзвон по этой карте прекращён.
Сотрудник №1 · круг 1 · попытка 1
ID звонка: 31656

РЕЗУЛЬТАТ

Действия

1

Звонок по карте эскалации

Начало: 16.09.2026 12:57:29 МСК
Finished: OK

Никита Царёв 16.09.26 в 11:41 AM

KubePodNotReady

👁 1

3 ответа

Notifica БОТ 16.09.26 в 12:57 PM

📞 Звоню: Никита Царёв.

Notifica БОТ 12:58 PM

✅ Дозвонился: Никита Царёв.

Notifica БОТ 12:58 PM

✅ Обработка завершена.

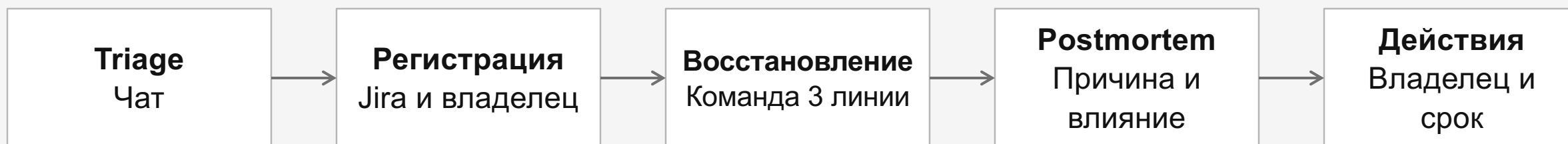
Открыть мониторинг «Квоты облака»

Восстановление и закрытие инцидента выполняет ответственная команда

MAGNIT
TECH

14

Инцидент: владелец, восстановление и действия



Инфраструктура

Движок / фреймворк

Данные / правило

DevOps / DataOps

Команда платформы

Владелец данных

S0 / S1: postmortem за 2 рабочих дня после закрытия, если не согласовано иное.

Интересно ли в это инвестировать в 2026?

НЕТ

В 2026 интересно инвестировать в AI

Rhetorica

+

СЕГОДНЯ

◇ Помоги подготовить недельный

◇ Помоги подготовить недельный

РАНЕЕ

◆ Расскажи динамику заказов в М

◇ Безалкогольное пиво стало сам

МИХАИЛ ФИЛИМОНОВ 11:55

Расскажи динамику заказов в Москве за неделю

RHETORICA ASSISTANT 11:55

✓ Answer ready

SYSTEM LOG 1 tool call

EXECUTIVE SUMMARY

Количество завершённых заказов в Москве за неделю 03.08.2026 — 09.08.2026:

Дата	Заказы
2026-08-03	
2026-08-04	
2026-08-05	
2026-08-06	
2026-08-07	
2026-08-08	
2026-08-09	

Итого за неделю: заказов.

Задайте вопрос о метрике, тренде, аномалии или причине проблемы. Rhetorica сохранит SQL и весь процесс выполнения запроса на виду

Rhetorica

МФ Михаил Филимо...
АНАЛИТИК

Rhetorica: СМЫСЛ ДАННЫХ В ОТВЕТЕ АГЕНТА

Подготовка	Что нужно агенту и команде
Смысл данных	Определения метрик, зерно витрины, связи, фильтры и исключения
Контекст	Каталог через MCP и синхронизированная документация
Методика	Skill с порядком работы и понятным условием применения
Проверка	Ответы только по данным с положительным DQ

Польза descriptions видна в ответе на вопрос бизнеса

Аналитик уточняет определения. Агент получает обновлённый контекст.

