

Как мы с минимальными затратами создали каталог данных над хранилищем

Magnit Data

2023



Мы занимаемся всем, что связано с понятностью данных, развитием инструментов и сервисов, которые помогают бизнес-пользователям быстрее найти и получить нужные данные в корпоративной платформе, сориентироваться в нюансах их использования и происхождения.

Олег Молчанов

Руководитель направления по метаданным



Масштаб платформы данных

>1 ПБ

Данных

10 000

Пользователей отчетов
и ad-hoc аналитики

200+

Систем источников

Данные в «Магните»

420+ Тб

Данных

125 млн

Запросов в месяц

2 000+

Уникальных пользователей
ежемесячно

Для них старались

>5 000

Витрин данных

О них речь и пойдет дальше

>80

Песочниц, предоставленных
бизнесу для прототипиро-
вания решений своих задач

Информация о КХД

Пользовательские сценарии

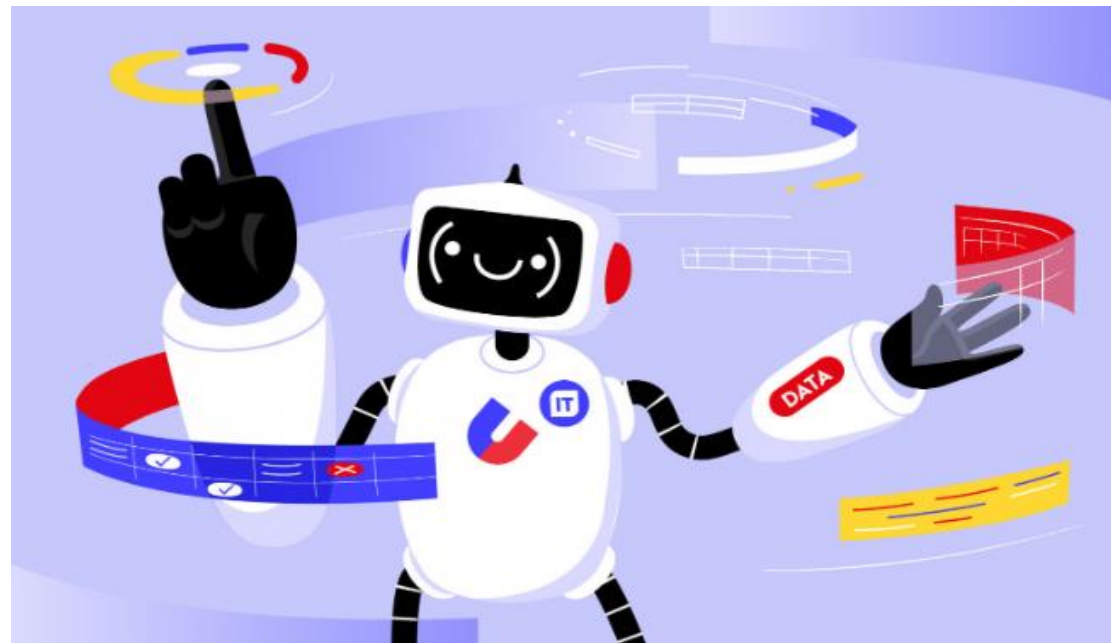
Найти

Понять

Получить

Научиться
пользоваться

Следить за
изменениями



