



СПЕЦИАЛИСТ ПО БОЛЬШИМ ДАННЫМ 10.0

МАРАФОН ДЛИНОЮ В 12 НЕДЕЛЬ

Благодарность



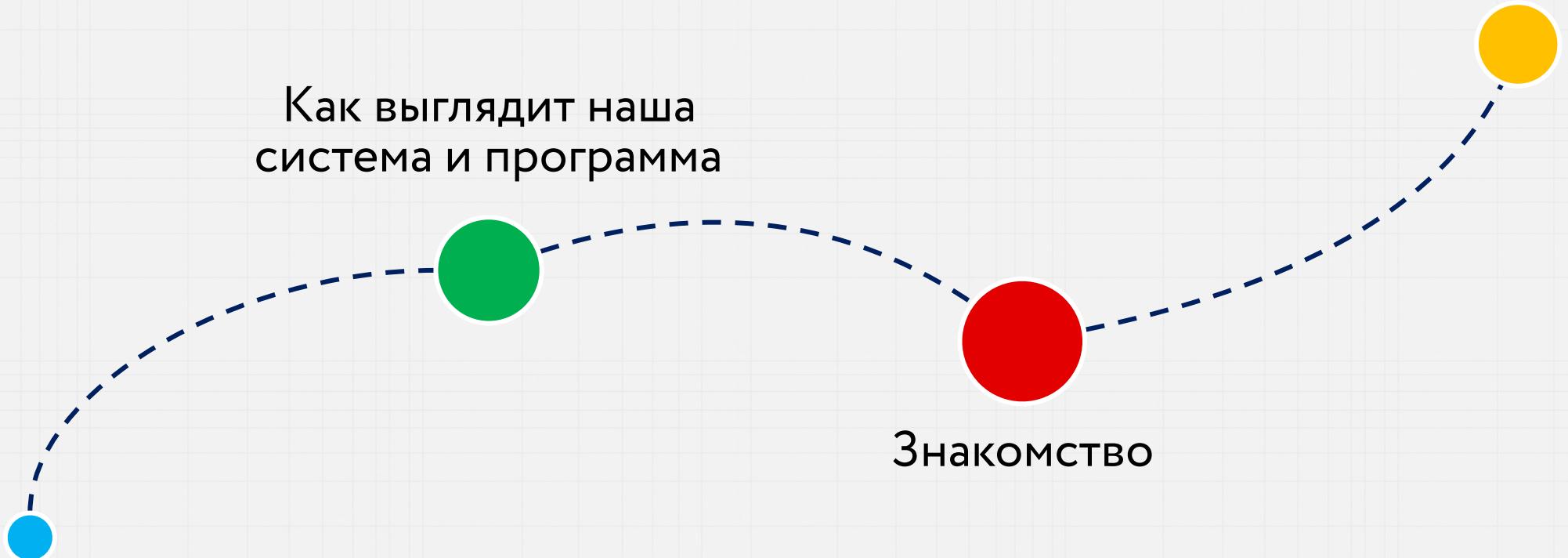
МЕГАФОН



План

Введение в
большие данные

Как выглядит наша
система и программа



Кто мы такие

Кто мы такие



New Professions Lab

Школа обучения работе с данными.

Мы собираемся покрыть всю вселенную ролей и навыков, необходимых в мире работы с данными.



Наша линейка программ

Управление всей
цепочкой целиком

CDO

Сырые
данные

Data Engineer

Обработанные
данные

Deep Learning
Специалист по
большим данным

Знания

Цифровизация
бизнеса

Стратегия

Анализ данных
на Scala

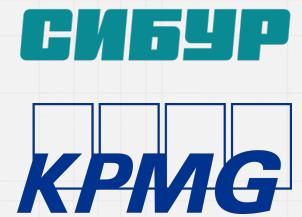
Продукт,
процесс



Развитие

Выпускники:

