

Что такое DevOps в мире хранилищ данных

@me

Руководитель группы автоматизации
хранилищ данных в Почтатехе



@Почтатех

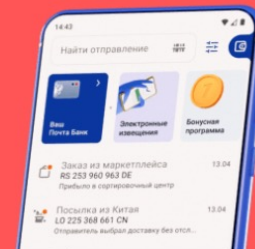
Почтатех – IT-дочка Почты России

100+ продуктовых команд

Классный IT-ландшафт

Мобильное приложение ¹²⁺

6 миллионов пользователей ежемесячно и 7-е место в топе российских приложений



Портал pochta.ru ¹²⁺

10 миллионов уникальных посетителей ежемесячно



Почтовые сервисы для бизнеса

100 тысяч компаний-клиентов



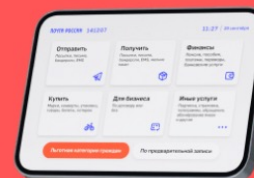
Логистические системы

1,5 миллиона посылок ежедневно



Автоматизированная система для отделений

38 тысяч отделений по всей России



@DataCloud

6,5 Пбайт активных данных

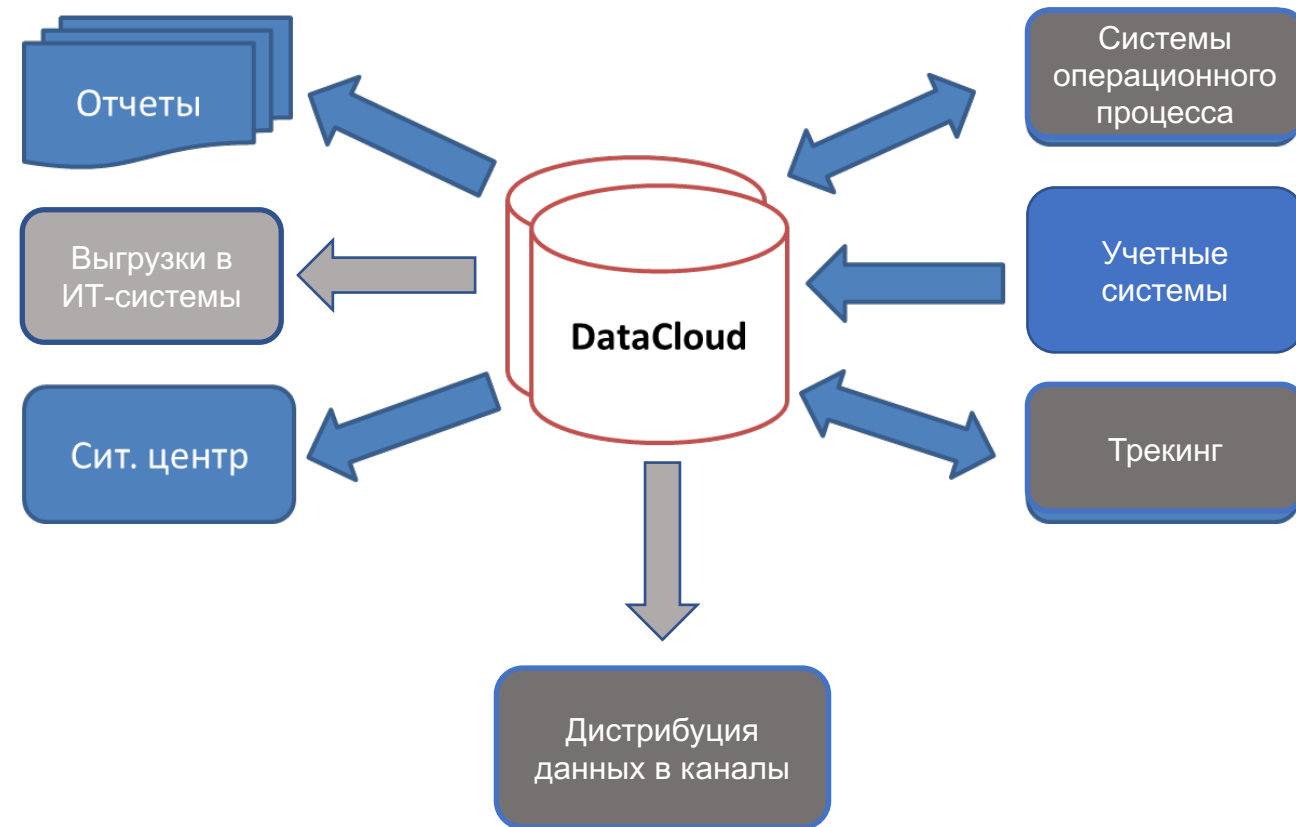
Развитая аналитика

Множество кластеров БД

Несколько вариантов ETL

Каталоги данных, мониторинги и не только

Каждый день внедряем в прод



```
dwhs = { 'Alfa', 'GPB', 'CiB', 'Post', 'C...' }
```

```
print(', '.join(dwhs))
```

Out: Alfa, GPB, CiB, Post C...

Обработка входных требований и постановка задач

Обработка входных требований и постановка задач
Реализация функционала/разбор проблем

Обработка входных требований и постановка задач
Реализация функционала/разбор проблем
Тестирование функционала

Обработка входных требований и постановка задач
Реализация функционала/разбор проблем
Тестирование функционала
Установка кода и его сопровождение

Обработка входных требований и постановка задач
Реализация функционала/разбор проблем
Тестирование функционала
Установка кода и его сопровождение
Поддержка платформы

Аналитик

BRD

SRS

Happy!

Разработчик

```
srs = 'Abra-kadabra'
understand = False

while not understand:
    if srs == 'Got it!':
        understand = True
    srs = 'Got it!'

write_code(srs)
```

Разработчик

```
srs = 'Abra-kadabra'
understand = False

while not understand:
    if srs == 'Got it!':
        understand = True
    srs = 'Got it!'

write_code(srs)
```

Разработчик

```
srs = 'Abra-kadabra'
understand = False

while not understand:
    if srs == 'Got it!':
        understand = True
    srs = 'Got it!'

write_code(srs)
```

Тестировщик



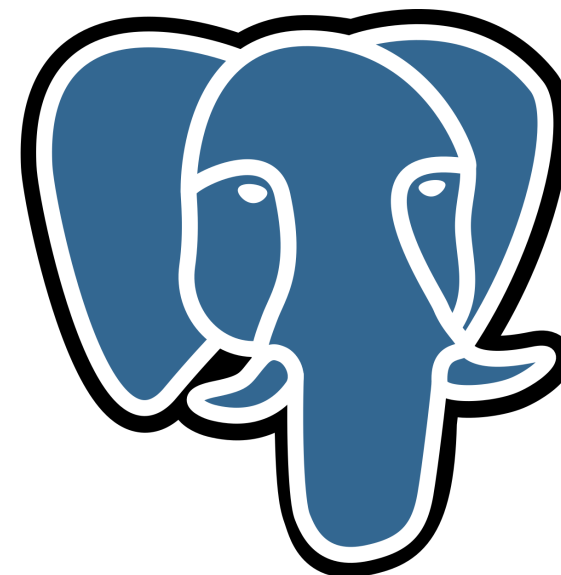
Саппорт

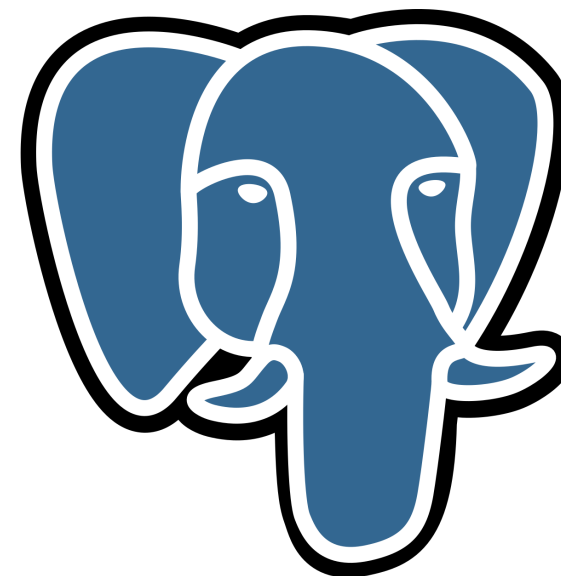
Документация

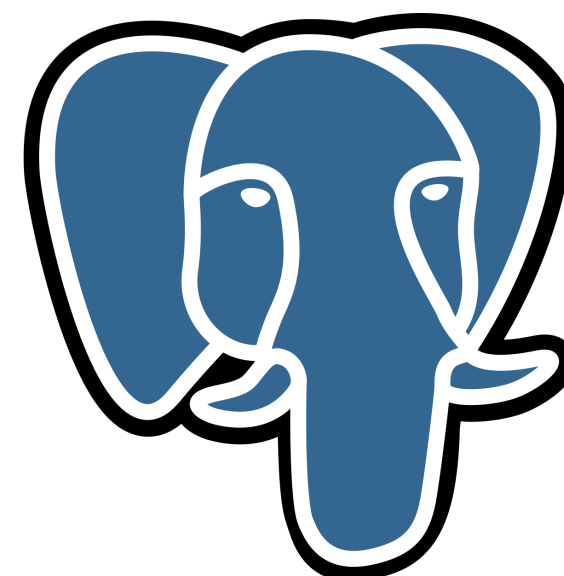
Выкат в production

Мероприятия по поддержке











```
SELECT 1;
```

```
def calculate():
```

```
    try:
```

```
        ultimate_question_of_Life_the_Universe_and_Everything = cursor.execute("select 1;").fetchone()
```

