



ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ТҰҢҒЫШ ПРЕЗИДЕНТІ - ЕЛБАСЫНЫҢ ҚОРЫ

«ҒЫЛЫМ ЖӘНЕ БІЛІМ – 2017»

студенттер мен жас ғалымдардың
XII Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ

СБОРНИК МАТЕРИАЛОВ

XII Международной научной конференции
студентов и молодых ученых
«НАУКА И ОБРАЗОВАНИЕ – 2017»

PROCEEDINGS

of the XII International Scientific Conference
for students and young scholars
«SCIENCE AND EDUCATION - 2017»



14th April 2017, Astana



**ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ**

**«Ғылым және білім - 2017»
студенттер мен жас ғалымдардың
XII Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ**

**СБОРНИК МАТЕРИАЛОВ
XII Международной научной конференции
студентов и молодых ученых
«Наука и образование - 2017»**

**PROCEEDINGS
of the XII International Scientific Conference
for students and young scholars
«Science and education - 2017»**

2017 жыл 14 сәуір

Астана

УДК 378

ББК 74.58

Ғ 96

Ғ 96

«Ғылым және білім – 2017» студенттер мен жас ғалымдардың XII Халықаралық ғылыми конференциясы = The XII International Scientific Conference for students and young scholars «Science and education - 2017» = XII Международная научная конференция студентов и молодых ученых «Наука и образование - 2017». – Астана: <http://www.eni.kz/ru/nauka/nauka-i-obrazovanie/>, 2017. – 7466 стр. (қазақша, орысша, ағылшынша).

ISBN 978-9965-31-827-6

Жинаққа студенттердің, магистранттардың, докторанттардың және жас ғалымдардың жаратылыстану-техникалық және гуманитарлық ғылымдардың өзекті мәселелері бойынша баяндамалары енгізілген.

The proceedings are the papers of students, undergraduates, doctoral students and young researchers on topical issues of natural and technical sciences and humanities.

В сборник вошли доклады студентов, магистрантов, докторантов и молодых ученых по актуальным вопросам естественно-технических и гуманитарных наук.

УДК 378

ББК 74.58

ISBN 978-9965-31-827-6

©Л.Н. Гумилев атындағы Еуразия
ұлттық университеті, 2017

существенно повышают точность сегментирования.

Список использованных источников

1. Y. LeCun, Y. Bengio, Convolutional networks for images, speech, and time series, the Handbook of Brain Theory and Neural Networks 3361 (1995).
2. Imagenet classification with deep convolutional neural networks. A Krizhevsky, I Sutskever, GE Hinton Advances in neural information processing systems, 1097-1105
3. Cosmin Cernazanu-Glavan, Stefan Holban. Segmentation of bone structure in X-ray images using convolutional neural network. Advances in Electrical and Computer Engineering, vol 1, 2013, pp 87-94.
4. Dan Ciresan, Alessandro Giusti, Luca M. Gambardella and Jürgen Schmidhuber. Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. Advances in Neural Information Processing Systems, vol 25, 2013, pp 2852-2860.
5. Fawcett, Tom (2006). "An Introduction to ROC Analysis". Pattern Recognition Letters 27 (8): 861–874

УДК 004.67

ИНСТРУМЕНТЫ ORACLE BI ДЛЯ АНАЛИЗА БОЛЬШИХ ДАННЫХ

Голенко Екатерина Сергеевна

golenko.katerina@gmail.com

Магистрант специальности Информационные системы – 6M070300 ЕНУ им.

Л.Н.Гумилева, Астана, Казахстан

Научный руководитель – к.ф.-м.н., доц. Сагнаева С.К

До наступления эпохи больших данных большинство коммерческих предприятий не придавало должного значения хранимой в базах данных информации, не имеющей непосредственной ценности в настоящий момент. С приходом эпохи больших данных ситуация кардинально изменилась: инвестиции в сбор и хранение данных с потенциальной ценностью возросли в несколько десятков раз, корпорации начали предпринимать усилия, чтобы сохранить каждый потенциальный бит данных. Стали появляться целые отделы анализа информации, целью которых было отыскание данных, представляющих ценность для организации. Успех подобных кампаний привел к сбору все большего и большего количества информации, результаты исследований показали, что чем больше количество хранимых данных, тем более точными могут быть прогнозы развития предприятия. Это привело к созданию и быстрому развитию технологий для хранения, обработки и анализа больших данных новыми и эффективными способами.