Rhetorica: создание агента

1. Коннекторы

LLM, DWH, Docs, MCP

Add agent

A cross-user chat profile with its own prompt, LLM preference, and connector allowlists

Slug

Title

Description

LLM connector

Default

deepseek-v4-flash

MAGNIT_GPT_COMMERCE

MAGNIT_KIMI

magnit-gpt

Active

SQL

Temperature

Max tokens

Database

clickhouse

clickhouse-monitoring

greenplum

impala-prod

m_cube

trino

trino-chd-commerce

trino-chd-data-governance

Documentation

platform_docs

2. Профиль

Инструкция, SQL, skills

Add agent

A cross-user chat profile with its own prompt, LLM preference, and connector allowlists

Database

clickhouse

clickhouse-monitoring

greenplum

impala-prod

m_cube

trino

trino-chd-commerce

trino-chd-data-governance

Documentation

platform_docs

GitLab

gitlab-offline

gitlab-omni

MCP

mcp-confluence-gk

mcp-confluence-omni

mcp-gitlab-offline

mcp-gitlab-omni

mcp-jira-corp-tander

mcp-jira-omni

mdm

openmetadata-mcp

Airflow

airflow-dq-prod

airflow-dwh-prod

Kubernetes

k8s-prod

VictoriaMetrics

No active connectors

3. Проверка

Задача и отказ

4. Каналы

Web и привязка TAG

Add agent

A cross-user chat profile with its own prompt, LLM preference, and connector allowlists

Chat skills

Аналитические запросы Магнит

Для любого аналитического запроса сначала прочти этот skill отдельным вызовом read_workspace_skill и дождись результата; не запускай параллельно OpenMetadata, SQL или profiling.

MDM MCP – рабочий протокол

Любые запросы, где нужно использовать MCP-коннектор mdm для навигации, поиска и анализа метаданных, записей справочников или зависимостей IBM Product Master.

OMNI vs Offline Routing

Любой вопрос, где нужно выбрать контур, базу, коннектор или источник данных.

GitLab OMNI Access Flow

Вопросы про доступы юзеров, ACL, grants, roles, owners, DAG/dbt implementation, GitLab repo docs/config/code.

Greenplum Trino OMNI SQL

OMNI SQL в Greenplum или Trino.

Mplus Loyalty Canonical

M+, mplus, loyalty, m_cube, бонусы, программы лояльности.

МДМ НСИ Общие вопросы – маршрутизация на Confluence

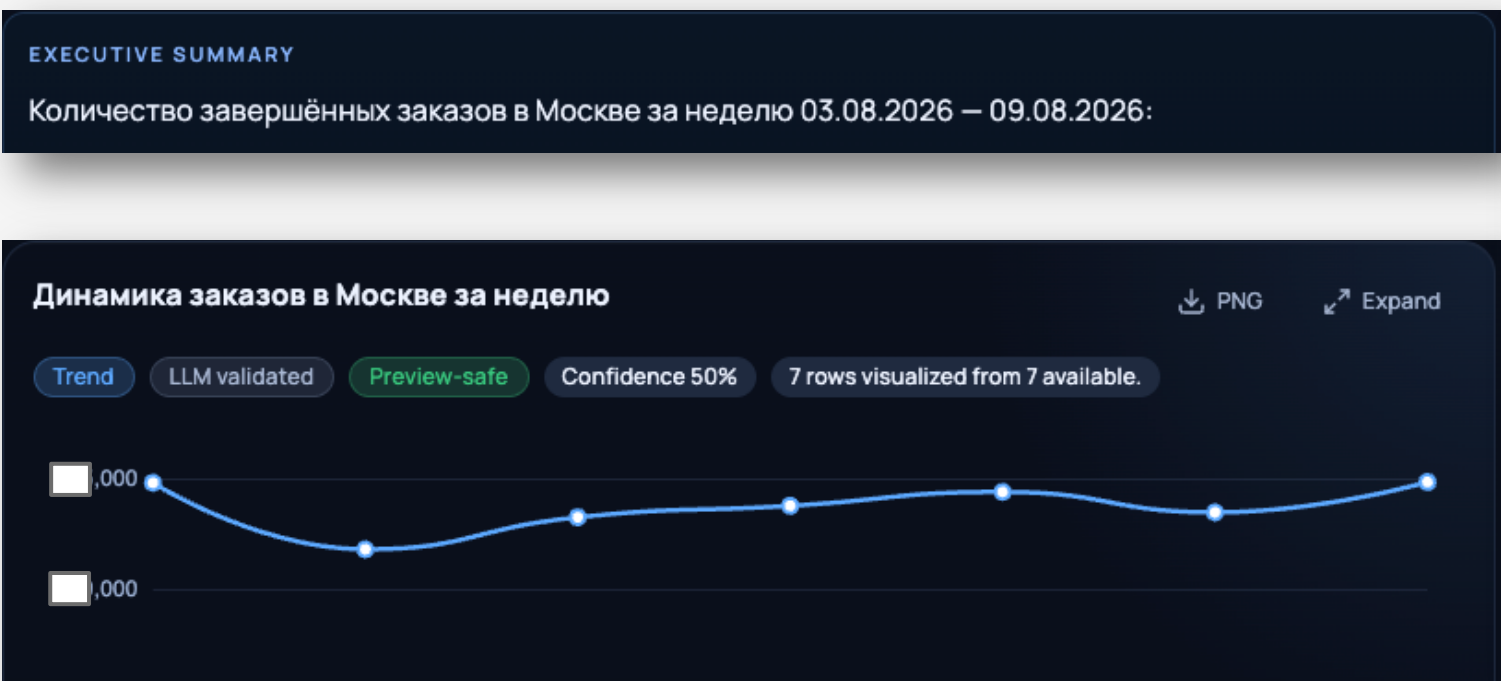
Вопросы по НСИ, МДМ, мастер-данным, справочникам, классификаторам, ServDB, PIM, ОУНСИ, ОСКУС, кодировке, репликации, интеграциям МДМ, состояниям подразделений, атрибутам товаров, операциям

