

Организатор



21 октября
Москва, отель «Корстон»



конференция
БОЛЬШИЕ ДАННЫЕ
в национальной экономике

При поддержке



РОССИЙСКИЙ
ФОНД
ФУНДАМЕНТАЛЬНЫХ
ИССЛЕДОВАНИЙ





Важнейшим условием успешного развития мировой экономики на современном этапе становится возможность фиксировать и анализировать огромные массивы и потоки информации. Существует точка зрения, что страны, которые овладеют наиболее эффективными методами работы с Большими Данными, ждет новая индустриальная революция. У России с ее колоссальным научным и образовательным потенциалом есть все шансы занять достойное место среди тех национальных экономик, где извлечение полезных знаний из больших объемов данных различной природы поставлено на службу индустриальному прогрессу.

Цель второй конференции **«Большие Данные в национальной экономике»** — предоставить экспертам в области работы с данными из научных организаций и индустрии площадку для обсуждения результатов своих проектов и перспектив их практического применения. Организаторы конференции также видят своей задачей популяризацию наиболее интересных работ, способствующих развитию Больших Данных как самостоятельного научного направления, а также расширению использования на практике методов работы с большими массивами информации в других областях научных исследований и в различных секторах российской экономики.

ОТ ИМЕНИ ОРГАНИЗАТОРОВ КОНФЕРЕНЦИИ ЖЕЛАЮ ЕЕ УЧАСТНИКАМ ПРОДУКТИВНОЙ РАБОТЫ.

НАТАЛЬЯ ДУБОВА,

зам. председателя оргкомитета конференции **«Большие Данные в национальной экономике»**,
научный редактор журнала **«Открытые системы.СУБД»**

Организационный комитет:

Гуляев Ю.В.	академик РАН, директор ИРЭ им. В.А. Котельникова РАН, председатель оргкомитета
Дубова Н.А.	научный редактор журнала «Открытые системы.СУБД», зам. председателя оргкомитета
Арлазаров В.Л.	чл.-корр. РАН, зав. отделом ИСА РАН
Будзко В.И.	д.т.н., зам. директора по научной работе ИПИ РАН
Волков Д.В.	с.н.с. ИПМ им. М.В. Келдыша РАН, гл. редактор журнала «Открытые системы.СУБД»
Гергель В.П.	д.т.н., декан факультета ВМК ННГУ им. Н.И. Лобачевского
Кузнецов Н.А.	академик РАН, президент Международного союза приборостроителей
Никитов С.А.	чл.-корр. РАН, зам. директора ИРЭ им. В.А. Котельникова РАН, зав. кафедрой МФТИ
Сухомлин В.А.	д.т.н., профессор, МГУ им. М.В. Ломоносова

СЕРГЕЙ КУЗНЕЦОВ,

д.т.н., главный научный сотрудник, ИСП РАН

БОЛЬШИЕ ДАННЫЕ: РАЗДЕЛЯЙ И ВЛАСТВУЙ

Сравнительно установившейся идеей является то, что горизонтально масштабируемые системы обработки данных можно основывать только на подходе shared nothing (грубо говоря, на распределенных системах без наличия общих ресурсов между узлами). Легко видеть, что этому подходу в той или иной мере следуют как массивно-параллельные СУБД, так и решения NoSQL (в частности, map/reduce). Однако, как нетрудно заметить, подобные системы хорошо работают только в том случае, когда данные хорошо разделены по узлам системы (иными словами, разделение соответствует текущей рабочей нагрузке). И обещанное горизонтальное масштабирование реально возможно только в том случае, когда данные можно эффективно переразделить. Так что реально проблемой Больших Данных является проблема их разделения. Научимся разделять — проблемы почти и не будет.

ФУАД АЛЕСКЕРОВ,

д.т.н., руководитель Департамента математики факультета экономики, НИУ ВШЭ

МОДЕЛИ ВЫБОРА ДЛЯ АНАЛИЗА БОЛЬШИХ ДАННЫХ

Предлагаются модели выбора для задачи поиска в анализе Больших Данных. Эти модели включают процедуры, обобщающие модели, предложенные Г. Саймоном, и другие классические и новые модели выбора, такие как q-Паретовские правила, экстремизационные, экстремизационные с погрешностью, надпороговое и минимаксное правила и многие другие. Рассматриваются модели суперпозиции этих правил, и исследуется их сложность.

