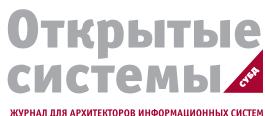




ПРАКТИЧЕСКИЙ СЕМИНАР

*Hadoop, Mapreduce и другие.
Технологии и проекты*

Организаторы



21 ноября
Бизнес-центр Japan House

MapReduce: от грез к реальности

Чего ожидали от этой технологии восторженные юноши, как их пытались пристыдить «старшие товарищи» и как был найден компромисс



СЕРГЕЙ КУЗНЕЦОВ, профессор, научный руководитель отдела информационных систем, **ИСП РАН**

Сергей Кузнецов — один из ведущих российских экспертов в области баз данных и ОС UNIX, представитель IEEE Computer Society в Москве. Является заместителем председателя Московской секции ACM SIGMOD, членом ACM, членом программных комитетов международных конференций DEXA, SOFSEM, ADBIS, ISD, BulticDB, SYRCoDIS. Основные результаты научной деятельности: операционная система ЦП АС-6, кластерная операционная система (КЛОС), GNU SQL-сервер, система виртуальной интеграции данных BizQuery, система управления базами XML-данных Sedna. Автор свыше 100 научных работ и двух учебников.

Обработка Больших Данных: модели и технологии

В последние годы сформировался стек технологий для эффективного решения многих задач обработки Больших Данных (Big Data). Доклад посвящен обзору этого стека. Особое внимание уделяется проекту Hadoop, который де-факто стал промышленным стандартом в данной области. Рассматривается модель и основные элементы проекта Hadoop, текущие границы применения этой технологии, а также альтернативные решения, направленные на расширение этих границ.



ДЕНИС ТУРДАКОВ, заведующий отделом информационных систем, **ИСП РАН**

Денис Турдаков окончил факультет ВМиК Московского государственного университета в 2006 году и аспирантуру того же факультета в 2009 году. Кандидат физико-математических наук. В настоящее время заведует отделом информационных систем Института системного программирования (ИСП РАН) и является научным сотрудником кафедры системного программирования ВМиК МГУ. Руководит работой спецсеминара «Распределенные объектно-ориентированные системы», является членом программного комитета и рецензентом ряда международных конференций. Область научных интересов: обработка естественного языка, машинное обучение, интеллектуальный анализ данных, анализ социальных сетей, распределенная обработка данных.

MapReduce в деталях

В докладе будут раскрыты ключевые концепции выполнения MapReduce-программ, запускаемых на кластере Hadoop

SQL вокруг Hadoop

Данное выступление будет построено на практических примерах работы с кластером Hadoop с использованием SQL-диалектов



АЛЕКСЕЙ ФИЛАНОВСКИЙ, Cloudera Certified Developer for Hadoop, старший консультант, **Oracle**

Алексей Филановский закончил МГТУ им. Баумана в 2011 году. С 2012 года работает в российском представительстве компании Oracle. Имеет шестилетний опыт работы с технологиями хранения и обработки данных (СУБД Oracle, Hadoop, базы данных NoSQL).

Обзор коммерческих продуктов на базе Hadoop

Стек технологий вокруг Hadoop постоянно расширяется. На рынке появляется все больше коммерческих продуктов, либо нацеленных на расширение стандартных возможностей Hadoop, либо дополняющих эту технологию. В докладе будут рассмотрены основные классы задач и функциональные возможности, на «покрытие» которых, как правило, нацелены данные продукты, и предпосылки к появлению таких решений. Для разных классов задач будут приведены примеры продуктов, позволяющих расширить возможности Hadoop и построить функционально более полное решение.



ПЕТР БОРИСОВ, менеджер по развитию услуг и решений, **DIS Group**

Петр Борисов отвечает в DIS Group за развитие новых направлений. Имея более чем десятилетний опыт работы в ИТ-индустрии, Борисов в течение последнего года занимается направлением Big Data, включая проработку как технических аспектов применения технологий данного класса, так и позиционирования подобных решений в инфраструктуре заказчиков. До работы в DIS Group Борисов занимался развитием системы корпоративной отчетности. Закончил Московский государственный технический университет СТАНКИН, факультет «Информационные технологии».

Использование языка JAQL для программирования над Hadoop и интеграции с другими компонентами платформы Big Data

В докладе дается обзор основных характеристик языка программирования над Hadoop JAQL, а также сравнение с другими популярными языками программирования над Hadoop (Pig, Hive). Возможности языка демонстрируются на примере его применения для анализа слабоструктурированных данных большого объема. Рассматриваются средства интеграции с другими компонентами платформы Больших Данных, среди которых: базы данных класса NoSQL, реляционные СУБД, параллельная база данных Netezza, язык программирования для статистической обработки данных R, средства текстового поиска Lucene, средства текстовой аналитики.



АЛЕКСЕЙ ВОВЧЕНКО, старший научный сотрудник, **ИПИ РАН**

Алексей Вовченко окончил факультет ВМиК МГУ им. М. В. Ломоносова в 2007 году и аспирантуру того же факультета в 2010 году. Кандидат технических наук. С 2006 года — сотрудник ИПИ РАН. В настоящее время является старшим научным сотрудником лаборатории методов композиционного проектирования информационных систем ИПИ РАН. Автор спецкурса «Управление разнотипными Большими Данными», читаемого им на факультете ВМиК МГУ для студентов старших курсов и студентов магистратуры. Область научных интересов: виртуальная и материализованная интеграция неоднородных информационных ресурсов, базы данных и базы знаний, Большие Данные, облачные вычисления, проблемно-ориентированные базы данных, извлечение знаний.

