



APACHE SPARK ДЛЯ АНАЛИЗА ДАННЫХ 11.0

О программе



1. КОРОТКО О НАС



Newprolab

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

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.

2. ФОРМАТ ПРОГРАММЫ

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

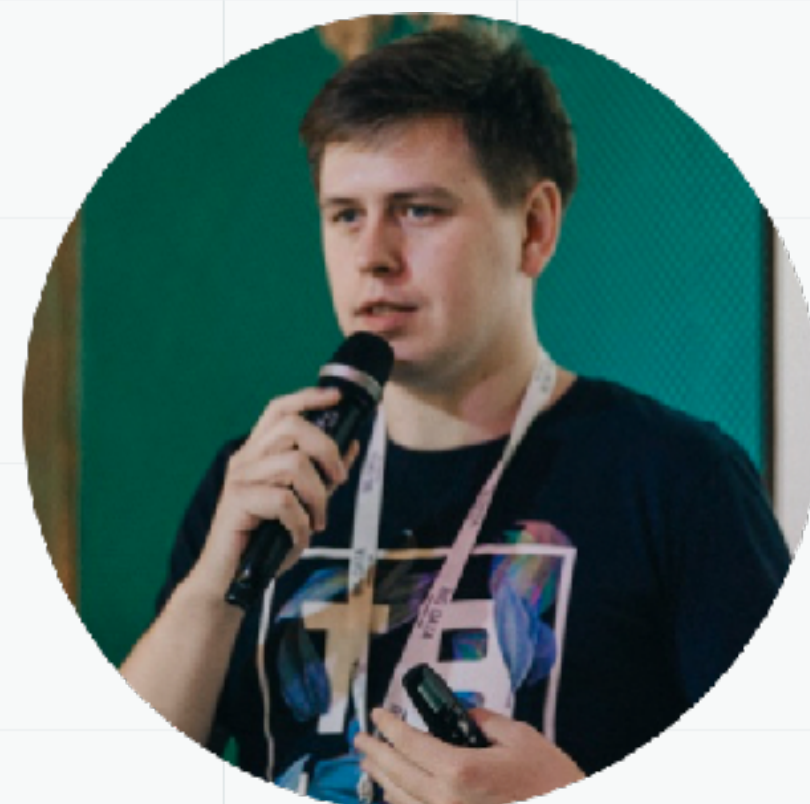


Входные требования

Что нужно знать и уметь:

- Знание Python, SQL
- Базовые знания Linux
- ML как плюс

Преподаватели



Егор Матешук

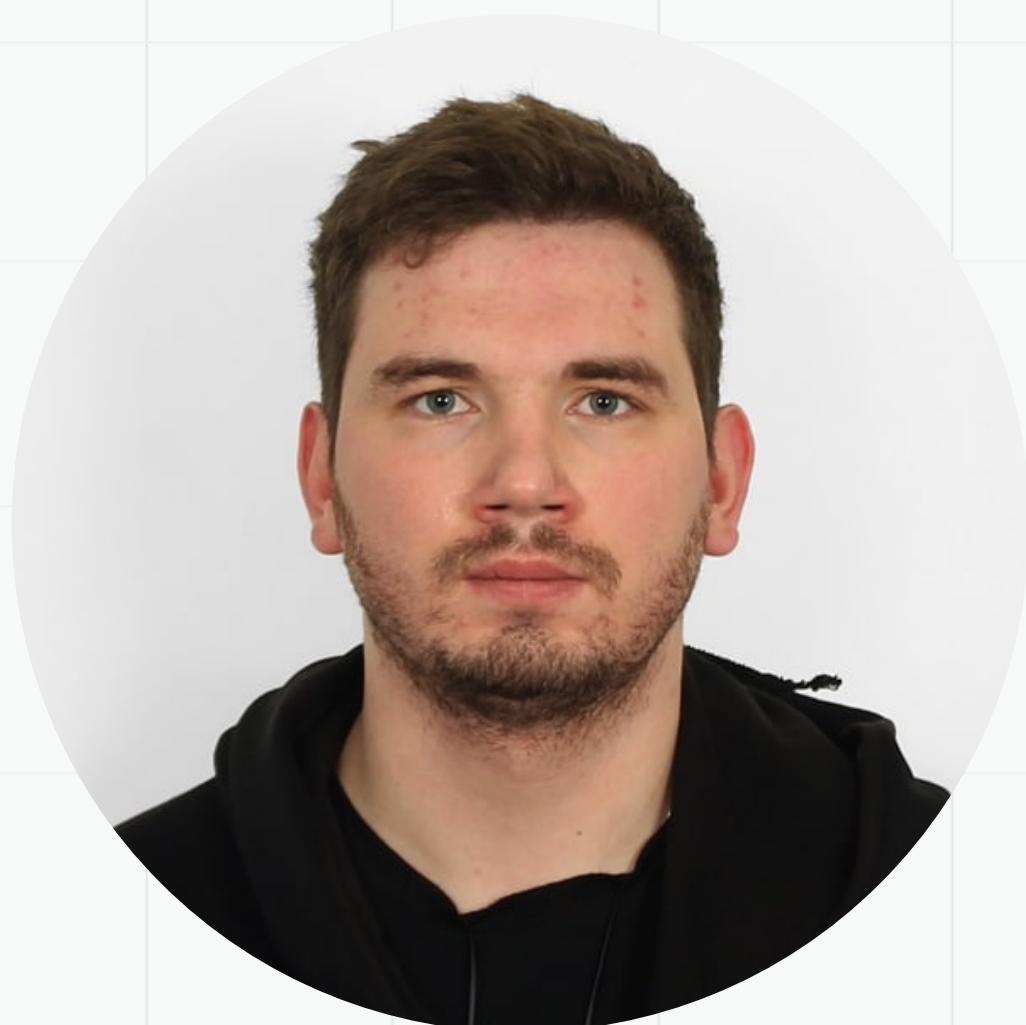
Технический директор
ГПМ Дата



Сергей Гришаев

Architect
Сбермаркет

Координаторы



Эскендер Джемадинов

Data Scientist



Назим Джавадов

Data Science Manager

Менеджер



Ольга Сафронова

Коммуникация с участниками



Формат программы

- 6 недель
- 2 занятия в неделю в зуме с 19:00 до 22:00
- В середине программы – неделя без занятий
- 5 лаб (практических домашних задач)
- 8 тестов
- Study buddy



Study Buddy

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

Можно делать индивидуально. Но вместе интереснее!



Ресурсы

1. Общий на всех кластер со Spark 2.4.7. Конфигурация: 18 нод по 16 CPU, 80GB RAM. 2 мастера с 32 ядрами и 256GB RAM.
2. Доступ к кластеру по SSH и через JupyterHub.
3. Личный кабинет с календарем занятий и чекерами для лаб.
4. GitHub с материалами занятий.
5. Канал и чат в Телеграме
6. Записи занятий с тайм-метками в СберУниверситете



Успешность

- ответить суммарно на 60% вопросов и более во всех тестах
- выполнить любые 4 лабы из 5
- при выполнении 5 лаб из 5 - наш сертификат с отличием

Удостоверение

О ПОВЫШЕНИИ КВАЛИФИКАЦИИ

Настоящее удостоверение свидетельствует о том, что

с 3 октября 2022 г. по 12 декабря 2022 г. прошел(а) обучение
в АНО ДПО «Корпоративный университет Сбербанка»

по дополнительной профессиональной программе повышения квалификации

Apache Spark для задач дата инжиниринга

в объеме 116 академических часов

ПК-200671
Регистрационный номер 22-219772
Дата выдачи 12.12.2022
Город Москва

Директор по учебно-методической работе –
начальник центра уровневых программ



Шаталов А.И.





Certificate of Completion

WITH HONOURS



this is to certify that

DMITRIY SOKOLOV

has demonstrated exceptional dedication to mastering the science and technology of

APACHE SPARK FOR DATA SCIENCE

a 5-week course that covers Apache Spark, PySpark, Spark ML, Spark Streaming, Spark Dataframes

10.11.2022

Date of Issue

A handwritten signature in blue ink, appearing to read "Anna Zhovtis".

