

Как мы обеспечиваем self-service разработки и деплоя витрин в Авито

Николай Огоров
Big Data Engineer

Айк Оганесян
Big Data Engineer



SmartData
2025

Витрины

Витрина — это таблица в базе данных.

- Структура и способ обновления витрины описывается SQL-кодом.
- Код витрины хранится в git репозитории.
- Шапка витрины — словарь параметров запуска витрины в yaml формате.

Шапка витрины

```
1  /* @header
2  datamart: DMA.test_dm_nvogorov
3  type: FULL_REFRESH
4  volatile: true
5  keys:
6    - event_date
7  params:
8    - name: launch_id
9    - name: first_date
10 description: тестовая витрина
11 scheduled: false
12 unit: Jobs
13 engine: Vertica
14 */
```

DDL

```
1  CREATE TABLE IF NOT EXISTS
   dev_DMA.test_dm_nvogorov
2      (
3          launch_id int not NULL,
4          version_id int not NULL ENCODING RLE,
5          event_date date,
6          user_id int,
7          new_column1 int,
8          new_column2 int,
9          new_column3 int
10     )
11 order by event_date, user_id
12 unsegmented all nodes
13 partition by version_id;
```

DML

```
1  delete /*+direct*/ from
   dev_dma.test_dm_nvogorov;
2
3  insert into dev_DMA.test_dm_nvogorov select
   :launch_id, 1, :first_date, 1, 1, 1, 1;
```

Регламент

Регламент — ациклический граф, узлами которого являются витрины.

- Запускается раз в сутки
- Поочередно выполняет код всех витрин



Немного истории

- **12** лет нашему DWH
- **4500** пользователей
- **4500+** витрин
- Из них **900+** критичных
- За сутки обрабатывается **30.000+** тасок



Внесение изменений в регламент

- **900** критичных витрин
- **1000** важных витрин
- **2500+** минорных витрин
- **665** авторов витрин (аналитики, BI, DS и другие)
- TTM витрины от **1 минуты**
- Любой сотрудник, умеющий писать на SQL — наш пользователь



Что мы контролируем

- Опасный код
- Performance
- Окружение
- Ручные апрувы
- Изменения DDL

Я ВАМ

запрещаю

DROP DATABASE



Путь витрины до регламента



Работа через stash



Nikolay Ogorov 09 Jul 2025 [↗](#)

dwh test

Reply · Edit · Resolve 😊 ...

👁️ 1 ✅ 1



Nikolay Ogorov 08 Apr 2025 [↗](#)

dwh merge

Reply · Edit · Resolve 😊 ...

👁️ 1 ✅ 1



DWH-teamcity Bot 09 Jul 2025 [↗](#)

Наблюдать за процессом выполнения тестов можно в этом сообщении
Оно будет обновляться автоматически при обновлении статуса билдов

datamarts/dma_test_dm_nvogorov2.sql

[Fast Tests](#): SUCCESS

[Limit0](#): SUCCESS

[Perf Tests](#): SUCCESS

[Report Tests](#): SUCCESS



DWH-teamcity Bot 08 Apr 2025 [↗](#)

Изменения успешно замержены!

- Оцени пожалуйста работу сервиса по [ссылке](#)

Reply 😊 ...

Тестовый отчет

Результат теста

[Требования к витринам](#)
[Ссылка на PullRequest](#)

