

Больше платных материалов бесплатно на openssource.biz / openssource.info

ЗАНЯТИЕ 0.2

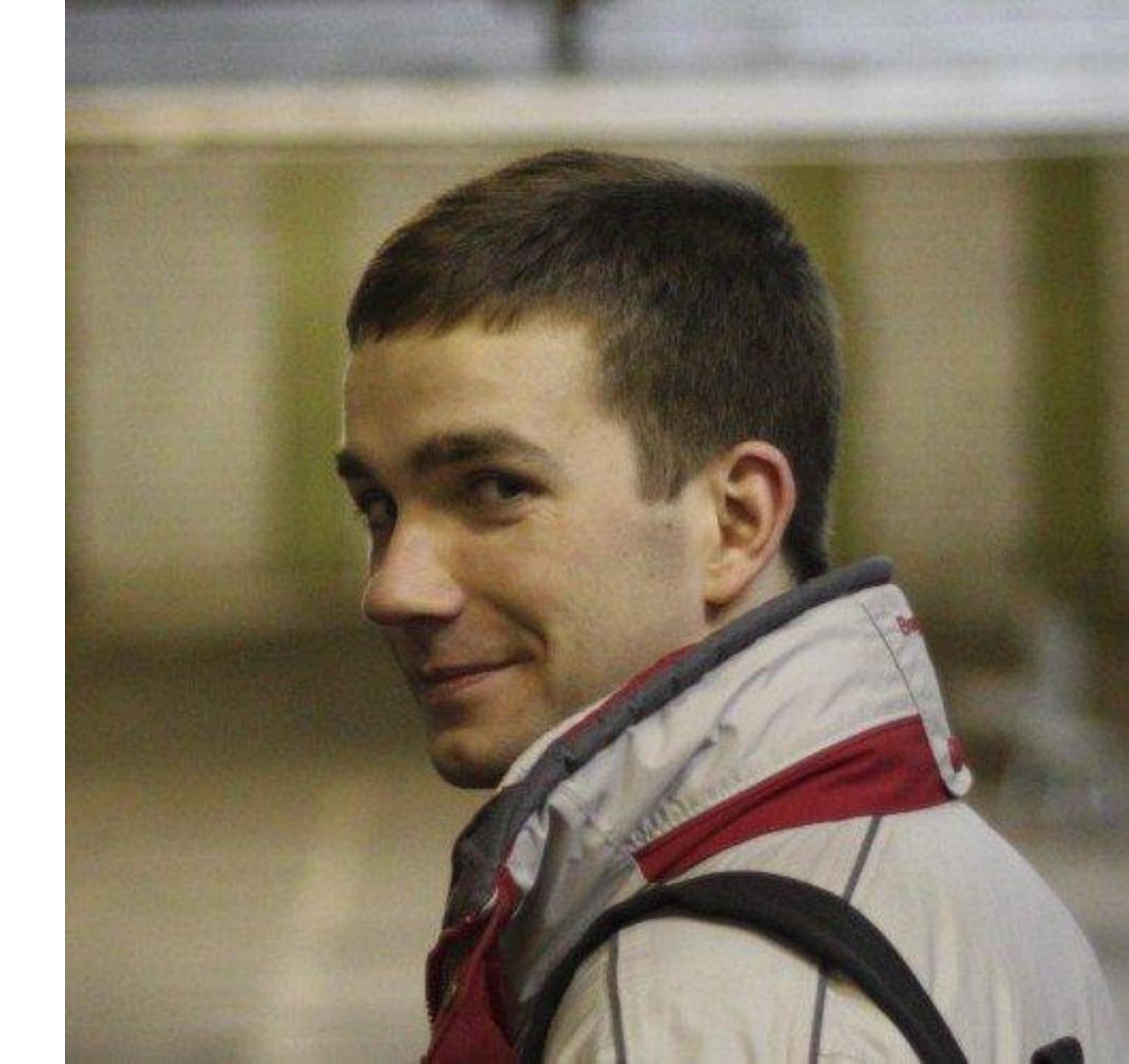
# GIT, PANDAS

### КОНСТАНТИН БАШЕВОЙ

«Ростелеком»

Habr

@kpi\_maker



# ЦЕЛИ ЗАНЯТИЯ

### В КОНЦЕ ЗАНЯТИЯ ВЫ СМОЖЕТЕ:

- создавать и обновлять репозитории;
- использовать ветки для эффективного управления версиями;
- использовать основные команды Pandas, чтобы не вспоминать про Excel.