Anna Zhovtis
CEO, New Professions Lab

Certificate number: 30770915629684379527919377470

3. ТЕМЫ И ЗАНЯТИЯ

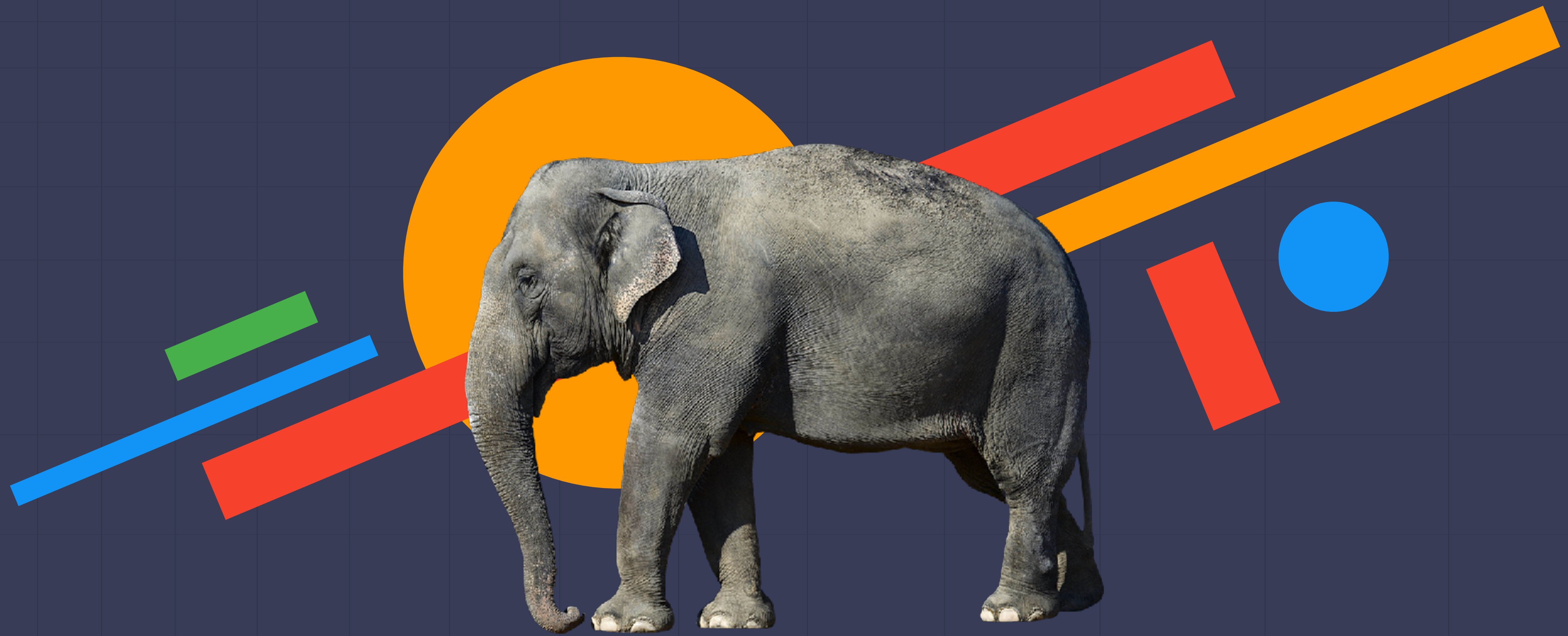


Занятия

- 
1. Hadoop для Spark-юзера
 2. Введение в Apache Spark
 3. Работа с DataFrames в Apache Spark
 4. Spark ML: основы и создание пайплайнов
 5. Практический ML на Spark. Решение кейсов с использованием Spark ML
 6. Spark Streaming
 7. Spark Scala API
 8. PySpark vs Python: Работа с графами
 9. Разбор лаб 1, 2 и 3. Q&A
 10. Разбор лаб 4 и 5. Q&A

4. ЛАБЫ

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



Big Data is Love