Показано, что предлагаемые модели работают эффективнее, чем многие известные правила, в частности метод опорных векторов.

Проведено сравнение эффективности различных процедур на данных компании Microsoft.

АЛЕКСЕЙ ВОВЧЕНКО,
к.т.н., с.н.с., ИПИ РАН

СЕРГЕЙ СТУПНИКОВ,
к.т.н., с.н.с., ИПИ РАН

МЕТОДЫ И ИНФРАСТРУКТУРЫ ИНТЕГРАЦИИ РАЗНОРОДНЫХ БОЛЬШИХ ДАННЫХ

В докладе рассматривается архитектура комбинированной виртуально-материализованной среды интеграции неоднородных коллекций данных различного вида (структурированных, слабоструктурированных и неструктурированных). Необходимость поддержки двух различных видов интеграции объясняется тем, что как виртуальный, так и материализованный подходы к интеграции имеют свои достоинства и недостатки. Виртуальная интеграция осуществляется с использованием технологии предметных посредников. Материализованная интеграция реализуется с использованием свободно распространяемой платформы распределенного хранения и обработки данных Hadoop, а также системы организации реляционных хранилищ данных над Hadoop, в качестве которой могут использоваться платформы Big SQL или Hive.

АНДРЕЙ ДМИТРИЕВ,
д.ф.-м.н., профессор, НИУ ВШЭ

ПРЕДСКАЗАТЕЛЬНАЯ АНАЛИТИКА НА ОСНОВЕ ПОТОКОВ БОЛЬШИХ ДАННЫХ

Дается краткий обзор методов предсказательной аналитики на основе потоков больших массивов данных, генерируемых в режиме реального времени. Более подробно обсуждаются возможности и ограничения методов детектирования кризисов и предкризисных режимов кризисов в данных потоках. Предложена нелинейно-динамическая модель, генерирующая временные ряды рыночных цен спроса и предложения с одним фундаментальным управляющим параметром. Калибровкой данного параметра по историческим данным и выделением интервалов его постоянства при определенных условиях удалось детектировать предкризисные режимы. Данная модель апробирована на биржевых данных по ценам спроса и предложения на драгоценные металлы и стальные билеты. Разработано приложение, позволяющее на основе данной динамической модели детектировать предкризисные режимы в потоках больших массивов данных, генерируемых в режиме реального времени.

МИХАИЛ ЦЫМБЛЕР,
к.ф.-м.н., доцент, ЮУрГУ

ПОИСК ПОХОЖИХ ПОДПОСЛЕДОВАТЕЛЬНОСТЕЙ ВРЕМЕННЫХ РЯДОВ НА СОПРОЦЕССОРАХ INTEL XEON PHI

Временной ряд (time series) представляет собой совокупность вещественных значений, каждое из которых ассоциировано с последовательными отметками времени. Задача поиска похожих подпоследовательностей (subsequence matching) заключается в нахождении подпоследовательностей временного ряда, имеющих ту же длину, что и заданный шаблон поиска, и минимальное отличие от шаблона в смысле используемой функции схожести рядов. Задача поиска похожих подпоследовательностей временного ряда возникает в широком спектре научных и прикладных задач из различных предметных областей: медицина, финансы, моделирование климата и др. В докладе представлен параллельный алгоритм решения данной задачи на сопроцессоре Intel Xeon Phi.

МИХАИЛ КОМАРОВ,
к.т.н., доцент, НИУ ВШЭ,

ЕВГЕНИЙ КУЧЕРЯВЫЙ, PhD, руководитель лаборатории
ET4NBIC Lab, Технологический университет Тампере
(Финляндия)

ИСТОЧНИКИ БОЛЬШИХ ДАННЫХ И СОВРЕМЕННЫЕ СПОСОБЫ ХРАНЕНИЯ ДАННЫХ

В рамках доклада будет показана тенденция технологического развития к мобильным сервисам реального времени, будут рассмотрены современные технологии, которые, в свою очередь, являются источниками Больших Данных, показана одна из ключевых проблем защиты персональных данных при обработке данных, а также кратко представлены современные способы их хранения.

