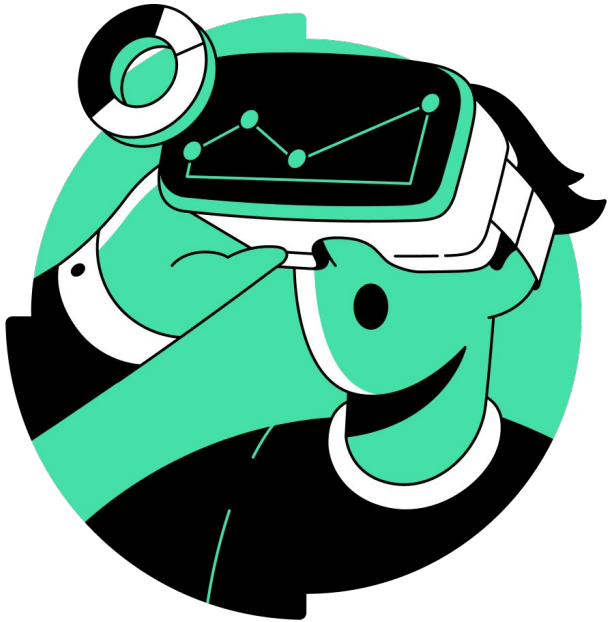
Дикий Запад

Кто такие дата-инженеры и почему они так сильно отличаются друг от друга.



SEMRUSH

Обо мне

Семён Осипов
Data Engineer, Semrush

@OsipovSimon

Труба Данных @ohmydataengineer

Обо мне

О чем мы поговорим?

1. Дата инженер это относительно новая специальность, и требования к ней слишком расплывчаты.

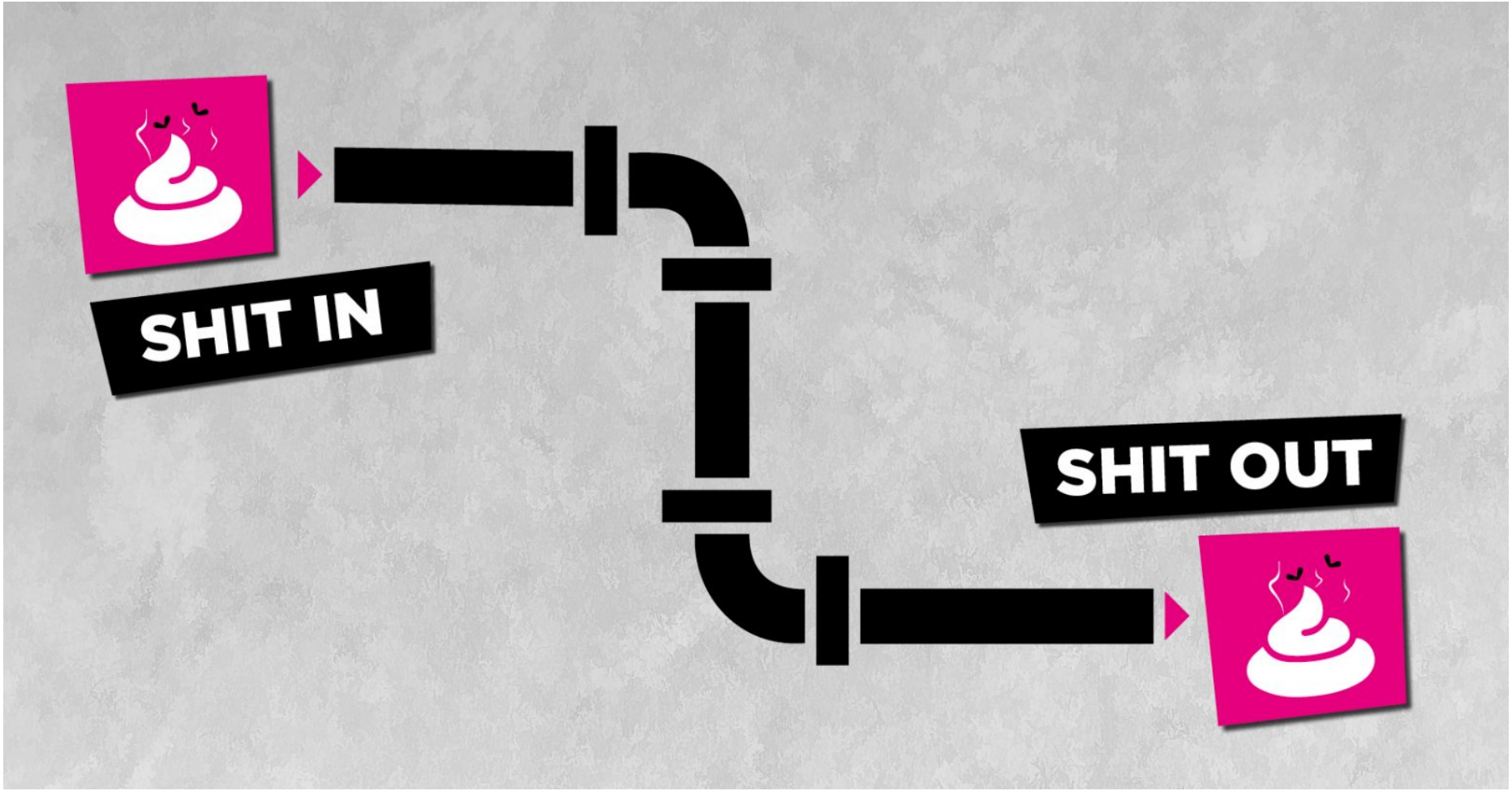
2. Для тех, кто хочет стать дата инженером не понятно, с чего начинать.