- 350 с открытых программ
 - 350 с корпоративных



3 Scala



1 Data Engineer

3 Цифра бизнеса

1 Deep Learning

3 Big Data

2 Big Data

2 Big Data

2 Data Engineer

Часть 2 Цифровизация бизнеса

1 Deep Learning

3 Data Engineer

3 Цифровизация бизнеса

2 Deep Learning





New Professions Lab



Мы школа обучения работе с данными, но **необычная...**



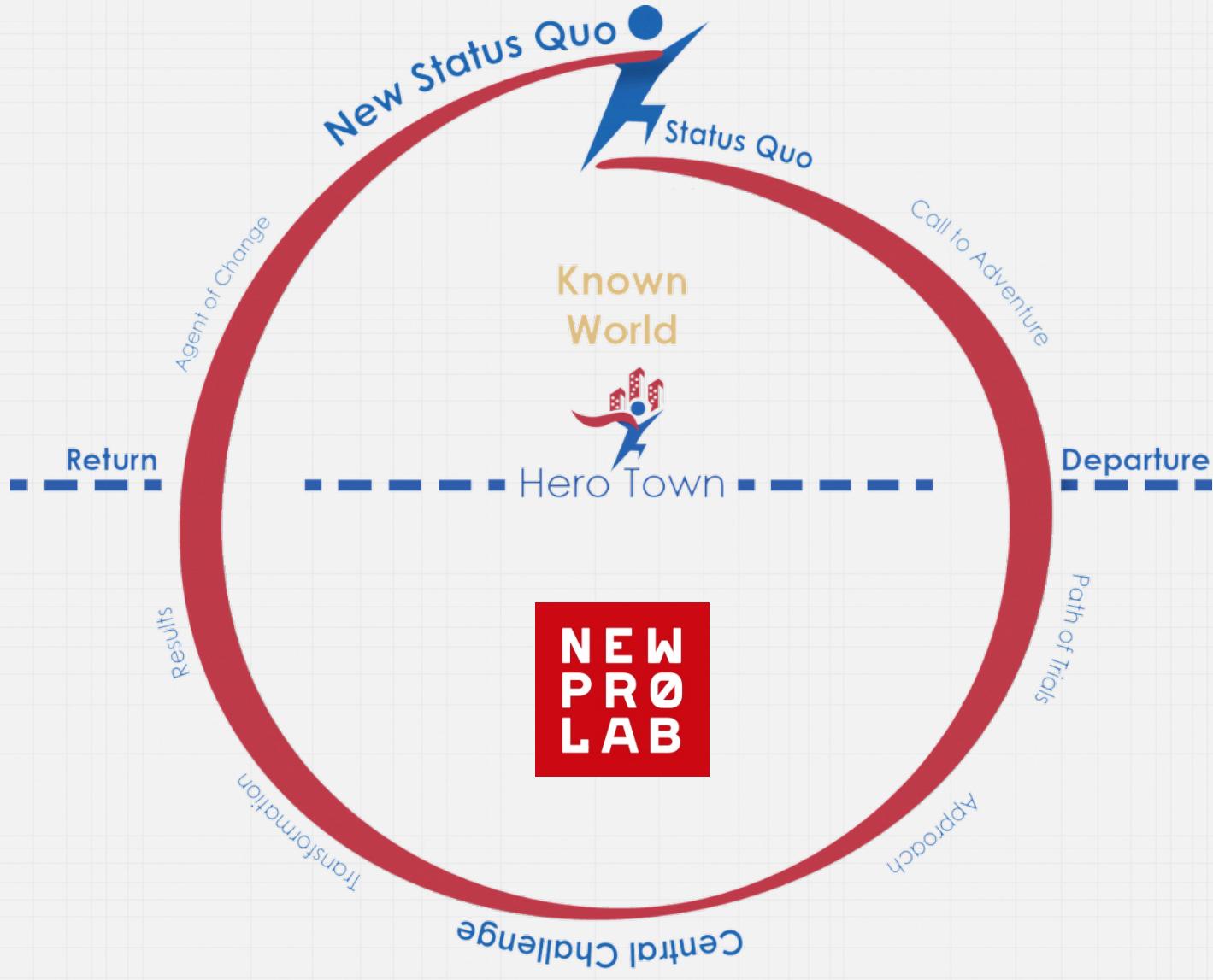
New Professions Lab

Наша цель лежит за пределами просто покрытия материала или передачи знаний и навыков.

Hero's journey



Hero's journey





Это уже давно так

За время программы я научился:

- “Терпению, усидчивости, самоконтролю, скрупулезности и веры в самого себя. Я крут, я всё могу и всё смогу.”
- “Упорству. То есть решать задачи, которые я сходу не знаю, как решить.”
- “Очень часто лабы у меня получались со 2-3 попытки. И это, как ни странно, дало больше уверенности в себе.”