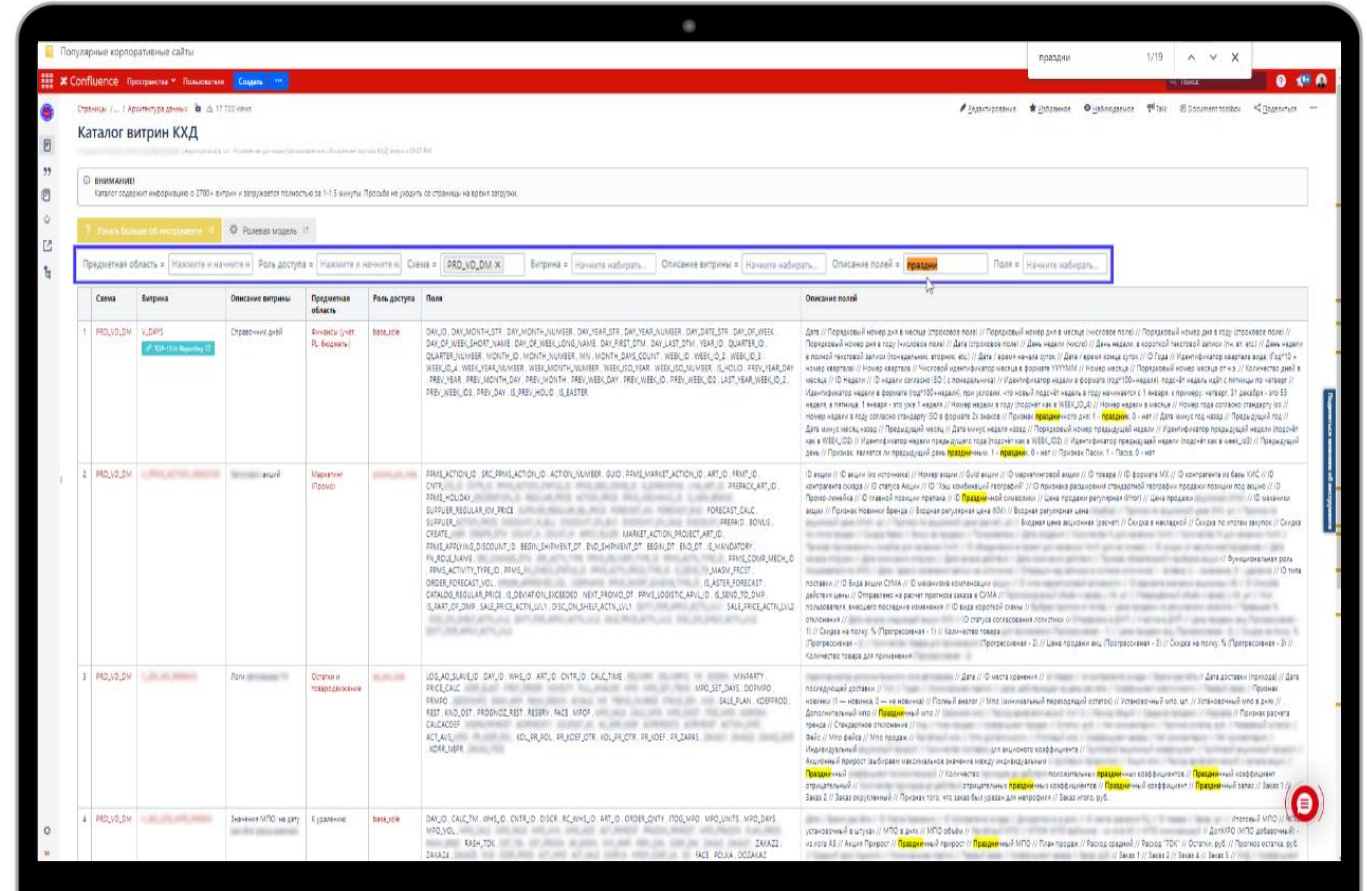
Основной кейс

Быстро найти данные, которым можно доверять

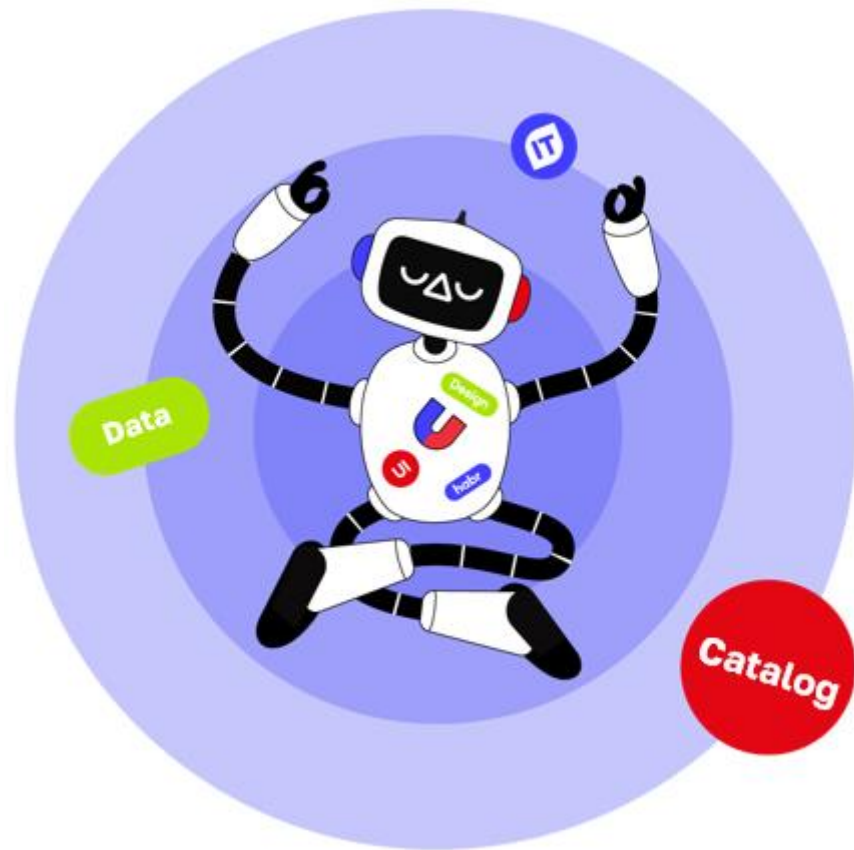
1. Поиск ВИТРИН по основным метаданным:

- имя и описание витрины
- поля витрины и их описания
- предметная область
- тэги

2. Релевантная поисковая выдача



На чём сделали каталог?



фронтенд на Confluence



бэкенд с метаданными



«Python-магия»

Паспорт витрин. Поля

Страницы / ... / PRD_VD_DM 620 views

PRD_VD_DM - V_DAYS

Создатель: икс: Управление данными (автоматическое обновление портала КХД), отредактировано 29 окт 2022

Справочник дней

Финансы (учет, PL, бюджеты) 15

TOP-10 in Reporting 15

Поля Происхождение данных Использование

Если необходимо найти все витрины, которые содержат заинтересовавшее вас поле, воспользуйтесь [Каталогом витрин КХД](#) или чат-ботом MagnitData на этой странице.

№	Поле	Описание поля	Тип данных
1	DAY_ID <small>UP</small>	Дата	DATE