ЕВГЕНИЙ ПАВЛОВСКИЙ,

к.ф.-м.н., старший преподаватель, НГУ, директор, «Исследовательские системы»

НАУЧНЫЕ ВЫЗОВЫ БОЛЬШИХ ДАННЫХ

В докладе предлагаются новые постановки задач в области математики, статистики, теории измерений, теории вероятностей, математической логики, математической лингвистики, возникающие из необходимости отвечать на вызовы Больших Данных.

ВИКТОР ТОПОРКОВ,

д.т.н., зав. кафедрой, НИУ МЭИ

БОЛЬШИЕ ДАННЫЕ И ВЫЧИСЛИТЕЛЬНАЯ НАУКА: МЕСТО И ВРЕМЯ ВСТРЕЧИ

Вопросы, связанные с большими задачами, для решения которых не хватает вычислительных ресурсов даже суперкомпьютерного класса, привлекают к себе внимание уже не один десяток лет. Однако попытки использования масштабируемых сред заставляют переосмыслить многие проблемы и взглянуть на них по-новому.

Масштабируемость подразумевает возможность наращивания количества процессоров, емкости памяти и независимость пропускной способности коммуникационной системы от числа процессорных узлов, участвующих в вычислениях. Распределенные среды, как известно, характеризуются потенциальной ненадежностью, но не с точки зрения возможного отказа или сбоя компьютеров. Здесь акцент делается на том, что у компьютеров нет никаких обязательств перед самой средой и в любой момент времени они могут отключиться от сети вне зависимости от того, завершена ли обработка выделенного задания. Как же в этом случае обеспечить согласование структуры задания с динамично изменяющимся составом, в общем случае, неоднородных ресурсов для эффективной организации вычислений? Иными словами, как обеспечить масштабируемость среды при соответствующих значениях показателей качества обслуживания.

В докладе основное внимание уделяется новым результатам, полученным в двух направлениях. Одно из них — построение моделей и парадигм распределенных вычислений, наиболее адекватных особенностям масштабируемых сред. Другое — выбор конфигурации ресурсов и планирование вычислений.

ВЛАДИМИР КОРЕНЬКОВ,

д.т.н., директор лаборатории информационных технологий, **ОИЯИ**, зав. кафедрой, Международный университет «Дубна»

АЛЕКСЕЙ КЛИМЕНТОВ,

к.ф.-м.н., **Брукхейвенская национальная лаборатория (США)**

РАСПРЕДЕЛЕННЫЕ ВЫЧИСЛЕНИЯ И БОЛЬШИЕ ДАННЫЕ В ЯДЕРНЫХ ИССЛЕДОВАНИЯХ (НА ОСНОВЕ ОПЫТА РОССИИ И ОИЯИ)

В докладе представлены концепция и эволюция глобальной компьютерной инфраструктуры для хранения, обработки и анализа данных экспериментов на Большом адронном коллайдере в ЦЕРНе. Дается краткая информация об участии России в этом процессе. Представлен обзор проектов в области распределенных вычислений и Больших Данных, выполненных Лабораторией информационных технологий (ЛИТ ОИЯИ) в России, ЦЕРНе, США, Китае, странах — участницах ОИЯИ.

Особое внимание уделено созданию центра уровня Tier 1 в России для хранения и обработки данных экспериментов на Большом адронном коллайдере, развитию облачной и гибридной инфраструктуры, а также модели компьютинга мегапроекта NICA в ОИЯИ.

Представлены результаты и планы развития платформы для управления Большими Данными.

АНДРЕЙ УСТЮЖАНИН,

к.ф.-м.н., руководитель проектов **CERN** и «Яндекс», приглашенный исследователь, **Лондонский имперский колледж**

ВОСПРОИЗВОДИМОСТЬ ЧИСЛЕННЫХ ЭКСПЕРИМЕНТОВ

Одним из больших «секретов» обработки Больших Данных является переход от восприятия отдельных событий к статистически значимым закономерностям системы. Работа с закономерностями позволяет строить модели происходящего и с их помощью прогнозировать развитие системы. Зачастую понимание таких закономерностей требует от исследователя экспериментального подхода — поиска наиболее значимых переменных модели, поиска метрик и подбора различных параметров модели.