3. Уже состоявшиеся специалисты находятся в сложном выборе, куда дальше развиваться.

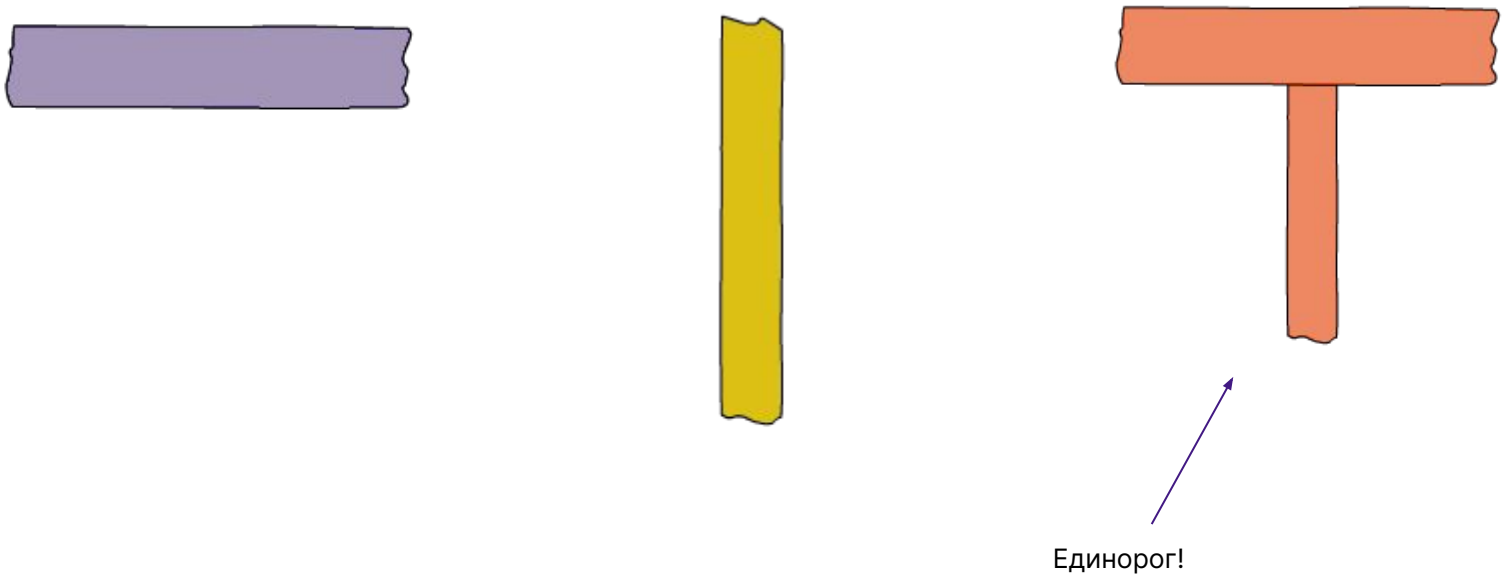
4. Работодатели закрывают вакансии очень долго, поиск релевантного специалиста осложнен требованиями к позиции.