Admin prompt

Rhetorica: недельный обзор заказов

«Расскажи динамику заказов в Москве за неделю»

Сводка для обсуждения с командой



SQL

EXECUTED SQL

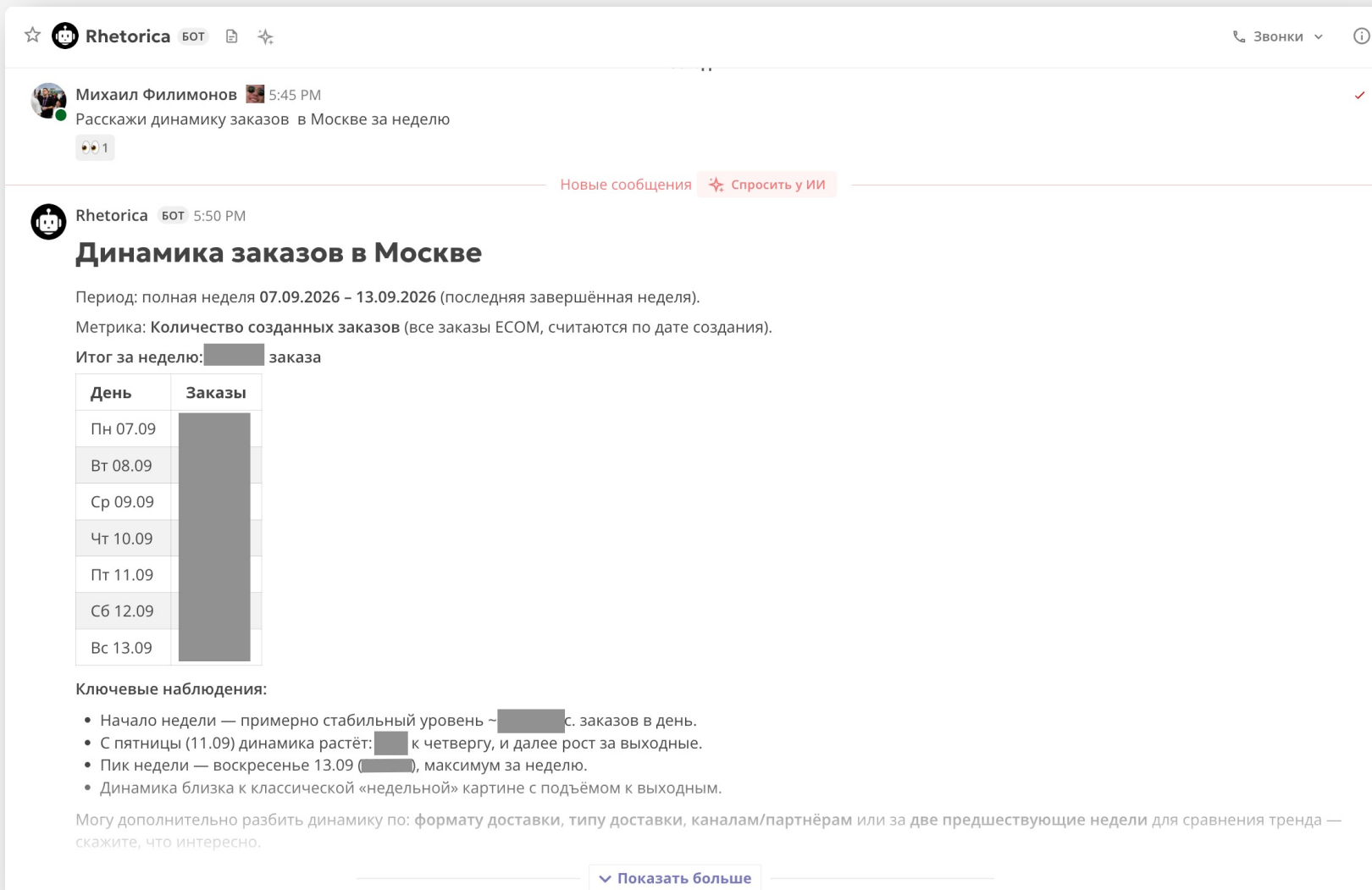
```
SELECT report_date AS dt, sum(orders) AS completed_orders FROM dm.orders_
```