2	DAY_MONTH_STR	Порядковый номер дня в месяце (строковое поле)	VARCHAR
3	DAY_MONTH_NUMBER	Порядковый номер дня в месяце (числовое поле)	INTEGER
4	DAY_YEAR_STR	Порядковый номер дня в году (строковое поле)	VARCHAR
5	DAY_YEAR_NUMBER	Порядковый номер дня в году (числовое поле)	INTEGER
6	DAY_DATE_STR	Дата (строковое поле)	VARCHAR
7	DAY_OF_WEEK	День недели (число)	INTEGER
8	DAY_OF_WEEK_SHORT_NAME	День недели, в короткой текстовой записи (пн, вт, etc.)	VARCHAR
9	DAY_OF_WEEK_LONG_NAME	День недели в полной текстовой записи (понедельник, вторник, etc.)	VARCHAR
10	DAY_FIRST_DTM	Дата / время начала суток	TIMESTAMP
11	DAY_LAST_DTM	Дата / время конца суток	TIMESTAMP

Паспорт витрины. Происхождение данных

Поля [Происхождение данных](#) [Использование](#)

Схема	PRD_VD_DM
Реестры требований	PT витрины PRD_VD_DM

ГЛУБИНА ХРАНЕНИЯ
Неизвестно
* для витрин, по которым данные вынесены в Hadoop

> Источники данных витрины в КХД 1-го уровня (1):

> Скрипт витрины:

▼ Таблицы-источники

Таблица	Размер, Гб	Кол-во строк	Дата обновления данных	Репликация	Процедуры	Потоки данных	Универсальный стейдж
ESB Корпоративная сервисная шина, Система управления и планирования							
PRD_DB_DM	0.61	810 тыс.	2023-04-01 18:33:06	TVA, DEV	<ul style="list-style-type: none">prddb_dm.prc_ [UPDATE] [MERGE]	<ul style="list-style-type: none">prdpc / MDM_MAPPINGS / WF_LOAD_MDM_FULLprdpc / WHS_MAPPINGS / WF_ALL_MAPPINGSprdpc / WHS_MAPPINGS / WF_DAILY_ETLprdpc / WHS_MAPPINGS / WF_LOAD_WHS_ARTICLE	<ul style="list-style-type: none">CDM.C05_ART_ATTRCDM.C05_ART_ATTR_NEWCDM.C05_ART_GRP_TAGCDM.C05_ART_UOMCDM.C33_ART_UOMLDV.T_ART_HLDV.T_ART_MOREINFO_HLDV.T_SRC_SYSTEM_HLDV.T_WHS_TYPE_HODS.XML05_ARTICLE

Если необходимо более детально разобраться в происхождении данных, то воспользуйтесь витриной [V_INFA_INFO](#)

Получиться мнение об инструменте

Паспорт витрины. Использование 1/2



Поля Происхождение данных **Использование**

Присутствие на серверах	PROD TVA DEV
Категория информации	КОРПОРАТИВНЫЕ ДАННЫЕ <small>Положение о коммерческой тайне, Регламент категорирования информации ПАО "Магнит"</small>
Роль доступа	<code>base_role</code> <div><p>ⓘ Обратите внимание!</p><ul style="list-style-type: none">Узнать, какие роли у вас уже есть, или запросить доступ к новой роли можно через чат-бота MagnitData на этой страницеРазобраться в ролевой модели можно здесь</div>


> Используется в отчетности (26):

> Витрина используется в объектах КХД (105):

Оценки пользователей

Choices	Your Vote	Current Result: (2 Total Votes)	Voters
☆☆☆☆	<input checked="" type="checkbox"/>	<div style="width: 100%;"><div style="width: 100%;"></div></div> 2 Votes , 100%	 ✕
☆☆☆☆	<input type="checkbox"/>	<div style="width: 0%;"><div style="width: 0%;"></div></div> 0 Votes , 0%	
☆☆☆	<input type="checkbox"/>	<div style="width: 0%;"><div style="width: 0%;"></div></div> 0 Votes , 0%	
☆☆	<input type="checkbox"/>	<div style="width: 0%;"><div style="width: 0%;"></div></div> 0 Votes , 0%	
☆	<input type="checkbox"/>	<div style="width: 0%;"><div style="width: 0%;"></div></div> 0 Votes , 0%	
All voters	<input checked="" type="checkbox"/>	<div style="width: 0%;"><div style="width: 0%;"></div></div>	 ✕

Посетить меню об инструменте



Паспорт витрины. Использование 2/2

① СТАТИСТИКА ВИТРИНЫ

- обращений пользователей непосредственно к витрине: Prod - 1 042 378; TVA - 134 630;
- уникальные пользователи: Prod - 2 210; TVA - 274;
- › [Топ-10 пользователей \(Prod\)](#)
- › [Топ-10 пользователей \(TVA\)](#)
- суммарно обращений с учётом дочерних витрин: Prod - 2 241 938; TVA - 1 334 190;
- уникальные пользователи с учётом дочерних витрин: Prod - 2 277; TVA - 2 275;

* за период 90 дней (📅 01.01.2023 - 📅 01.04.2023)

① Примечание

1 Краткое описание:

Справочник товаров, копия справочника товарных позиций ОУНСИ.