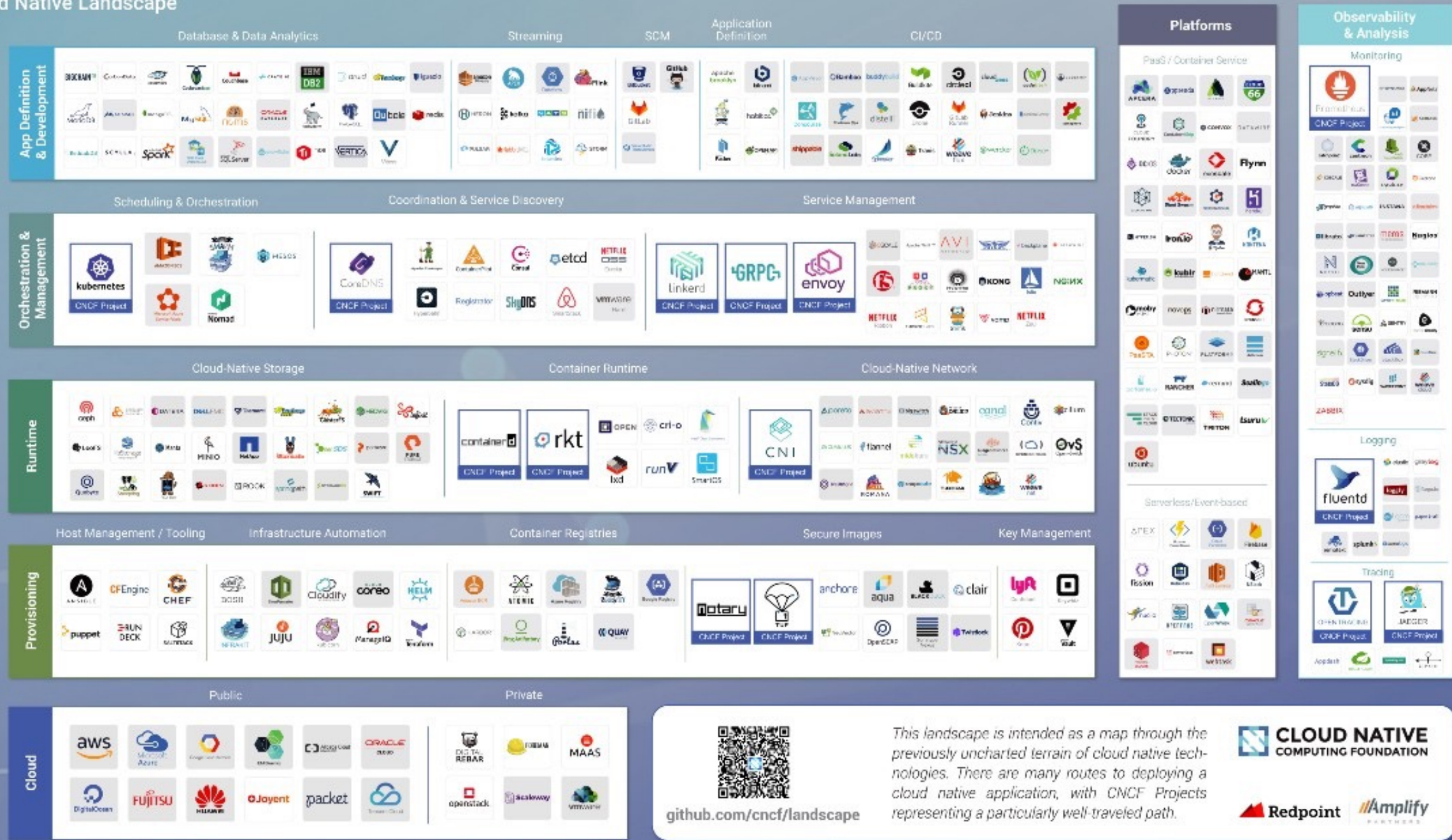
```
    except MultiverseExplode:
```


```
        ultimate_question_of_Life_the_Universe_and_Everything = 42
```

```
    return ultimate_question_of_Life_the_Universe_and_Everything
```

```
select 1;
```

	C1
1	1

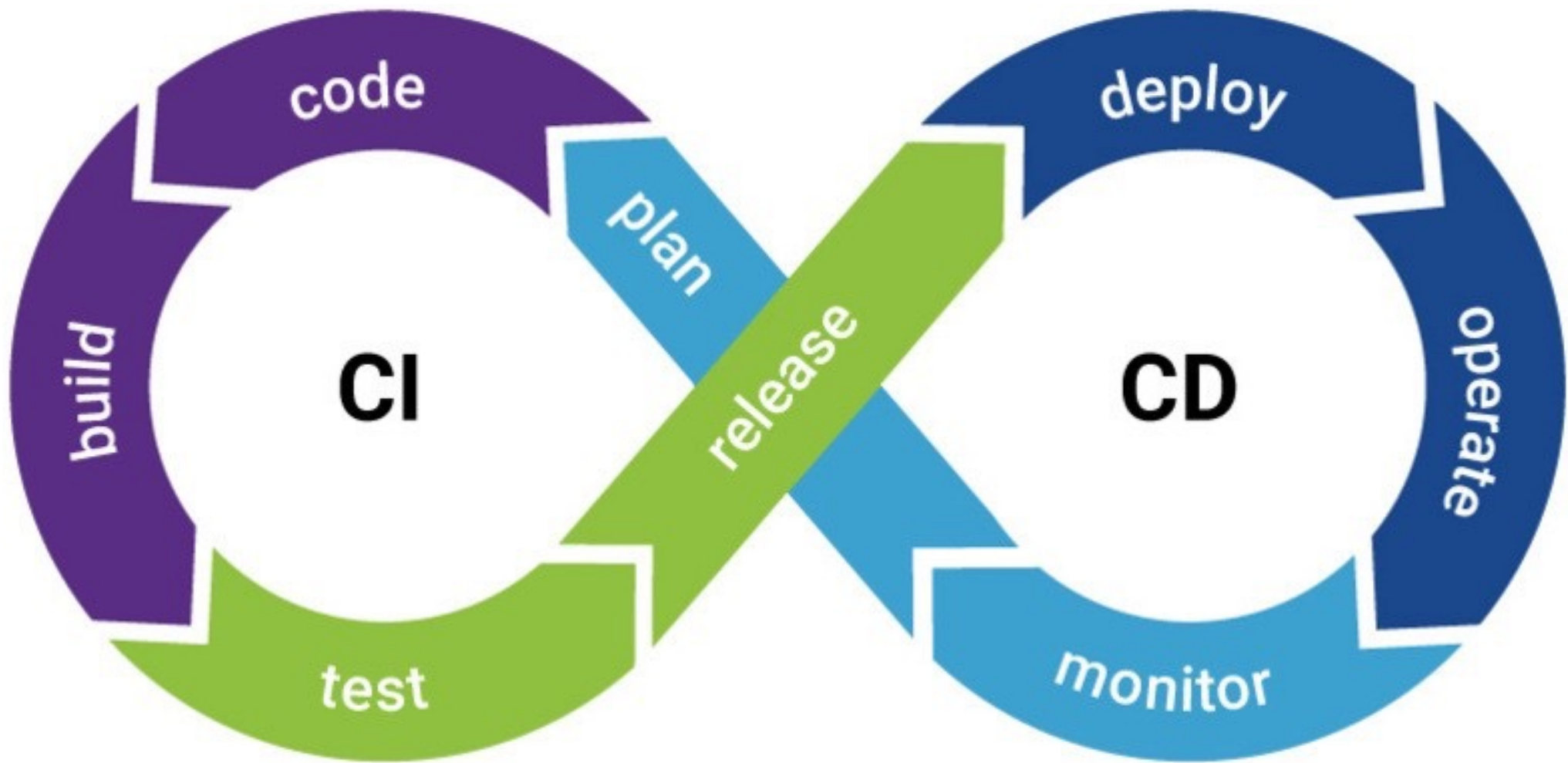



github.com/cncf/landscape

This landscape is intended as a map through the previously uncharted terrain of cloud native technologies. There are many routes to deploying a cloud native application, with CNCF Projects representing a particularly well-traveled path.

https://miro.medium.com/max/1200/1*wzt-mF3yXbhQZru2Tq61Zw.jpeg

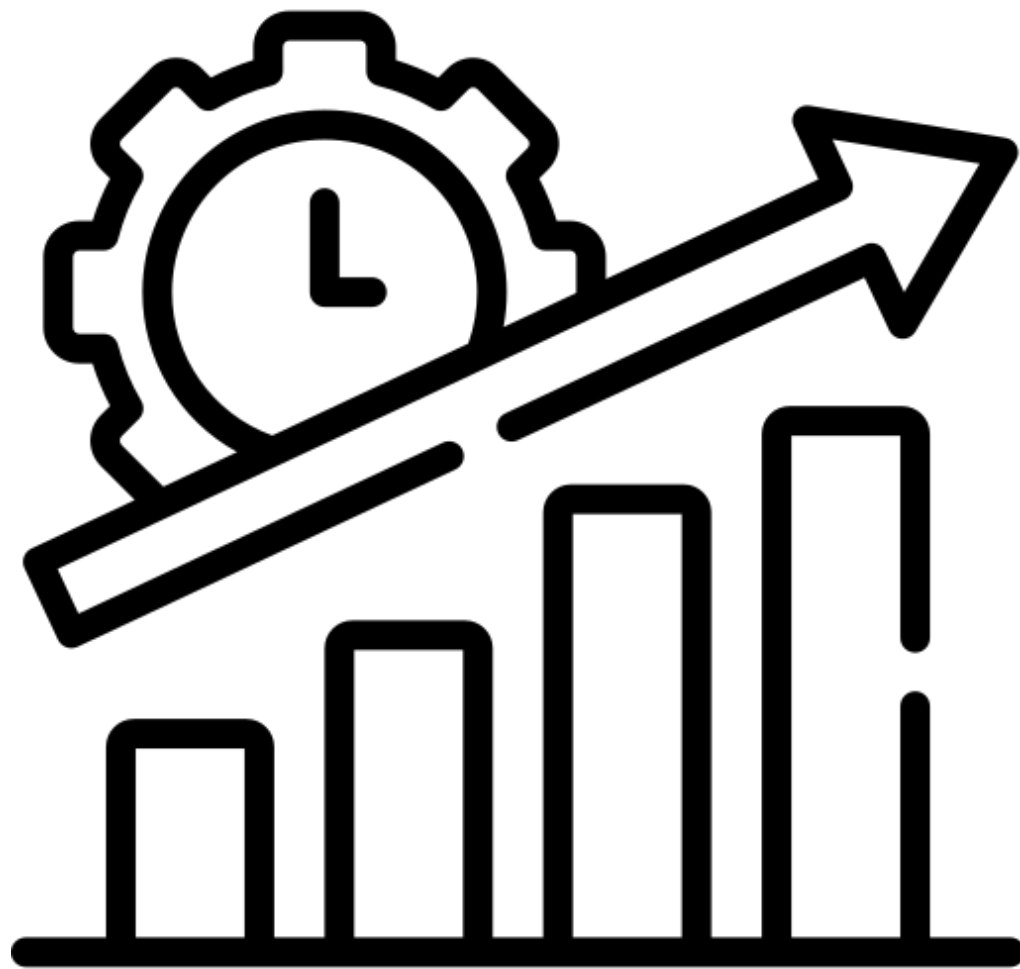












Зачем это все

Измеряемый и прозрачный процесс

Грамотное планирование релизов

Быстрые и стабильные релизы

ВІ и ІТ развивались параллельно



У SQL нет состояний

```
alter table table1 add column some_column  
varchar(100);
```

```
print('hello world!')
```

У SQL нет состояний

```
alter table table1 add column some_column  
varchar(100);
```

```
alter table table1 drop column some_column;
```

