Вопросы:

1. Рынок труда по профессии инженера данных: количество вакансий, отрасли компаний, размер компаний.
2. Требования: популярные языки, технологии, фреймворки, опыт.
3. Кратко опишите своими словами, для чего используется каждая из популярных технологий, фреймворков (не более 10).
4. Ваши цели на обучение. Перечислите интересующие Вас темы и технологии.

Ответы:

1. По данным [hh](https://hh.ru/) 365 вакансий «Data Engineering».   
     
   Отрасли компаний распределяются следующим образом:

Компании по численности варьируются от маленьких (стартапов) до огромных телеком провайдеров (ВымпелКом)

1. Основные требования, технологии, опыт:
   * Реляционные базы данных (GreenPlum, Oracle)
   * Кластер Hadoop
   * Python
   * Airflow
   * PySpark
   * Продвинутый уровень SQL
   * Хороший уровень Python
   * Хороший уровень Spark (PySpark)
   * Хороший уровень Airflow
   * Опыт работы 1-3 года
2. Кластер Hadoop - это озеро данных (Data Lake). Знание его необходимо для понимания, как храниться большой объем данных в их первоначальной форме и у нас есть необходимый и удобный доступ к данным.  
   Airflow – как инструмент необходим для запуска задач (контейнеров) по расписанию или реагированию на различные типы событий.

Spark – необходим для анализа или обработке больших объемов данных и формирования готовых витрин и не только и т.д.

1. Цель на данное обучение: получение необходимых знаний и опыта для перехода работы из сферы DA в DE. Т.к. давно интересуюсь этой специальностью и очень нравится специфика работы с данными