ORIGINAL SQL DRAFT

```
SELECT
    report_date AS dt,
    sum(orders) AS completed_orders
FROM dm.orders_daily_report
WHERE report_date BETWEEN toDate('2026-08-03') AND toDate('2026-08-09')
    AND group_city = 'Moscow'
GROUP BY dt
ORDER BY dt
```

В том же диалоге можно: уточнить разрез или период, выгрузить результат

Rhetorica: расчёт в обсуждении TAG



ВЫЗОВ

Личное сообщение
или @бот в канале

Продолжение

@бот с уточнением
в том же треде

Выдача

Текст и PNG графика SQL по настройке бота

Бот использует профиль выбранного агента

Что изменилось и как оценить эффект

Область	Q1 2026	Q3 2026	Доказательство эффекта
Создание DQ	DQ-аналитики Около 3 дней	YAML и единый runner Публикации в общем контракте	Время сейчас – час Доля self-service - 100%
Развитие каталога	Informatica Доработки недоступны	OMD 2.0 и RDF Свои сканеры и UI	100% владение инструментами
Поставка DQ	Релизы через переменные Airflow	Версионированный пакет CI и Artifactory	3.5 тысячи проверок за квартал
Работа с данными	Написание SQL	Rhetorica SQL и проверяемый ответ	Мультимодальный AI чат для работы с данными без кода

Сценарий: недельный обзор заказов

«Как менялись завершённые заказы в Москве за неделю?»

Результат: дневная динамика и сводка для обсуждения с командой

- | | | |
|----|----------|---|
| 01 | Данные | Витрина заказов и владелец показателя |
| 02 | Смысл | В OMD: статус заказа, дата, регион и исключения |
| 03 | Качество | VERA на Soda и маршрут реакции на нарушение |
| 04 | Агент | Rhetorica: ответ, график и сверка с эталонным SQL |

Описание витрины становится частью ответа бизнесу

Начать можно с одного показателя и одного проверенного сценария