История изменений: [ссылка]

История изменений: [ссылка]

[Период обновления]

Ежедневно

▼ Обращения пользователей (2)

🔍 DATAGOV - [ссылка] [vi] ЗАКРЫТ

? DATAGOV - new Запрос на информацию from [ссылка] ЗАВЕРШЕН

› Все роли доступа (11)

👍 Нравится Станьте первыми кому понравится это



Поделись мнением об инструменте

Отклик аудитории

1000+

MAU Каталога
витрин КХД

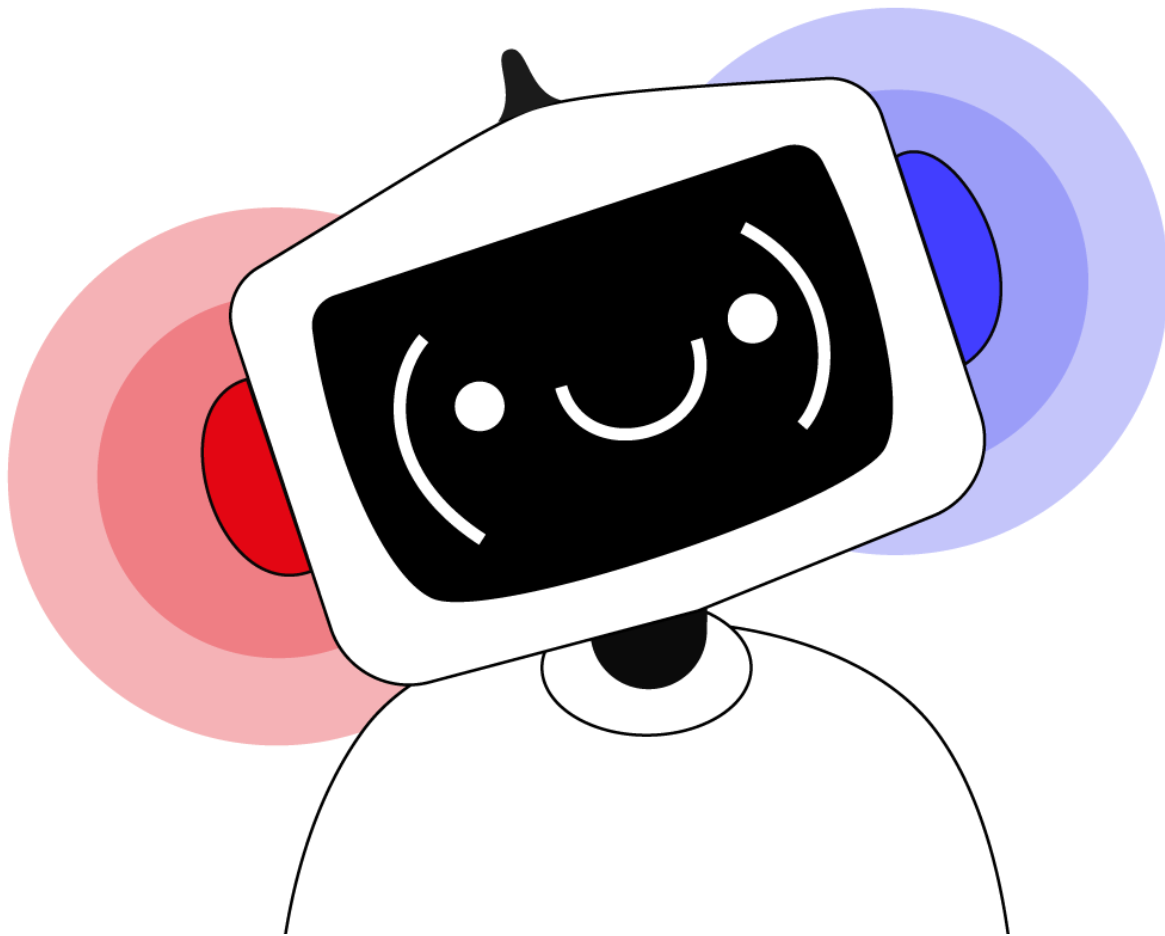
1500+

визитов
пользователей
ежедневно

450+

MAU стартовой
страницы

Вопросы-ответы



Блог Магнита на [Хабр](#)



@MOLCHANOV_OE

МОЛЧАНОВ ОЛЕГ

Руководитель направления по метаданным

+7 (995) 110 08 12

molchanov_oe@magnit.ru