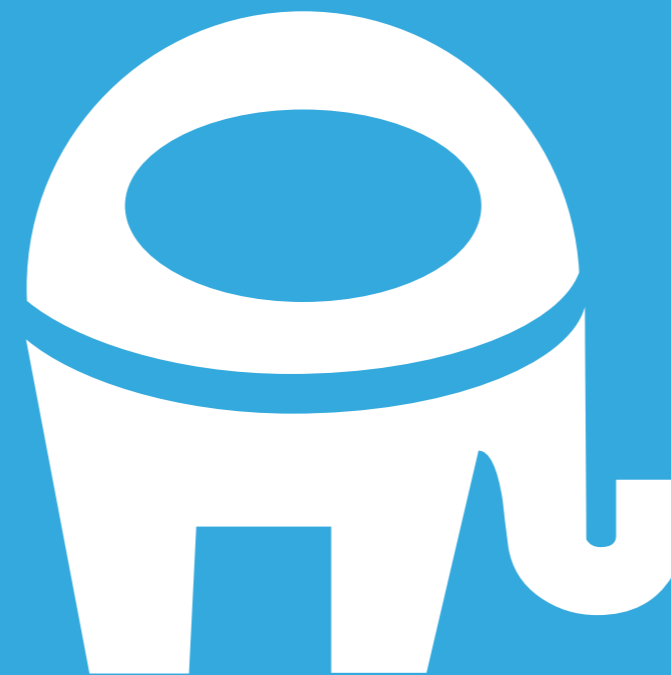


22 апреля
Отель Мандарин

	Открытые системы
2015	
	ЦИКЛ ПРАКТИЧЕСКИХ СЕМИНАРОВ



HADOOP

HADOOP НА ПРАКТИКЕ:
ПРОЕКТЫ И ИНСТРУМЕНТЫ

Организаторы





ДМИТРИЙ ВОЛКОВ,
ведущий цикла практических семинаров,
главный редактор, «Открытые системы.СУБД»

Зародившись как набор библиотек вокруг технологии MapReduce и постепенно превратившись в полнофункциональную систему масштабируемой обработки распределенных данных в условиях интерактивных и поточных разнопрофильных нагрузок, Hadoop сегодня претендует на роль фактического стандарта в мире распределенной обработки Больших Данных.

В России постоянно растет интерес к инструментам работы с большими объемами неструктурированных данных, ведутся проекты на базе технологий Hadoop, однако разработчики и архитекторы по-прежнему ощущают дефицит сведений о новых тенденциях и практике применения технологий из стека Hadoop — мало возможностей для живого общения с теми, кто уже работает с Hadoop. В результате многие проекты не реализуются, либо не оправдывают ожиданий. Этот пробел призван устранить четвертый практический семинар «Hadoop: проекты и инструменты» издательства «Открытые системы». Наши семинары — проводник в мир Hadoop для новых слушателей и щедрый источник практического и профессионального опыта для тех, кто уже принимал участие в них.





ЮРИЙ ПЕТРОВ,

Инновационный центр космической медицины

Практика миграции реляционных баз данных в экосистему Hadoop

Современный бизнес уже не устраивают реляционные базы данных, не позволяющие на лету выполнять сложные вычисления, обеспечивающие неограниченное линейное масштабирование, и одинаково эффективно работать как со структурированными, так и с неструктурированными данными. Кроме этого, решение важной для бизнеса задачи снижения стоимости владения ИТ определяется сегодня способностью СУБД работать на любом оборудовании. Набирающая популярность экосистема Hadoop потенциально способна решить эти и ряд других проблем, связанных с развертыванием и надежной эксплуатацией масштабируемых систем обработки и хранения больших объемов данных. В докладе анализируется практика построения решений на базе Hadoop и их способность удовлетворять требованиям современного бизнеса.





