

ПРОГРАММА РАБОТЫ СЕМИНАРА НА ВЕСНУ 2021 г.

17.02.2021

Р.К. Классен (КНИТУ-КАИ)

ВЫБОР ФУНКЦИИ ДИНАМИЧЕСКОГО СЕГМЕНТИРОВАНИЯ В CLUSTERIX-ПОДОБНЫХ СИСТЕМАХ

Обработка больших массивов данных требует применения вычислительных кластеров и распределения данных по его узлам. За распределение промежуточных данных при обработке запросов в известных Clusterix-подобных системах отвечает функция динамического сегментирования, значение которой – остаток от деления ключа отношения на количество ядер в кластере. Такая функция часто позволяет получить идеальное распределение. Но в случае, когда ключ для соединения отношений содержит несколько уникальных значений, применение хеш-функции, используемой для хеширования хранимых данных, может оказаться не эффективным. В докладе обсуждается применение разных хеш-функций для динамического сегментирования и первоначального распределения по узлам. Демонстрируются карты распределения данных по ядрам кластера для исходных и временных отношений. Приводится сравнение производительности при использовании тех или иных хеш-функций. Даются рекомендации по релевантному выбору хеш-функции для целей динамического сегментирования.

17.03.2021

КРУГЛЫЙ СТОЛ С УЧАСТИЕМ ПРЕДСТАВИТЕЛЕЙ ВЕДУЩИХ ИТ-ПРЕДПРИЯТИЙ г. КАЗАНИ

1. Краткие сообщения:

Р.К. Классен (КНИТУ-КАИ)

СУБД CLUSTERIX-N КЛАССА BIG DATA

Р.Ф. Гибадуллин (КНИТУ-КАИ)

СУБД КАРТОГРАФИЧЕСКИХ И ТЕКСТОВЫХ ДАННЫХ С АССОЦИАТИВНОЙ ЗАЩИТОЙ

Дается обзор разработок исследовательских версий сложных ИТ-систем, выполненных в последние годы на кафедре Компьютерных систем КНИТУ-КАИ им. А.Н. Туполева. Приводятся их технические характеристики и технические предложения по организации дальнейших совместных работ по доведению этих систем до уровня готовых программных продуктов широкого практического применения.

2. Дискуссия по материалам технических предложений на тему:

Практические возможности организации совместных работ КНИТУ-КАИ с ведущими ИТ-фирмами г. Казани ПО ДОВЕДЕНИЮ ПРЕДСТАВЛЕННЫХ СИСТЕМ ДО ЖЕЛАЕМОГО УРОВНЯ.

14.04.2021

Р.Х. Тукшаитов (КГЭУ)

О ТИПОВЫХ ОШИБКАХ, ДОПУСКАЕМЫХ ПРИ ОБРАБОТКЕ И ПРЕДСТАВЛЕНИИ РЕЗУЛЬТАТОВ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ В ОТЕЧЕСТВЕННОЙ И ЗАРУБЕЖНОЙ ЛИТЕРАТУРЕ

Рассмотрены типовые ошибки, допускаемые при обработке и представлении результатов моделирования и эксперимента в отечественной и зарубежной литературе. Проведена дифференциация этих ошибок и самих журналов по уровню применения методов статистической обработки в различных областях знаний (физика, техника, химия, биомедицина). В основу доклада положены материалы монографий за 2002 и 2006 гг. и нескольких последующих публикаций.

12.05.2021

А.Х. Рахматуллин (ОАО «СЕТЕВАЯ КОМПАНИЯ»)

СИНТЕЗ И АНАЛИЗ ЭЛЕМЕНТАРНЫХ АЛГОРИТМОВ ДЛЯ ДИФФЕРЕНЦИАЛЬНОГО НЕЙРОННОГО КОМПЬЮТЕРА

Предложенный исследовательской группой DeepMind подход, объединяющий механизм внимания и внешнюю память в рекуррентных сетях, стал серьезным шагом в развитии вычислений при помощи нейронных сетей. Архитектура, названная нейронной машиной Тьюринга, представляет собой усовершенствованную модель дифференциального нейронного компьютера. Она является расширением классической машины Тьюринга, все операции которой - дифференцируемые. В докладе предлагается метод исследования, основанный на использовании дифференциального нейронного компьютера для реализации основных элементарных алгоритмов. Проводится анализ характеристик построенной модели в сравнении с реализацией основных алгоритмов для машины Тьюринга. Полученные результаты позволяют дать практические рекомендации по использованию дифференциального нейрокомпьютера и возможному улучшению архитектуры нейронной сети.

26.05.2021

Р.К. Классен, И.А. Казанцев (КНИТУ-КАИ)

ВЫБОР ЛУЧШИХ ВАРИАНТОВ ИЗ СГЕНЕРИРОВАННЫХ МНОЖЕСТВ РЕГУЛЯРНЫХ ПЛАНОВ ВЫПОЛНЕНИЯ ЗАПРОСОВ И ФУНКЦИЙ ДИНАМИЧЕСКОЙ СЕГМЕНТАЦИИ CLUSTERIX-NEW

Ранее представленные доклады по теме автоматического претранслятора показали, что его эффективность оставляет желать лучшего. Для поиска более эффективного регулярного плана в претранслятор добавлена схема БД с указанием размера полей и количества строк в отношениях. Как показали экспериментальные исследования, учет этих параметров дает уменьшение влияния сетевых передач за счет более уместного выбора размерности полей и выполнения операций JOIN с порядком отношений от меньшего к большему. Дальнейший рост эффективности СУБД Clusterix-New связан с поиском лучшей функции хеширования для процессов динамической сегментации. Основной критерий выбора хеш-функции – возможность хеширования не только по целым числам, но и по строкам и байтам. В докладе рассматриваются популярные хеш-функции: CRC32, MurMurHash2 и MOD(целочисленный остаток от деления). Производится оценка качества хеширования с помощью специальной метрики, вычисляемая по данным экспериментов. Дается заключение по выбору хеш-функции.