Image credit: www.ceotodaymagazine.com/


ПИТОНИСТАМ МАЛО ПЛАТЯТ...
СТАНОВИСЬ ДАТА САЕНТИСТОМ КАК Я
ТЫ ЧЁ ПЁС, Я МАТЕМАТИК
ЭТО ЗАЧЕМ?
БУДЕШЬ РАЗРАБАТЫВАТЬ ИСКУССТВЕННЫЙ ИНТЕЛЕКТ ЗА 300К/СЕК
ТАК ТЫ ЖЕ ПРОСТО РАНДОМНО ПОДБИРАЕШЬ КОЭФФИЦИЕНТЫ ПОКА КРОСС-ВАЛИДАЦИЯ НЕ ДАСТ НОРМАЛЬНЫЙ РЕЗУЛЬТАТ

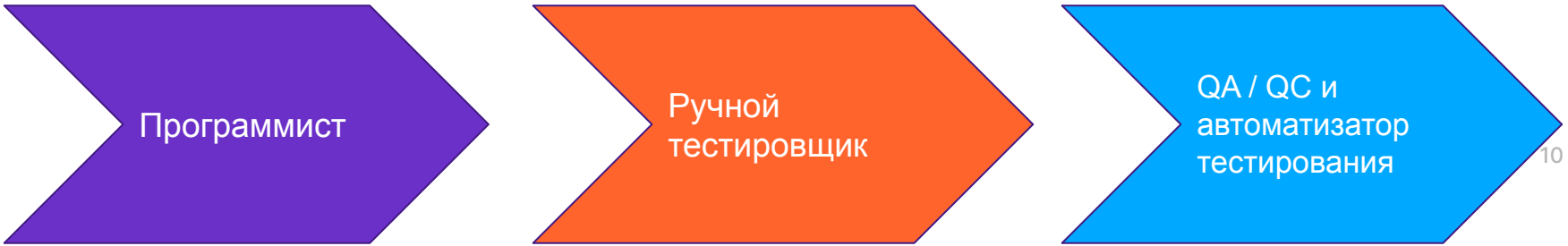

SHIT IN
SHIT OUT

Генерализация и специализация

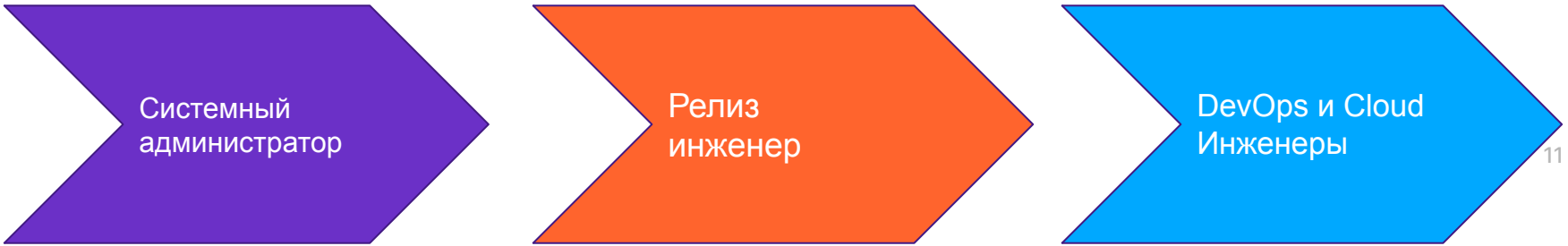


Немного истории

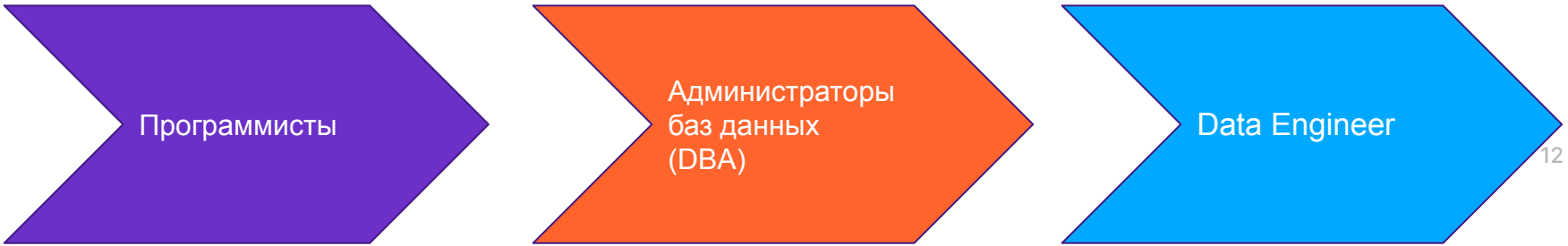
Помните, были такие...



А еще были...


Системный
администратор
Релиз
инженер
DevOps и Cloud
Инженеры

История повторяется..



Почему так происходит?

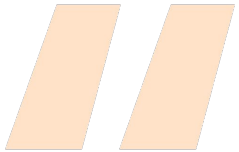