✔ В шапке должны присутствовать обязательные атрибуты и не должно быть дублирования	▼
✔ Значения параметров в шапке должны быть определенного для них типа.	▼
✔ Параметрам должно быть присвоено значение	▼
✔ Все параметры из шапки должны использоваться в коде	▼
✔ Параметр days_col должен соответствовать требованиям"	▼
✔ Параметр days_to_keep должен соответствовать требованиям	▼
✔ Проверка атрибута scheduled	▼
✔ Проверка заполненности поля engine	▼
✔ Проверка правильности заполнения параметра keys.	▼
✔ Проверка правильности заполнения параметра type для витрин.	▼
✔ Проверка sql синтаксиса кода витрины	▼
✔ Проверка DDL синтаксиса создания таблицы	▼
✔ Все имена колонок должны быть форматированы в snake_case	▼
✔ Для колонок типов varchar и numeric всегда должна быть указана размерность	▼
✔ Для колонок типов varchar и numeric размерность не должна быть слишком большой, типы Decimal с размерностью больше 37 запрещено использовать	▼
✔ В новых витринах launch_id должен быть первой колонкой	▼
✔ Колонки из ключа витрины должны существовать в таблице	▼
✔ Проверка безопасного обновления данных	▼
✔ Проверка таблиц, которые удаляются в коде	▼
✔ INCREMENTAL_REFRESH витрины должны быть партиционированы по date_col	▼
✔ Логика формирования витрины не должна зависеть от logical_category, vertical	▼
✔ В витринах нежелательно хранить атрибуты logical_category, vertical	▼
✔ В запросах должны использоваться только разрешенные схемы	▼

2025-10-03	08:54:15,859	::	INFO	::	run tests on recalc cluster
2025-10-03	08:54:15,860	::	INFO	::	BUILD_DURATION_FAIL: 3600
2025-10-03	08:54:15,861	::	INFO	::	BUILD_DURATION_WARN: 3200
2025-10-03	08:54:15,862	::	INFO	::	BUILD_DURATION_CONV: 60
2025-10-03	08:54:15,863	::	INFO	::	CPU_TIME_FAIL: 75000
2025-10-03	08:54:15,864	::	INFO	::	CPU_TIME_WARN: 72000
2025-10-03	08:54:15,864	::	INFO	::	CPU_TIME_CONV: 60
2025-10-03	08:54:15,865	::	INFO	::	DURATION_FAIL: 2100
2025-10-03	08:54:15,866	::	INFO	::	DURATION_WARN: 1800
2025-10-03	08:54:15,867	::	INFO	::	DURATION_CONV: 60
2025-10-03	08:54:15,916	::	INFO	::	Starting test CodeSyntax
2025-10-03	08:54:15,918	::	INFO	::	Test CodeSyntax completed
2025-10-03	08:54:15,947	::	INFO	::	Starting test DDLSyntax
2025-10-03	08:54:15,949	::	INFO	::	Test DDLSyntax completed
2025-10-03	08:54:15,980	::	INFO	::	Starting test CTASSyntax
2025-10-03	08:54:15,982	::	INFO	::	Test CTASSyntax completed
2025-10-03	08:54:16,015	::	INFO	::	Starting test TableName
2025-10-03	08:54:16,017	::	INFO	::	Test TableName completed
2025-10-03	08:54:16,049	::	INFO	::	Starting test VerticaComplexTypes
2025-10-03	08:54:16,051	::	INFO	::	Test VerticaComplexTypes completed
2025-10-03	08:54:16,087	::	INFO	::	Starting test VerticaFunctions
2025-10-03	08:54:16,088	::	INFO	::	Test VerticaFunctions completed
2025-10-03	08:54:16,120	::	INFO	::	Starting test PartitionExpressionFR
2025-10-03	08:54:16,121	::	INFO	::	Test PartitionExpressionFR completed
2025-10-03	08:54:16,153	::	INFO	::	Starting test PartitionFormat
2025-10-03	08:54:16,154	::	INFO	::	Test PartitionFormat completed
2025-10-03	08:54:16,185	::	INFO	::	Starting test PartitionCount
2025-10-03	08:54:16,187	::	INFO	::	Test PartitionCount completed
2025-10-03	08:54:16,217	::	INFO	::	Starting test FindDeletion
2025-10-03	08:54:16,219	::	INFO	::	Test FindDeletion completed
2025-10-03	08:54:16,253	::	INFO	::	Starting test CheckIncludingProjections
2025-10-03	08:54:16,254	::	INFO	::	Test CheckIncludingProjections completed
2025-10-03	08:54:16,288	::	INFO	::	Starting test VarcharNumericSizeExists
2025-10-03	08:54:16,289	::	INFO	::	Test VarcharNumericSizeExists completed
2025-10-03	08:54:16,319	::	INFO	::	Starting test VarcharNumericSizeNotLarge
2025-10-03	08:54:16,320	::	INFO	::	Test VarcharNumericSizeNotLarge completed
2025-10-03	08:54:16,351	::	INFO	::	Starting test Timestamptz
2025-10-03	08:54:16,353	::	INFO	::	Test Timestamptz completed
2025-10-03	08:54:16,385	::	INFO	::	Starting test ForbiddenFunctions
2025-10-03	08:54:16,386	::	INFO	::	Test ForbiddenFunctions completed
2025-10-03	08:54:16,419	::	INFO	::	Starting test AllowedFunctions
2025-10-03	08:54:16,420	::	INFO	::	Test AllowedFunctions completed

