СБЕРБАНК

APACHE SPARK ДЛЯ АНАЛИЗА ДАННЫХ

NEW PRO LAB

КОРОТКО О НАС

Newprolab

- Занимаемся обучением работе с данными с 2015 года
- Самая широкая линейка программ про данные на рынке
- Используем андрагогику и сжатые практикоориентированные программы

История

Выпускники:

- 700+ с открытых программ
- 700 с корпоративных

3 Big Data

1 Deep Learning

2 Big Data

1 Data Engineer

3 for Executives

3 Deep Learning

2 Big Data

2 Data Engineer

2 for Executives

1 Deep Learning

2 Big Data

2 Scala

2 Data Engineer

3 for Executives

1 Deep Learning

2 Big Data

1 Scala

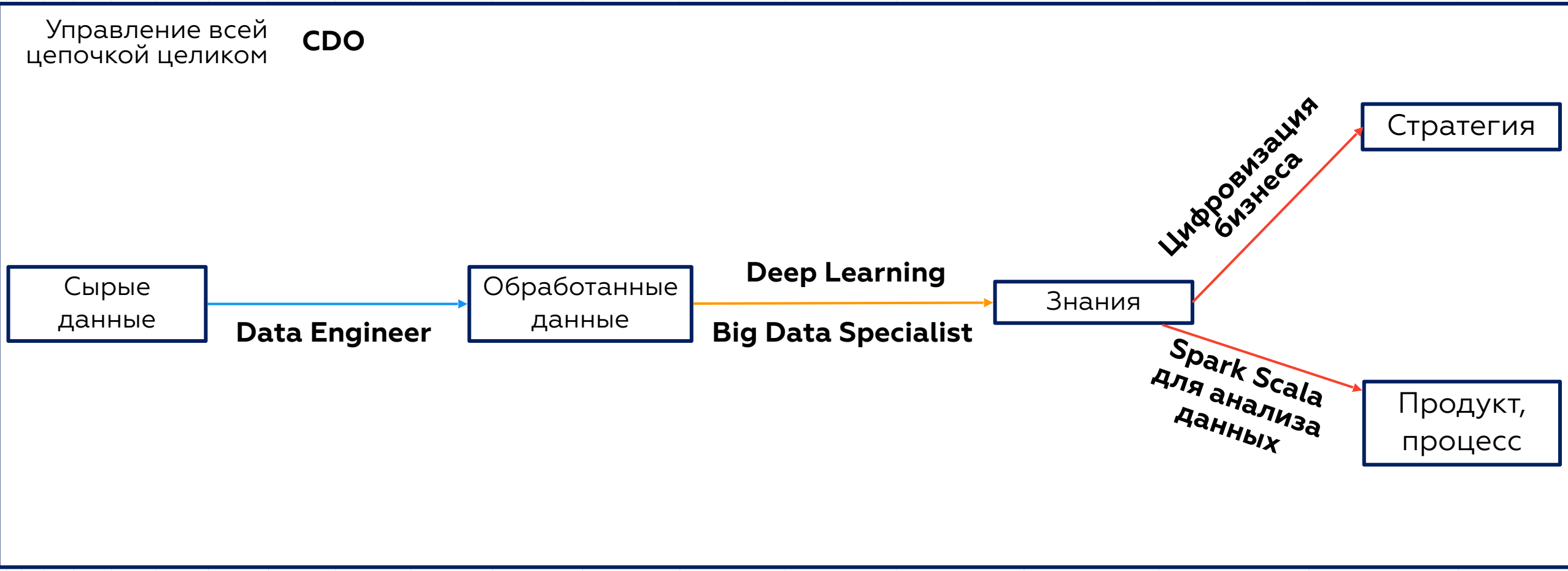
2 Data Engineer

Клуб **CDO**

2 Deep Learning

2 Big Data

Линейка программ



ПРИНЦИПЫ

Hero's journey

Hero’s journey

Главная идея

Баланс между самостоятельностью и поддержкой.

Другие принципы

- Занятия: задавайте вопросы (не существует глупых вопросов).
- Лабы: просите помощи у сокурсников и координатора, но вначале попробуйте решить сами.
- Делайте лабы заранее. В последний момент можно не успеть.
- Делайте лабы самостоятельно.

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
- 5 лаб (практических домашних задач)
- 3 теста

Занятия

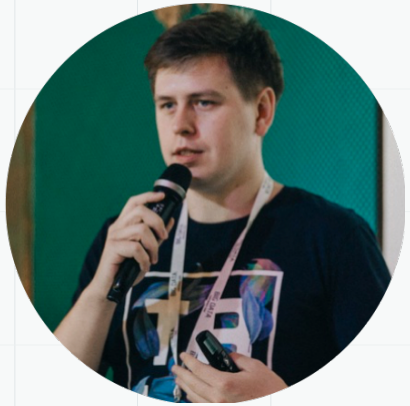
Неделя 1	Неделя 2	Неделя 3	Неделя 4	Неделя 5
Hadoop	Spark Dataframes	Практический ML	Labs & Q&A	Python vs PySpark на примере работы с графами
Введение в Spark	Spark ML	Spark Streaming	Spark Scala API	Labs & Q&A

Лабы

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.