Как выглядит наша
система



New Professions Lab

Держа в уме нашу цель, мы создали пространство – лабораторию, в которой объединили:

1. реальные задачи из индустрии,
2. современные инструменты,
3. технологии, помогающие в обучении,
4. старших товарищей (преподаватели и координатор),
5. методологию, удерживающую баланс между самостоятельностью и поддержкой.

1. Реальные задачи

Модуль 1. Построение
DMP-системы

Модуль 2. Разработка
рекомендательной

1. Реальные задачи

Лаба 1. Деплой кластера
в облаке и запуск
MapReduce

Лаба 2. Фильтрация
логов из HDFS в HBase

Лаба 3. Классификация
пользователей по логам

Лаба 4.
Прогнозирование
оттока клиентов банка

Лаба 5. Классификация
 отзывов в интернете

Проект 1.
Прогнозирование пола
и возраста по логам

1. Реальные задачи

Лаба 6.
Неперсонализированные
рек. системы

Лаба 7. Content-based
рек. системы

Лаба 8.
Коллаборативная
фильтрация

Лаба 9. Гибридная рек.
система фильмов

Лаба 10. Рекомендации
фильмов по
телесмотрению

Проект 2. Рек. система
для e-commerce

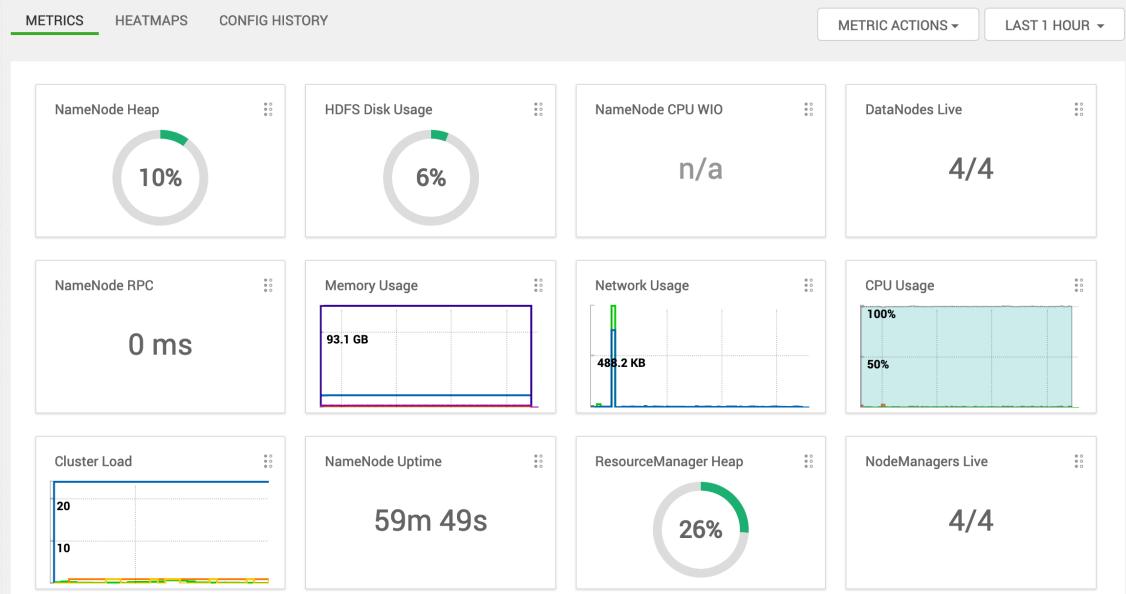
2. Инструменты



2. Инструменты

Кластер из:

- 4 ноды с 24 vCPU и 128GB
- 1 нода с 56 vCPU и 256GB



3. Технологии

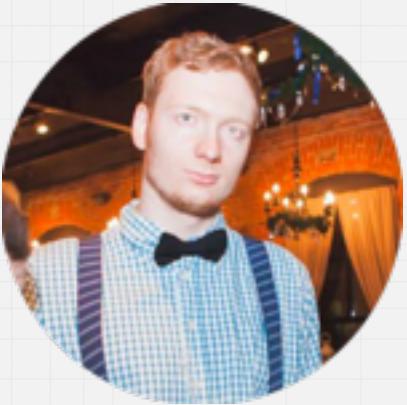
1. Чекеры с автоматической проверкой лаб
2. Записи занятий
3. Slack
4. GitHub (инвайт!)
5. Google Calendar