# О ЧЁМ ПОГОВОРИМ И ЧТО СДЕЛАЕМ

- 1. Создание и обновление репозитория;
- 2. Создание веток и merge
- 3. Практикум по Pandas

GIT

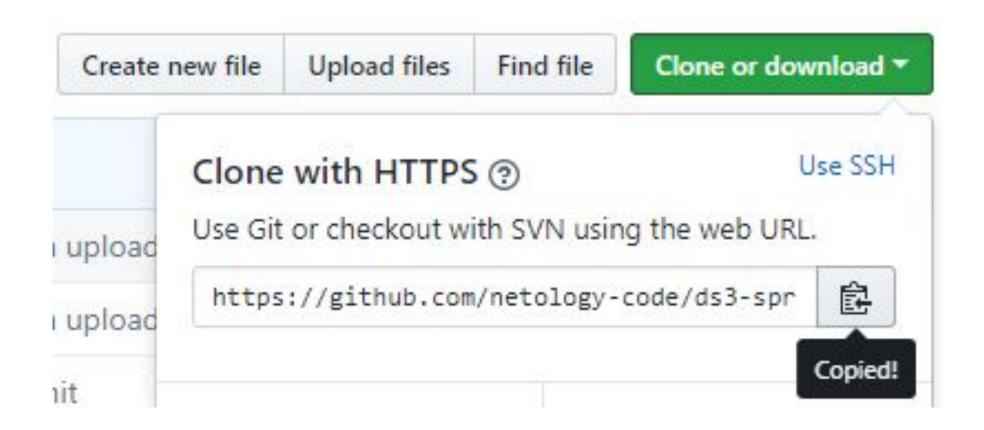
### СИСТЕМА КОНТРОЛЯ ВЕРСИЙ







### КЛОНИРУЕМ РЕПОЗИТОРИЙ СЕБЕ



# из командной строки Git
git clone <a href="https://github.com/netology-code/ds3-spring-2018.git">https://github.com/netology-code/ds3-spring-2018.git</a>
# заходим в скачанную папку
cd ds3-spring-2018/

