



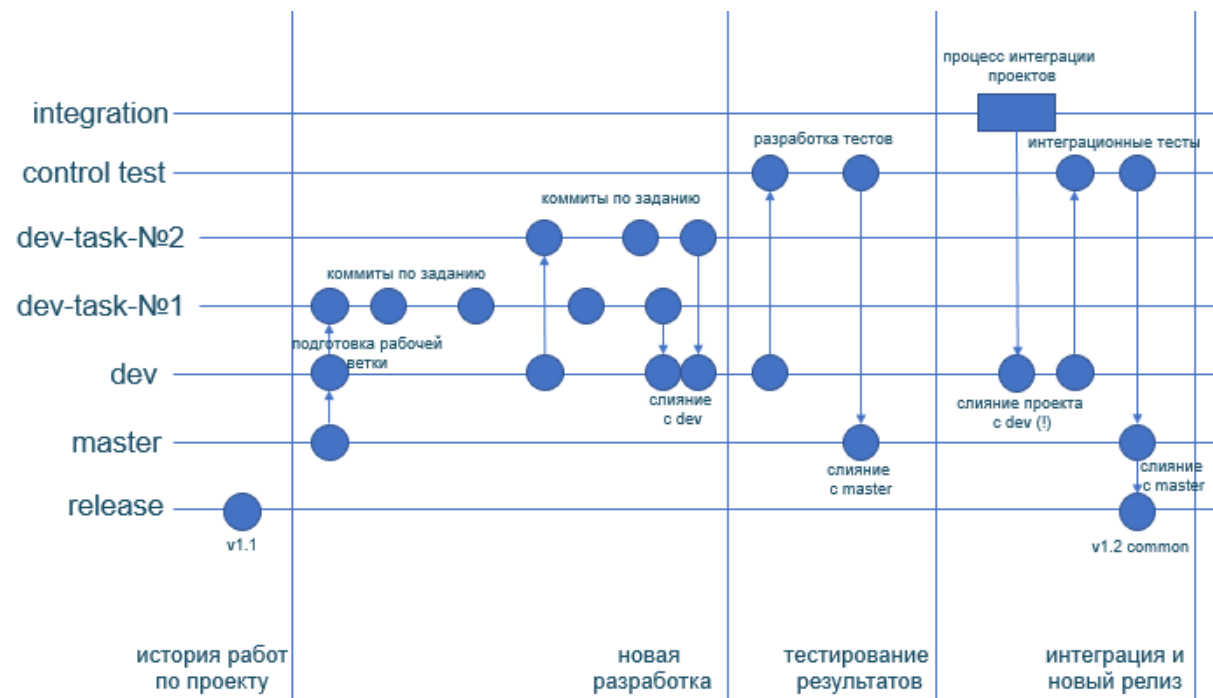
VERSIONING ML

SELEZNEV ARTEM
SENIOR DE @ SBER

ГДЕ МОИ ВЕРСИИ,...?

