



ПОЛИТЕХ

Санкт-Петербургский
Политехнический Университет
Петра Великого

**Санкт-Петербургский политехнический университет Петра Великого
Институт компьютерных наук и технологий
Высшая школа программной инженерии**

Системы анализа больших данных (САБД)

Весенний семестр

Направление: *09.04.04 – «Программная инженерия»*

Аспирант, ассистент ВШПИ

Доцент, к.т.н.

Аспирант, консультант

Ковалев Артем Дмитриевич

Никифоров Игорь Валерьевич

Излев Владислав Александрович

Осенний семестр

- В осеннем семестре были выполнены
 - 4 лабораторные работы
 - выбран инструмент анализа данных
 - сделана презентация с докладом о выбранном инструменте

Весенний семестр

- В этом семестре вам предоставляется выбор:
 1. внести вклад в разработку инструмента анализа больших данных с открытым исходным кодом (не больше 20 человек)
 2. выполнить лабораторные работы посвященные стеку ELK (ElasticSearch, LogStash, Kibana) (не больше 20 человек)
 3. разработать 5-6 лабораторных работ с использованием Grafana + Prometheus (не больше 5 человек)

Бонусы отличникам

- Те, кто в прошлом семестре получили оценку отлично, могут выбрать 2-е направление (ELK-стек) и выполнить только первые 3 пункта лабораторных
- Всем остальным для получения оценки «отлично» в этом семестре необходимо выполнить задание полностью

Задание 1

1. Выбрать программный инструмент, средство, фреймворк или библиотеку с открытым исходным кодом, цель которой (одна или несколько)
 - обработка
 - хранение
 - передача
 - анализ данных
2. Исследовать подход к разработке выбранного инструмента
 - определить систему контроля версий
 - определить систему трекинга ошибок
 - определить наличие документации
 - определить способ сборки продукта и поддерживаемую систему CI
3. Выбрать несколько открытых задач для инструмента, реализовать их
4. Продемонстрировать презентацию, покрывающую сделанную работу

Пример 1 – Hadoop YARN

- система трекинга ошибок JIRA -
<https://issues.apache.org/jira/projects/YARN>
- ИСХОДНЫЙ КОД
 - <https://github.com/apache/hadoop>
- КАК ВНОСИТЬ ВКЛАД
 - <https://cwiki.apache.org/confluence/display/HADOOP/How+To+Contribute>

Пример 2 – Active MQ

- система трекинга ошибок JIRA -
<https://issues.apache.org/jira/projects/AMQ/>
- ИСХОДНЫЙ КОД
 - <https://github.com/apache/activemq>
- КАК ВНОСИТЬ ВКЛАД
 - <https://activemq.apache.org/contributing>

iResolver

- Вхождение в разработку нового инструмента, особенно с открытым исходным кодом, сложная задача
- Для упрощения вхождения в разработку и поиска похожих задач, предлагается воспользоваться инструментом iResolver (issue Resolver)
- Исходный код:
 - <https://github.com/kovaloid/iresolver>

Задание 2

1. Установка ELK стека
2. Скачивание набора данных для анализа
3. Импорт данных в Elasticsearch с помощью Logstash
4. Построение графиков и визуализаций
5. Создание рабочих столов (dashboards)
6. Разработка программного средства для поиска в Elasticsearch на основе REST API

Задание 3

- нужно проработать структуру лабораторных работ (установка инструментов, загрузка данных, визуальный мониторинг, автоматизация развертывания решения)
- сделать описание лабораторных работ в виде текстового документа или презентации
- командная и творческая работа

Правила оценки успеваемости

1. Внесение вклада в open-source
 - Создан pull request и начато ревью – 4
 - PR принят и смержен с мастером – 5
2. Выполнение лабораторных по ELK-стеку
 - Для тех кто получил 5 за прошлый семестр
 - Выполнен 1 пункт – 3
 - Выполнено 2 пункта – 4
 - Выполнены 3 пункта – 5
 - Для тех кто НЕ получил 5 за прошлый семестр
 - Выполнено 4 пункта – 3
 - Выполнено 5 пунктов – 4
 - Выполнены все 6 пунктов – 5
3. Разработка лабораторных работ по Grafana + Prometheus
 - Оценивается индивидуально для каждого участника в зависимости от вклада (3-5)

НОВЫЙ КОНСУЛЬТАНТ

- *Ивлев Владислав Александрович*
- *Можно обращаться по любым вопросам в MS Teams*

К следующему занятию

- Я
 - вышлю все необходимые материалы
- Вы
 - подготовите вопросы по заданиям
 - предоставите ссылку на google таблицу с распределением студентов по заданиям

Спасибо за внимание!

Вопросы?