Опасный код

- Синтаксические ошибки
- Лишние удаления
- Плохие параметры в шапке
- Опасные функции
- Неоптимальное хранение или изменение данных



Все параметры из шапки должны использоваться в коде
В шапке есть неиспользуемые параметры: last_date.



Проверка sql синтаксиса кода витрины

При парсинге кода витрины возникли ошибки. Возможно где-то забыты "," или ";". Для ускорения прохождения тестов рекомендуем прогонять код витрины локально.

[Документация](#)

```
QUERY:
insert into dev_DMA.test_dm_nvogorov select1 :launch_id, 1,
:first_date, 1, 1, 1, 1
```

```
ERRORS:
line 1 pos 37: Illegal character 'select1'
```



Проверка таблиц, которые удаляются в коде

Добавьте параметры в код или удалите их из шапки. Проверка таблиц, которые удаляются в коде. Разрешено удаление только таблиц из DEV_TMP% схем

```
DMA.ALL_CONTACTS
```

Проверка перформанса

Результат теста

[Требования к витринам](#)

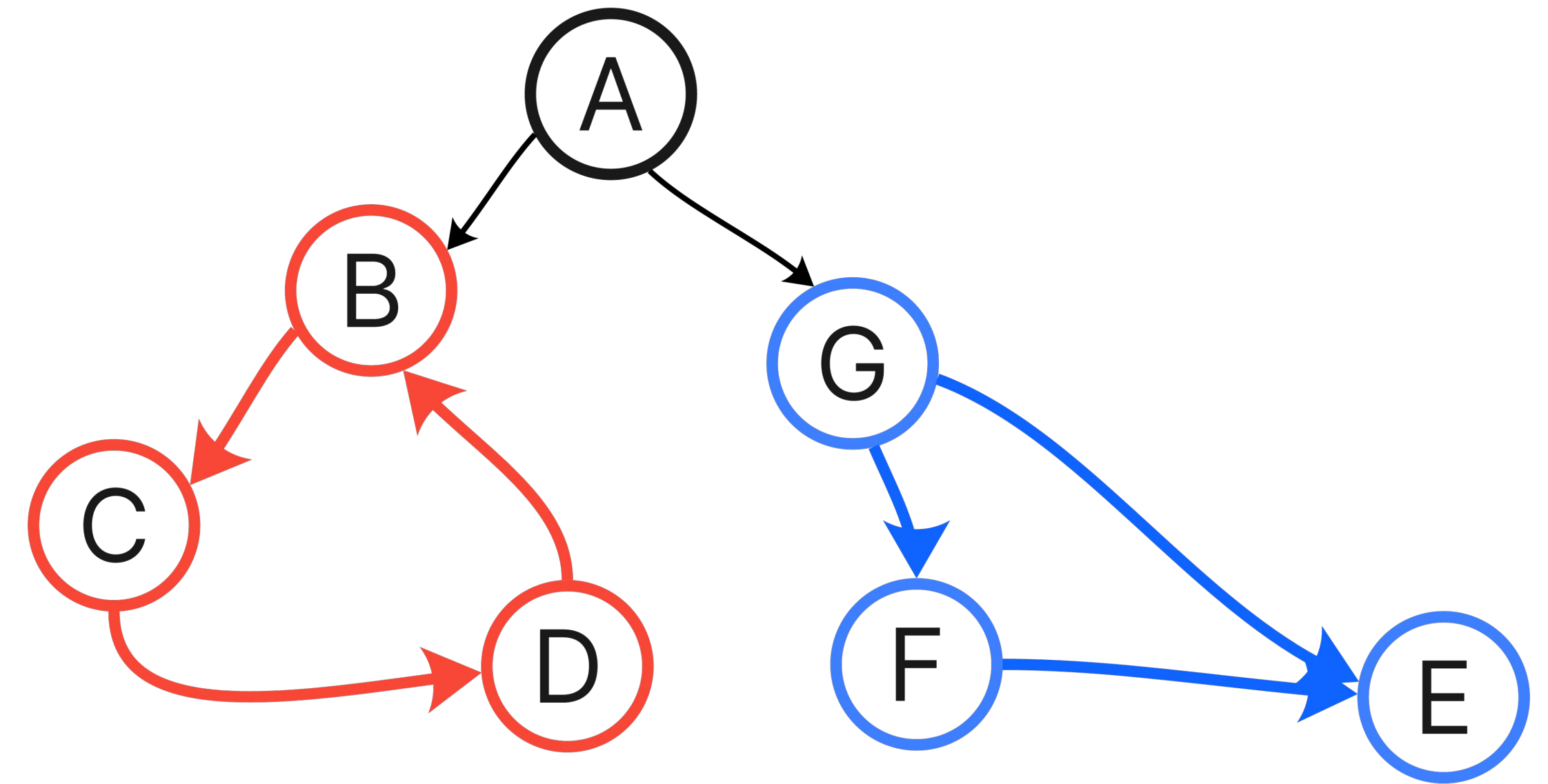
[Ссылка на PullRequest](#)