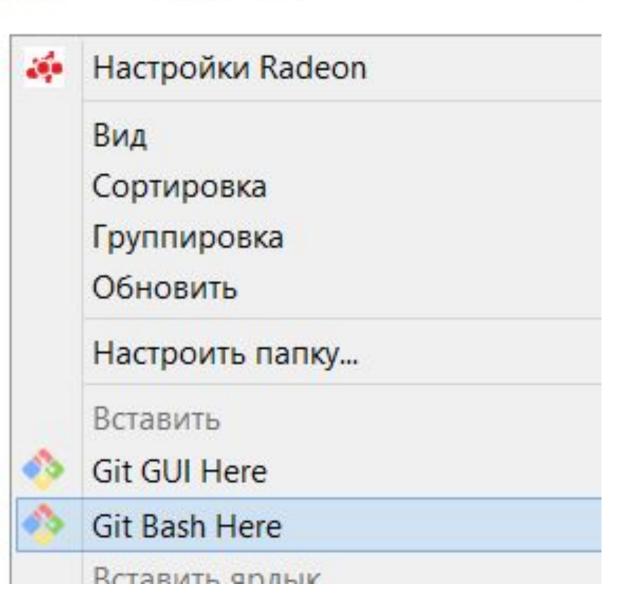
### ОБНОВЛЯЕМ РЕПОЗИТОРИЙ

В папке, где есть папка .git

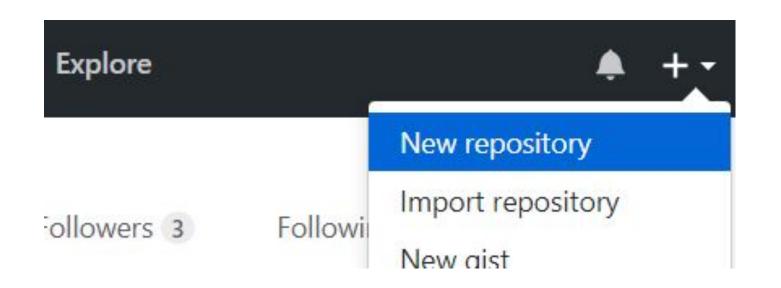
git pull

## СОЗДАЕМ СВОЙ РЕПОЗИТОРИЙ

my awesome code.py	07.10.2017 20:05	Файл "РҮ"	1
readme.md	07.10.2017 20:05	Файл "MD"	1

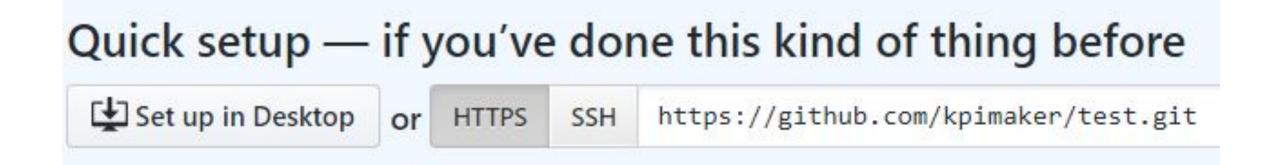


### СОЗДАЕМ СВОЙ РЕПОЗИТОРИЙ



# из командной строки Git
git init – создаем папку репозитория на диске
git add . – добавляем все файлы из папки
git commit –m "Test commit test version" – создаем коммит
git status – проверяем, что нужные файлы попали в коммит