- Появляются новые возможности
- С новыми возможностями приходят новые проблемы
- Для решения новых проблем нужны новые подходы и инструменты

Почему так происходит?

- Инструментов, подходов и задач становится все больше
- Это все тяжело уместить в голову одного специалиста
- Появляется специализация на решение определенного круга новых задач

Да придут
Дата Инженеры

Так ли это?



Новое модное название для разработчика хранилищ?

Обслуга для DS

Поддерживает наши базы данных и грузит туда из разных источников.

Витриноделы!

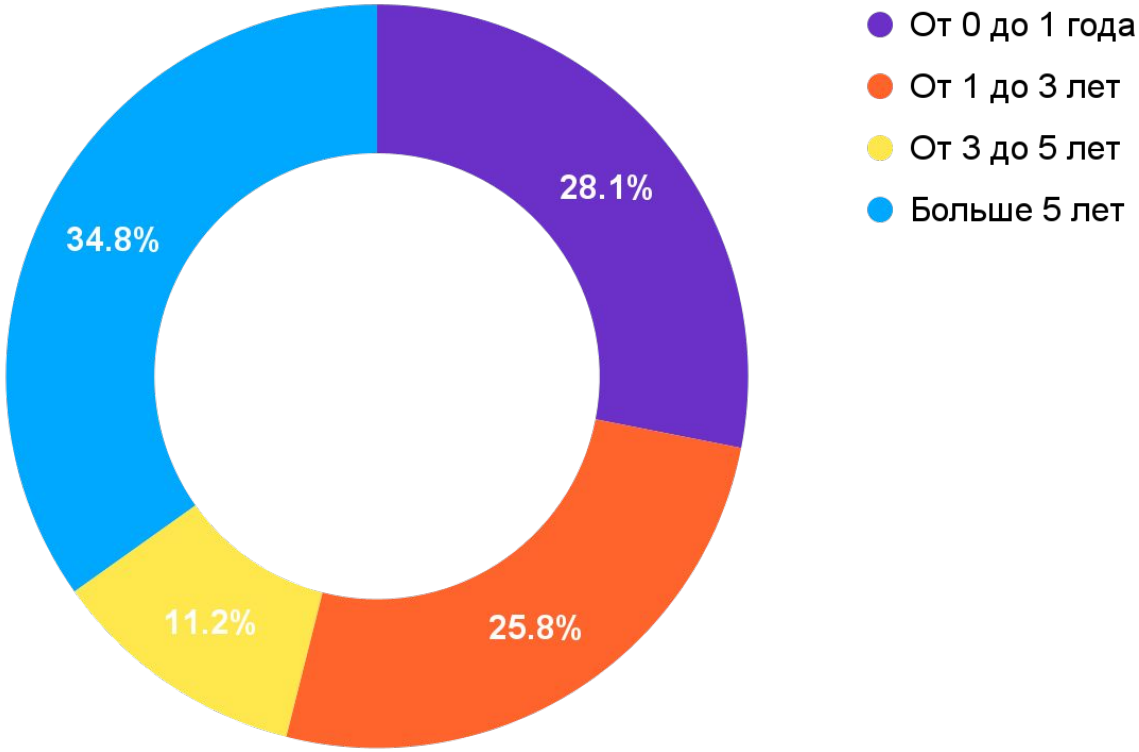
Если я приложение написал многопоточное, которое данные туда-сюда гоняет, я DE или Разработчик?