ГЕРАРД КОСТИН, «Мобильные ТелеСистемы»

От «песочницы» к реальным проектам: инструменты реального времени анализа трафика в телекоммуникационных компаниях

В общем случае задача сбора и сопоставления информации о сессии передачи данных сети обслуживания мобильных абонентов предполагает интеграцию данных, поступающих в режиме реального времени с интенсивностью 50 – 200Мбайт/с от таких источников (DATA-активностей) как: сервер на сигнальных интерфейсах; системы тарификации в режиме реального времени мобильных сервисов передачи данных для всех категорий абонентов; система клиринга, организации и учета обменов данными между операторами связи при нахождение абонентов в международном роуминге; система биллинга; управляющий узел сети мобильной передачи данных. Очевидно, что форматы данных, генерируемых каждой такой системой, не сопоставимы, имеют разную структуру, а идентификаторы сессии и варианты агрегации транзакций отличаются от источника к источнику. В докладе анализируется опыт использования инструментов Cloudera и Impala для трассировки трафика разнообразных данных, позволяющей сформировать требования к системе управления данными и определить расхождения в данных между источниками для их последующей корректной настройки.





СЕРГЕЙ ЖЕМЖИЦКИЙ,
CleverDATA

Spark: сегментация пользовательской аудитории в онлайн-рекламе

Одной из задач проведения рекламной кампании является формирование предложения для посетителя какого-либо веб-ресурса, в соответствии с его текущим потребностями.

Интересы и потребности посетителей с достаточно высокой степенью вероятности удастся определить не всегда, из-за чего может возникнуть ряд вопросов: как найти тех, кто кликнул на баннер, но так и не совершил никакой покупки; как определить тех, кто видел баннер несколько раз, но так на него и не отреагировал и т.п. В докладе будет рассказано про практический опыт решения таких и схожих задач при помощи Apache Spark.





АЛЕКСАНДР ПРОЗОРОВ,

Лаборатория информационных систем реального времени, МФТИ

Персональная медицина — построение массивно-параллельного хранилища данных медицинской диагностики в экосистеме Hadoop Hortonworks и R

Все большее распространение в обществе получают медицинские информационные системы, учитывающие персональные особенности каждого пациента, что предполагает необходимость обработки огромных массивов данных и развертывание масштабируемой инфраструктуры хранения. В докладе на примере проекта системы мониторинга пациентов, развернутой на базе Hadoop Hortonworks (Avro, HDFS, Hive, HBase, Storm) и среды моделирования R (+RStudio), рассмотрен конвейер по сбору данных о пациентах, позволяющих получить их индивидуальные биометрические показатели, их загрузке в массивно-параллельное хранилище и выполнению сложных аналитических операций на больших массивах данных. Кроме этого, будут продемонстрированы средства компактной визуализации больших массивов данных и возможности создания интерфейса для интерактивного взаимодействия врача с набором доступных данных о здоровье пациента.





АЛЕКСАНДР ЕРМАКОВ, EMC

Решение средствами Pivotal HAWQ задач обеспечения оперативно-розыскных мероприятий

Традиционный подход к анализу больших объемов данных основан на реляционных СУБД, жестко привязанных к процессам загрузки и преобразования данных, что не позволяет обрабатывать запросы в реальном времени. Наряду с MPP-платформами, позволяющими ускорить обработку, все более популярны становятся решения базе Hadoop, которые, однако, не позволяли использовать преимущества SQL. Сегодня на рынке появились среды обработки данных в Hadoop, предоставляющие полноценные средства SQL и производительность, типичную для параллельных MPP-платформ. В докладе анализируется опыт использования фреймворка HAWQ, воплотившего лучшие черты Impala, Stinger, BigSQL при решении задач, типичных для системы обеспечения функций оперативно-розыскных мероприятий.