4. Старшие товарищи



Антон Пилипенко

Big Data Engineer,
Conundrum



Николай Марков

Senior Data Engineer,
Aligned Research Group

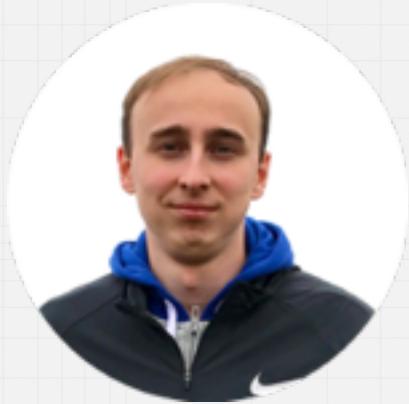


Павел Клеменков

Chief Data Scientist,
NVIDIA

Технологический стэк

4. Старшие товарищи



Владимир Опанасенко

Исполнительный директор,
Департамент анализа данных и моделирования
Газпром банк

Андрей Зимовнов

Старший разработчик,
Яндекс.Дзен

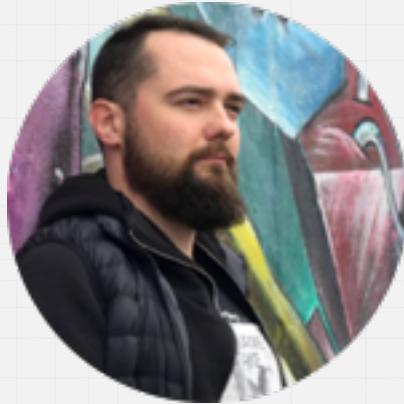
Алгоритмический стэк

4. Старшие товарищи



Александр Ульянов

Data Science Executive Director,
Сбербанк



Олег Хомюк

Руководитель R&D,
Lamoda



Александр Филатов

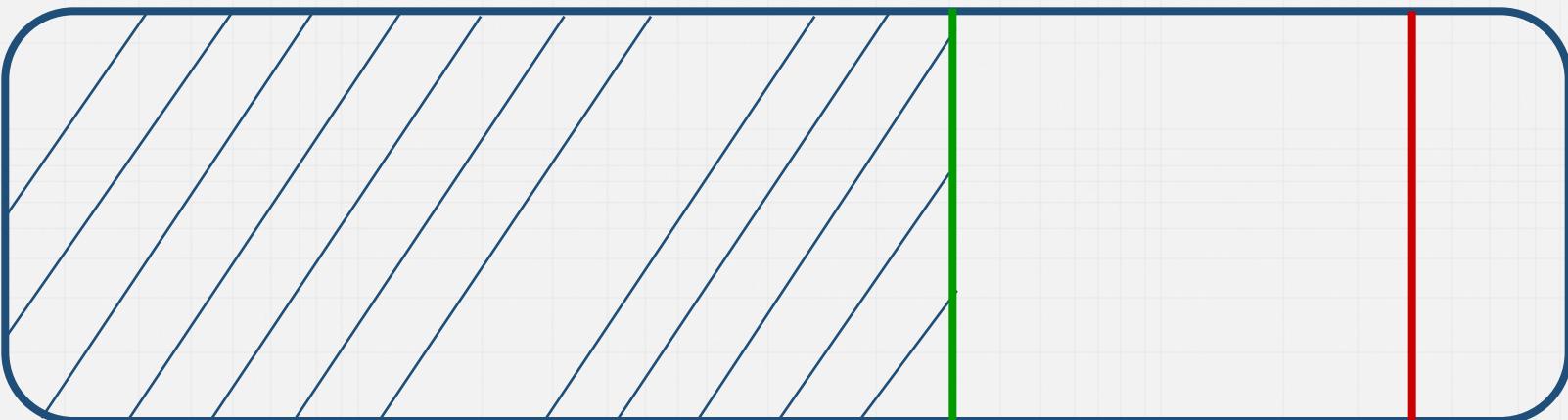
Product Analytics Manager,
VISA

Бизнес стэк

5. Методология

1. Есть лабы. Есть суперачивки.
2. В лабах не все будет разжевываться. Часто придется гуглить.
3. Лабы будут часто выкладываться до занятий.
4. В Slack создаем возможность для p2p-обучения.