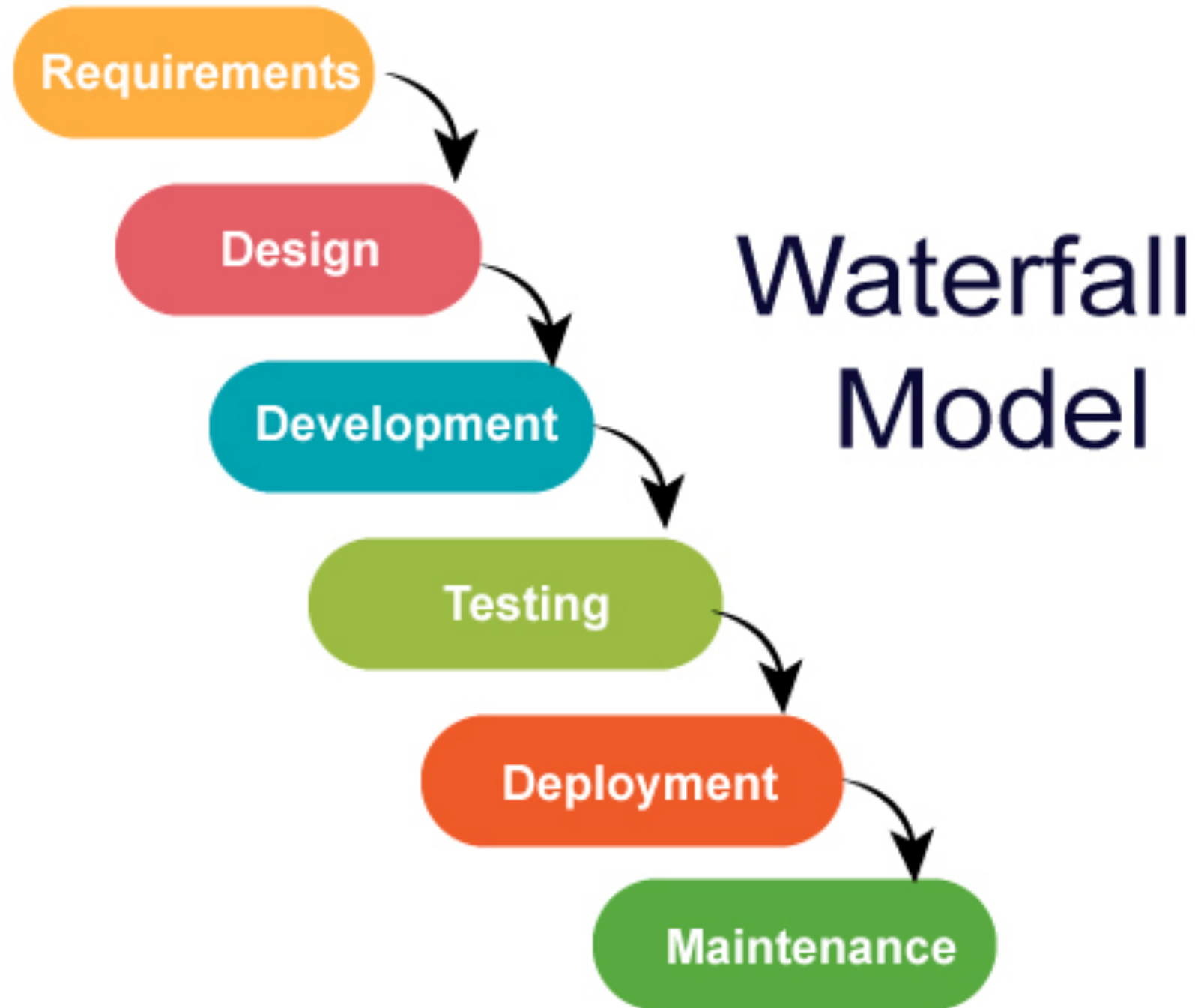
Главный спойлер

Реализуем состояние SQL = Трюб-DevOps



- Слышали ли вы, Холмс, про кота Баскервилей?
- Нет.
- Это ужасный зверь, Холмс. В пять раз больше собаки, полосатый, с огромными когтями и клыками, невероятно прыгучий.
- Этот зверь называется Идемпотентность, Ватсон.