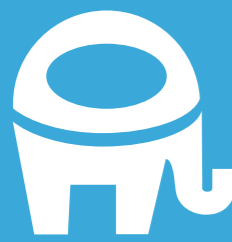
АЛЕКСЕЙ КОСТАРЕВ, «Невод»

Spark и YARN: аналитика реального времени для городского ситуационного центра

Число и разнообразие данных телеметрии, жизнь современного общества уже немыслима без различных средств мониторинга, поток данных от которых постоянно растет. Вместе с тем, кроме выполнения своих первоначальных функций, например фиксации фактов нарушения ПДД транспортными средствами и выписки штрафа, из потока сырых данных можно попутно выделить массу другой полезной информации (вычисление средней скорости потока, обнаружение фактов применения на транспортных средствах несанкционированных устройств слежения и т. п.). В докладе рассматриваются вопросы создания аналитической системы ситуационного центра субъекта федерации на базе технологий Apache Hadoop (YARN) и Spark, работающих как в пакетном (MapReduce, Spark), так и в потоковом (Spark Streaming) режиме с использованием брокера сообщений Apache Kafka. Для анализа потоков данных мониторинга применяются библиотеки статистики и машинного обучения Mllib, а также система распределенной обработки больших графов GraphX. Кроме этого, проект ситуационного центра предусматривает масштабирование вплоть до уровня нескольких субъектов федерации, что становится возможным с привлечением ЦОД провайдеров телекоммуникационных услуг и облачных сред (Azure HDInsight) для решения задач распределенного хранения и обработки данных.

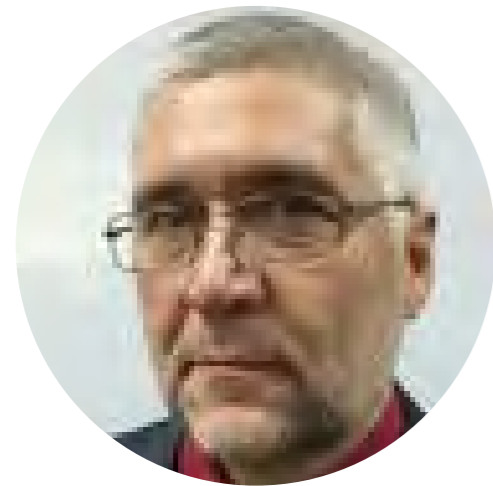


	Открытые системы
2015	
	ЦИКЛ ПРАКТИЧЕСКИХ СЕМИНАРОВ



HADOOP

HADOOP НА ПРАКТИКЕ:
ПРОЕКТЫ И ИНСТРУМЕНТЫ



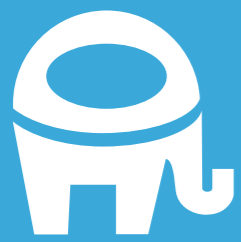
ВАЛЕРИЙ СЕЛЕЗНЕВ,
«Эксклюзивные бизнес-технологии»

Hadoop в задачах онлайн-мониторинга транспортных потоков и телеметрии служб ЖКХ

Сегодня уже никого не удивляют различные сервисы, построенные на базе навигационно-временной информации, широко используемой, например, для мониторинга здоровья пациентов или контроля функционирования оборудования и регистрирующих устройств в инфраструктуре жилищно-коммунального хозяйства страны. В докладе анализируется практика реальных проектов в транспортной отрасли и в ЖКХ, использующих технологии обработки Больших Данных в реальном и отложенном времени.



ДЛЯ ЗАМЕТОК



ЕВГЕНИЙ СТЕПАНОВ,
HP

Расширенная аналитика в Hadoop

В докладе рассмотрены возможности расширенной аналитики для Hadoop на базе инструментов портфеля HP Big Data, включающего платформу поиска и анализа неструктурированных данных IDOL, систему HP Vertica SQL on Hadoop и Distributed R — язык R для использования в распределенной архитектуре.



АНДРЕЙ ОРЛОВ,
IBM CIS

Практика использования инструментов визуализации и анализа в Hadoop от IBM

В рамках доклада будут рассмотрены инструменты из дистрибутива Hadoop от компании IBM: BigSheets, BigSQL, BigR, а также варианты их применения в конкретных задачах телекоммуникационной и финансовой отраслей.