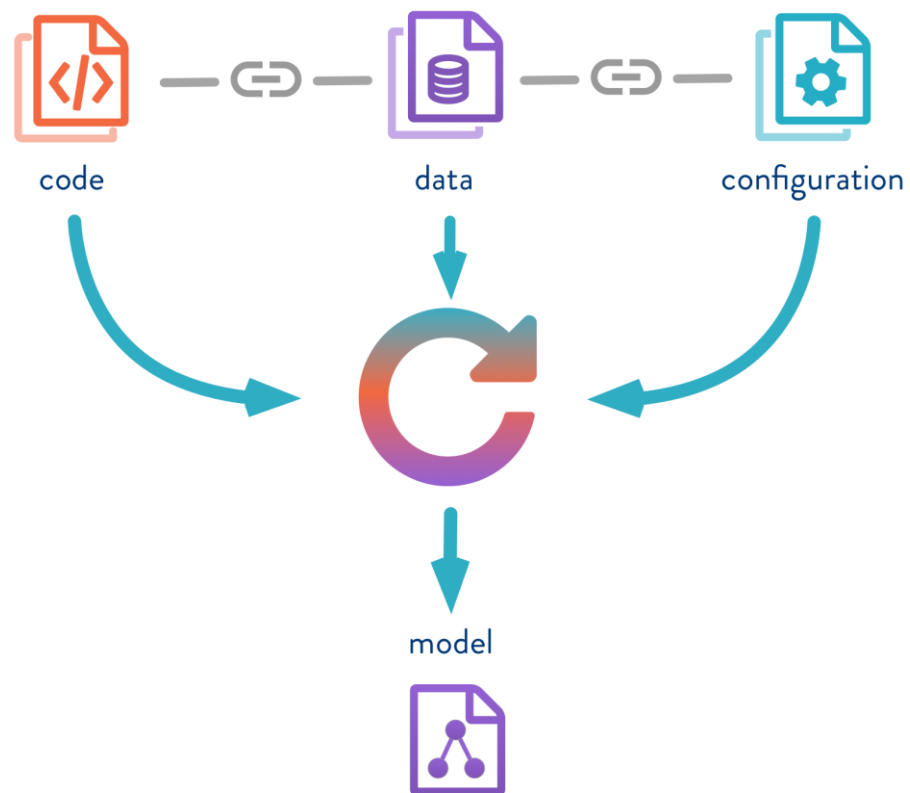
НА СЕГОДНЯ

- Git для DS



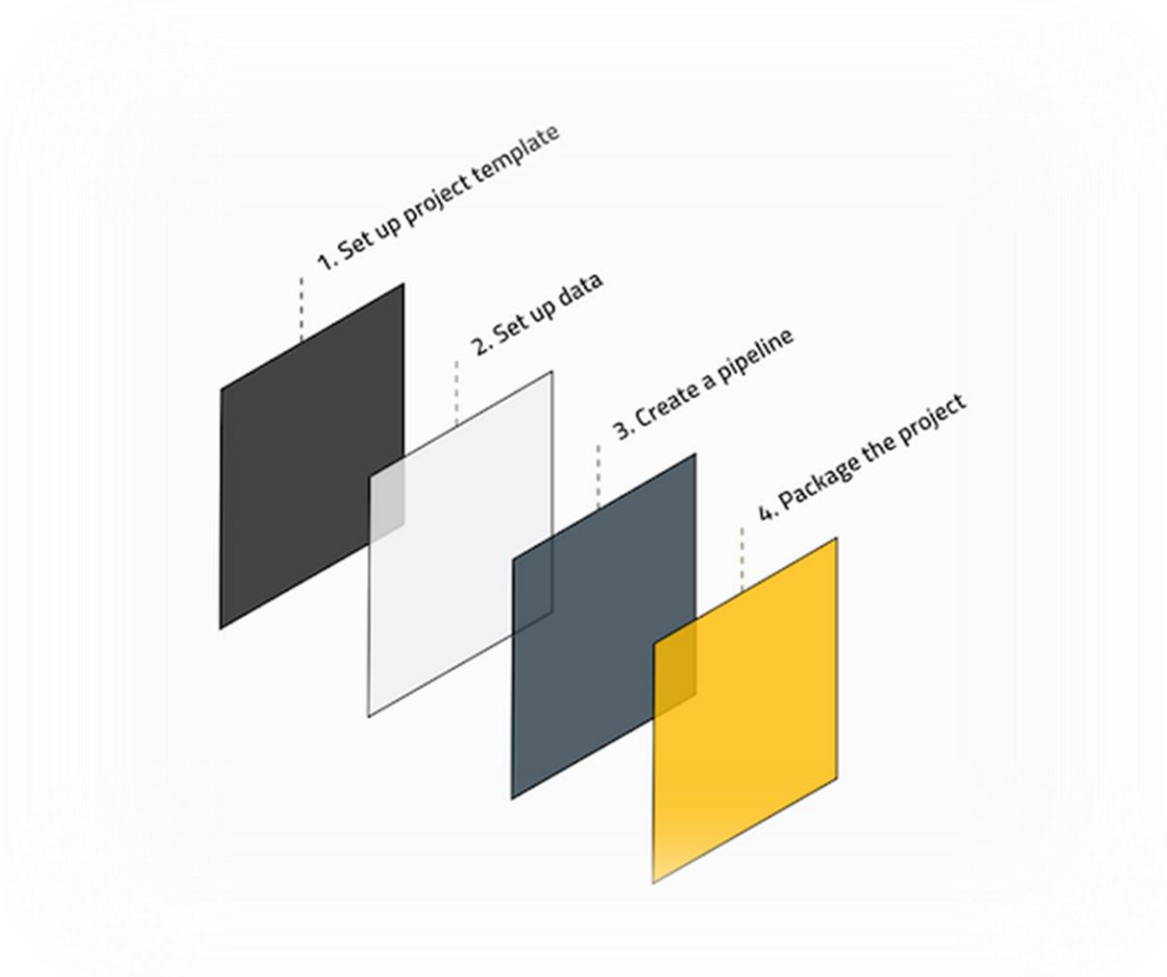
НА СЕГОДНЯ

- Git для DS
- DVC



НА СЕГОДНЯ