Несмотря на то, что многие организации достигли высокого уровня профессионализма в использовании своих данных путем их анализа, они все еще находятся на ранних стадиях создания аналитической модели, которая может принести реальную ценность для бизнеса с использованием больших данных. Основными препятствиями являются медленные и скрытые процессы для обеспечения прямого и своевременного доступа к корпоративным данным. Однако новые технологии сокращают разрыв между ИТ-специалистами и аналитиками данных, предоставляя расширенную аналитику в самой базе данных, снижая необходимость перемещения больших объемов данных.

В то же время новые типы данных дополняют традиционные источники данных и BI-операции. Например, отслеживается движение посетителей на веб-сайтах, обрабатывается информация из социальных сетей, записей в блогах, индексов поиска, потоков кликов, датчиков оборудования и всех типов мультимедийных файлов, включая аудио, видео и фотографии.

Эти данные могут собираться не только на компьютерах, но и на миллиардах

мобильных телефонов, постоянно расширяющемся массиве сетевых датчиков от автомобилей, коммунальных счетчиков, морских контейнеров, оборудования цехов и многих других источников. Большая часть этих данных очень неоднородна, а также не несет полезной информации.

Некоторые данные лучше размещать в распределенной файловой системе Hadoop (HDFS) или в нереляционных базах данных, которые обычно называются базами данных NoSQL. Во многих случаях это отправная точка для анализа больших данных [1].

На сегодняшний день одним из мощных средств для анализа больших данных является Oracle Business Intelligence Enterprise Edition. Данный программный продукт представляет собой довольно обширный набор инструментов для хранения и обработки больших данных. Существуют следующие традиционные инструменты для анализа данных, представленные корпорацией Oracle: инструмент Oracle Endeca Information Discovery для исследования данных корпоративного класса, BI-инструменты, такие как Oracle Exalytics, и инструменты управления принятием решений, например, Oracle Real Time Decisions.

Для проведения анализа больших данных необходимо соблюдать следующие требования [2]:

- Сведение к минимуму перемещений данных.

Сведение к минимуму перемещений данных заключается в сохранении вычислительных ресурсов. В традиционных сценариях анализа данные передаются на компьютер, обрабатываются, а затем отправляются следующему адресату. Например, производственные данные могут быть извлечены из систем электронного бизнеса, преобразованы в реляционный тип данных и загружены в оперативное хранилище данных, структурированное для отчетности. Но по мере роста объема данных такая архитектура ETL становится все менее эффективной, так как накапливается слишком много данных для перемещения. Имеет смысл хранить и обрабатывать данные в одном и том же месте.

- Использование существующих навыков.

С появлением новых данных и новых источников данных возникает потребность в приобретении новых навыков. Однако, большинство организаций имеют больше людей, которые могут анализировать данные с использованием SQL, чем при использовании MapReduce. По этой причине Oracle BIEE имеет возможность поддерживать оба типа обработки больших данных.

- Обеспечение защиты данных.

Безопасность данных важна для многих корпоративных приложений. Пользователи хранилища данных, как правило привыкли не только к тщательно определенным метрикам, размерам и атрибутам, но также к надежному набору политик администрирования и средств управления безопасностью.

Существует пять основных подходов к анализу больших данных в Oracle BIEE [2]:

1. Инструменты обнаружения (Discovery tools) полезны на протяжении всего жизненного цикла информации для быстрого, интуитивного изучения и анализа информации из любой комбинации структурированных и неструктурированных источников. Эти инструменты позволяют проводить анализ наряду с традиционными системами источников BI. Поскольку нет необходимости в предварительном моделировании, пользователи могут получить новые идеи, прийти к содержательным выводам и быстро принять обоснованные решения.

2. Инструменты BI (BI tools) важны для отчетности, анализа и управления производительностью, в первую очередь, с транзакционными данными из хранилищ данных и производственных информационных систем. Инструменты BI предоставляют комплексные возможности для бизнес-аналитики и управления производительностью, включая корпоративные отчеты, информационные панели, специальный анализ, оценочные карты и анализ «что-если» на интегрированной корпоративной платформе.

3. Аналитика внутри базы данных (In-Database Analytics) включает в себя множество методов поиска шаблонов и отношений в данных. Поскольку эти методы применяются

непосредственно в базе данных, исключается перемещение данных на другие аналитические серверы и с них, что ускоряет время информационного цикла и снижает общую стоимость анализа.

4. Hadoop полезен для предварительной обработки данных для макросов идентичности или для поиска образцов информации, таких как значения вне диапазона. Это позволяет компаниям открывать потенциальную ценность из новых данных с помощью недорогих товарных серверов. Организации в основном используют Hadoop как представление передовых форм аналитики.