Сертификат



60% сертификат

90% с отличием



Баллы

- Проект – 20 баллов
- Лаба – 10 баллов
- Суперачивка – 5 баллов
- Тест – 1 балл
- Упражнения – 0 баллов :)



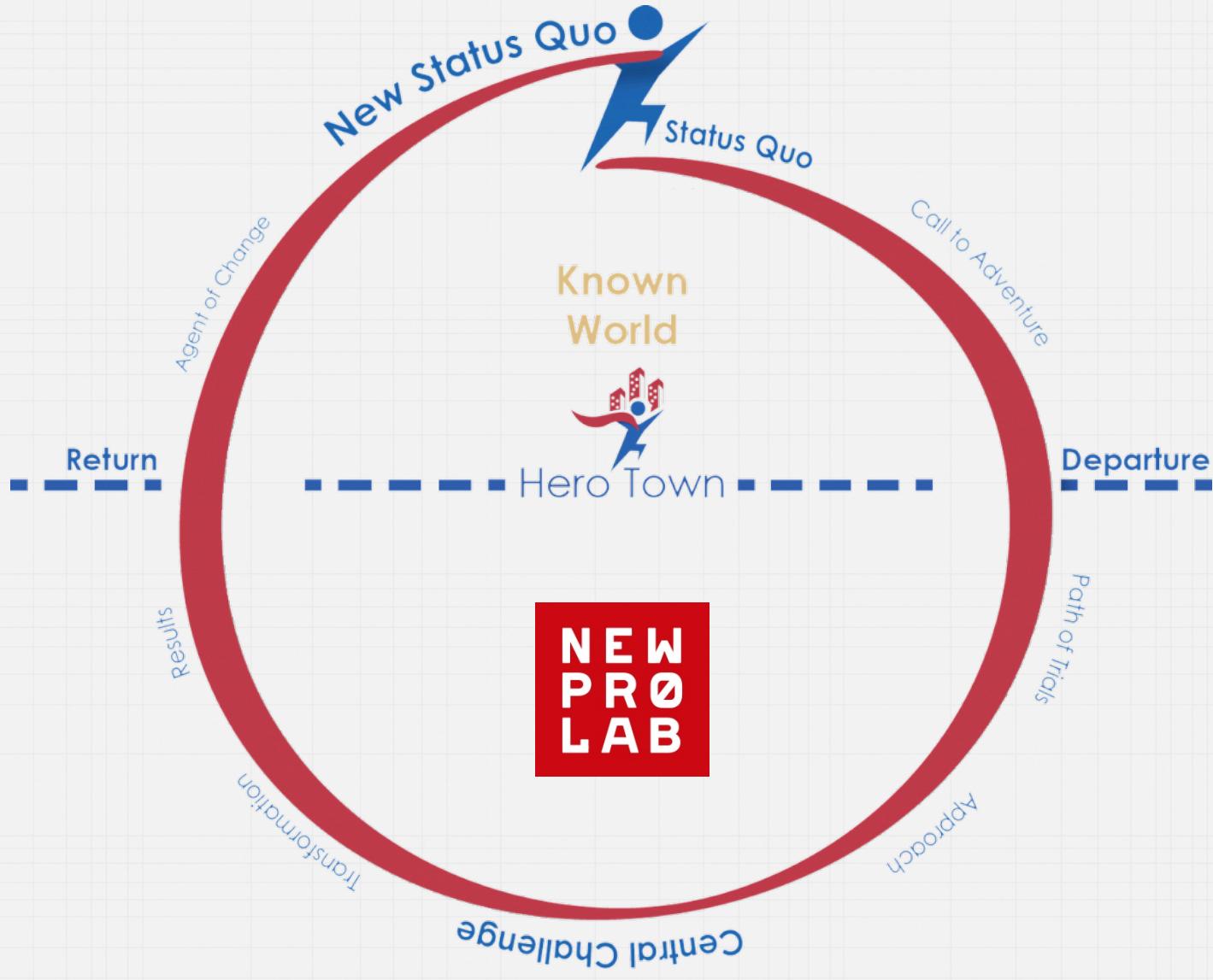
Тесты

- Сформировать контекст занятия.
- Проверить свое усвоение материала.
- Вскрыть неявные вещи.
- Получить дополнительные знания.

Упражнения

- Получить базовые навыки по теме.
- Содержат в себе «подсказки» к лабам.