- **Git для DS**
- **DVC**
- **Kedro**



GIT (ДЛЯ DS/DE)



GIT

! Определитесь с терминологией !

Проект - ?

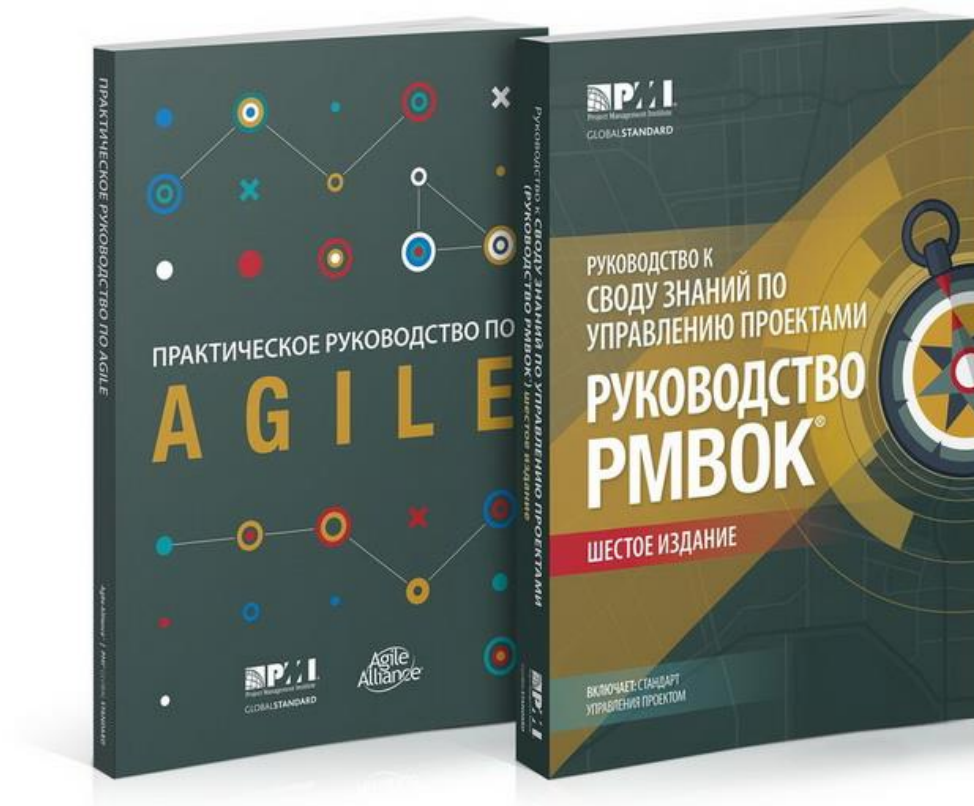
Задача - ?