QUIZ





День X

Правила

Соглашения

Интерфейсы

Автоматизация



DevOps = коммуникация

COMMUNICATION

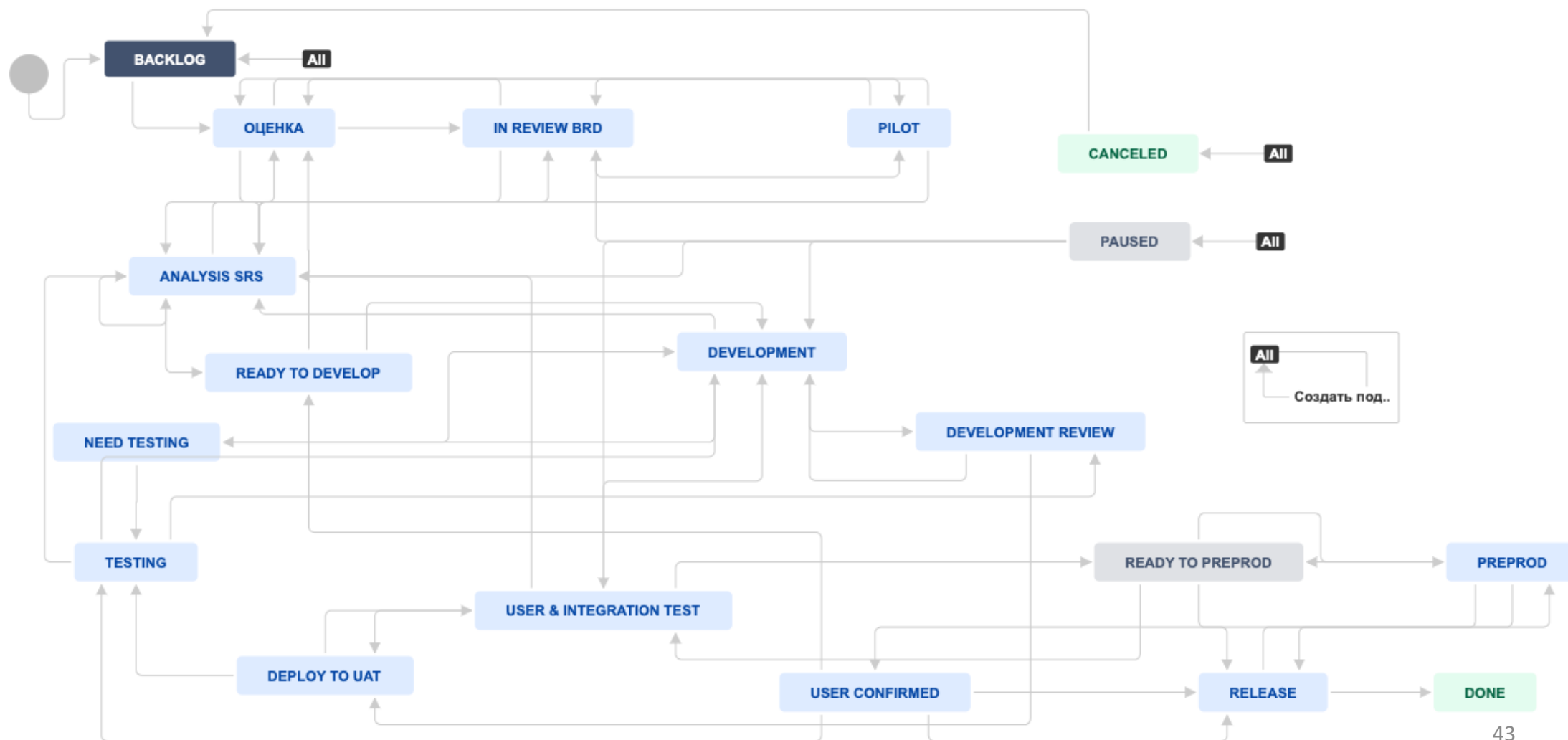


THE KEY IS

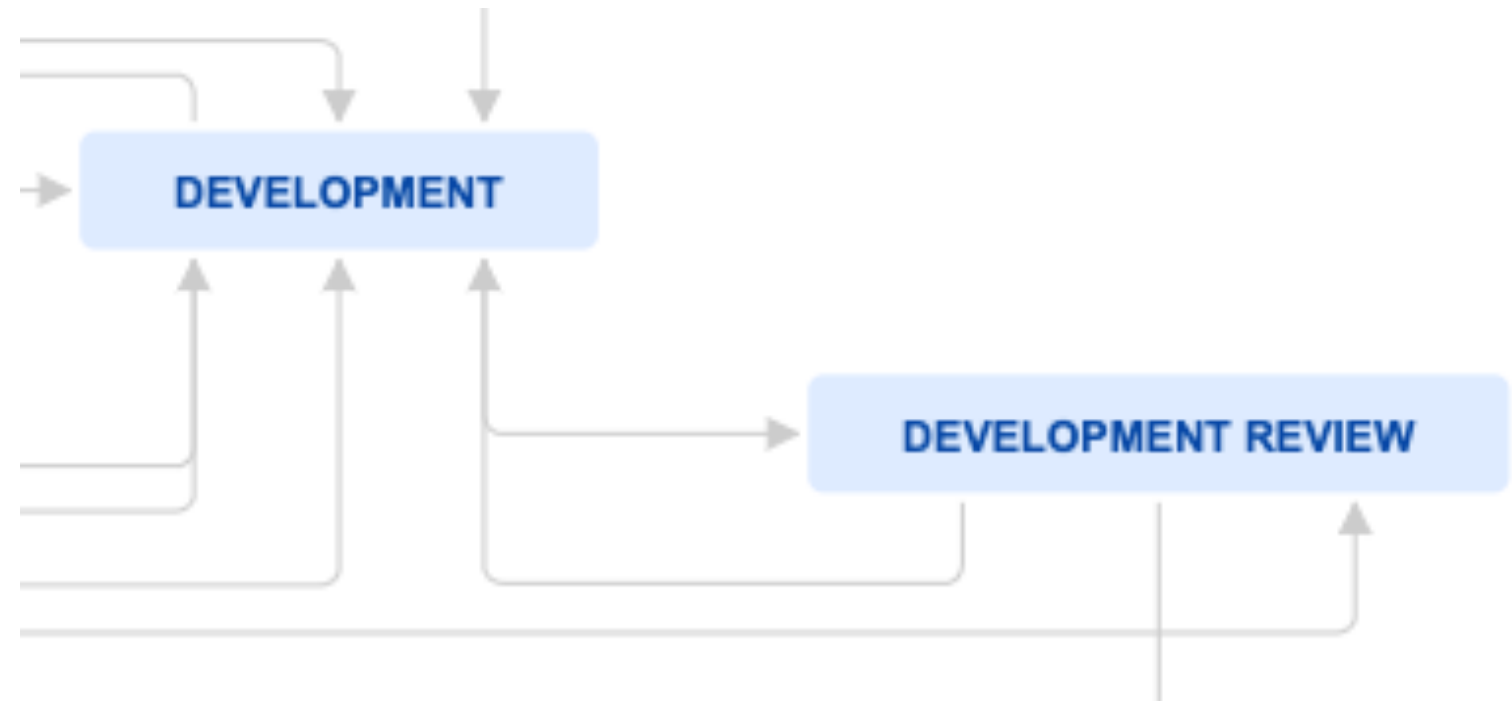
Пайплайн



Пайплайн

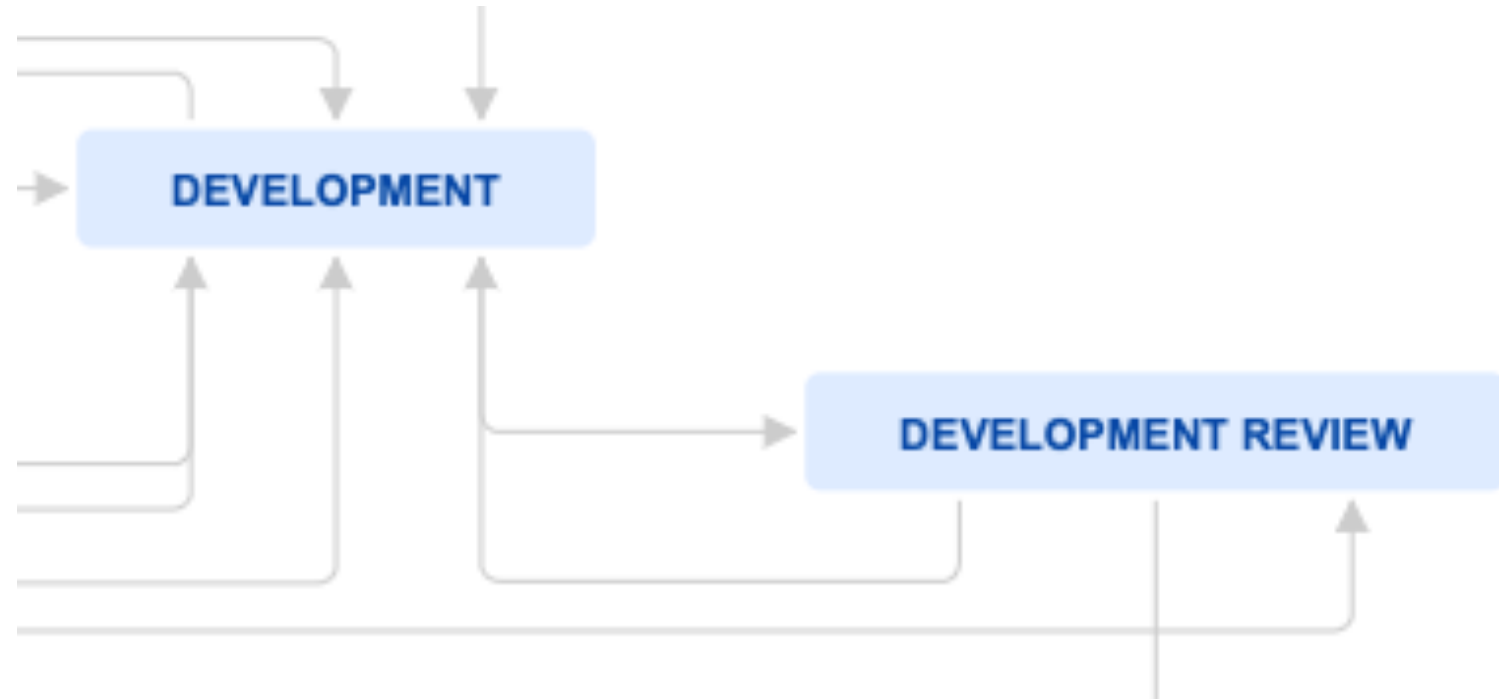


T3



T3

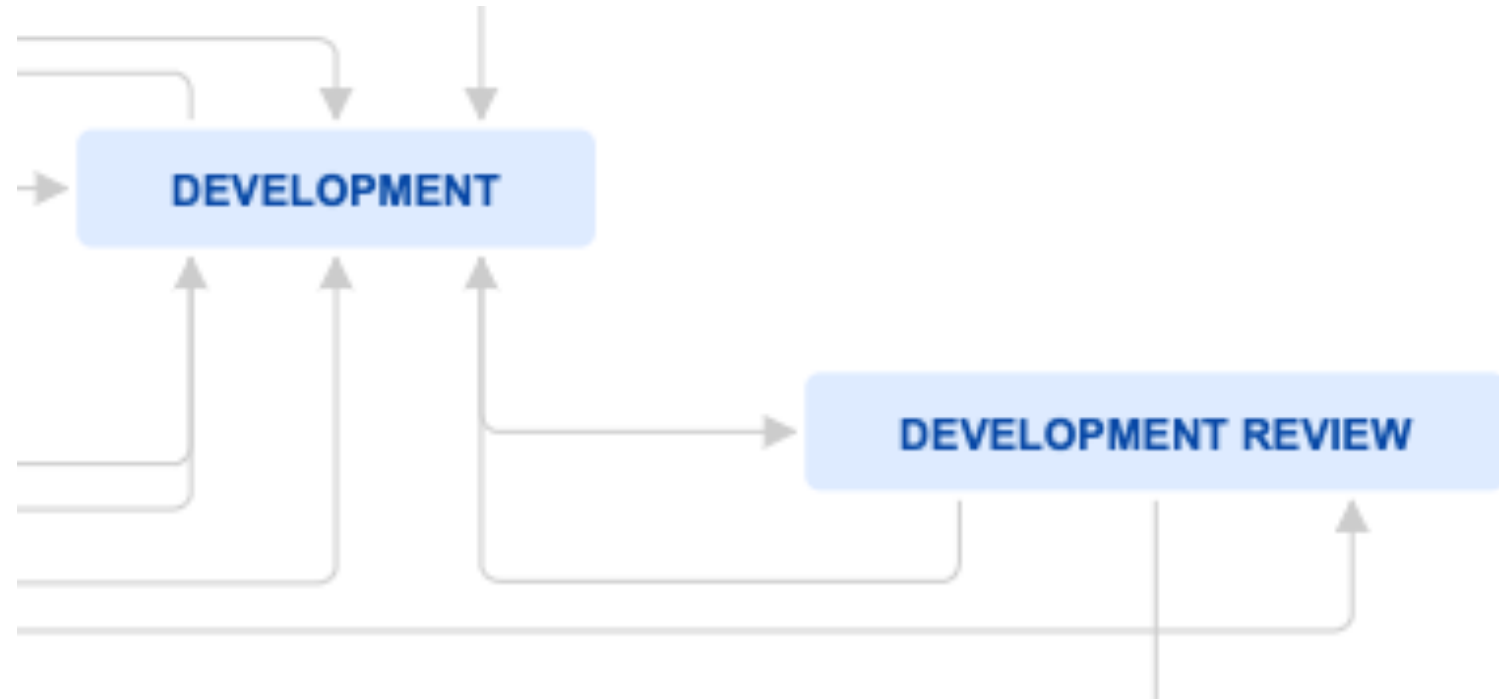
VCS



T3

VCS

Стандарт хранения кода

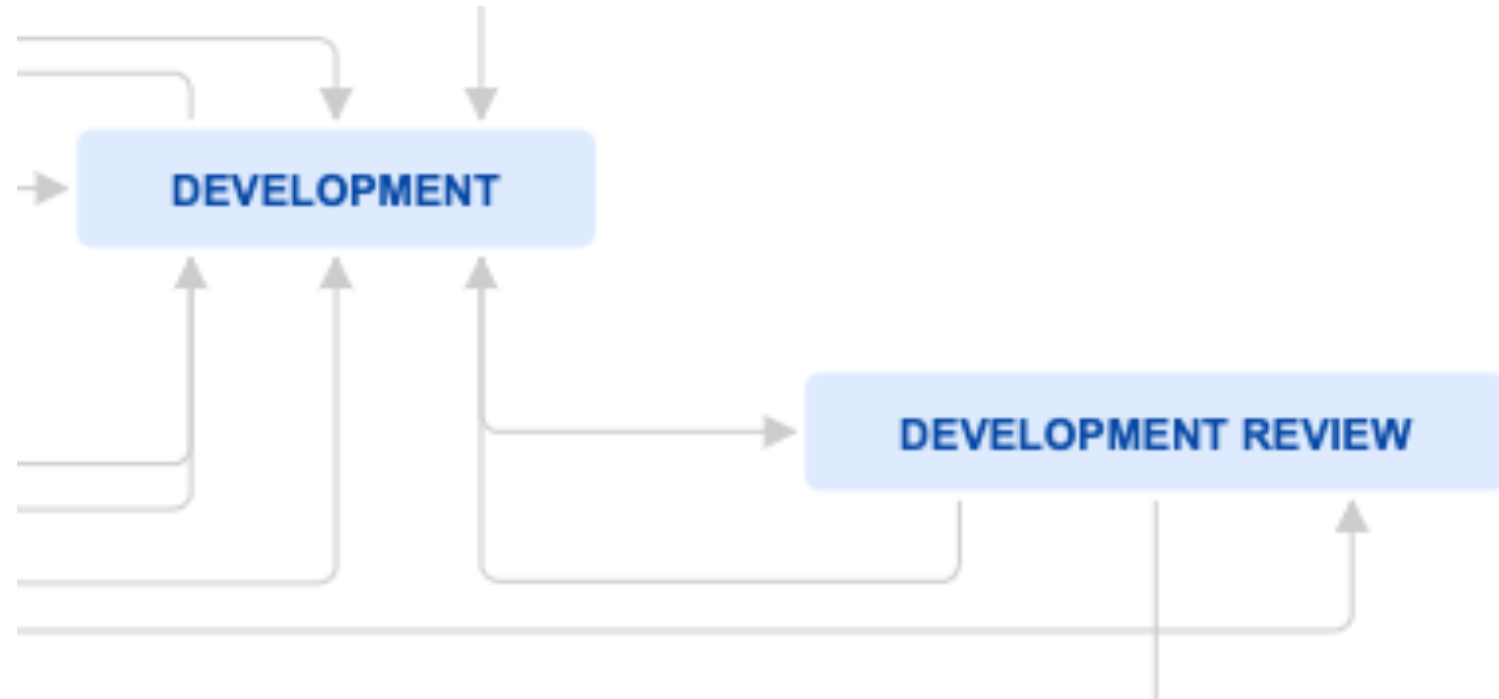


