



ВТОРОЙ РОССИЙСКИЙ ПРАКТИЧЕСКИЙ СЕМИНАР

***Hadoop, MapReduce и другие.
Опыт реальных проектов***

Организаторы

**Открытые
системы**
ЖУРНАЛ ДЛЯ АРХИТЕКТОРОВ ИНФОРМАЦИОННЫХ СИСТЕМ

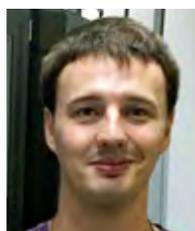
 **OSPCON**
BRINGING TOGETHER



АЛЕКСАНДР ЛЕБЕДКИН,
руководитель, malerex.com

Как Facebook оптимизировала систему машинного обучения?

Facebook использует системы, основанные на обработке Больших Данных и машинном обучении, во многих проектах, таких как распознавание лиц или рекламная биржа Facebook Exchange (FBX). В своем докладе, основанном на опыте работы в Facebook, я расскажу о том, как порой весьма простые оптимизации при работе с Большими Данными способны ускорить работу всей системы машинного обучения в несколько раз.



АЛЕКСЕЙ ФИЛАНОВСКИЙ,

Cloudera Certified Developer for Apache Hadoop, старший консультант отдела продаж, **Oracle**

Тестирование технологий Big Data для телекома. Задачи, решения, результаты

В докладе будет рассказано о тестировании технологий обработки Больших Данных для решения задач одного из операторов связи. Будут показаны и прокомментированы результаты для каждого из кейсов. Обсуждения и дискуссии на тему применимости Hadoop и его окружения в рамках задач заказчиков по ходу доклада всячески приветствуются.

Cloudera Impala. Опыт реального тестирования

Доклад посвящен описанию концептуальных основ продукта Cloudera Impala, а также опыту его практического применения.



ПАВЕЛ МЕЗЕНЦЕВ,
руководитель отдела
Больших Данных, Банк
«Тинькофф Кредитные
системы»



ВАСИЛИЙ КАТРАЕВ,
старший разработчик
проектов Big Data, Банк
«Тинькофф Кредитные
системы»

Matching big data sets: умное «пересечение» независимых множеств

Один из главных сценариев использования Больших Данных в компаниях — это предсказание интересов и поведения клиентов и выстраивание на основе этого персонального подхода к каждому из клиентов.

К сложным задачам, которые приходится решать в ходе построения клиентской модели, является задача метчинга, но задача нахождения одних сущностей в основном наборе данных и в добавляемых к нему данных. Традиционные подходы и инструменты для подобного метчинга не годятся, так как не масштабируются на терабайты данных.

В докладе мы расскажем о нашем подходе к метчингу, в котором используется связка hBase + Solr, о тех проблемах, с которыми мы столкнулись в процессе реализации задачи, и предложенных нами способах их решения.



БОРИС ВАСИЛЕВСКИЙ,
руководитель разработки «Поиска по людям, «Яндекс»

«Яндекс. Поиск по людям» и YaMapReduce

Доклад состоит из двух частей. В первой мы посмотрим на разработческий кластер с одной из «яндексских» реализаций MapReduce, обсудим ее особенности и запустим пару задач. Во второй части мы рассмотрим действующую систему подготовки базы данных для поиска по людям.



ВЯЧЕСЛАВ МУРАШКИН,
ведущий аналитик, **RetailRocket**

Как измерить слона? Оценка производительности кластера Hadoop

В докладе описан подход к оценке производительности Hadoop-кластера. Рассмотрены наиболее вероятные узкие места, влияющие на производительность кластера. Приведен пример сравнения производительности двух кластеров с разными аппаратными конфигурациями.



АНДРЕЙ РИВКИН,
Cloudera Certified Developer for Apache Hadoop, эксперт-консультант,
«ФОРС Дистрибуция»

Использование Cloudera Search в качестве корпоративной поисковой системы

В выступлении будет представлен опыт компании «Форс» в области внедрения бесплатной поисковой системы на основе решения Cloudera Search для поиска внутри документов. Докладчик объяснит, зачем нужно такое решение, каковы его плюсы и минусы. Будут рассмотрены «подводные камни», обнаружившиеся в ходе реализации данного проекта. Наконец, будет рассмотрен философский вопрос: – с чего начать и где искать помощь при знакомстве с Big Data?



