

**Государственное автономное образовательное учреждение  
высшего образования города Москвы  
«Московский городской педагогический университет»**

Институт цифрового образования  
Департамент информатики, управления и технологий

**ЗАДАНИЕ  
НА КУРСОВОЙ ПРОЕКТ**

по дисциплине  
Инструменты для хранения и обработки больших данных  
студентке 3-го курса АДЭУ-201 учебной группы  
Петровой Анастасии Олеговне

**ТЕМА**

**ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА ХРАНИЛИЩА МЕТРИК  
ОБРАЗОВАТЕЛЬНОГО КОНТЕНТА МОСКОВСКОЙ ЭЛЕКТРОННОЙ  
ШКОЛЫ В АРАСНЕ HADOOP**

**СОДЕРЖАНИЕ:**

**1. Целевая установка:**

спроектировать и разработать хранилище больших данных на основе программного продукта АРАСНЕ HADOOP в ОС Ubuntu 20.04 с возможностью выгрузки метрик образовательного контента МЭШ.

**2. Основные вопросы, подлежащие разработке:**

провести анализ предметной области;  
провести сравнительный анализ существующих систем по хранению и организации доступа к большим данным;  
спроектировать архитектуру хранилища больших данных на основе программного продукта АРАСНЕ HADOOP;  
реализовать выгрузку метрик образовательного контента из информационной системы МЭШ, используя озеро данных PostgreSQL, в хранилище данных;  
провести анализ качества данных в хранилище больших данных.

**3. Исходные данные:**

озеро данных PostgreSQL информационной системы МЭШ, установочные файлы АРАСНЕ HADOOP, ОС Ubuntu 20.04.

**4. Перечень разрабатываемых материалов (отчет, расчеты, презентация, текст компьютерной программы, скриншоты, рисунки, графики и др.):**  
Презентация, отчет.

**5. Перечень необходимых проектных и (или) экспериментальных работ и расчетов:**

представить:

- инфологическую схему базы данных метрик образовательного контента из информационной системы МЭШ;
- инфологическую схему хранилища больших данных в системе APACHE HADOOP;
- скрипты настройки хранилища больших данных в ОС Ubuntu 20.04;
- инструкцию пользователя «Использование хранилища метрик МЭШ для анализа данных».

**6. Общий объем отчета и основные требования к структуре и оформлению отчета по курсовой работе (проекту):**

согласно требованиям МГПУ (не более 30-40 стр.).

**7. Основные рекомендуемые литературные источники (учебники, руководства, пособия, интернет-ресурсы и др.):**

7.1. Босенко, Т. М. Использование OLAP-технологии в процессе обучения студентов специальности 38.03.05 - бизнес-информатика / Т. М. Босенко, П. К. Григорьев // Актуальные проблемы теории и практики обучения физико-математическим и техническим дисциплинам в современном образовательном пространстве : Сборник избранных статей VI Всероссийской (с международным участием) научно-практической конференции, Курск, 15–16 декабря 2022 года / Ответственный редактор: В.Н. Фрундин. – Курск: Курский государственный университет, 2022. – С. 146-150.

7.2. Настройка Hadoop, HDFS и Yarn вручную (автономно)

[https://github.com/BosenkoTM/BigDataAnalytic\\_Practice/blob/main/exercises/winter\\_semester\\_2021-2022/01\\_hadoop/Excercise\\_1.pdf](https://github.com/BosenkoTM/BigDataAnalytic_Practice/blob/main/exercises/winter_semester_2021-2022/01_hadoop/Excercise_1.pdf).

7.3. Джеймс Р. Грофф. SQL: полное руководство, 3-е издание/SQL: The Complete Reference, Third Edition. — М.: Вильямс, 2014. — 960 с.

7.4. Универсальный инструмент для работы с базами данных DBeaver [Электронный ресурс] URL <https://dbeaver.io/about/>.

**8. Сроки представления руководителю отчета о выполненной работе:**

16.06.2023

**9. Дата защиты курсовой работы (проекта) 16.06.2023**

Руководитель \_\_\_\_\_ (Т.М. Босенко)  
«10» февраля 2023г.

Задание получил:

Студентка уч. группы АДЭУ-201, 3 курс \_\_\_\_\_  
«10» февраля 2023г.