T3

VCS

Стандарт хранения кода

Модель ветвления



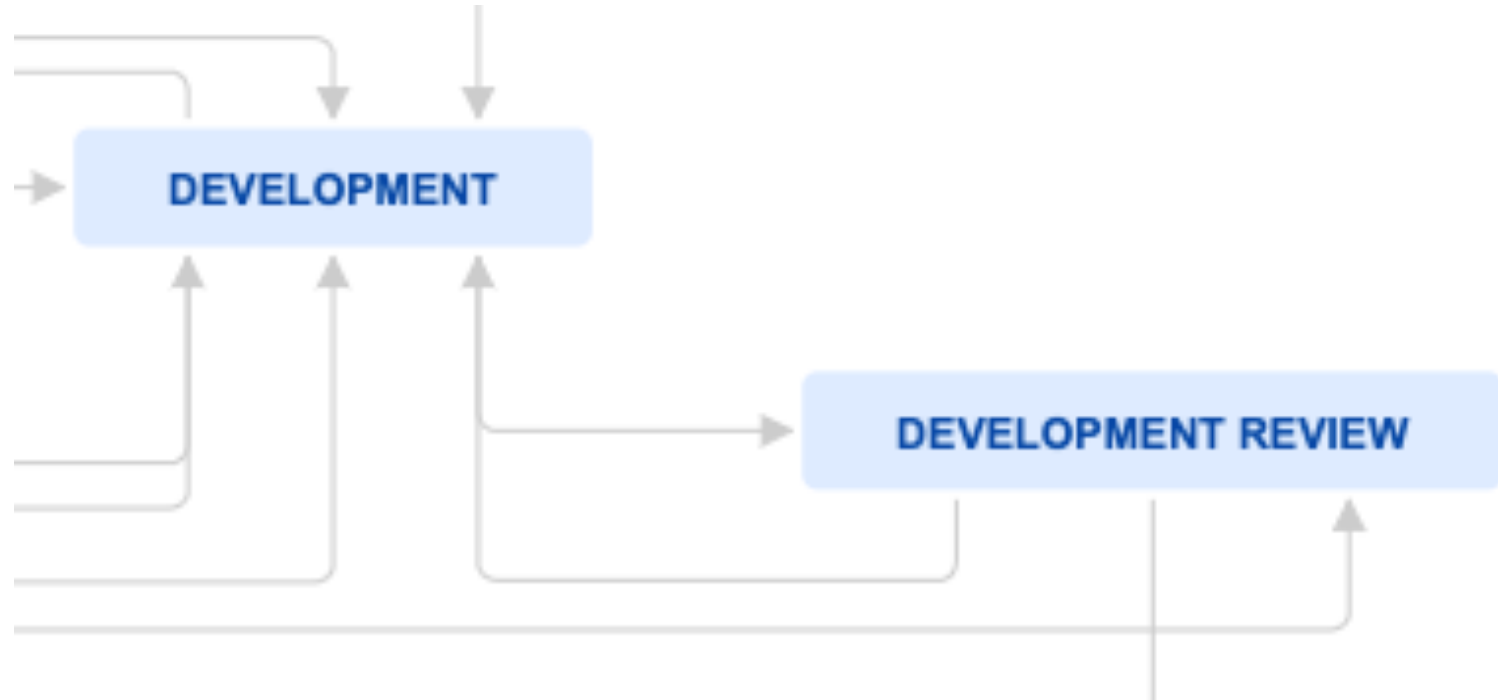
T3

VCS

Стандарт хранения кода

Модель ветвления

Merge/Pull Requests



T3

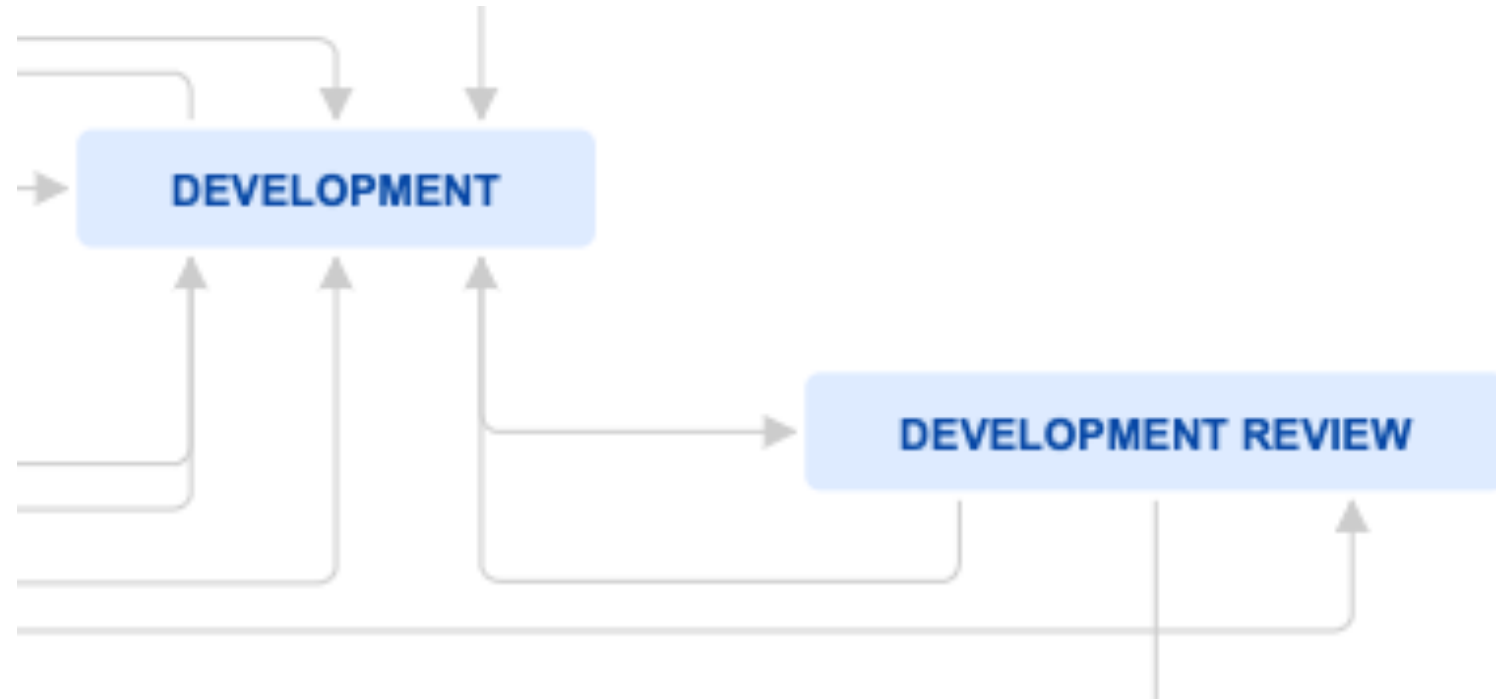
VCS

Стандарт хранения кода

Модель ветвления

Merge/Pull Requests

Нотификации



T3

~~VCS~~

коммуникация

~~Стандарт хранения кода~~

коммуникация