GIT

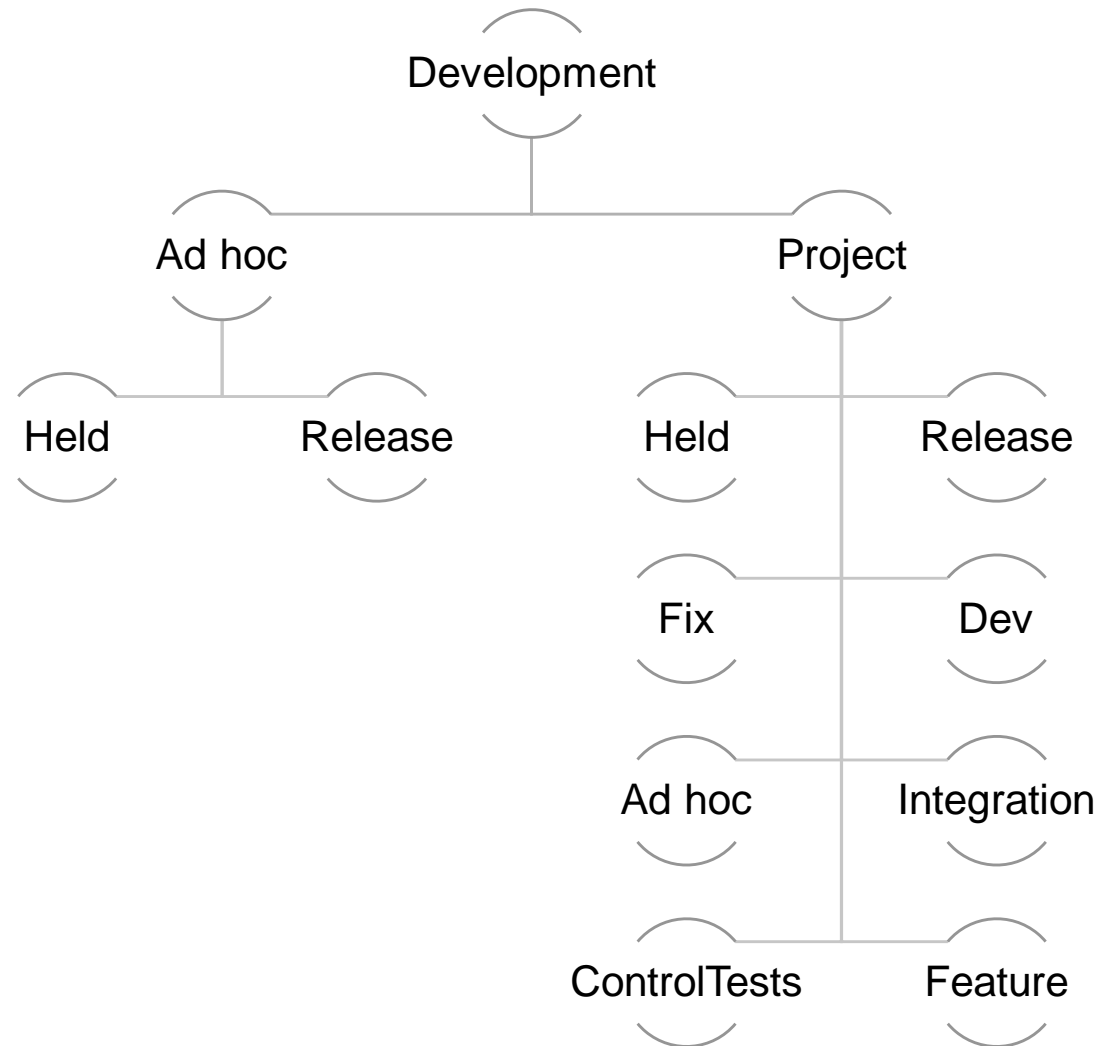
! Определитесь с терминологией !

Проект - ?

