



NEW  
PRO  
LAB

# СБЕРБАНК

## APACHE SPARK ДЛЯ АНАЛИЗА ДАННЫХ



# КОРОТКО О НАС

- Занимаемся обучением работе с данными с 2015 года.
- Предоставляем широкую линейку программ про работу с данными.
- Используем андрагогику и интенсивные практикоориентированные программы.

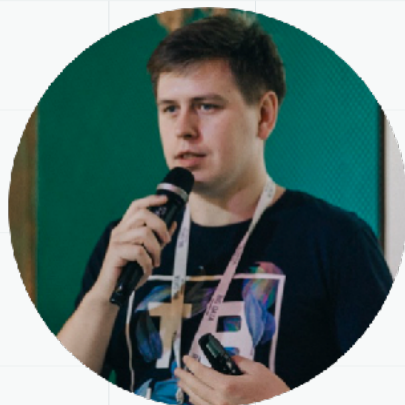
# ПРИНЦИПЫ НАШИХ ПРОГРАММ

Мы – школа по работе с данными и лаборатория:

- Project Based Education: обучаем работе с инструментами и технологиями на реальных задачах – лабах.
- Предоставляем работу с облачным кластером.
- Соблюдаем баланс между самостоятельностью и поддержкой

# КОМАНДА ПРОГРАММЫ

# Преподаватели



**Егор Матешук**

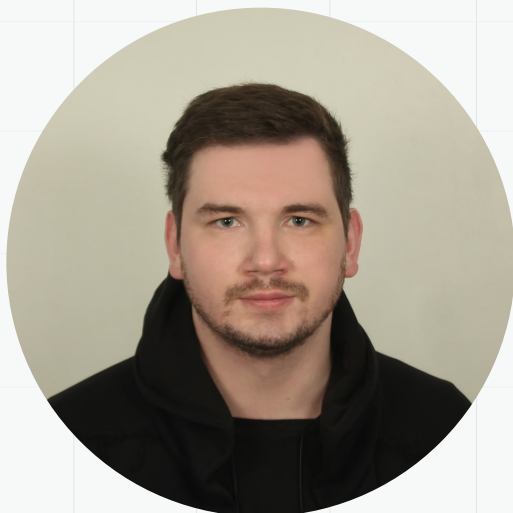
Технический директор  
**ГПМ Дата**



**Сергей Гришаев**

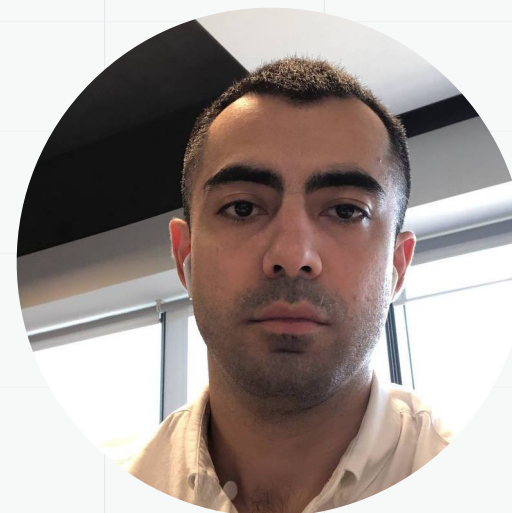
Senior Data Scientist  
**Сбермаркет**

# Координаторы



**Эскендер Джемадинов**

Data Scientist  
CDEK IT



**Назим Джавадов**

Data Science Manager  
Visa



# Менеджер



**Ольга Сафронова**

Коммуникация с участниками

# О ПРОГРАММЕ

# Цели

Научить работе со Spark для различных задач по анализу данных: от их предобработки до построения моделей машинного обучения и использования в real-time приложениях

# Формат

- 5 недель, 2 занятия в неделю
- Каждое занятие: 3 часа в zoom с 19:00 до 22:00
- Лабы и тесты

1. Общий на всех кластер со Spark 2.4.7.
2. Конфигурация: 18 нод по 16 CPU, 64GB RAM. 2 мастера с 32 ядрами и 256GB RAM.
3. Доступ к кластеру по SSH и через JupyterHub.
4. Личный кабинет с календарем занятий и чекерами для лаб.
5. GitHub (**доступ!**).
6. Slack.

1. Дескриптивный анализ рейтингов фильмов на RDD.
2. Content-based рекомендательная система с использованием датафреймов.
3. Рекомендация фильмов на основе данных об истории телесмотрения (соревнование).
4. Прогнозирование пола и возрастной категории пользователей на основе логов посещений в real-time.
5. Прогнозирование оттока клиентов.