### СОЗДАЕМ СВОЙ РЕПОЗИТОРИЙ

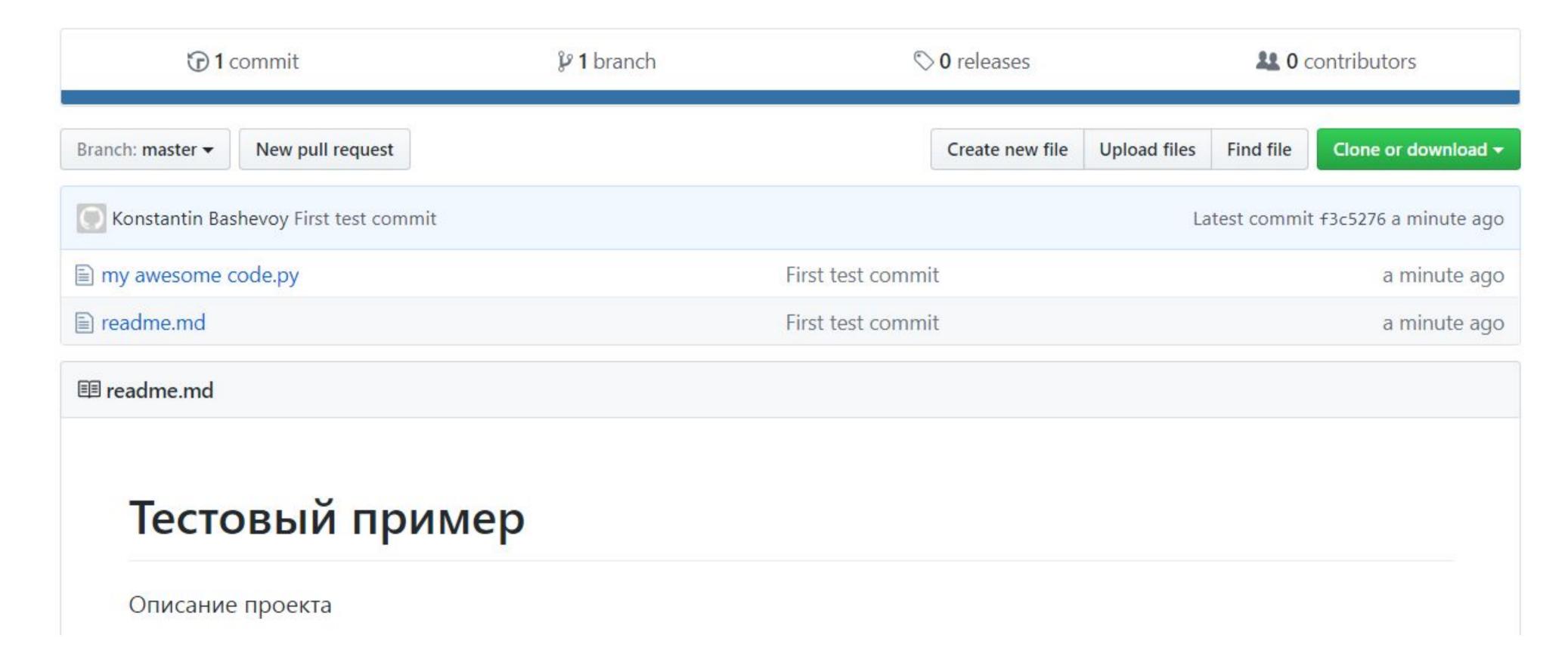


git remote add origin https://github.com/kpimaker/test.git - указываем какой репозиторий хотим обновить