КОМАНДА ПРОГРАММЫ

Преподаватели



Егор Матешук

CDO, Qvant

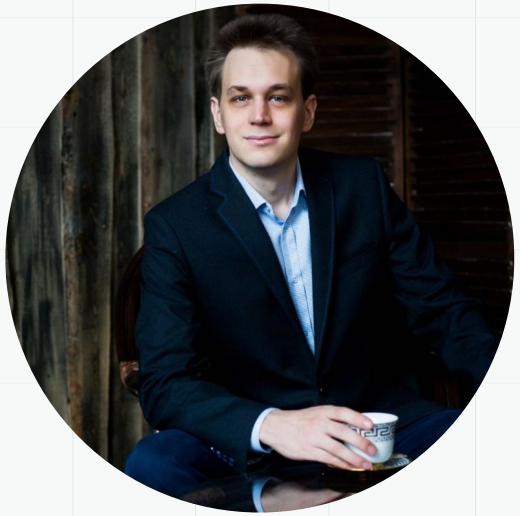
Наталья Притыковская

Senior Data Scientist,
Mechanica AI

Андрей Титов

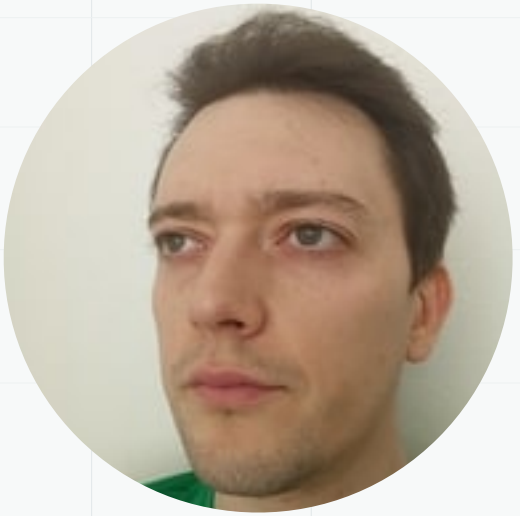
Senior Spark Engineer,
NVIDIA

Координаторы



Михаил Новиков

Middle DS, выпускник курса
Специалист по большим данным



Александр Николаев

Data Engineer, Сбербанк

ИНФРАСТРУКТУРА

Ресурсы

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.

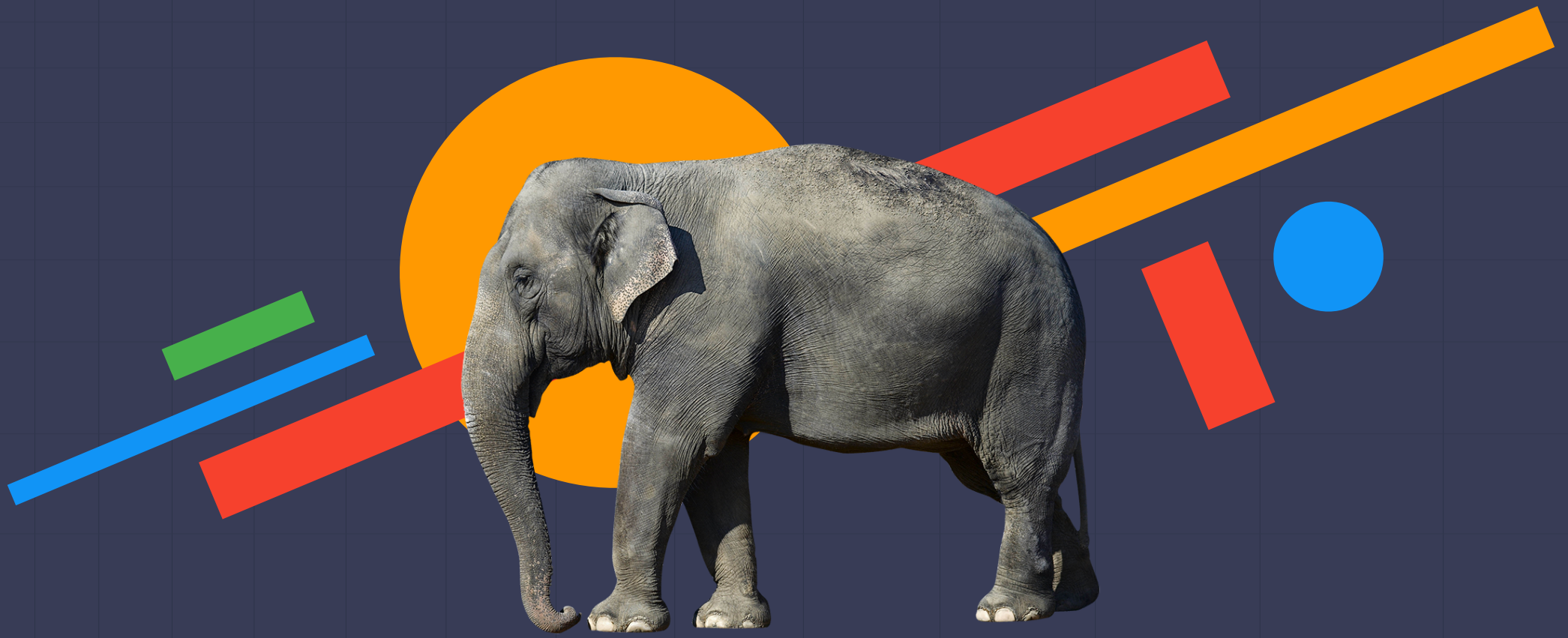
Ресурсы

Slack

УСПЕШНОСТЬ ПРОХОЖДЕНИЯ

Успешность

1. 4 из 5 лаб сданы успешно и в срок.
2. 50% и выше за тесты. Всего 2 теста по 10 вопросов + 1 тест на 5 вопросов. Иными словами, нужно ответить правильно суммарно минимум на 13.



Big Data is Love

NEWPROLAB.COM