

**САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ПЕТРА ВЕЛИКОГО
Физико-механический институт**

УТВЕРЖДАЮ

Руководитель образовательной программы

_____ К.Н. Козлов

« _____ » _____ 202_ г.

**ЗАДАНИЕ
на выполнение выпускной квалификационной работы**

студенту Хамидуллину Ильясу Ильназовичу гр. 5030102/10201

1. Тема работы: Автоматическая генерация конфигураций элементов инфраструктуры программных систем для работы с большими данными.
2. Срок сдачи студентом законченной работы: июнь 2025 г.
3. Исходные данные по работе:
 - Декларативные конфигурационные файлы в формате YAML, задаваемые пользователем (инженером данных) для описания инфраструктуры обработки данных
 - Демонстрационный датасет, включающий тестовые данные, хранящиеся в различных источниках (PostgreSQL, S3) и обрабатываемые в системе
 - Автоматически сгенерированные конфигурационные файлы
4. Инструментальные средства:
 - Языки программирования: Python
 - Форматы конфигурационных файлов: YAML, JSON
 - Среда разработки: VS Code
 - Система управления версиями: Git
 - Средства контейнеризации и оркестрации: Docker, Docker Compose
 - Платформы потоковой обработки данных: Apache Kafka, Kafka Connect (Debezium)
 - Системы управления базами данных (СУБД): PostgreSQL, ClickHouse
 - BI-инструмент: Apache Superset

5. Содержание работы (перечень подлежащих разработке вопросов):

5.1. Введение.

5.2. Постановка задачи.

5.3. Обзор существующих решений.

5.4. Введение в предметную область.

5.5. Разработка инструмента

5.6. Проектирование и реализация инфраструктуры для работы с большими данными

5.7. Исследование разработанного продукта

5.8. Заключение

Ключевые источники литературы:

- Альфред Ахо, Рави Сети, Джефффри Ульман. Раскрутка // Компиляторы: принципы, технологии и инструменты = Compilers: Principles, Techniques, and Tools. — М.: Вильямс, 2003. — С. 681—684. — 768 с. — ISBN 5-8459-0189-8.
- Фаулер М. Непрерывная поставка: Надежная автоматизация сборки, тестирования и развертывания программного обеспечения = Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation. — М.: Вильямс, 2011. — 432 с. — ISBN 978-5-8459-1739-3
- Таненбаум Э., ван Стин М. Распределенные системы: принципы и парадигмы = Distributed Systems: Principles and Paradigms. — 2-е изд. — М.: ДМК Пресс, 2021. — 584 с. — ISBN 978-5-97060-708-4

6. Дата выдачи задания: 03.02.2025.

Руководитель ВКР

Ф.А. Новиков

Консультант ВКР

Д.Ю. Иванов

Задание принял к исполнению

Студент

И.И. Хамидуллин