Кто такой дата инженер?

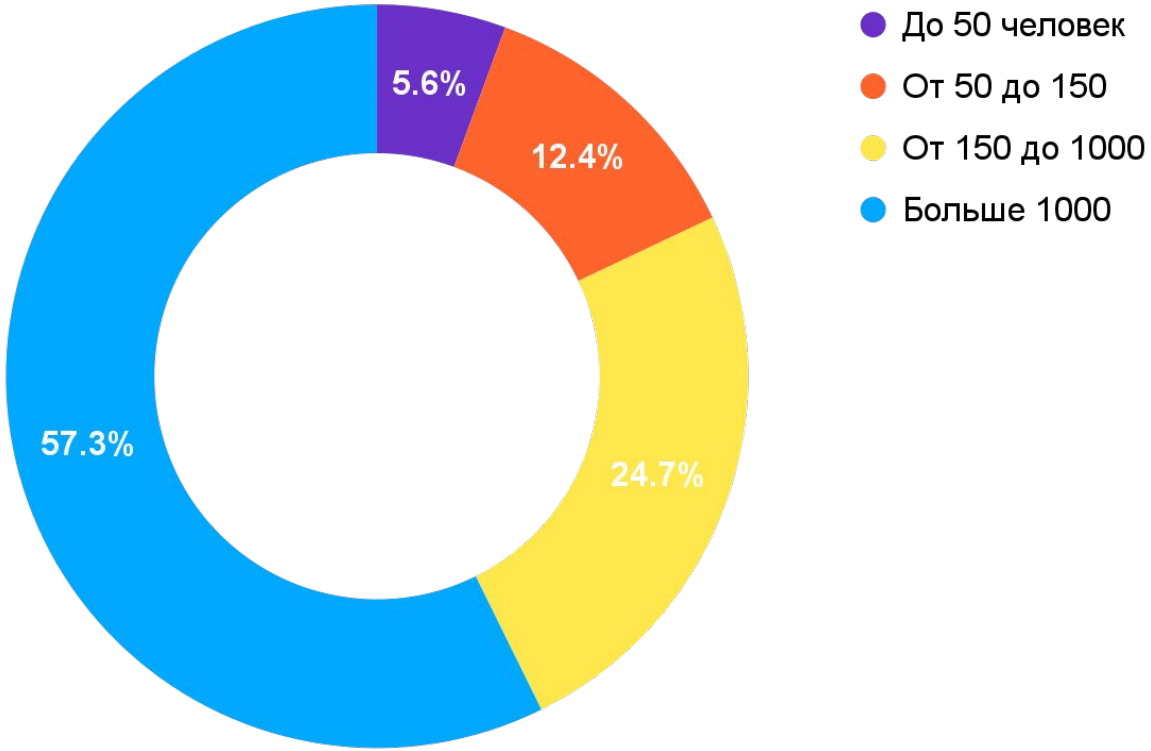
Инженер, который специализируется на работе с системами хранения и загрузки данных.