# УСПЕШНОСТЬ ПРОХОЖДЕНИЯ

# Тесты

Ответить верно суммарно на 60% и более всех вопросов (всего 75 вопросов – **ответить на 45 и более**)

На прохождение каждого теста дается 10-15 минут и 2 попытки (тесты на платформе [flexiquiz.com](https://flexiquiz.com), логин и пароль как от лк [lk-spark.newprolab.com](https://lk-spark.newprolab.com))



# Лабы

Выполнить 60% и более лаб в срок – три любых лабы

**Сертификат об успешном прохождении программы**

Выполнено 4 из 5 (набрать не **менее всего 85 баллов**)

**Сертификат о прохождении с отличием**

Выполнено 5 из 5 (набрать **95 баллов и более**)



# ЛИДЕРБОРД

# Лидерборд

По итогам программы будет общий рейтинг по баллам  
(тесты + лабы)



# СТАДИ БАДДИ

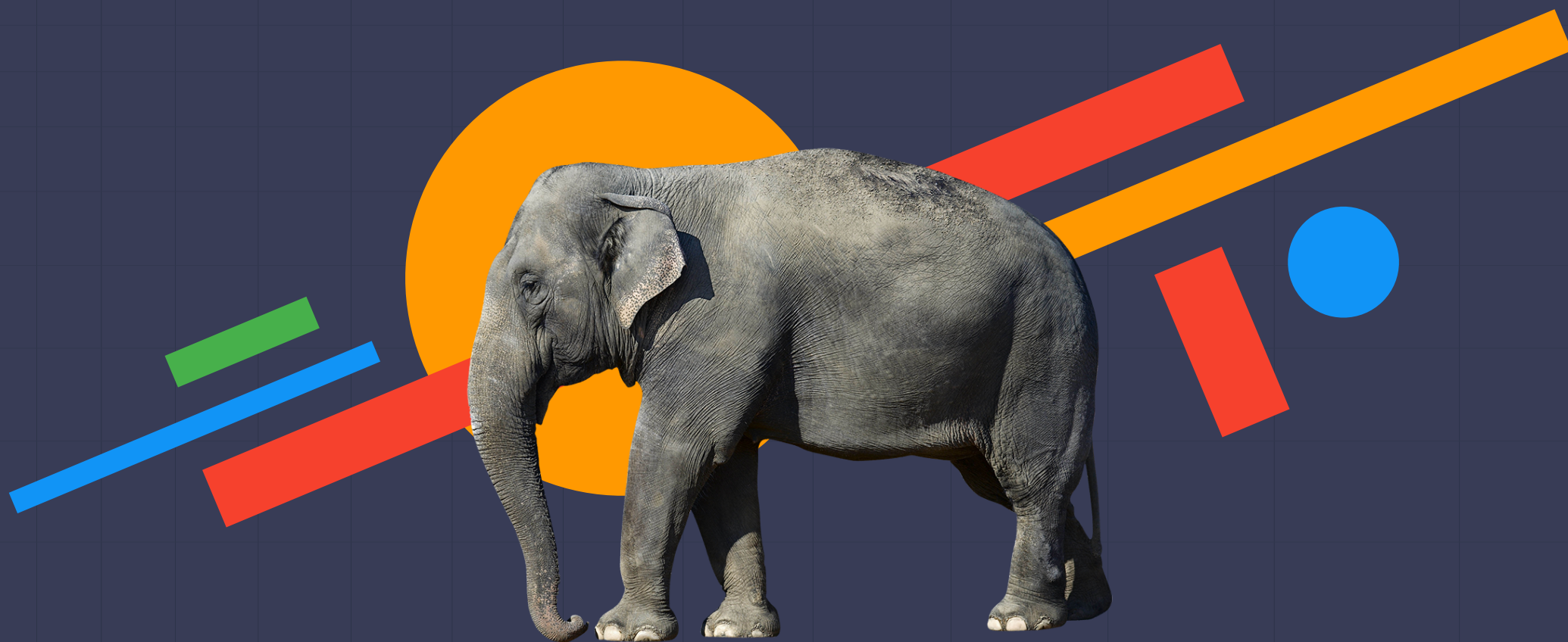
# Стади Бадди

Можно объединиться в команду до 3-х человек и выполнять лабы вместе (чекером проверяем индивидуально) – **обсуждать, помогать, поддерживать**

Можно делать индивидуально. Но вместе интереснее!



# ВАШИ ВОПРОСЫ



# Big Data is Love

NEWPROLAB.COM