**Тестовое задание для дата инженера**

**(обязательное)**

Сгененировать и наполнить два источника - БД (postgre) и кафка.

В бд справочники, в кафке сами данные в виде json.

Необходимо построить пайплайн с загрузкой данных раз в 15 минут. Обновление справочников раз в сутки.

данные надо партицировать по неделям.

Инструменты - NiFi, запись в hdfs (для теста попробовать в бд), оркестрацию можно сделать в самом nifi, либо oozie либо airflow

Вопросы

Как будете партицировать данные?

Что будете мониторить и как?

Что будете логгировать?

Что будете делать с оффсетом?

Сериализация данных?