git push -u origin master

Больше платных материалов бесплатно на openssource.biz / openssource.info

### СОЗДАЕМ СВОЙ РЕПОЗИТОРИЙ

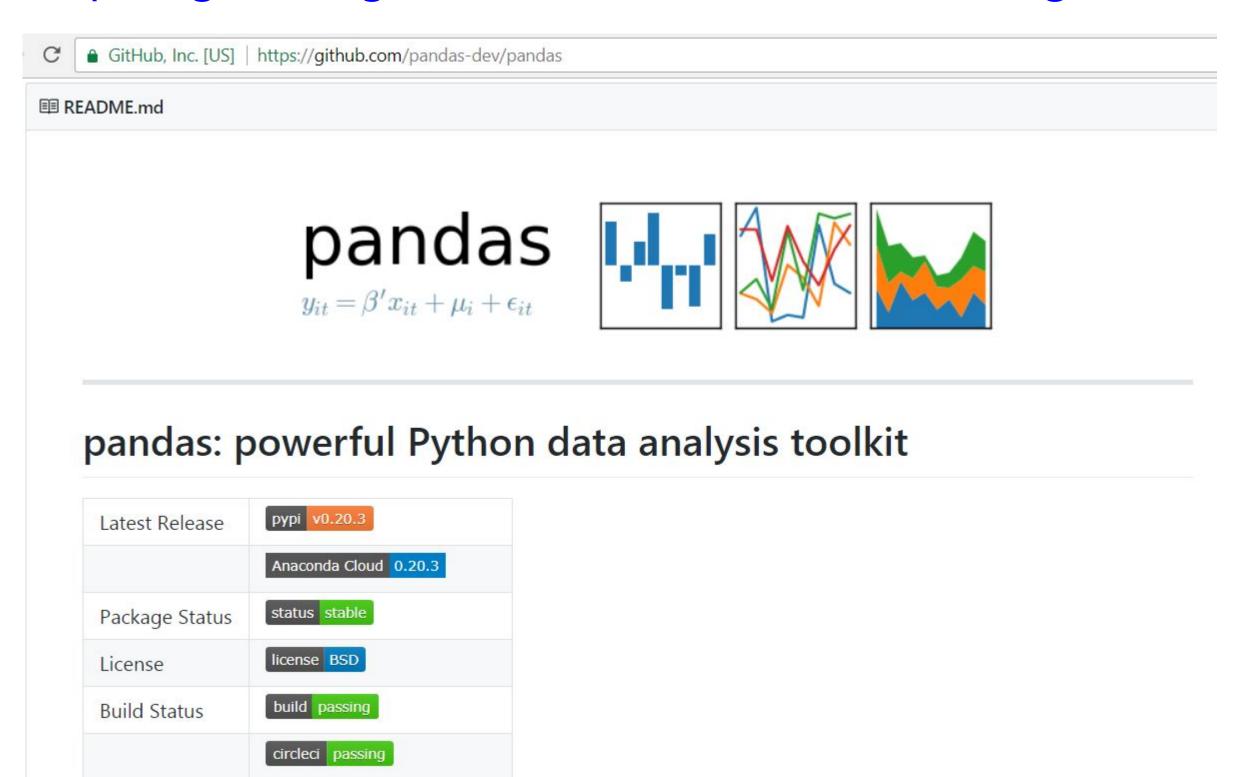


# README.MD

### README.MD

o build passing

#### https://quides.github.com/features/mastering-markdown/



### README.MD

III readme.md

#### Отправка сообщений и стикеров от бота в Телеграме

#### Перед началом работы

- 1. Внесите токен существующего Телеграм-бота в файл config.yaml (раздел telegram --> bot1). Нового бота можно завести, написав в Телеграме сообщение @BotFather.
- 2. Установите библиотеку для работы с Telegram API (https://github.com/python-telegram-bot/python-telegram-bot). В командной строке выполните: pip install python-telegram-bot --upgrade
- 3. Узнайте параметр chat\_id и запишите его в файл config.yaml (раздел users, пользователя "default" c chat\_id 0 лучше оставить). Для этого напишите любое сообщение вашему боту. Затем запустите файл get\_chat\_id.py:

```
python get_chat_id.py
Last message chat ID: 121314151 Ivan Ivanov
```

#### Например:

```
users:
me: 121314151
default: 0
```

#### Быстрый старт:

#### Что очень желательно

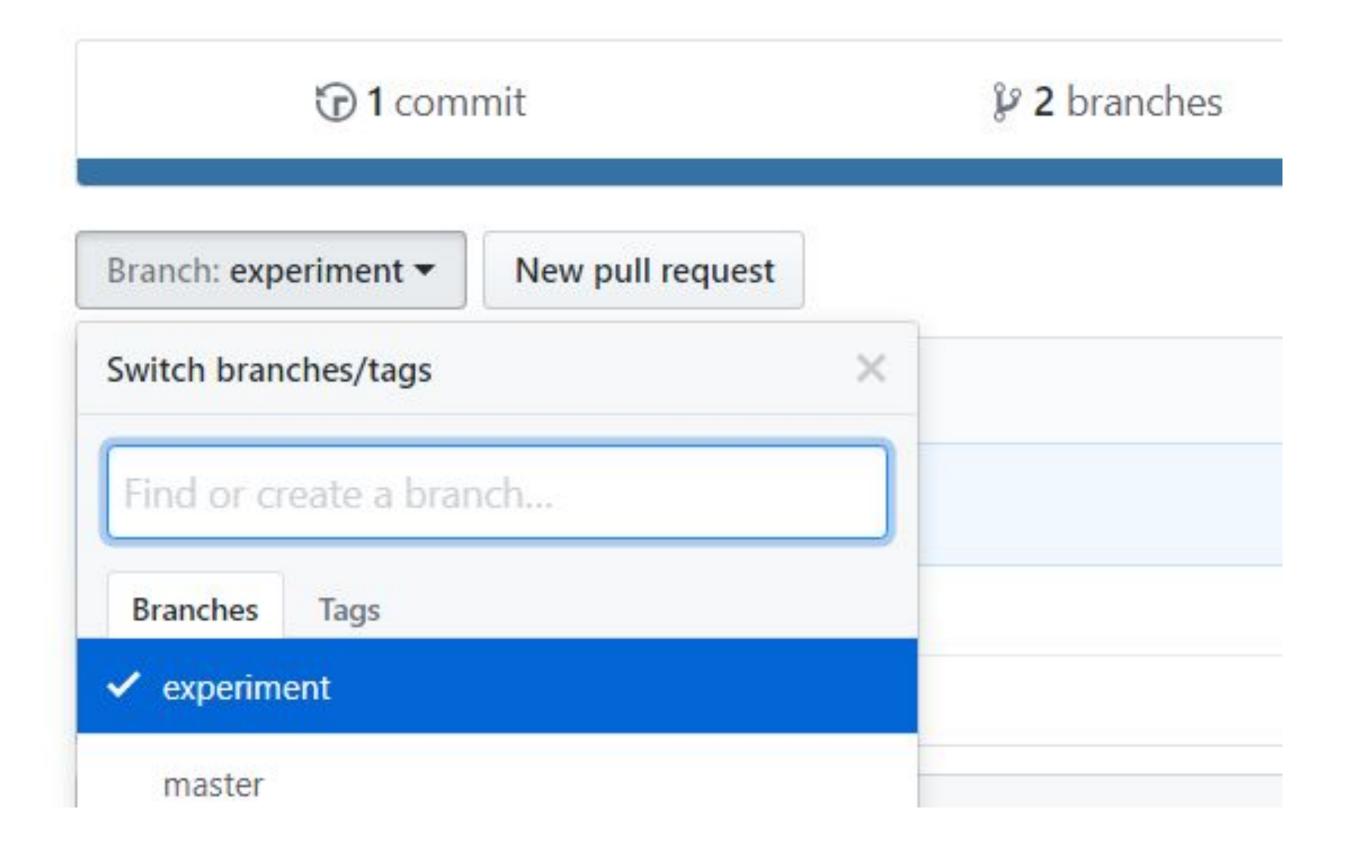
- что делает проект
- быстрый старт
- описание методов