Такие проверки, которые в общем виде принято называть экспериментами, требуют огромного количества времени и ресурсов. Эффективная работа с экспериментами является ключевым фактором эффективности работы с Большими Данными. Одним из важных путей экономии ресурсов, повышения эффективности и прозрачности работ является воспроизводимость экспериментов — возможность получить те же самые результаты на тех же самых входных данных. В докладе будут рассмотрены примеры численных экспериментов, с которыми сталкиваются исследователи «Яндекса» и ученые ЦЕРНа, и технические способы повышения воспроизводимости экспериментов.

СЕКЦИЯ.

Большие Данные в научных исследованиях

СЕРГЕЙ БЕРЕЗИН,

к.ф.-м.н., доцент, руководитель совместного исследовательского центра МГУ — Microsoft Research, МГУ

ОБЛАЧНЫЕ ТЕХНОЛОГИИ В ЕСТЕСТВЕННЫХ И ГУМАНИТАРНЫХ НАУКАХ

В докладе будут рассмотрены облачные приложения FetchClimate и «Хронозум», совместно разработанные Microsoft Research и МГУ им. М.В. Ломоносова. FetchClimate ориентирован на специалистов в области вычислительного моделирования и предоставляет пользовательский интерфейс и веб-сервис для получения климатических данных. «Хронозум» — ориентированное на историков и представителей гуманитарных наук приложение для визуализации Универсальной истории. На примере приложений FetchClimate и «Хронозум» будут рассмотрены вопросы создания, практического применения и возможного развития облачных приложений для работы с научными данными.

СВЕТЛАНА МАЛЬЦЕВА,
д.т.н., профессор, НИУ ВШЭ

ВЛИЯНИЕ ТЕХНОЛОГИЙ БОЛЬШИХ ДАННЫХ НА ДИЗАЙН ОРГАНИЗАЦИИ

Большинство компаний сегодня проявляют интерес к возможностям использования технологий Больших Данных, многие компании уже используют эти технологии. Фазы зрелости самой технологии при этом различны, в большинстве случаев это начальные фазы. Повышение зрелости используемых в организации технологий Больших Данных связано с новыми возможностями оптимизации бизнеса, превращением данных в ценные информационные активы предприятия, трансформацией самой организации. Критически важными являются изменения в управлении данными, ИТ-инфраструктуре в целом, структуре и компетенциях персонала. Важной задачей является оценка эффективности внедрения технологий Больших Данных.

МИХАИЛ СЛИВИНСКИЙ,
руководитель отдела маркетинговой и поисковой аналитики, «Викимарт»

ВОЗМОЖНОСТИ ИЗВЛЕЧЕНИЯ МАРКЕТИНГОВОЙ ИНФОРМАЦИИ В E-COMMERCE

Мы не занимаемся фундаментальными исследованиями. Наша задача — быстро создавать полезные для e-commerce решения на основе собственной статистики и внешних источников данных. Соответственно, исследовательская часть заведомо небезупречна, и мы знаем об этом. Однако, надеюсь, наши идеи и подходы могут быть интересны.

Основные темы доклада:

1. Особенности маркетинга online и offline.
2. Общая задача — понять пользователя на всех стадиях.
3. Что мы хотим знать о покупателе?
4. Собственная статистика и внешние источники данных.
5. Некоторые наши решения:
 - задачи нашей Олимпиады (<http://olymp.wikimart.ru/>);
 - рекомендательные системы, ранжирование товаров;
 - измерение потребности пользователя в различных типах контента;
 - обнаружение большого количества неочевидных поисковых запросов в заданном кластере.

Программа конференции «Большие Данные в национальной экономике»