Семён Осипов
Semrush, Data Engineer

Опыт в IT до прихода в DE



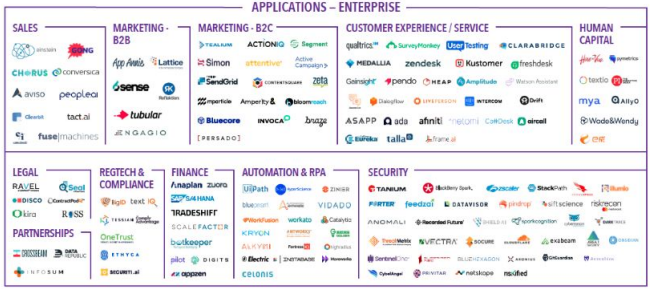
Размер компании, в которой работает DE

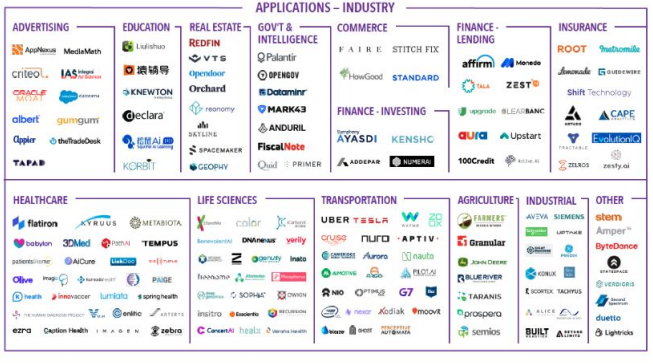


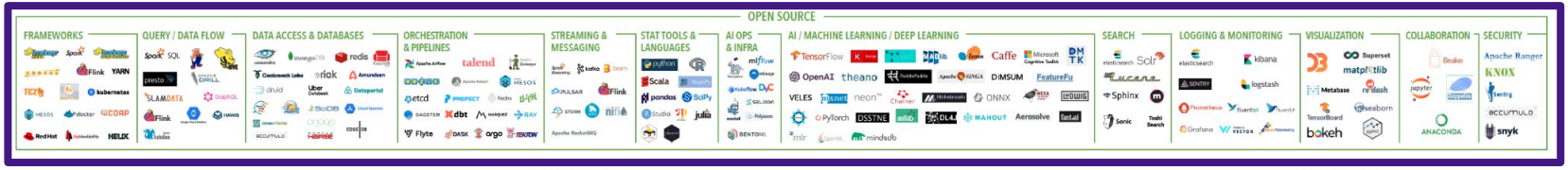
Топ 5 навыков для DE

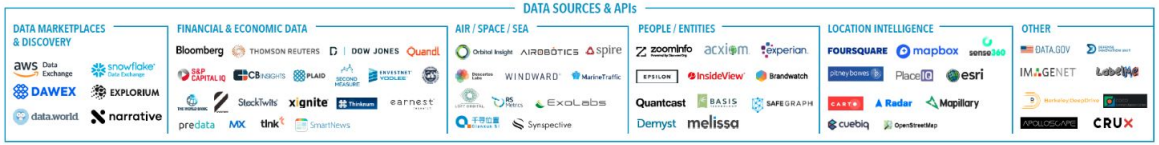
Если бы вы сейчас нанимали DE, что бы вы искали в кандидате?