КИРИЛЛ ЧИХРАДЗЕ,

научный сотрудник, отдел информационных систем, **ИСП РАН**

Распределенная генерация социальных графов на Amazon EC2 с использованием Apache Spark

В области анализа социальных сетей возникает проблема оценки степени масштабируемости алгоритмов и необходимости тестирования на Больших Данных. Однако сбор реальных данных из интерфейсов социальных сервисов в объеме, сопоставимом с размером сети, сопряжен с существенными временными и материальными затратами. Поэтому необходим инструмент, способный генерировать случайные социальные графы размером более миллиарда вершин. Эти графы должны обладать реалистичной структурой сообществ и свойствами реальных сетей, так как от этого зависит достоверность тестирования на сгенерированных графах.

В докладе будет представлен генератор, удовлетворяющий вышеописанным требованиям и реализованный с помощью системы Apache Spark.

Spark — фреймворк, позволяющий эффективно выполнять итеративные алгоритмы за счет поддержки кэширования результатов в памяти и базирующийся на модели для организации распределенных вычислений с использованием понятия устойчивой к сбоям распределенной коллекции данных (RDD). В ходе тестирования данного алгоритма на Amazon EC2 было установлено, что масштабируемость генератора близкая к линейной. Также был произведен распределенный расчет графовых статистик, с помощью которого подтверждена корректность как самой модели, так и ее реализации.

EMC²

ПЕРВЫЙ РУССКОЯЗЫЧНЫЙ ОНЛАЙН КУРС

Специалисты Центра исследований и разработок EMC Россия в сотрудничестве с преподавателями кафедры Общей информатики ФИТ НГУ поддержали разработку первого русскоязычного онлайн-курса по большим данным. Курс "Введение в аналитику больших массивов данных" рассчитан на аудиторию, имеющую высшее образование и не знакомых с понятием <большие данные>.

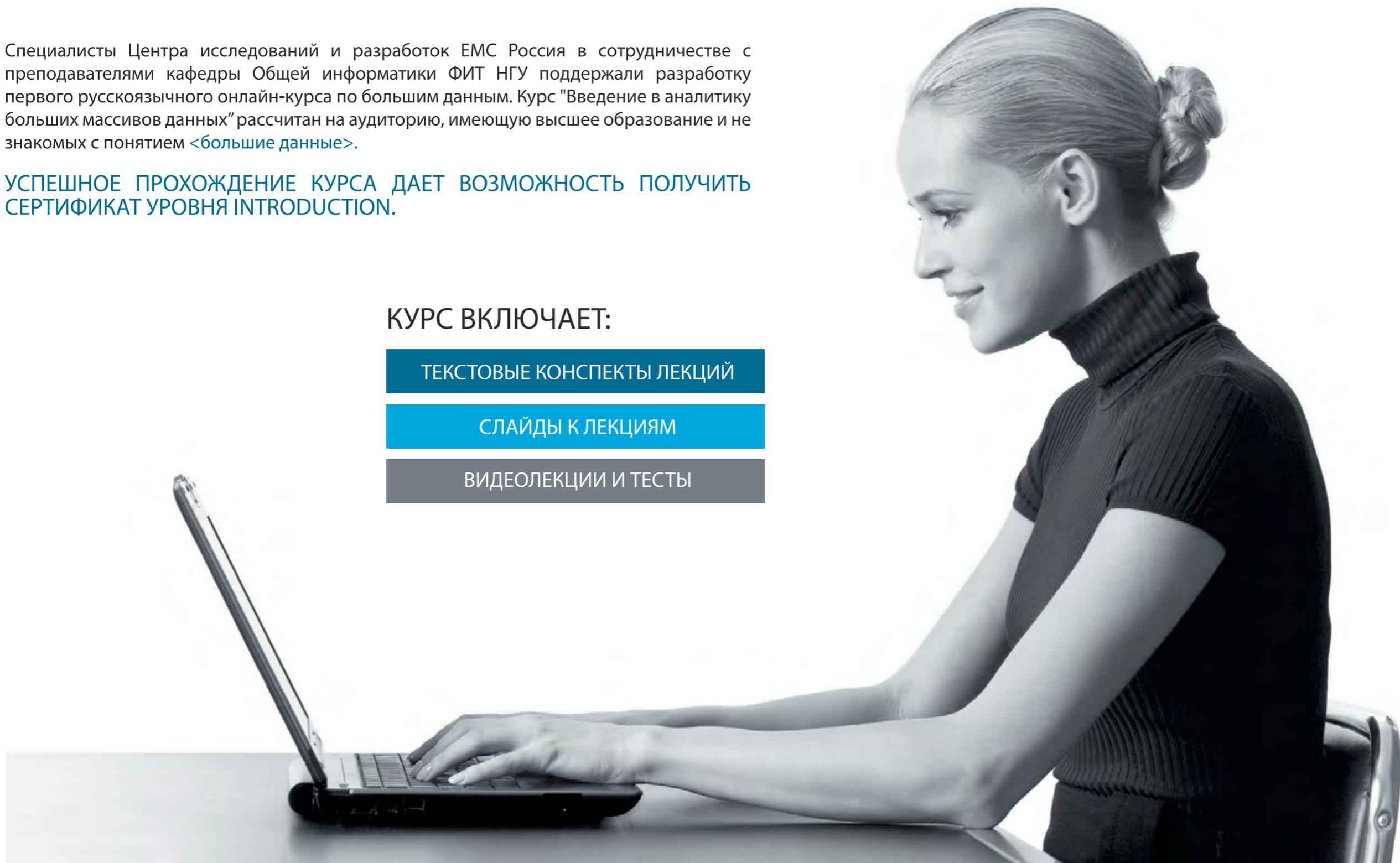
УСПЕШНОЕ ПРОХОЖДЕНИЕ КУРСА ДАЕТ ВОЗМОЖНОСТЬ ПОЛУЧИТЬ
СЕРТИФИКАТ УРОВНЯ INTRODUCTION.