5. Управление принятием решений (Decision Management) включает в себя интеллектуальное моделирование, бизнес-правила и самообучение для принятия информированных действий на основе текущего контекста. Этот тип анализа составляет индивидуальные рекомендации по нескольким каналам, максимизируя ценность каждого взаимодействия с клиентом. Оценки Oracle Advanced Analytics могут быть интегрированы для ввода в действие сложных прогнозирующих аналитических моделей и создания процессов принятия решений в реальном времени.

Hadoop является хорошим выбором, когда необходимо фильтровать, сортировать или предварительно обрабатывать большое количество новых данных на месте и перегонять его для создания более «плотных» данных, которые теоретически содержат больше «информации». Предварительная обработка включает в себя фильтрацию новых источников данных, чтобы сделать их пригодными для дополнительного анализа в хранилище данных. Например, промоутер концерта может захотеть проанализировать twitter-каналы, чтобы определить, понравились ли посетителям постановка, сет-лист, костюмы, связанные с новым туром Lady Gaga. Компания может начать со сбора постов, связанных с художником, с использованием хэш-меток, таких как «#Gaga», «#concert», «#Palladium» и т. д. Сентиментальность каждого поста может быть определена путем анализа текста и сравнения его с положительными и отрицательными словами в словаре английского языка. В сочетании с MapReduce Hadoop может обрабатывать огромное количество данных параллельно на нескольких серверах, а затем объединять их в единый набор ответов или интегрировать его с другими типами корпоративных данных. Результирующий набор данных можно импортировать в хранилище данных для интеллектуального анализа данных и принятия решений.

Любая коммерческая организация, которая имеет базу данных клиентов, может продолжить анализ, определив, как положительное и отрицательное отношение влияет на общий объем продаж, запросы на поддержку и другие ключевые показатели. Предварительная обработка данных в социальных сетях с помощью Hadoop часто является первым шагом к прогнозированию поведения клиентов, предвидению возможностей перекрестного сбыта, повышению эффективности ответов на маркетинговые кампании, предотвращению оттока и анализу корзин покупок для выявления ассоциаций, шаблонов и отношений. Hadoop также отлично подходит для фильтрации и предварительной обработки данных в файлах weblog.

Большинство клиентов Oracle хорошо знакомы с SQL как языком для запросов, отчетов и анализа структурированных данных. Это де-факто стандарт для анализа, а также технология, лежащая в основе большинства BI-инструментов. R - популярный язык программирования с открытым исходным кодом для статистического анализа. Аналитики, ученые, исследователи и ученые обычно используют R, что приводит к увеличению числа программистов R.

После загрузки данных в Oracle Database пользователи могут использовать Oracle Advanced Analytics (OAA), чтобы выявить скрытые связи в данных. Oracle Advanced Analytics, вариант Oracle Database Enterprise Edition, предлагает сочетание мощных алгоритмов в базе данных и алгоритмов с открытым исходным кодом R, доступных через языки SQL и R. Он сочетает в себе высокопроизводительные функции интеллектуального анализа данных с открытым исходным языком R, что позволяет выполнять

интеллектуальную аналитику, интеллектуальный анализ данных, интеллектуальный анализ текста, статистический анализ, расширенные численные вычисления и интерактивную графику - все это в базе данных [2].

Oracle Advanced Analytics предоставляет все основные аналитические возможности и языки на мощной архитектуре в базе данных. Эти аналитические возможности включают алгоритмы интеллектуального анализа данных, реализованные в базе данных, собственные функции SQL для основных статистических методов и интеграцию с открытым исходным кодом R для статистического программирования и доступ к более широкому набору статистических методов.

Инструменты интеллектуального анализа данных в ОАА позволяют аналитикам данных работать непосредственно с данными внутри базы данных, графически анализировать данные, строить и оценивать несколько моделей интеллектуального анализа данных, а также развертывать прогнозы и идеи на всех предприятиях. ОАА включает 15 алгоритмов интеллектуального анализа данных для классификации, кластеризации, анализа корзины рынка, обнаружения мошенничества и интеллектуального анализа текста, которые могут применяться для решения широкого круга проблем, связанных с данными. Также включает в себя десяток алгоритмов, которые можно использовать для создания и развертывания интеллектуальных приложений, которые автоматически обрабатывают данные схемы «звезда» для получения результатов и предсказаний в реальном времени [3]. Поскольку данные, модели и результаты остаются в базе данных Oracle, перемещение данных устраняется, информационная латентность сводится к минимуму и сохраняется безопасность. Используя стандартные команды SQL, пользователь может обращаться к высокоэффективным алгоритмам базы данных для получения таблиц, представлений, схем, транзакционных и неструктурированных данных. Любой пользователь, имеющий доступ к данным, хранящимся в базе данных Oracle, может использовать результаты ОАА, прогнозы, рекомендации и открытия, используя стандартные отчеты и инструменты BI.