- ❌ Проверка чекеров витрины ▼
- ⚠️ Попадание метрик выполнения запроса в лимиты ▼
- ⚠️ Отчет со сравнением перф-метрик ^
Отчет доступен по ссылке - [redash](#)
- ⚠️ Количество строк в партици для MERGE_REFRESH и INCREMENTAL_REFRESH должно быть от 100 млн до 10 млрд



Проверка окружения

- Удаление витрины, столбцов
- Циклические зависимости



- ✓ [GRAPH ACTUAL] - Проверка на наличие цикла ✓
- ✓ [GRAPH ACTUAL] - Проверка критического пути витрины ✓
- ✓ [GRAPH ACTUAL] - Проверка изменения критического пути big10 ✓
- ✓ [GRAPH ACTUAL] - Изменения витрины из апстрима big10 ✓

Ручные аппрувы

- Изменение чужого кода
- Изменение критичных витрин

dwh merge

Reply · Edit · Resolve 🗨️ ...

👁️ 1 ❌ 1



DWH-teamcity Bot 9 mins ago 🔗

Мерж запрещен. Причины:

- Вы меняете код витрины dma.all_contacts, владельцем которой вы не являетесь.
Просим вас получить approve от кого-то из владельцев витрины: anoganesyan или сотрудников юнита: nvogorov.
- В пулл-реквесте есть витрины требующие аппрува:
витрина(ы) из апстрима big_10, необходимо получить 3 аппрув(а): dma.all_contacts

Reply 🗨️ ...



Миграции и их проверка

Автогенерация миграций

 **DWH-teamcity Bot** 39 mins ago 

Идёт генерация миграции для файла datamarts/dma_web_push_tokens.sql, это может занять некоторое время...

Reply  ...

 **DWH-teamcity Bot** 39 mins ago 

Идёт проверка миграции для файла datamarts/dma_web_push_tokens.sql, это может занять некоторое время...

Reply  ...

 **DWH-teamcity Bot** 39 mins ago 

Проверка миграции для файла dma_web_push_tokens завершилась успешно.

Reply  ...