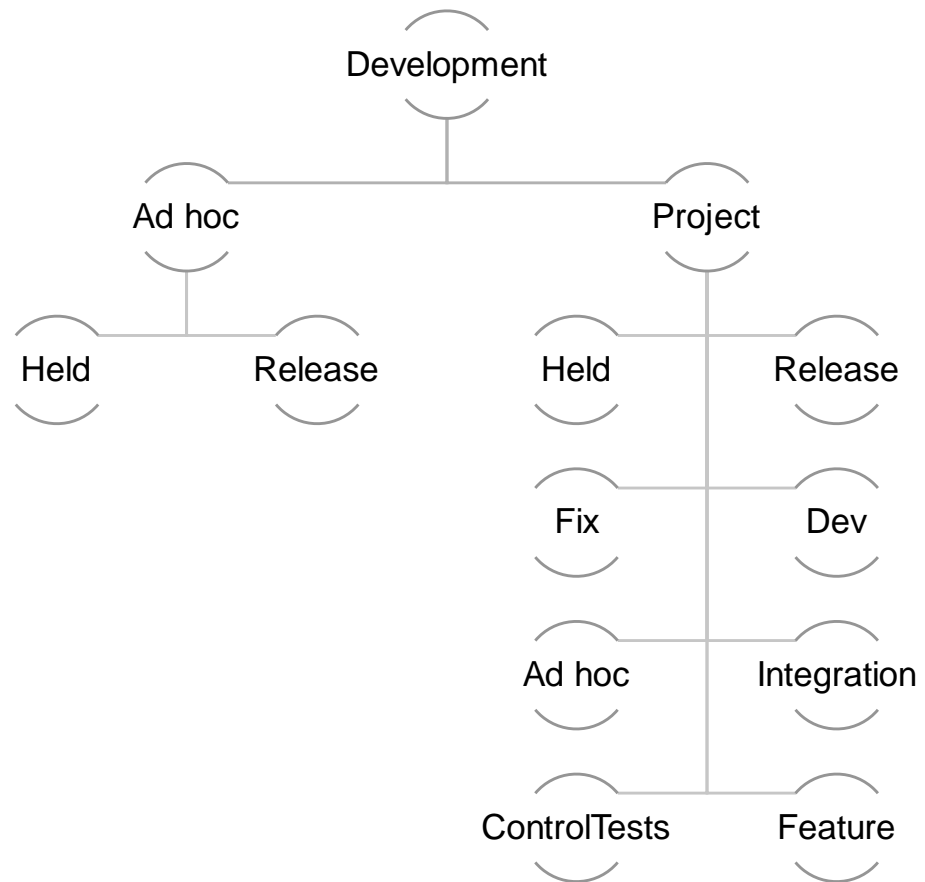
Задача - ?



