1. Взять назначенную Jira Task в работу
2. Авторизоваться под своей учетной записью <https://gitlab.prod.dmp.xxx.ru>
3. Отписаться в задаче о выполнении п.2
4. После авторизации на вас заведу заявку на доступ к репозиториям и платформе DMP

Для каждого будет создано 2 репозитория:

Репозиторий для spark приложения: https://gitlab.prod.dmp.xxx.ru/school-de/datamarts-<AD>

Репозиторий для dags airflow: https://gitlab.prod.dmp.xxx.ru/school-de/dags-<AD>

где AD – ваша учетная запись

1. Склонировать репозитории на свою локальную машину
2. Создать в каждом репозитории ветку для разработки – develop
3. Переключиться на ветку develop
4. Добавить файл .gitignore, содержащий:

.idea/\*

.bsp/\*

target/\*

project/\*

\*.iml

!project/build.properties

!project/plugins.sbt

1. Написать код в каждом репозитории, согласно заданиям ниже.
2. datamarts-<AD>: Написание спарк-приложения:

Используемые версии зависимостей:

***ThisBuild / scalaVersion := "2.11.8"***

***val sparkVersion = "2.3.2"***

***libraryDependencies ++= Seq(***

***"org.apache.spark" %% "spark-core" % sparkVersion % "provided",***

***"org.apache.spark" %% "spark-sql" % sparkVersion % "provided",***

***"org.postgresql" % "postgresql" % "42.3.3"***

***)***

Написать два приложения (scala класса)

1. Загрузка данных из реляционной БД на кластер Hadoop.

Необходимо загрузить данные из базы

postgresql://ingress-1.prod.dmp.xxx.ru:5448/demo

в hive базу school\_de шмртаблицы:

bookings.seats -> school\_de.seats\_<AD>

bookings.flights\_v -> school\_de.flights\_v\_<AD> (загрузить с разбивкой на партиции по актуальной **дате** вылета))

***В коде для сдачи доступ к postgres осуществлять из под ТУЗ tech\_repbgd\_local***

1. Выполнение запросов из практического задания 1 используя распределенные вычисления spark, записав результат в таблицу school\_de.results\_<AD>

Для решения задачи использовать Spark DataFrame API, без использования spark.sql(“….”)

Для разработки и тестирования использовать базу school\_de\_stg

1. dags-<AD>: Создать два дага
   1. Для загрузки данных задачи п. 8.1 (***доступ к postgres осуществлять из под ТУЗ tech\_repbgd\_local, получив креды из Vault***)
   2. Для расчёта данных задачи п. 8.2

***Оба дага должны иметь регламент запуска None и запускаться из под ТУЗ tech\_schoolde\_bgd\_ms***

1. Закоммить свои изменения веток develop в удаленные репозитории
2. Создать Merge Request в ветку мастер для каждого репозитория
3. Закрыть Jira Task до срока сдачи решения

P.S. ***Доступ к postgres для разработки производить из под личной УЗ, база керберезированая.***