**Выпускник**

**Ваше полное ФИО**

Полное наименование вашего института/факультета/центра

e-mail: Место для ввода текста.

**Научный руководитель**

**Полное ФИО Вашего научного руководителя**

Степень, звание, место работы научного руководителя

e-mail: Место для ввода текста.

УДК 004.04

**НАИМЕНОВАНИЕ работы**

Работа выполнена в рамках темы НИР №619296 «Исследование и применение методик обработки промышленных данных».

**Ключевые слова**

Киноа, разработка рецептур, белковая составляющая продукта, аминокислотный состав, мороженое.

**Цели и задачи**

Современная промышленность характеризуется использованием высокотехнологичного оборудования и передового программного обеспечения (ПО). Оборудование снабжено большим количеством датчиков, которые собирают различного рода информацию, а ПО работает с данными, а также генерирует их. Вся информация хранится на серверах. В связи с этим, встает задача анализа и обработки данных с целью построениях систем оповещения, принятия решений и рекомендательных систем.

**Основная часть (методология, результаты)**

Компания NVIDIA разработала и предложила набор библиотек RAPIDS, который дает возможность выполнять аналитику и обработку данных исключительно на графических процессорах (graphics processing unit, GPU). Данный подход позволяет ускорять этапы подготовки данных и машинного обучения за счет представления информации в колоночном виде на основе технологии Apache Arrow [2]. Технологии NVIDIA делают возможным объединение нескольких GPU в кластер, что дает значительные преимущества, такие как более высокая скорость обработки, экономия на аппаратной части и на операционных расходах (узел из нескольких GPU ведет себя гораздо более эффективно, чем классический сервер, с точки зрения энергопотребления и тепловыделения).

RAPIDS разработана на основе обширного опыта работы с аппаратным обеспечением и областью науки о данных [3]. Данная технология использует инструмент NVIDIA CUDA для низкоуровневого взаимодействия с аппаратными средствами, что дает возможность выполнять параллельные вычисления, и, с другой стороны, использует язык программирования Python, как интерфейс. На рисунке в общем виде представлено то, как связаны между собой Python, CUDA и GPU для организации эффективной обработки данных и удобства использования.



Рисунок. Упрощенное представление взаимосвязи Python, CUDA и GPU

В состав набора RAPIDS входят такие библиотеки, как cuGraph, cuIO, cuDF и cuML. Каждая из этих библиотек имеет свой аналог в классических Python-библиотеках, а, следовательно, имеет практически идентичный синтаксис, чтобы обеспечить пользователю максимально простой и незаметный переход на данную технологию. Преимущество этих библиотек в том, что они позволяют работать с графическими процессорами, а также в них добавлены новые и более быстрые методы работы с данными.

**Выводы и дальнейшие перспективы исследования**

В ходе работы была рассмотрена технология для ускорения обработки и анализа данных на графических ускорителях. Рассмотренный эксперимент подтвердил эффективность использования ускорения на GPGPU для простых вычислительных операций, что должно обеспечить существенный прирост при использовании методов машинного обучения. Эксперимент проводился с использованием стенда на основе трех графических ускорителей 1080TI.

В дальнейшей работе планируется применение статических и вероятностных методов обработки данных технологического назначения с ускорением на GPU.

**Список использованных источников**

1. Sisyukov A.N., Yulmetova O.S., Kuznecov V.A., «GPU accelerated industrial data analysis in private cloud environment» // Proceedings of the 2019 IEEE Conference. 2019. P. 348–352.
2. Ускорение анализа данных на GPU с RAPIDS [Электронный ресурс]. Режим доступа: https: // info.nvidia.com/accelerate-gpu-data-analysis-with-rapids-reg-page.html (дата обращения: 20.12.2019).
3. Open GPU Data Science | RAPIDS [Электронный ресурс]. Режим доступа: https://rapids.ai/ (дата обращения: 08.01.2020).