GIT

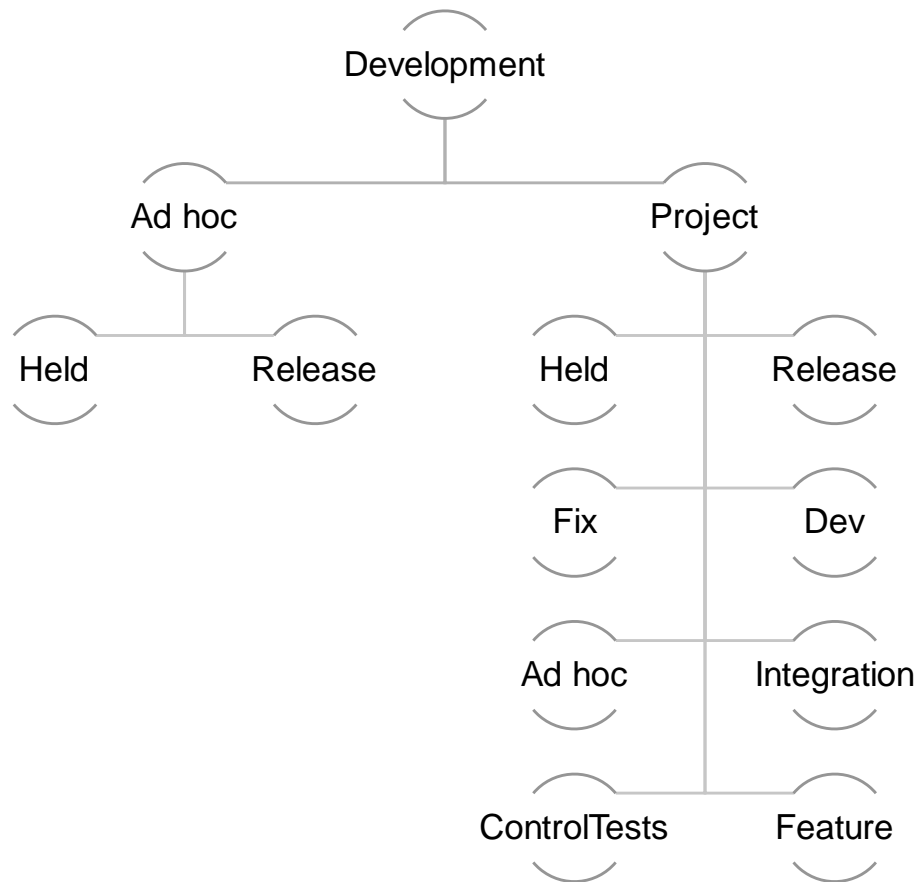


GIT



- ▶ config
- ▶ data
- ▶ docs
- ▶ notebooks
- ▶ profiling
- ▶ reports
- ▶ src
- gitignore
- __init__.py
- main.py
- requirements.txt
- tox.ini
- VERSION