DATA & AI LANDSCAPE 2020









FIRSTMARK
EARLY STAGE VENTURE CAPITAL

Разработчик DWH

Задачи

Разработка хранилища данных внутри компании (таблицы, слои и витрины данных)

Типовой набор навыков

SQL
Язык программирования
Базы данных

Когда нужен

Когда компания решает строить платформу данных:
Data Lake, Data Warehouse, Data Lakehouse, Data Swamp и т. д. 😀

Разработчик ETL

Задачи

Получение и хранение данных из различных источников.

Типовой набор навыков

SQL
Язык программирования
Базы данных
Оркестратция запусков

Когда нужен

Когда нужно автоматизировать загрузку данных

Инфраструктурный инженер

Задачи

Поддержание и масштабирование систем получения, обработки и хранения данных.

Типовой набор навыков

Стек Big Data (Hadoop / Spark)
Базы данных
Облака

Когда нужен

Когда данных становится настолько слишком много.

ML инженер

Задачи

Вывод в прод, поддержание и масштабирование систем и моделей машинного обучения

Типовой набор навыков

Стек ML (PyTorch, TensorFlow)
Kubernetes
Docker
Язык программирования

Когда нужен

Когда важна скорость и стабильность работы ML моделей

Data Quality инженер

Задачи

Отслеживание качества данных для работы всех систем

Типовой набор навыков

SQL
Язык программирования
Системы мониторинга

Когда нужен

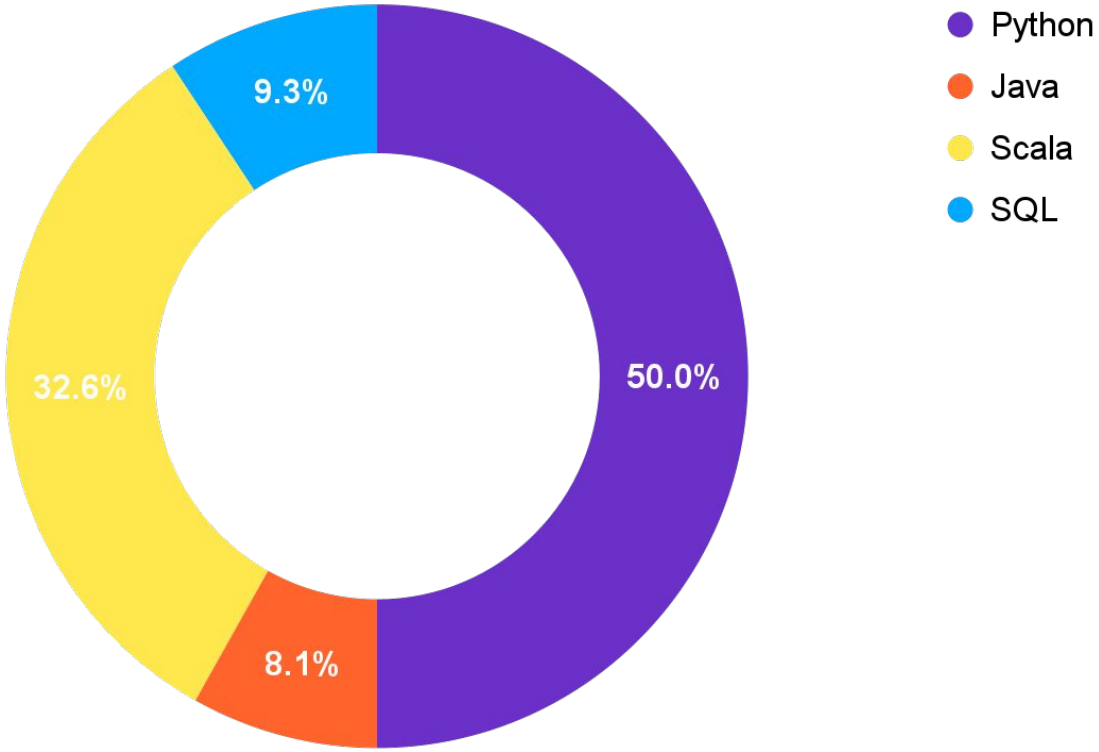
Когда обожглись, работая без проверки данных 😀