Oracle Advanced Analytics была разработана, чтобы позволить статистикам использовать R для очень больших наборов данных. Аналитические модели могут быть записаны в R. Связанные таблицы и представления в Oracle Database отображаются как R-объекты. Таким образом, нет необходимости писать SQL-операторы. Аналитики могут писать код R для манипулирования данными в базе данных. Запустив R-программы прямо в базе данных, нет необходимости перемещать данные. Эта интегрированная архитектура обеспечивает исключительную безопасность и производительность, поскольку пользователь может применять массивные масштабируемые аппаратные ресурсы для решения сложных проблем. ОАА поддерживает также существующие R-скрипты и сторонние пакеты. Тесная интеграция между R, Oracle Database и Hadoop позволяет аналитикам написать один R-скрипт, который может работать в трех различных средах: ноутбук с открытым исходным кодом R, Hadoop с Oracle Big Data Connectors и Oracle Database. Легко связать результаты анализа с инструментами бизнес-аналитики, такими как Oracle Business Intelligence и Oracle Exalytics.

Oracle имеет три разработанных системы (рисунок 1), которые решают различные проблемы большой данных. Каждая платформа включает в себя все необходимое аппаратное и программное обеспечение, необходимое для экстремальной обработки данных. Все компоненты предварительно интегрированы и готовы к развертыванию и эксплуатации. Oracle проделала большую работу по связыванию этих сконструированных систем, чтобы пользователи могли извлекать выгоду из данных с помощью передовой платформы больших данных со встроенной аналитикой. Это комплексное решение включает в себя несколько систем обработки данных, загрузки, хранения, управления, анализа, интеграции и представления, что позволяет быстро извлекать выгоду из больших данных с помощью интегрированной аналитики.

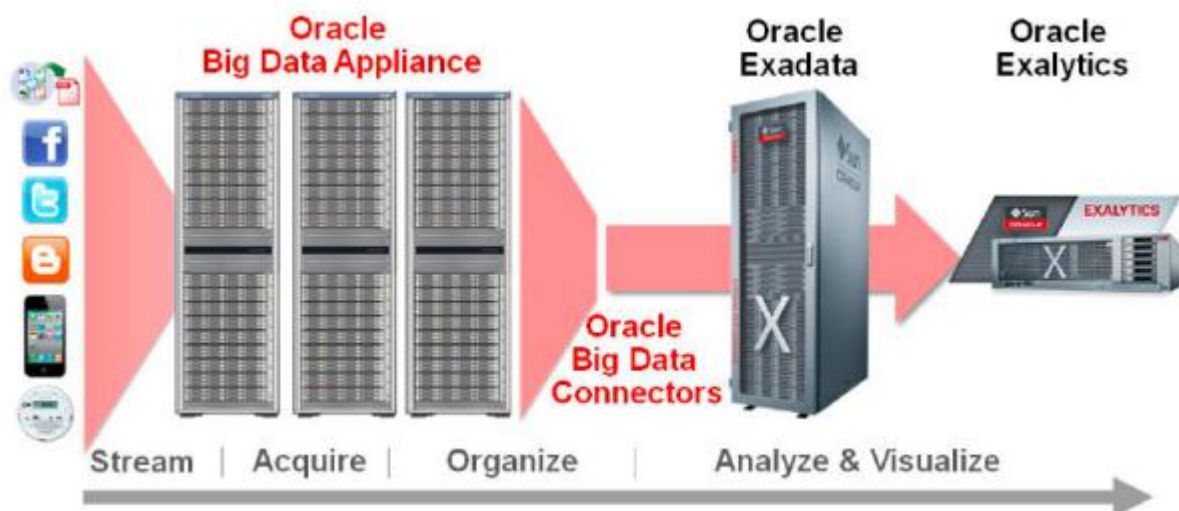


Рисунок 1 Платформа Oracle Big Data (примечание: составлен на основе источника [2])

Oracle Big Data Appliance включает в себя сочетание программного обеспечения с открытым исходным кодом и специализированного программного обеспечения, разработанного Oracle для удовлетворения потребностей в больших данных. Находясь в начале жизненного цикла больших данных, он предназначен для эффективного хранения и организации больших данных и является самой экономичной платформой для запуска Hadoop [4].

