

ВТБ

APACHE SPARK ДЛЯ ДАТА ИНЖИНИРИНГА



КОРОТКО О НАС

- Занимаемся обучением работе с данными с 2015 года.
- Самая широкая линейка программ про данные на рынке.
- Используем андрагогику и hands-on подход для создания сжатых практикоориентированных программ.

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

2015

2016

2017

2018

2019

2020

Линейка программ

Управление всей
цепочкой целиком

CDO

Сырые
данные

Data Engineer

Обработанные
данные

Deep Learning
Big Data Specialist

Знания

**Big Data for
Executives**

Стратегия

**Анализ данных
на Scala**

Продукт,
процесс

КОМАНДА ПРОГРАММЫ

Команда



Андрей Титов
Преподаватель

Senior Spark Engineer,
NVIDIA



Андрей Качетов
Координатор

DS/DE, Управление технологий EPM
Сбербанк



О ПРОГРАММЕ

Цели

Приобретение уверенных навыков использования Apache Spark для задач дата инжиниринга.

Формат

- 4 занятия и 2 недели (по понедельникам и средам).
- Каждое занятие: 3 часа с перерывом.
- 2 лабы (практические домашние задачи).



Расписание



Занятия будут проходить в Zoom.

18:00 – 21:00.

1. Spark Dataframes I.
2. Spark Dataframes II.
3. Spark Dataframes III.
4. Мониторинг и оптимизация Spark.

1. Подбор топ-350 релевантных URL для сегмента пользователей.
2. Создание витрины данных из разных источников: файлы, NoSQL-хранилища, реляционные базы данных.

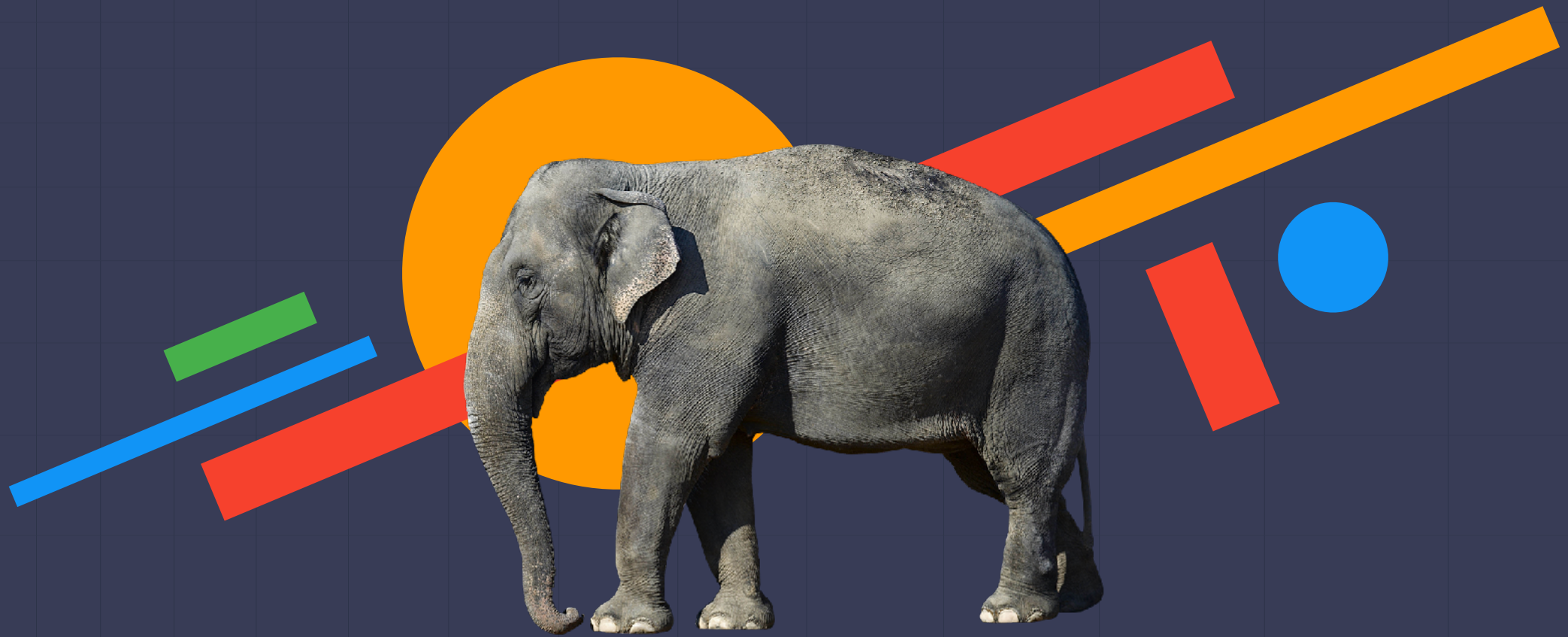
Важные принципы

- 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.



ИНФРАСТРУКТУРА

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 (**доступ!**).
5. Slack.



Big Data is Love

NEWPROLAB.COM