Ручное создание миграции



Nikolay Ogorov 01 Aug 2025 

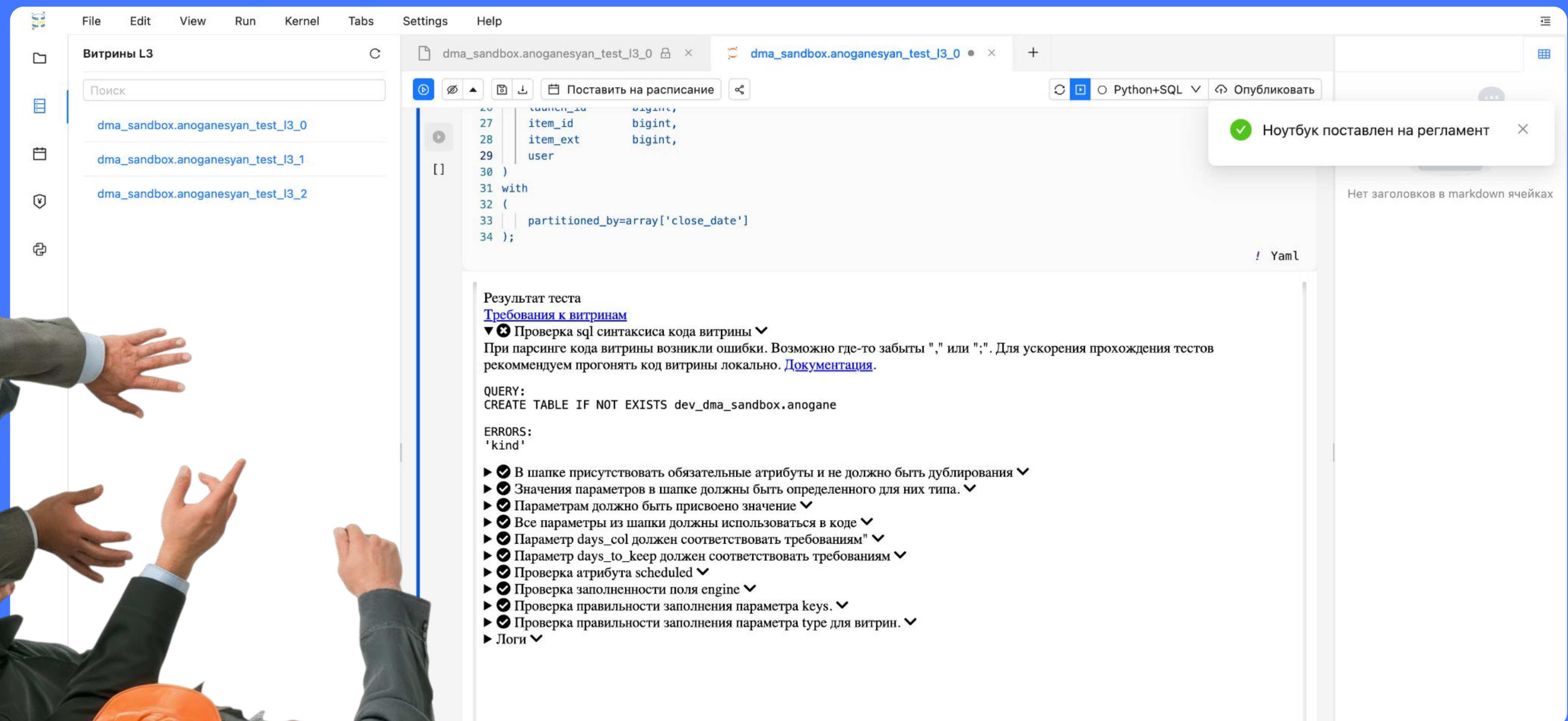
```
use migration dma_test_dm_nvogorov2
CREATE TABLE IF NOT EXISTS dev_DMA.test_dm_nvogorov2
(
  launch_id int not NULL,
  version_id int not NULL ENCODING RLE,
  event_date date,
  user_id int,
  new_column1 int,
  new_column2 int,
  new_column3 int
)
order by event_date, user_id
unsegmented all nodes
partition by launch_id;
```

Reply · Edit · Resolve  ...

 1  1

Планы

- Самостоятельное изменение данных в витринах пользователями
- Админка и гибкая настройка системы тестирования
- Система тестов as platform
- Интеграция в IDE



Вопросы