~~Модель ветвления~~

коммуникация

~~Merge/Pull Requests~~

коммуникация

~~Нотификации~~

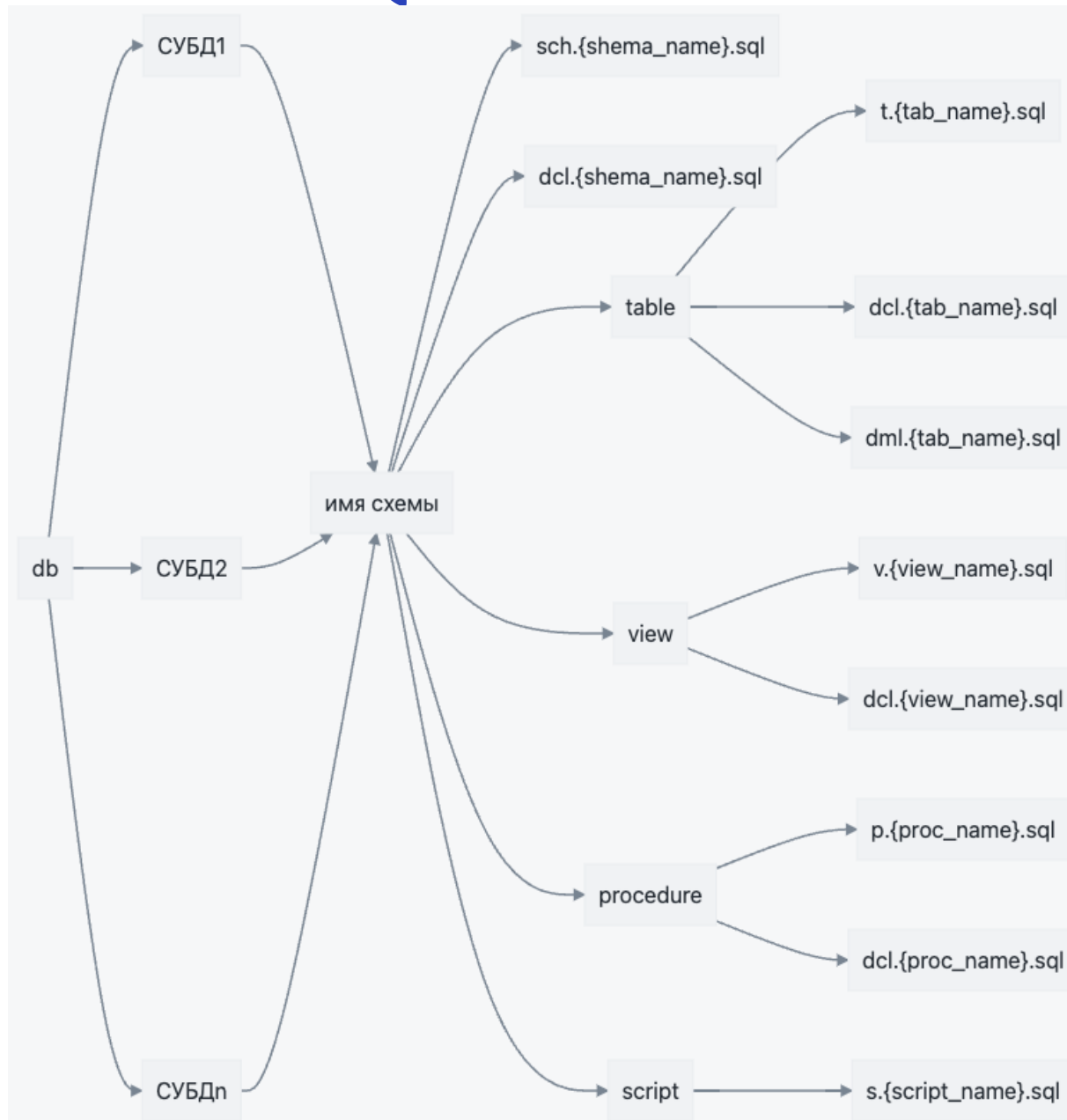
коммуникация

Ключевые принципы

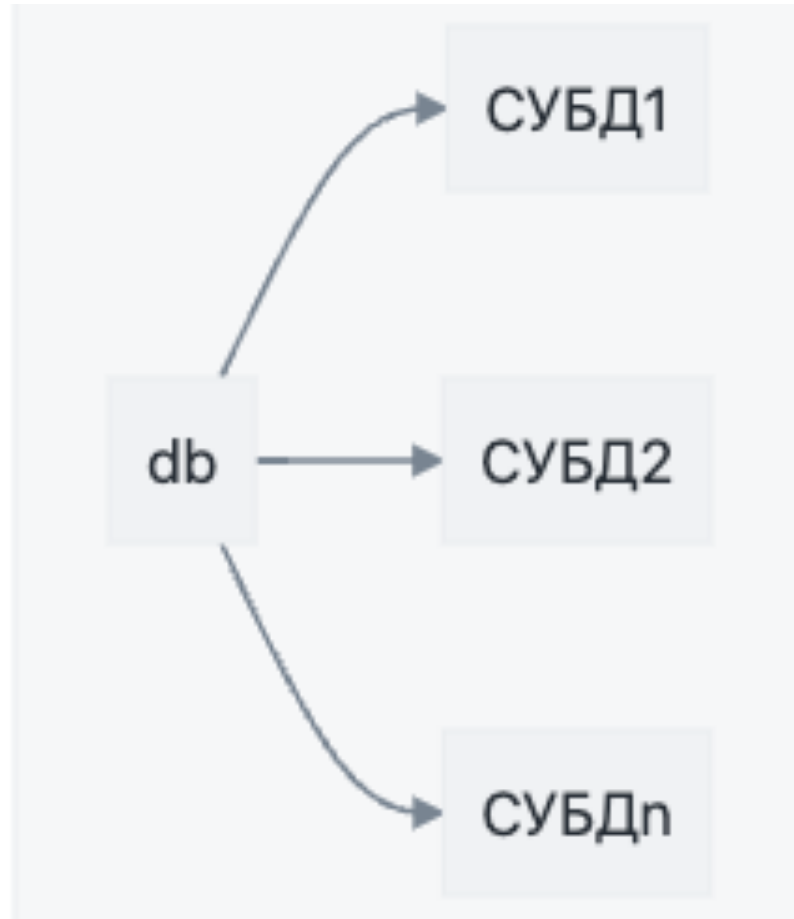
Git – единственный источник правды

Объектная модель по SQL-файлам

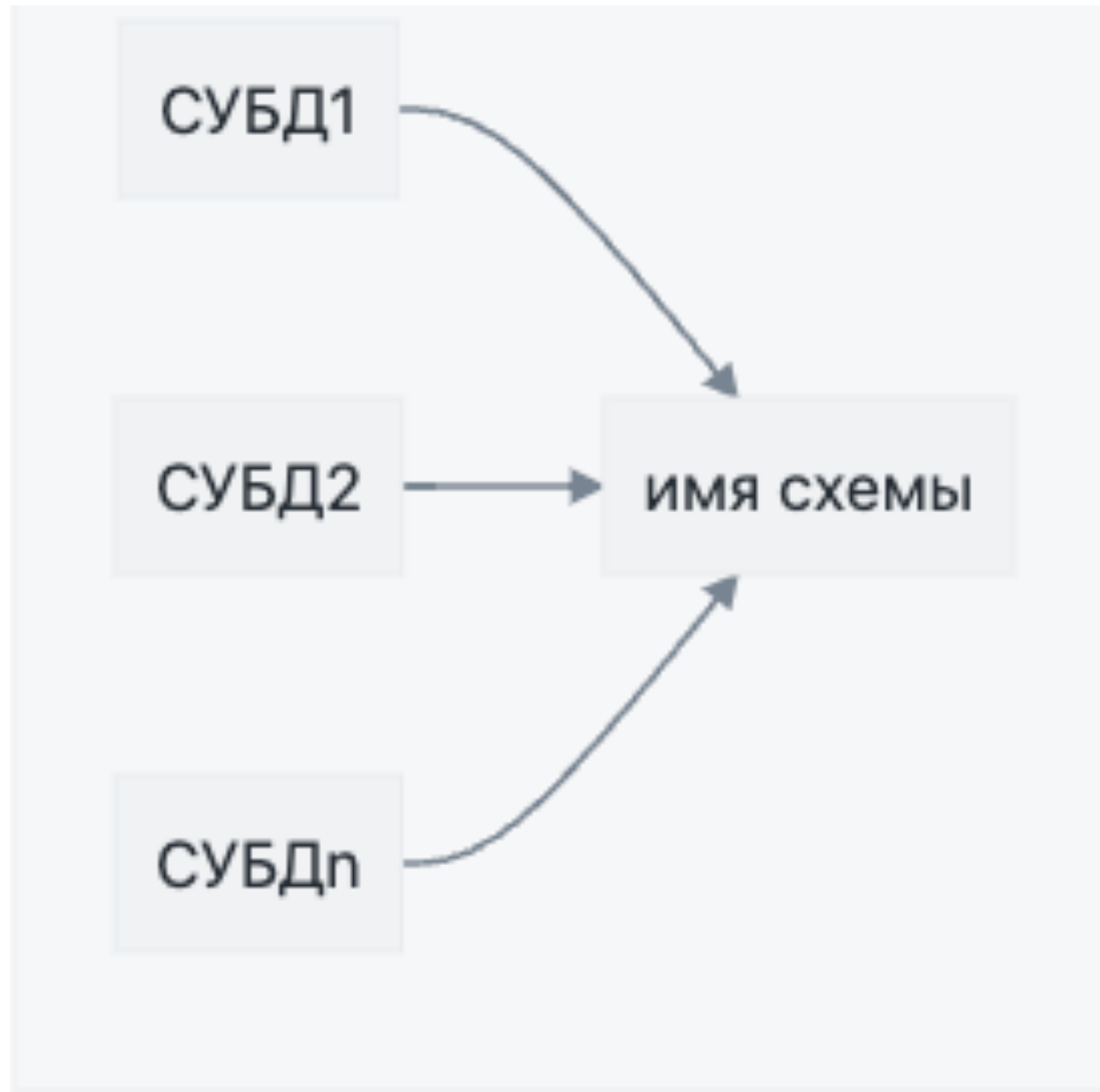
Версионирование SQL



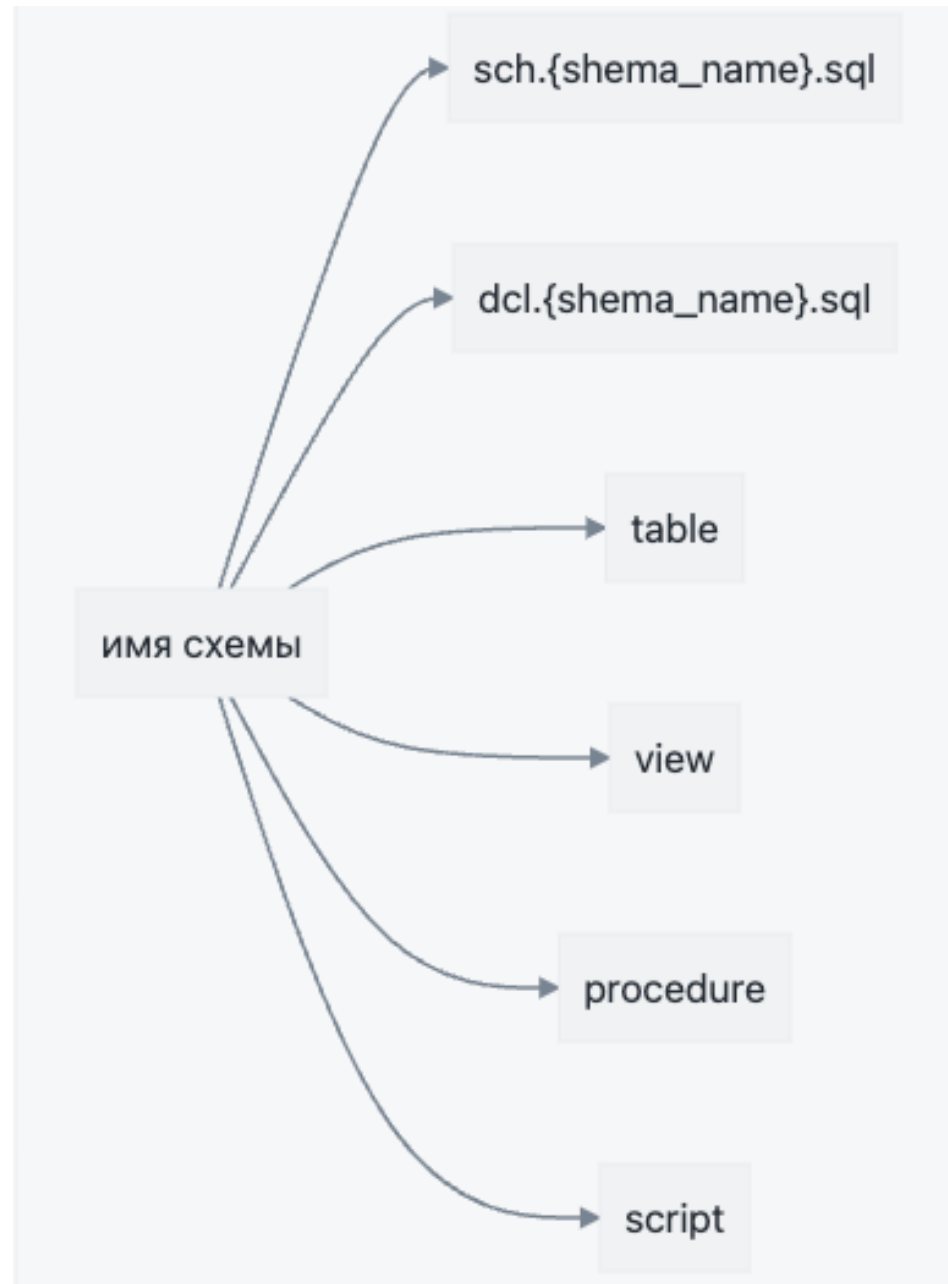