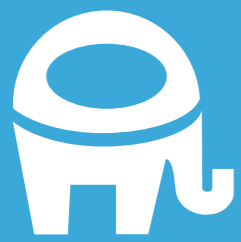


АЛЕКСАНДР БОГДАНОВ,
AT Consulting

Лямбда-архитектура в Hadoop. Анализ потоковых данных в реальном времени

На сегодняшний день отсутствует общий инструмент построения масштабируемой информационной системы, способной справиться как с пакетными данными больших объемов, так и с данными реального времени, которые вынуждены обрабатывать, например, провайдеры сотовой связи. В докладе рассматривается λ -архитектура, предназначенная для уменьшения сложности системы и обеспечения эффективного хранения разнородных данных путем использования разнообразных инструментов в рамках одной системы.





АНДРЕЙ ЕГОРОВ,
Motorola Solutions



АЛЕКСАНДР СМИРНОВ,
Motorola Solutions

Конвейер анализа параметров функционирования телекоммуникационной системы

Эффективность анализа функционирования телекоммуникационного оборудования определяется возможностями автоматизированной системы по сбору и обработке постоянно растущих объемов данных мониторинга. В докладе представлен опыт построения системы конвейера статистического анализа данных от телекоммуникационного оборудования, использующей инструменты оркестровки Spring Framework, загрузки данных в HDFS, обработки и анализа средствами Hadoop, сохранения и индексации результатов в системе полнотекстового поиска Elasticsearch и их визуализации в системе обработки логов Kibana.

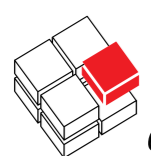


	Открытые системы
2015	
	ЦИКЛ ПРАКТИЧЕСКИХ СЕМИНАРОВ



HADOOP

HADOOP НА ПРАКТИКЕ:
ПРОЕКТЫ И ИНСТРУМЕНТЫ



ОТКРЫТЫЕ СИСТЕМЫ
Open Systems Publications

Издательство «Открытые системы», созданное в 1993 году для комплексной информационной поддержки профессионалов, отвечающих за построение масштабных компьютерных систем, своей основной целью видит предоставление полной и качественной информации для различных категорий читателей. Сегодня «Открытые системы» – ведущее издательство в России, производящее высокопрофессиональные издания для специалистов и руководителей из таких отраслей, как информационные технологии, компьютерные системы, нефтегазовая отрасль, телекоммуникации, сети передачи данных, полиграфия и медицина.

Подробнее: www.osp.ru

Открытые системы
ЖУРНАЛ ДЛЯ АРХИТЕКТОРОВ ИНФОРМАЦИОННЫХ СИСТЕМ

«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», 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 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

Планируйте. Приходите. Участвуйте!




УСЛОВНЫЕ ОБОЗНАЧЕНИЯ

 ДЛЯ ЗАМЕТОК


 ССЫЛКА ДЛЯ ПЕРЕХОДА

13 мая



ЭКОСИСТЕМА
И ТЕХНОЛОГИИ
OPENSTACK

- Операционная система для облаков
- Секреты успеха OpenStack
- OpenStack — инструмент сервис – провайдеров
- Опыт расширения и будущее OpenStack



13 мая



МИРЦОД
[2015]

- Модульные ЦОД 2.0
- Системы управления DCIM. Время пришло
- Катастрофоустойчивые решения
- Программируемые ЦОД – SD-DC
- Гибридные облака



www.osp.ru/dcworld

4 июня



IT Management Forum
XII Russian

- Антикризисные рецепты управления ИТ
- Оценка эффективности процессов
- ITSM на практике
- Agile ITSM
- Универсальное управление сервисами



www.osp.ru/itsm

ВСЕ МЕРОПРИЯТИЯ



По вопросам участия: Ольга Пуркина

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