9:30	10:00	РЕГИСТРАЦИЯ УЧАСТНИКОВ	
ПЛЕНАРНАЯ СЕССИЯ. ПЕРСПЕКТИВНЫЕ МЕТОДЫ АНАЛИЗА БОЛЬШИХ ДАННЫХ			
10:00	10:25	Большие Данные: разделяй и властвуй	Сергей Кузнецов, д.т.н., МГУ им. М.В. Ломоносова, МФТИ, ИСП РАН
10:25	10:50	Модели выбора для анализа Больших Данных	Фуад Алескеров, д.т.н., руководитель Департамента математики факультета экономики, НИУ ВШЭ
10:50	11:15	Методы и инфраструктуры интеграции разнородных Больших Данных	Алексей Вовченко, к.т.н., с.н.с., Сергей Ступников, к.т.н., с.н.с., ИПИ РАН
11:15	11:35	Предсказательная аналитика на основе потоков Больших Данных	Андрей Дмитриев, д.ф.-м.н., профессор, НИУ ВШЭ
11:35	11:55	Поиск похожих подпоследовательностей временных рядов на сопроцессорах Intel Xeon Phi	Михаил Цымблер, к.ф.-м.н., доцент, ЮУрГУ
11:55	12:15	Источники Больших Данных и современные способы хранения данных	Михаил Комаров, к.т.н., доцент, НИУ ВШЭ, Евгений Кучерявый, PhD, руководитель лаборатории ET4NBIC Lab, Технологический университет Тампере (Финляндия)
12:15	12:30	КОФЕ-БРЕЙК	
СЕКЦИЯ. БОЛЬШИЕ ДАННЫЕ В НАУЧНЫХ ИССЛЕДОВАНИЯХ			
12:30	12:50	Научные вызовы Больших Данных	Евгений Павловский, к.ф.-м.н., старший преподаватель, НГУ, директор, «Исследовательские системы»
12:50	13:10	Большие Данные и вычислительная наука: место и время встречи	Виктор Топорков, д.т.н., зав. кафедрой, НИУ МЭИ
13:10	13:30	Распределенные вычисления и Большие Данные в ядерных исследованиях (на основе опыта России и ОИЯИ)	Владимир Кореньков, д.т.н., директор лаборатории информационных технологий, ОИЯИ, зав. кафедрой, Международный университет «Дубна», Алексей Климентов, к.ф.-м.н., Брукхейвенская национальная лаборатория (США)
13:30	13:50	Воспроизводимость численных экспериментов	Андрей Устюжанин, к.ф.-м.н., руководитель проектов CERN и «Яндекс», приглашенный исследователь, Лондонский имперский колледж
13:50	14:10	Облачные технологии в естественных и гуманитарных науках	Сергей Березин, к.ф.-м.н., доцент, руководитель совместного исследовательского центра МГУ – Microsoft Research, МГУ им. М.В. Ломоносова

Программа конференции «Большие Данные в национальной экономике»

СЕКЦИЯ. ПРИКЛАДНЫЕ АСПЕКТЫ БОЛЬШИХ ДАННЫХ			
14:10	14:30	Влияние технологий Больших Данных на дизайн организации	Светлана Мальцева , д.т.н., профессор, НИУ ВШЭ
14:30	14:50	Возможности извлечения маркетинговой информации в e-commerce	Михаил Сливинский , руководитель отдела маркетинговой и поисковой аналитики, «Викимарт»
14:50	15:00	ЗАКРЫТИЕ КОНФЕРЕНЦИИ	



Об Издательстве «Открытые системы»

Издательство «**Открытые системы**» – ведущее российское издательство, выпускающее широкий спектр журналов для профессионалов и активных пользователей в сфере ИТ, цифровых устройств, телекоммуникаций, медицины и полиграфии, а также журналы для детей. Издательство является организатором крупных и значимых конференций и форумов по ИТ-тематике: IT Management Forum, BIG Data, Мир ЦОД, Бизнес-Видео, Ethernet Forum, МСКФ и других.

Тесное сотрудничество с крупнейшим международным издателем технической литературы и профессиональной прессы International Data Group (IDG) позволяет поддерживать высочайший профессиональный уровень и обеспечивать актуальность публикаций и мероприятий.

Подробнее: www.osp.ru

Планируйте. Приходите. Участвуйте!

28 октября

ETHERNET FORUM 2014

- Сетевая инфраструктура для ЦОД и «облаков»
- Сетевые решения для современного офиса
- Сеть для умного города
- Industrial Ethernet

Wi-Fi

InterDC

Fabrics

SDN

25 ноября



- Выбор облачной платформы
- Где разместить «облако»
- Безопасность в «облаке»
- Гибридная модель

12 ноября



HADOOP НА ПРАКТИКЕ.
НОВЫЕ ИНСТРУМЕНТЫ И ПРОЕКТЫ

- С чего начать собственный проект больших данных
- Критический взгляд на hadoop
- Уроки реальных проектов
- Инструменты дополняющие hadoop

26 марта



По вопросам участия: Ольга Пуркина

+7 495 725 47 80 kon@osp.ru www.ospcon.ru

 **OSP CON**
BRINGING TOGETHER

 **ОТКРЫТЫЕ СИСТЕМЫ**
Open Systems Publications