Методы и средства текстовой аналитики над Большими Данными в среде IBM BigInsights

В докладе представлен обзор средств текстовой аналитики в среде IBM InfoSphere BigInsights, позволяющих извлекать информацию из неструктурированных и слабоструктурированных данных. Рассматривается декларативный язык AQL (Annotation Query Language) как язык спецификации программ извлечения информации. Возможности языка AQL демонстрируются в среде разработки InfoSphere BigInsights Text Analytics на базе Eclipse.



ДМИТРИЙ БРЮХОВ, старший научный сотрудник, **ИПИ РАН**

Дмитрий Брюхов закончил факультет ВМиК МГУ им. М. В. Ломоносова, кандидат технических наук. В настоящее время является старшим научным сотрудником лаборатории методов композиционного проектирования информационных систем ИПИ РАН. Имеет более 35 научных публикаций. Область научных интересов: интероперабельные среды неоднородных информационных ресурсов, объектные методы проектирования систем, Большие Данные, текстовая аналитика над Большими Данными.

Case Study: проект Spark и анализ социальных графов

В Университете Беркли была разработана модель для организации распределенных вычислений, основанная на понятии устойчивой к сбоям распределенной коллекции данных (RDD). На основе данной модели была реализована система Spark, позволяющая эффективно выполнять итеративные алгоритмы за счет поддержки кэширования результатов в памяти. В докладе будет дан обзор системы Spark и ее преимуществ при распределенной обработке графов.



РОМАН ПАСТУХОВ, научный сотрудник, ИСП РАН

Область научных интересов включает хранение и анализ данных пользователей онлайн-социальных сетей, разработку пользовательских сервисов и аналитических приложений. Ведущий разработчик проекта по распределенной обработке социальных графов. Занимается исследованием технологий баз данных, является одним из разработчиков XML-СУБД Sedna.

Как выбрать слона? Тестируем платформы Hadoop от разных поставщиков

Технология Hadoop все быстрее набирает популярность и все больше компаний начинают рассматривать ее как платформу для решения части своих задач в области хранения и обработки больших объемов данных. В то же время сегодня на рынке Больших Данных присутствует значительное количество поставщиков, которые предлагают промышленные решения на базе технологии Hadoop. Данное выступление посвящено проблеме выбора заказчиком коммерческой Hadoop-платформы для решения существующих задач. Будут рассмотрены основные критерии, которые помогают сделать такой выбор, а также представлены результаты нашего тестирования Hadoop-решений от компаний Cloudera, MapR и IBM по таким критериям, как удобство настройки и администрирования, производительность обработки данных и отказоустойчивость кластера.



ПЕТР БОРИСОВ,
менеджер по
развитию услуг и
решений,
DIS Group



**АЛЕКСАНДР
СИВОХИН,**
ведущий консультант
по направлению Big
Data, **DIS Group**



**ВАЛЕНТИН
УШАКОВ,**
разработчик, консуль-
тант по направлению
Big Data, **DIS Group**

Петр Борисов отвечает в DIS Group за развитие новых направлений. Имея более чем десятилетний опыт работы в ИТ-индустрии, Борисов в течение последнего года занимается направлением Big Data, включая проработку как технических аспектов применения технологий данного класса, так и позиционирования подобных решений в инфраструктуре заказчиков. До работы в DIS Group Борисов занимался развитием системы корпоративной отчетности. Закончил Московский государственный технический университет СТАНКИН, факультет «Информационные технологии».

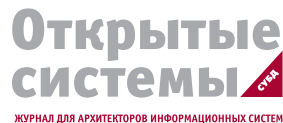
Александр Сивохин является одним из ведущих консультантов DIS Group по направлению Big Data, ездил на обучение в Германию по технологии Hadoop и прошел сертификацию компании Cloudera. До работы в DIS Group Сивохин на протяжении семи лет работал инженером-программистом.

Валентин Ушаков в компании DIS Group является одним из ключевых членов группы Big Data и принимает активное участие как во внешних проектах компании, так и во внутреннем тестировании и изучении технологий Big Data. Ушаков непосредственно занимается тестированием и освоением таких Hadoop-платформ, как IBM BigInsights и MapR. Имеет довольно большой опыт программирования, работает в ИТ-сфере более семи лет.



Агентство корпоративных коммуникаций OSP-Con – подразделение издательства «Открытые системы», в которое в 2004 году был выделен весь event-бизнес. Сегодня OSP-Con является организатором многочисленных конференций для отечественного ИКТ- и бизнес-сообщества. Основной акцент в своей деятельности OSP-Con делает на детальную проработку контентной составляющей проводимых мероприятий, начиная с определения тематики конференции и кончая предоставлением ее участникам собственной аналитической экспертизы. Именно это обстоятельство сделало конференции, проводимые OSP-Con, заметным явлением на российском рынке и позволило агентству занять на нем ведущие позиции.

Подробнее: www.ospcon.ru



Журнал «Открытые системы. СУБД» издается с 1993 года и сегодня является одним из самых авторитетных в России периодических изданий, посвященных построению гетерогенных информационных систем корпоративного уровня; архитектурам современных компьютерных систем и микропроцессоров; операционным системам; СУБД и хранилищам информации, а также веб-технологиям. Большое внимание на страницах журнала уделяется освещению практического опыта разработки и внедрения конкретных решений на промышленных предприятиях, в научных и государственных учреждениях, на транспорте, в медицине и в телекоммуникационных компаниях. В 2007 году журнал вошел в список изданий, рекомендованных Высшей аттестационной комиссией (ВАК).

Подробнее: www.osmag.ru