# BRANCH

### ВЕТКИ



### СОЗДАНИЕ ВЕТКИ



### ПЕРЕКЛЮЧАЕМСЯ НА ВЕТКУ

Пока локальная копия ничего не «знает» о новой ветке.

git pull – обновляем локальную копию репозитория
git checkout experiment – переключаемся на ветку experiment
git branch – проверяем, что сейчас мы в ветке experiment
git add –u – добавить в коммит все обновленные файлы
git commit -m 'Новый клевый алгоритм' – делаем коммит в ветку experiment
git push origin experiment – публикуем новый код

### ОБЪЕДИНЯЕМ КОД С MASTER

Убедились, что новый алгоритм работает.

Тогда:

git checkout master – переключаемся на основную ветку

git merge experiment – объединяем ветки

git push origin master – обновляем код на master

Больше платных материалов бесплатно на openssource.biz / openssource.info



# PANDAS



городской округ Первомайск, Нижегородская область, Россия

54.926792, 44.064449









Искать рядом:









Рестораны Кафе Музеи Отели



### ПЕРЕСЕЧЕНИЕ ТАБЛИЦ

Партнер 1 - 71 951 запись

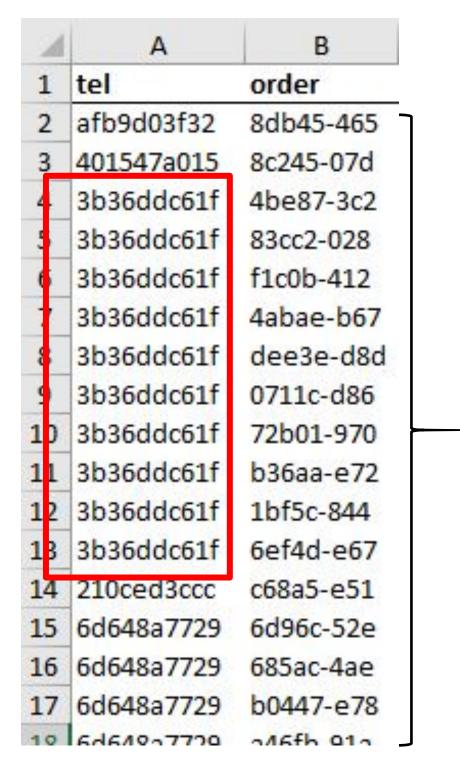
Γ	Тартнер	2 –	340	296	записей	1
	Idpilicp			200	Janvicev	

