**Домашнее задание. Практическая работа**

**Дисциплина**

ETL-процессы.

**Тема**

Тема 9. Работа с Big Data и NoSQL

**Форма проверки**

Домашнее задание с самопроверкой.

Совет: выполняйте его сразу, как только изучите тему.

**Имя преподавателя**

Артём Озерков.

**Время выполнения**

2 часа.

**Цель задания**

Практика работы с Hadoop и Spark в облаке

**Инструменты для выполнения ДЗ**

Apache Hadoop, Apache Spark, Yandex Data Proc

**Правила приёма работы**

Сделайте скриншот изначальных и изменённых данных. Прикрепите в LMS файл или ссылку на GitHub со скриншотами.

Важно:

* убедитесь, что доступ к материалам открыт;
* название файла должно содержать фамилию и имя студента, номер и название ДЗ.

**Чек-лист самопроверки**

**Задание считается выполненным:**

* если прикреплены файл с выполненным заданием или ссылка на GitHub;
* в файле или на GitHub содержится выполненное задание;
* доступ к материалам открыт.

**Задание не выполнено:**

* если файл или ссылка на GitHub не прикреплены или отсутствует доступ к ним;
* файл или страница на GitHub не содержит выполненное задание.

**Дедлайн**

6 дней после открытия задания на платформе.

**Описание задания**

Повторите работу из демонстрации вебинара «Работа с Big Data и NoSQL».

* Разверните кластер Hadoop и Spark с помощью Yandex Data Processing
* Загрузите данные для обработки
* Проведите трансформацию и запись в Yandex Data Processing