Analytics Engineer

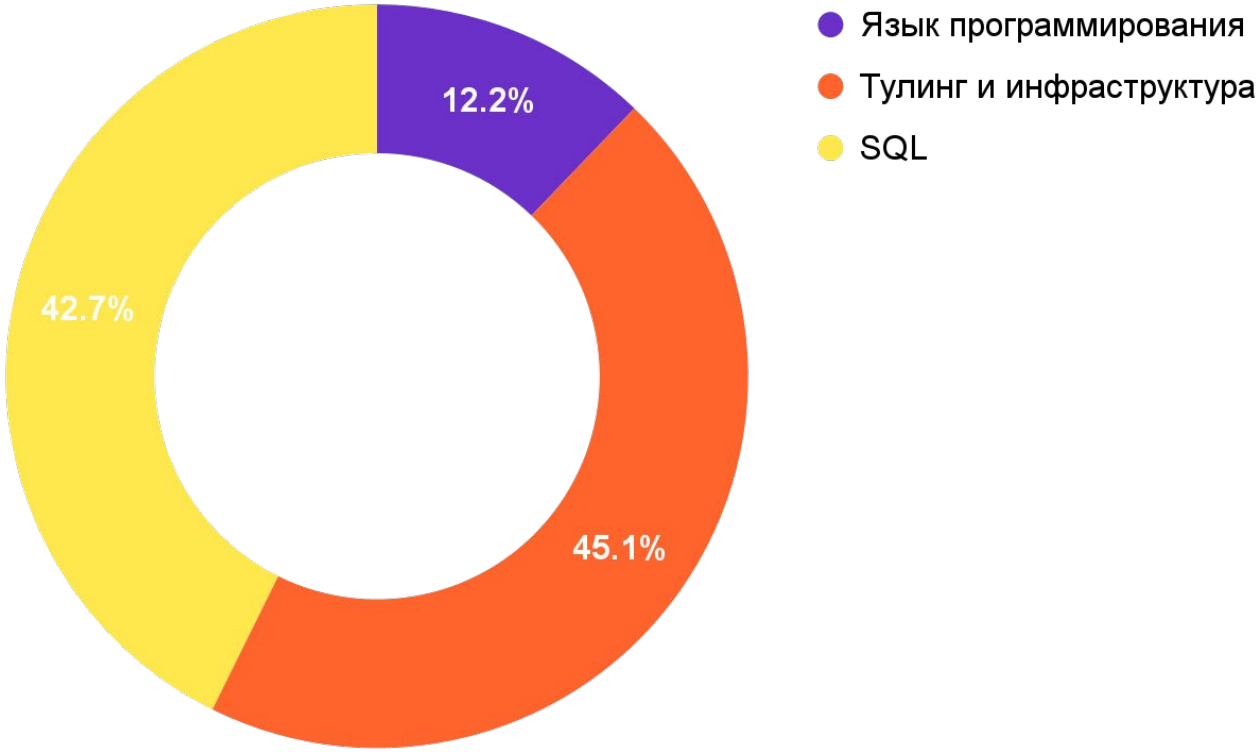
Задачи	Типовой набор навыков	Когда нужен
Создание и развитие инструментов и платформ для аналитики	Комбинация навыков DE + Аналитик + DS	Тот самый единорог 😀

Главный холивар - язык программирования!

Главный холивар - язык программирования!



Но язык программирования не главное



Что же делать компаниям?

1. Понять, кто им действительно нужен
2. Не пихать все возможные технологии в одно описание вакансии
3. Развивать культуру изнутри, обучать разработчиков или выращивать джунов

Где искать?

1. Очевидные места: HH и LinkedIn
2. ODS, #jobs / #jobs_hr канал
3. TG @hadoopusers, а точнее его ответвление @DataJobs

Задать вопрос?

@OsipovSimon

@SimonOsipov