Версионирование SQL



Версионирование SQL

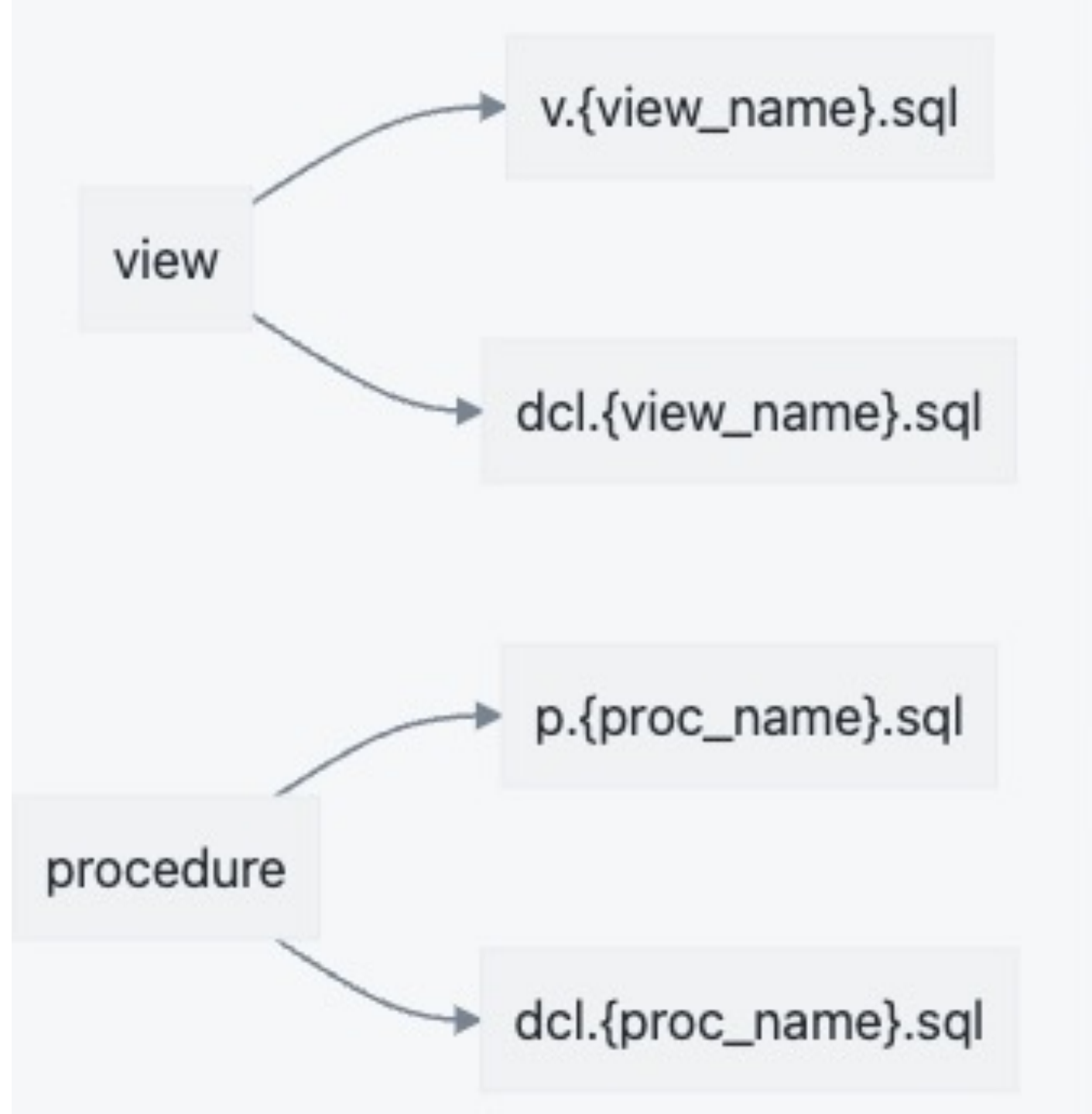


Версионирование SQL



Накат вью и процедур

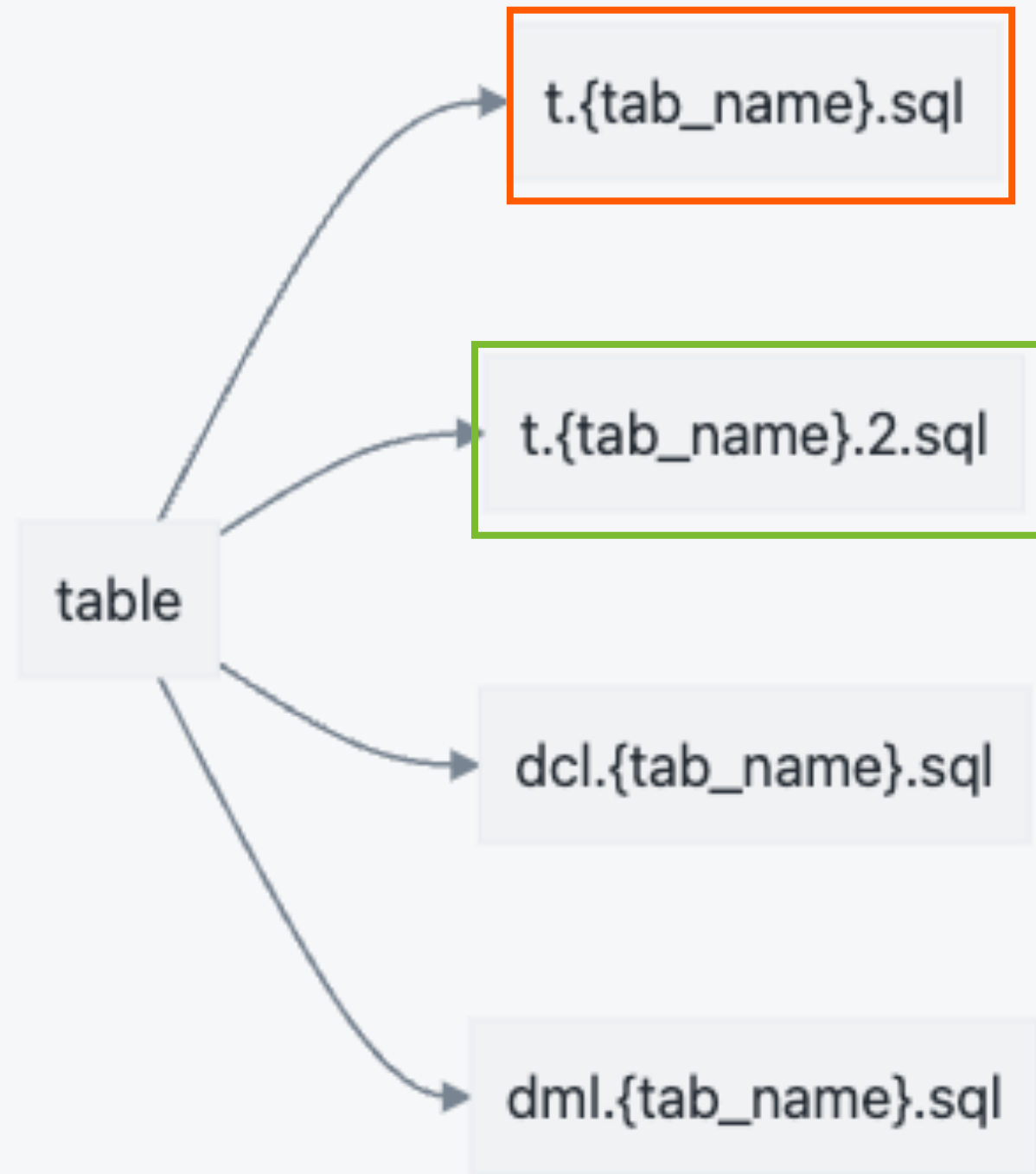
```
create or replace  
view view1 as  
select id1 from  
table1;
```



Накат таблиц

```
create table table1  
(id integer);
```

```
alter table table1  
add column  
some_column  
varchar(100);
```