Hero's journey



Ближайшие шаги

- 27.03 – выкладываем Лабу 1 и суперачивку
- 28.03 – занятие по Linux
- 30.03 – введение в Python

Напутствия и лайфхаки от выпускников



Знакомство



Команда



Оксана Страшинская

координирует программу



Анастасия Кошель

отвечает за площадку



Марина Михайлова

отвечает за документы



Команда



Андрей Булатов

чтобы не было ни
единого разрыва



Ксения Хизова

отвечает за маркетинг



Елена Третьякова

отвечает за всё

Ваша очередь

1. Имя, фамилия, компания
2. Профессиональное достижение
3. Любопытный факт о себе

1 минута

Разомнем мозги



Ответ

\$ ssh -i bigdata10.pem user@newprolab.com

\$ ccr -s lsqnkdk10.zow ecob@xogzbyvkl.myw



Ответ

Шифр Цезаря со смещением 10.

```
NUM LETTERS = 26
```

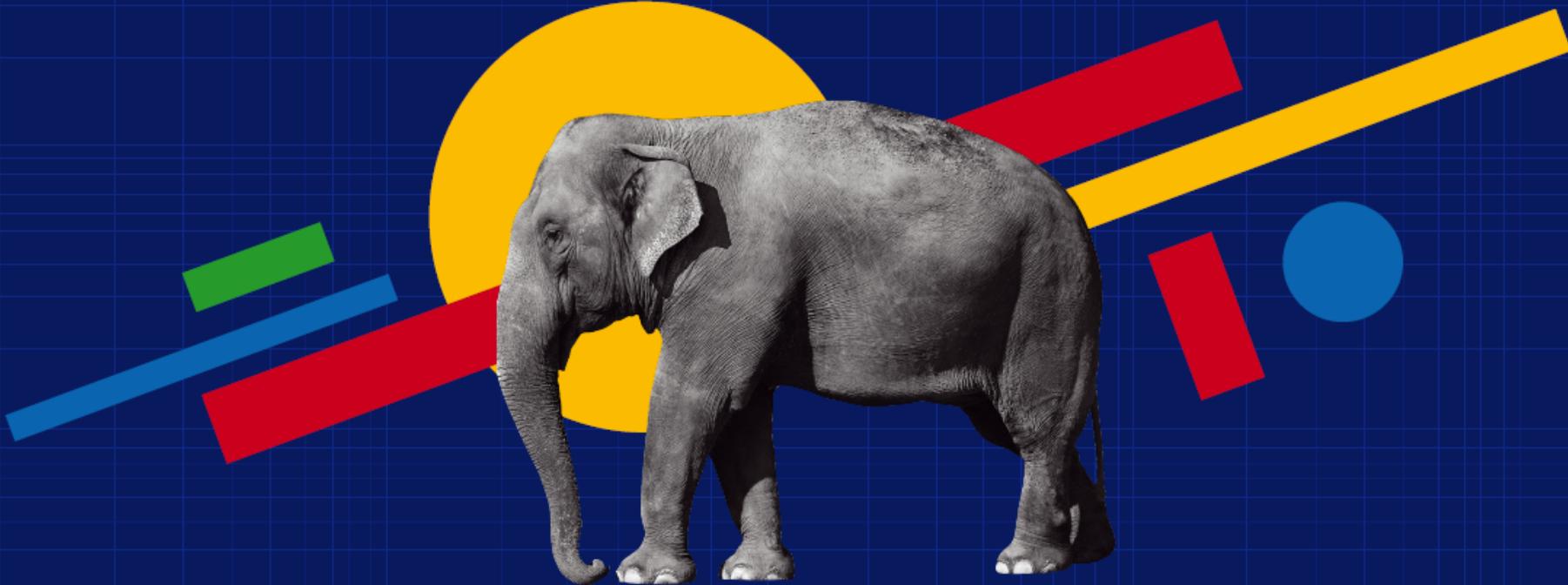
```
# ord('a') = 97  
A_ASCII_SHIFT = 97
```

```
# Welcome, Newprolab BigData 10!  
SHIFT = 10
```

```
def encode(letter, shift):  
    if not letter.isalpha():  
        return letter
```

```
letter_code = ord(letter) - A_ASCII_SHIFT  
shifted_letter_code = (letter_code + shift) % NUM LETTERS
```

```
return chr(shifted_letter_code + A_ASCII_SHIFT)
```



BIG DATA IS LOVE

NEWPROLAB.COM