

Министерство науки и высшего образования Российской Федерации

Национальный исследовательский университет ИТМО

Инфраструктура больших данных

Весна

2024

Лабораторная работа №2

ВЗАИМОДЕЙСТВИЕ С ИСТОЧНИКАМИ ДАННЫХ

Цель работы:

Получить навыки выгрузки исходных данных и отправки результатов модели с использованием различных источников данных согласно варианту задания.

Ход работы:

1. Создать репозитории-форк модели на GitHub, созданной в рамках лабораторной работы №1, регулярно проводить commit + push в ветку разработки, важна история коммитов.
2. Реализовать взаимодействие сервиса модели и базы данных, согласно варианту задания.
3. Обеспечить процессы аутентификации/авторизации при обращении сервиса модели к базе данных в момент отправки результата работы модели. В исходном коде **не должно быть явно** прописаны пары логин/пароль, адрес/порт сервера базы данных, токены доступа.
4. Возможно наполнить базу данных наборами для обучения/валидации модели.
5. Переиспользовать CI pipeline (Jenkins, Team City, Circle CI и др.) для сборки docker image и отправки их на DockerHub.
6. Переиспользовать CD pipeline для запуска контейнеров и проведения функционального тестирования по сценарию, запуск должен стартовать по требованию или расписанию или как вызов с последнего этапа CI pipeline.

7. Результаты функционального тестирования и скрипты конфигурации CI/CD pipeline приложить к отчёту.

Результаты работы:

1. Отчёт о проделанной работе;
2. Ссылка на репозиторий GitHub;
3. Ссылка на docker image в DockerHub;
4. Актуальный дистрибутив модели в zip архиве.

Обязательно использование docker-compose.

Варианты задания

| Номер | Источник данных |
|-------|---|
| 1 | PostgreSQL |
| 2 | MySQL |
| 3 | Oracle |
| 4 | MS SQL Server |
| 5 | MongoDB (в случае недоступности использовать Neo4J) |
| 6 | Apache HBase |
| 7 | Redis |
| 8 | Cassandra |
| 9 | Greenplum |
| 10 | ClickHouse |
| 11 | HDFS |