КУРС ВКЛЮЧАЕТ:

ТЕКСТОВЫЕ КОНСПЕКТЫ ЛЕКЦИЙ

СЛАЙДЫ К ЛЕКЦИЯМ

ВИДЕОЛЕКЦИИ И ТЕСТЫ



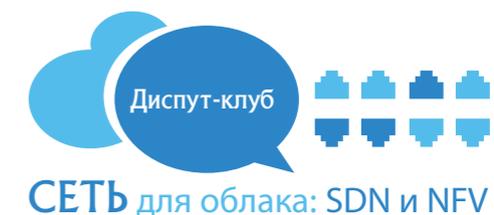
14 мая

- Бизнес-выгоды мобильных технологий
- Инфраструктура мобильного предприятия: построение и эксплуатация
- Подводные камни концепции BYOD
- Мобильные бизнес-приложения и их администрирование
- Защита мобильной корпоративной среды
- Корпоративная мобильность и облака



Реклама 16+

15 мая



- Виртуализация сети и сетевых функций (NFV)
- SDN: философия или технология
- Сеть для приложений
- Миграция к сети NG



27-28 мая

Ежегодный международный форум
«МИР ЦОД – 2014» (DCWorld Forum 2014)

мир цод
[2014]



МИР ТЕХНОЛОГИЙ — технологии и решения, определяющие развитие индустрии ЦОД и гарантирующие компаниям бизнес-преимущества

МИР ПРОЕКТОВ — уникальный опыт проектирования, построения и эксплуатации ЦОД разных отраслей экономики

МИР ЗНАНИЙ — мастер-классы и практические семинары

Открой свой «МИР ЦОД»

www.osp.ru/dcworld



5 июня

XI Russian
IT Management
Forum

- Как построить сервисную стратегию
- Управление организационными изменениями
- Процессы ITSM: отраслевая практика внедрения
- ITSM «для маленьких» - управление ИТ-сервисами в небольших компаниях
- Универсальный ITSM: сервисный подход вне ИТ
- Управление ИТ в эпоху облаков, консьюмеризации ИТ, Больших Данных и мобильности



www.osp.ru/itsm/



«Открытые Системы. СУБД» – издание для архитекторов информационных систем. Подробные обзоры современных технологий и анализ тенденций развития информационных инфраструктур предприятий. Журнал о том, что профессионалам в области ИТ надо знать сегодня про Большие Данные, мобильные предприятия, безопасность и бизнес-аналитику и к чему готовиться завтра.

Подробнее: www.osmag.ru



Агентство корпоративных коммуникаций OSP-Con – подразделение издательства «Открытые системы». OSP-Con является организатором многочисленных конференций для отечественного ИКТ- и бизнес-сообщества. Основной акцент 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.

Подробнее: www.ospcon.ru



ВТОРОЙ РОССИЙСКИЙ ПРАКТИЧЕСКИЙ СЕМИНАР