GIT



Задание из task-manager.
В Jira: jiraspace.ru/some/**YOU_COMMAND_SPACE-999**

you_command_space – выше пространство
you_command_space == task

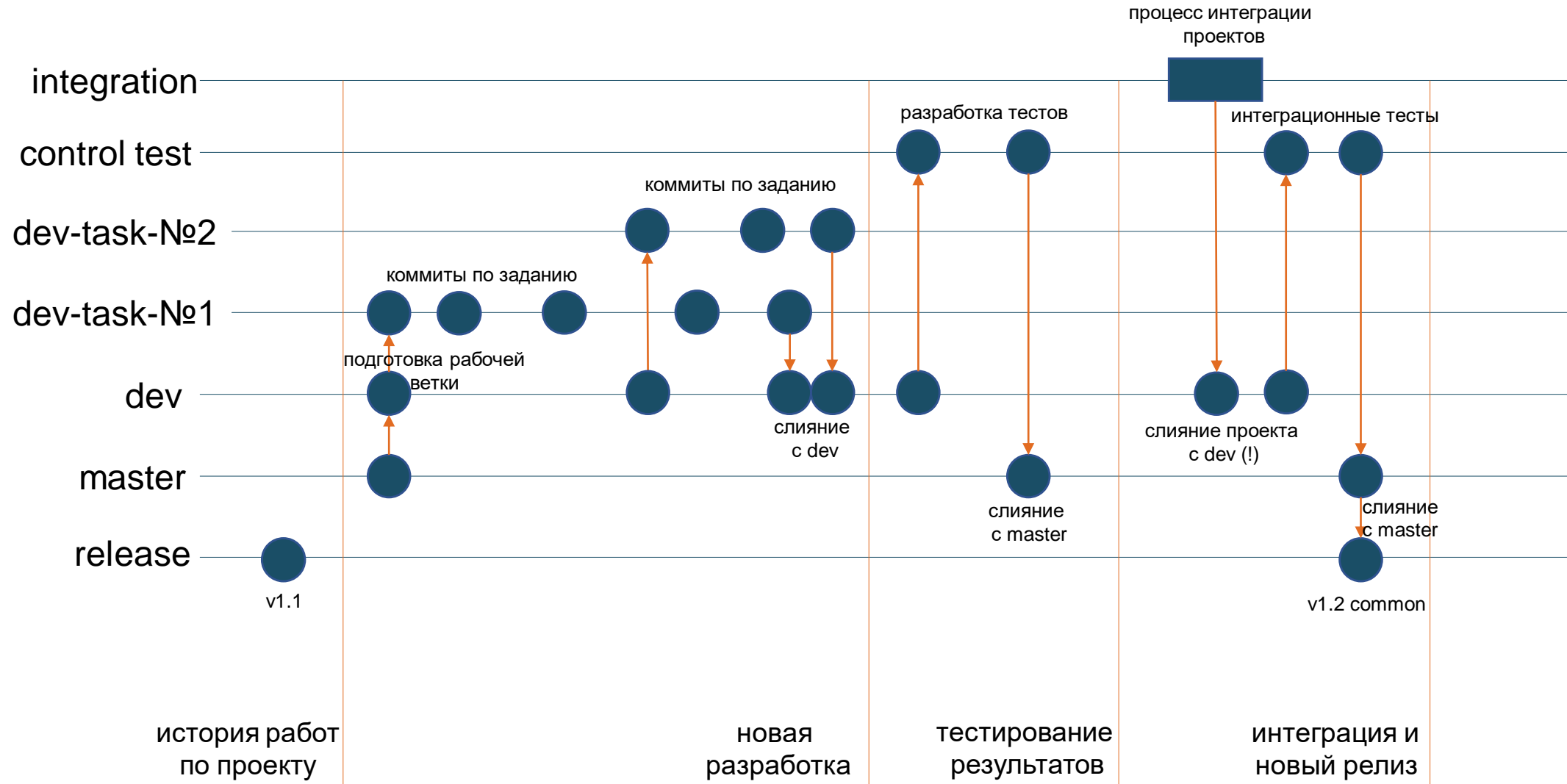
dev-task-№

ИМЯ ВЕТКИ

YOU_COMMAND_SPACE-999

999 – номер задания
999 == №

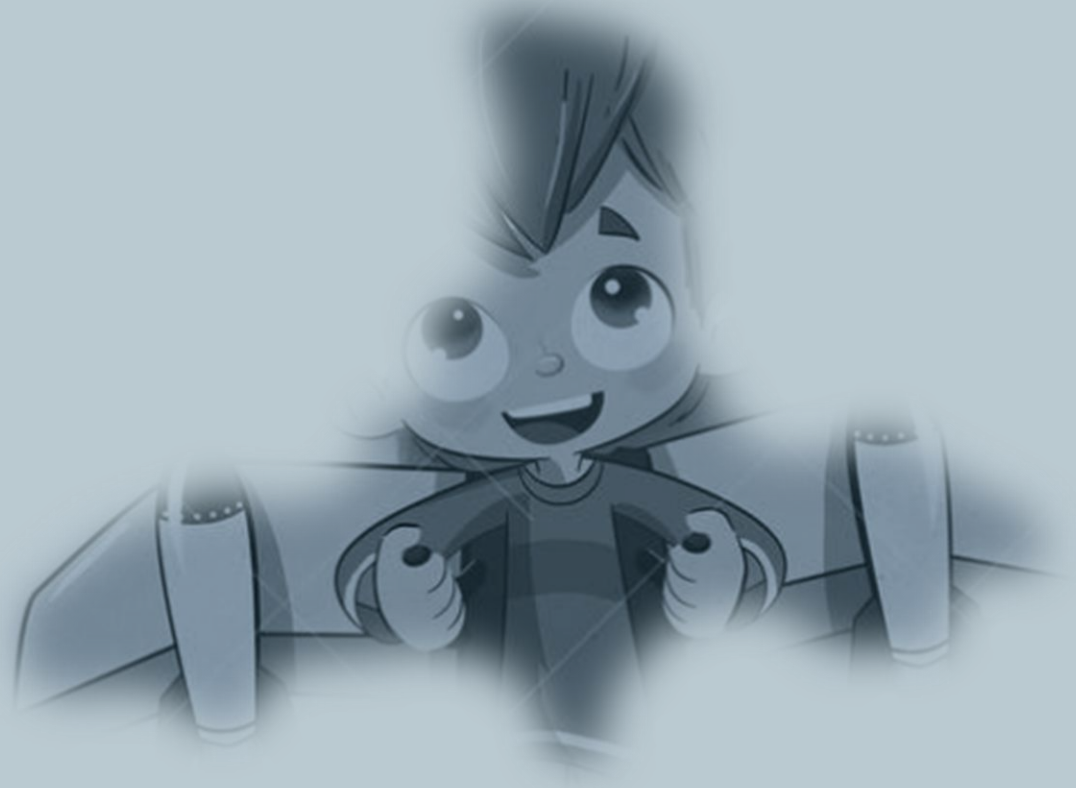
GIT





ВЕРСИОНИРОВАНИЕ ПРОЦЕССОВ

ЗАДАНИЕ



MODEL -> KEDRO

- Выбрать один из видов ml pipeline для дальнейшей работы
- Выбранный pipeline сделать согласно правилам Kedro
- Реализовать data engineer pipeline в Kedro
- Реализовать data science pipeline в Kedro
- Связать их работу и получить такой же скоринг, как от базового pipeline, который был взят за основу
- Результат:
 - Kedro визуализирует ваш pipeline
 - ML pipeline в Kedro показывает одинаковый скоринг с базовым pipeline для модели