git

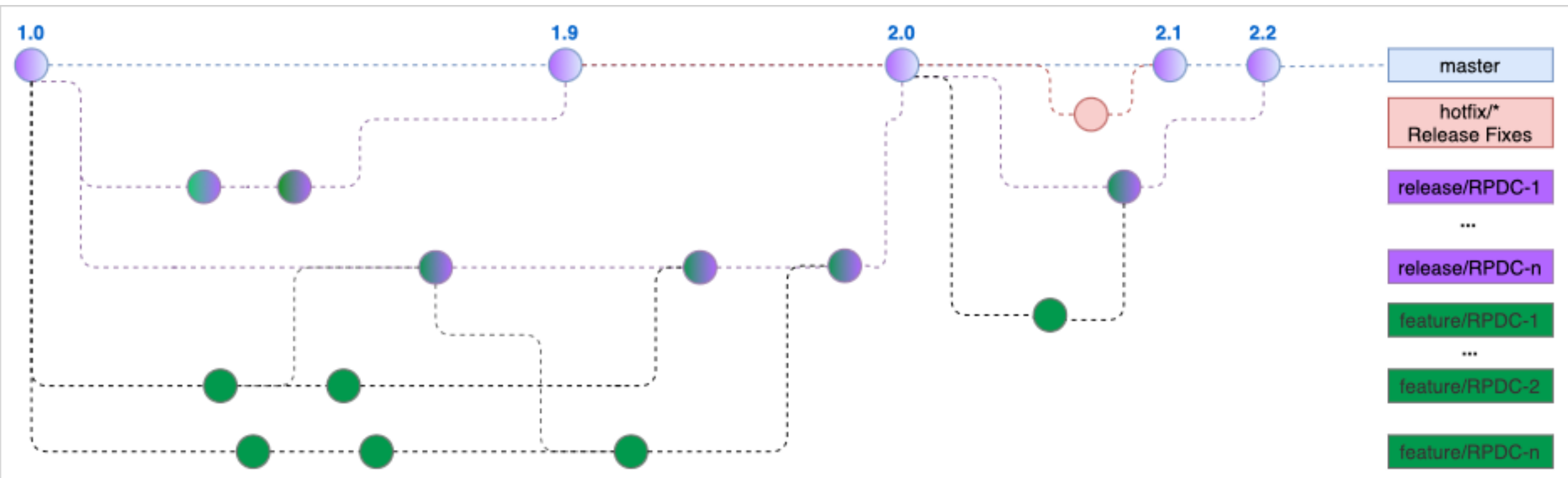
magic

Сборка

Нам нужны миграции

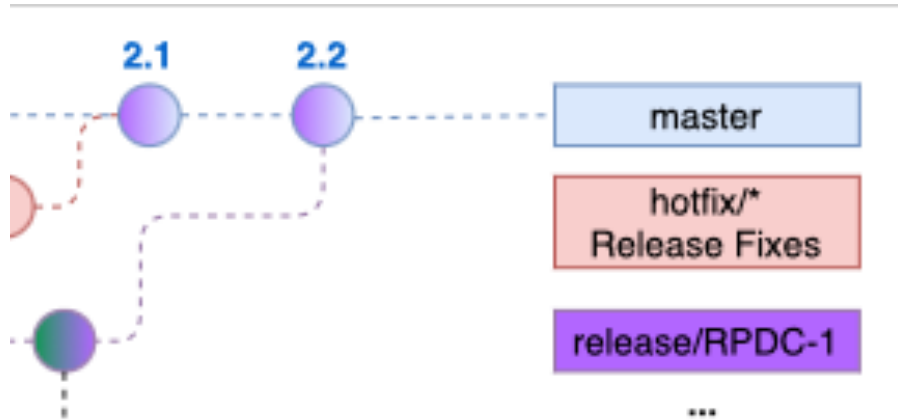
Миграции = патчи

Сборка

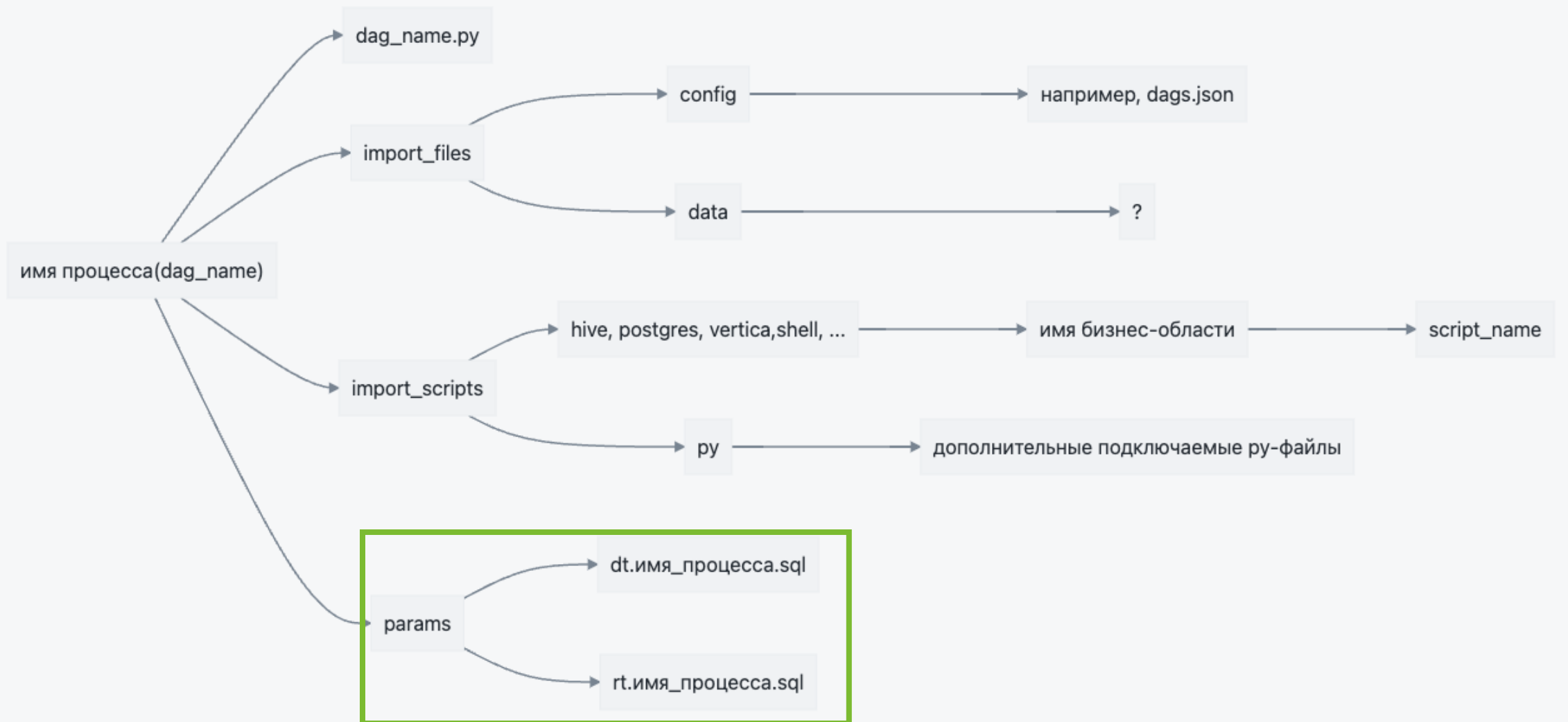


Сборка

`git diff <COMMIT HASH> HEAD --name-only`



Структура ETL



Deploy. Release Notes

[instruction]

```
etl/airflow/dags/business_space/dag01 -- af_1 auto unpause | Установить dag инсталлятором и снять с паузы. -- --
DAG1 -- запустить технический ДАГ DAG1, для генерации json-файла с design-time параметрами
etl/pentaho/business_space/test_job/params/rt.params.sql -- postgres_dwhmeta auto | Запустить скрипт с параметрами. -- --
db/hive/business_space/table/t.hive_table.1.sql -- hive_dwh auto | Запустить скрипт с параметрами. -- ${hive_table_2}=test_value1 --
db/vertica/business_space/table/t.vertica.1.sql -- vertica_dwh auto | Запустить скрипт с параметрами. -- --
etl/pentaho/business_space/test_job -- pentaho_dwh auto | Установить etl инсталлятором. -- --
DAG2 -- Включить все добавленные ДАГи: DAG2 ,DAG3
```

Deploy. Release Notes

`${var_name}="value"`

Цель релиза:

%Описать кратко суть доработки: добавление новых атрибутов в сущность такую-то/создание новой витрины такой-то/и т.п.%

Версия dataops-automation

%Версия модуля dataops_automation%

Документация

Указать ссылки из Jira, Confluence

ссылка	версия документа и страницы
BRD	0.x conf_v.x
SRS	0.x conf_v.x
Интеграция	0.x conf_v.x
Admin Guide	0.x conf_v.x
JIRA_ID	ссылка

Риски

здесь требуется отразить критичность меняемой функциональности, возможное влияние:

- обозначить информацию о возможных задержках расчетов в случае сбоя
- указать риск порчи данных, которые затем используются зависимыми системами/далее по pipeline
- перечислить зависимые модули/системы/сервисы

Артефакты разработки

Указать номера/наименования всех передаваемых на установку артефактов Патча, с указанием версий, путей размещения и т.д.

code-review*	ревью
хеш коммита для релизной ветки*	хеш коммита для релизной ветки
релизная ветка*	релизная ветка
merge_request	merge_request
названия потоков	названия потоков

для доработок в репозитории view-rpdc - не обязательно

Состав доработки (патча)

› [Click here to expand...](#)

Пререквизиты:

Стратегия вывода и проверки на DEV

✓ [Click here to expand...](#)

Период загрузки данных (справочная информация):



Сущность	Начало периода	Конец периода	Режим (ИЗ/регулярный)	Если в цикле, то к-во циклов	Тестовые/реальные данные	Объемы относительно промышленных
****заполни****	****заполни****	****заполни****	****заполни****	****заполни****	****заполни****	
****заполни****	****заполни****	****заполни****	****заполни****	****заполни****	****заполни****	
****заполни****	****заполни****	****заполни****	****заполни****	****заполни****	****заполни****	


Последовательность шагов



#	file	instruction	vars	info
---	------	-------------	------	------



Deploy. Кнопки. Пайплайн

Test


 auth 

 build_deploy_patch

 generate-ri 



 pipeline_generator 



Downstream



 build_deploy_patch
#513457
Child

<



Build



 build 



 cleaner_check 



 intersections 

Deploy-dev

 1-1-auto-dev 

 2-2-manual-dev 

 3-6-auto-dev 

 7-7-manual-dev 

Deploy. Кнопки. Пайплайн

Последовательность шагов

#	file	instruction	vars	info
1	etl/airflow/dags/business_space/dag01	af_1 auto unpause Установить dag инсталлятором и снять с паузы.		
2	DAG1	Не менять вручную! Ческий ДАГ DAG1, для генерации json-файла с design-time параметрами		
3	etl/pentaho/business_space/test_job/params/rt.params.sql	postgres_dwhmeta auto Запустить скрипт с параметрами.		
4	db/hive/business_space/table/t.hive_table.1.sql	hive_dwh auto Запустить скрипт с параметрами.	\${hive_table_2}=test_value1	
5	db/vertica/business_space/table/t.vertica.1.sql	vertica_dwh auto Запустить скрипт с параметрами.		
6	etl/pentaho/business_space/test_job	pentaho_dwh auto Установить etl инсталлятором.		
7	DAG2	Включить все добавленные ДАГи: DAG2 ,DAG3		

Deploy-dev

1-1-auto-dev

2-2-manual-dev

3-6-auto-dev

7-7-manual-dev

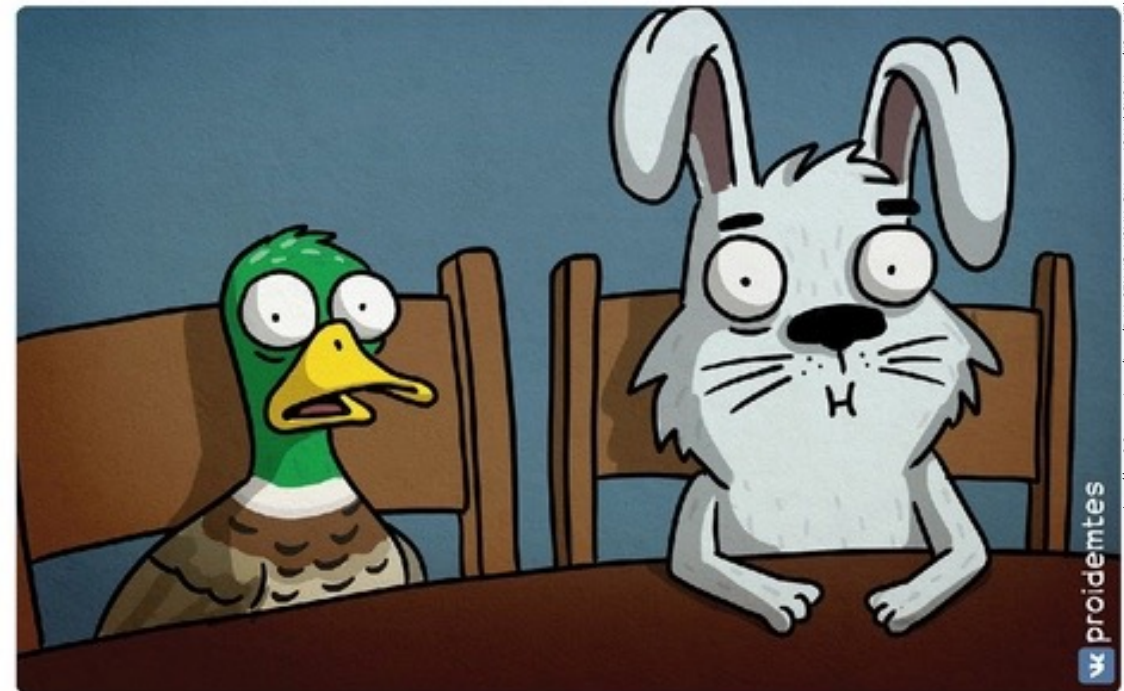
Про DS

Модели на python

Python в Docker

Docker в Helm

Helm в kubernetes



Сегодня мы многое поняли

Стандарт хранения кода

Параметры и переменные

git diff – друг

Release Notes

Кнопки в gitlab-ci

Сегодня мы многое поняли

Стандарт хранения кода

Параметры и переменные

git diff – друг

Release Notes

Кнопки в gitlab-ci

Нафига, т.е. как бы зачем?

КОММУНИКАЦИЯ

Грабли

Автоопределение последовательности скриптов – зло

Грабли

Автоопределение последовательности скриптов – зло

Для release notes лучше yml, json, toml

Грабли

```
>>> with open("pyproject.toml", mode="rb") as fp: ...  
    tomlib.load(fp) ...
```

```
{'build-system': {'requires': ['flit_core>=3.2.0,<4'], 'build-backend': 'flit_core.buildapi'}, 'project': {'name': 'tomli', 'version': '2.0.1', 'description': 'A lil' TOML parser', 'requires-python': '>=3.7', 'readme': 'README.md', 'keywords': ['toml'], 'urls': {'Homepage': 'https://github.com/hukkin/tomli', 'PyPI': 'https://pypi.org/project/tomli'}}}
```

Грабли

Автоопределение последовательности скриптов – зло

Для release notes лучше yml, json, toml

Разработка в ООП

Грабли

Автоопределение последовательности скриптов – зло

Для release notes лучше yml, json, toml

Разработка в ООП

Git flow без develop – хорошо

Грабли

Автоопределение последовательности скриптов – зло

Для release notes лучше yml, json, toml

Разработка в ООП

Git flow без develop – хорошо

Маленькие улучшайзеры – хорошо!

Грабли

Автоопределение последовательности скриптов – зло

Для release notes лучше yml, json, toml

Разработка в ООП

Git flow без develop – хорошо

Маленькие улучшайзеры – хорошо!

Тесты – must!

Грабли

Автоопределение последовательности скриптов – зло

Для release notes лучше yml, json, toml

Разработка в ООП

Git flow без develop – хорошо

Маленькие улучшайзеры – хорошо!

Тесты – must!

Прямая совместимость – must!

Грабли

Автоопределение последовательности скриптов – зло

Для release notes лучше yml, json, toml

Разработка в ООП

Git flow без develop – хорошо

Маленькие улучшайзеры – хорошо!

Тесты – must!

Прямая совместимость – must!

Парсить SQL на квери – плохо!

Грабли

Автоопределение последовательности скриптов – зло

Для release notes лучше yml, json, toml

Разработка в ООП

Git flow без develop – хорошо

Маленькие улучшайзеры – хорошо!

Тесты – must!

Прямая совместимость – must!

Парсить SQL на квери – плохо!

Про людей

Регулярные публичные встречи

Про людей

Регулярные публичные встречи

Открытый бэклог/роадмап

Про людей

Регулярные публичные встречи

Открытый бэклог/роадмап

Регулярные новости/блогпосты/рассылки

Про людей

Регулярные публичные встречи

Открытый бэклог/роадмап

Регулярные новости/блогпосты/рассылки

Постоянное обучение/мастер-классы

Про людей

Регулярные публичные встречи

Открытый бэклог/роадмап

Регулярные новости/блогпосты/рассылки

Постоянное обучение/мастер-классы

Общая точка для консультаций (чат/почта и т.п)

Про людей

Регулярные публичные встречи

Открытый бэклог/роадмап

Регулярные новости/блогпосты/рассылки

Постоянное обучение/мастер-классы

Общая точка для консультаций (чат/почта и т.п)

Больше прозрачности богу прозрачности!

Что такое DevOps в мире хранилищ данных

@me