Oracle Exadata Database Machine обеспечивает высокую производительность и масштабируемость для всех типов приложений баз данных. Это самая быстрая платформа для работы с базой данных Oracle и связанной с ней аналитикой [5].

Oracle Exalytics - это инженерная система, которая включает платформу бизнес-аналитики, программное обеспечение для анализа в оперативной памяти и аппаратное обеспечение, оптимизированное для крупномасштабной аналитики. Благодаря инструментам для расширенной визуализации и исследования данных, он позволяет пользователям получать эффективную информацию из больших объемов данных. Когда Oracle Exalytics используется вместе с Oracle Advanced Analytics, у пользователя есть платформа, которая обеспечивает понимание ключевых бизнес-задач, таких как прогнозирование оттока, рекомендации по продуктам, анализ настроений и оповещение о мошенничестве [6].

Организации в каждой отрасли пытаются использовать массовый приток больших данных, а также разработать аналитические платформы, которые могут синтезировать традиционные структурированные данные с полуструктурированными и неструктурированными источниками информации. При правильном хранении и анализе большие данные могут предоставить уникальную информацию о тенденциях рынка, сбоях оборудования, структуре покупок, циклах обслуживания и многих других бизнес-задачах, снижении затрат и обеспечении более целенаправленных бизнес-решений.

Для получения ценности из больших данных необходим комплексный набор решений для сбора, обработки и анализа данных, получения данных и выявления новых возможностей принятия повторяющихся решений и масштабирования соответствующих информационных систем.

Oracle BI является вполне достойным решением для выявления скрытых отношений в источниках больших данных. Независимо от того, нужно ли предсказывать поведение клиентов, предвидеть возможности перекрестного сбыта, повысить уровень ответов на маркетинговые кампании, предотвратить отток, проанализировать «рыночные корзины», чтобы обнаружить ассоциации, шаблоны и отношения, использовать влиятельные лица в социальных сетях, уменьшить мошенничество или предвидеть будущее Oracle Advanced Analytics может помочь. При использовании в сочетании с инструментами с открытым

исходным кодом, такими как Hadoop и MapReduce, это аналитическое решение предоставляет необходимые инструменты для приобретения, организации, анализа и максимизации ценности больших данных внутри предприятия, при выполнении фундаментальных требований минимизации перемещения данных, использования существующих навыков сотрудников и обеспечения высокого уровня безопасности.

Список использованных источников:

1. Jared Dean Big data, data mining, and machine learning: value creation for business leaders and practitioners. - New Jersey, John Wiley & Sons Inc., 288 с.
2. Mark Rittman Big Data Analytics&Advanced Analytics in Oracle Database – 2013 <http://www.oracle.com/technetwork/database/options/advanced-analytics/bigdataanalyticswpoaa-1930891.pdf>
3. Foster Provost, Tom Fawcett Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. – CA, O'Reilly Media Inc., 2013, 369 с.
4. Бабурин В. А., Яненко М. Е. Технологии Big Data в сервисе: новые рынки, возможности и проблемы // Техничко-технологические проблемы сервиса (СПбГЭУ), № 1 (27), 2014, с.100-105
5. <http://www.oracle.com/technetwork/database/exadata/overview/index.html>
6. <http://www.oracle.com/ru/engineered-systems/exalytics/overview/index.html>

УДК 004.774.6

КЛАССИФИКАЦИЯ СТИЛЕЙ В WEB-ДИЗАЙНЕ

Едигеева Мариям Сериковна

ms_edigeeva@mail.ru

магистрант кафедры Информатики и информационной безопасности

ЕНУ им.Л.Н.Гумилева, Астана, Казахстан

Научный руководитель – А. Кинтонова

Интернет технологии стремительно развиваются в современном обществе и охватывают все сферы деятельности человека. При такой сильной интернете-зависимости, у каждой крупной компании, организации, ИП возникает потребность в создании собственного сайта.

При разработке любого Интернет-ресурса одним из важнейших моментов является правильный выбор дизайна. Каждый проект имеет свой стиль, направление и имидж. Поэтому необходимо использовать дизайн, который передавал бы фирменный стиль и уникальность компании, настраивая посетителя на более полное и качественное понимание информации, размещенной на сайте.[1]

Web-дизайн – отрасль web-разработки и разновидность дизайна, в задачи которой входит проектирование пользовательских web-интерфейсов для сайтов или web-приложений[2]. Стили web-дизайна – это основные концепции создания сайтов, по которым сайты можно классифицировать. (Схема 1)