1	Α	В	С	D	E F
1	tel	order	region	city	address
2	d76182a41a	dbd28-2a6	Α	Сусуман	Кирова 10
3	d4d482c479	5f375-8c9	В	Якутск	Турку 3
4	113ab84955	b9ce9-caf	C	Пионерский	Лермонтовская 5
5	92bb06cf14	9bad8-af4	A	Владивосток	Красноармейская 7
6	41a27123c7	77c85-fd6	D	Донецк	Солнечный 8
7	41a27123c7	db644-dc0	E	Нальчик	Набережная 11
8	19faa65c0e	6fb0b-eaf	F	Магнитогорск	Первых 9
9	19faa65c0e	6fb0b-eaf	F	Шемышейка	Хабаровская 4
10	bee22d0e4e	738d4-08f	G	Находка	Солнечный 1
11	162a7ce84a	db20e-61f	G	Александровск-Сахалинский	Фрунзе 10
12	162a7ce84a	adb44-889	F	Александровск-Сахалинский	Бумажная 7
13	466ccef8c8	dbf71-f5d	С	Большой	Приморская 1
14	4af51d12a0	d5f99-9e5	F	Вешкайма	Коллонтай 3
15	289ab94588	a3e6b-7e5	F	Биробиджан	Коллонтай 8
16	e3e6ff9214	d2dcd-1d4	F	Сусуман	улица 5
17	e3e6ff9214	d2dcd-1d4	F	Сергиев	Первых 7
12	20h/1762112	/11020-5cc	F	Науолиа	Коллонтай 3

4	Α	В
1	tel	order
2	afb9d03f32	8db45-465
3	401547a015	8c245-07d
4	3b36ddc61f	4be87-3c2
5	3b36ddc61f	83cc2-028
6	3b36ddc61f	f1c0b-412
7	3b36ddc61f	4abae-b67
8	3b36ddc61f	dee3e-d8d
9	3b36ddc61f	0711c-d86
10	3b36ddc61f	72b01-970
11	3b36ddc61f	b36aa-e72
12	3b36ddc61f	1bf5c-844
13	3b36ddc61f	6ef4d-e67
14	210ced3ccc	c68a5-e51
15	6d648a7729	6d96c-52e
16	6d648a7729	685ac-4ae
	6d648a7729	
10	64640-7720	2/6fh 912

### EXCEL – 20 MИНУТ

Строки имеют дубликаты



340 тысяч строк за 2 недели Надо брать 3 месяца за раз

### PANDAS – 5 КОМАНД И 2 СЕКУНДЫ

```
%%time
import pandas as pd

channel1 = pd.read_csv( 'channel_1_data.txt', sep = '\t' )
    channel2 = pd.read_csv( 'channel_2_data.txt', sep = '\t' )

channel1.drop_duplicates( subset = 'tel', keep = 'first', inplace = True )
    channel2.drop_duplicates( subset = 'tel', keep = 'first', inplace = True )

intersect = pd.merge( channel1, channel2, on = 'tel', how = 'inner', suffixes = ['_channel_1', '_channel_2'] )
```

Wall time: 1.16 s



### LIFETIME

userld	movield	rating	timestamp
1	31	2.5	1260759144
1	1029	3.0	1260759179
1	1061	3.0	1260759182
1	1129	2.0	1260759185
1	1172	4.0	1260759205

26 000 000 строк 693 Мб

Больше платных материалов бесплатно на openssource.biz / openssource.info

### 20 СЕКУНД

#### Подсчет LT для 26млн записей

```
"""
import pandas as pd

ratings = pd.read_csv( 'ratings.csv' )

ltData = ratings.groupby( by = ['userId'] ).agg( [min, max] ).reset_index()
ltData['diff'] = ltData['timestamp']['max'] - ltData['timestamp']['min']

LT = ltData['diff'].mean()

CPU times: user 15.7 s, sys: 2.79 s, total: 18.5 s
Wall time: 20.4 s

LT / 3600 / 24
```

126.22761148027355



# Спасибо за внимание!

### КОНСТАНТИН БАШЕВОЙ

Habr: @kpi\_maker