

Big Data: технологии, решения, услуги



Сегодня феномену Больших данных придаётся особое значение. Речь идет не просто о том, что данных накопилось слишком много и их стало сложно хранить и обрабатывать, а о том, что они стали другими. Десятилетиями компании принимали управленческие решения, обладая лишь той информацией, которая хранилась в реляционных базах данных. Традиционный бизнес-анализ имел дело с выверенными, очищенными и полностью достоверными данными. Их могло быть очень много, но все они содержались во внутренних источниках – хранилищах данных и корпоративных информационных системах. Сейчас к ним добавился огромный поток неструктурированных или слабо-структурированных данных – блоги, социальные сети, Интернет-ресурсы, показатели различных датчиков, изображения и фотографии. Содержащаяся в них информация потенциально несет в себе огромные возможности для бизнеса, если суметь извлечь из неё пользу.

Какие данные считаются большими

По подсчетам аналитиков, объем данных возрастает каждый год примерно на 40%. Это означает, что в период с 2009 по 2020 год их станет больше в 44 раза. Однако, дело не только в количестве. Большие данные принято характеризовать следующими важными свойствами:

- **Объем.** Никогда ранее человечество не сталкивалось с необходимостью обработки и хранения информации, измеряемой тысячами терабайт и петабайт
- **Скорость.** Трафик генерируется все быстрее, на обработку данных и получение результатов анализа требуется всё меньше времени
- **Многообразие.** Постоянно появляются новые виды информации, большая часть которой неструктурированные данные — тексты, изображения, видео, потоковые данные, графы связей и другие
- **Ценность.** Значимость тех или иных данных сильно различается в зависимости от целей исследования и ряда других факторов, поэтому здесь главный вызов состоит в том, чтобы определить, какие данные полезны, суметь правильно преобразовать их и извлечь для дальнейшего анализа

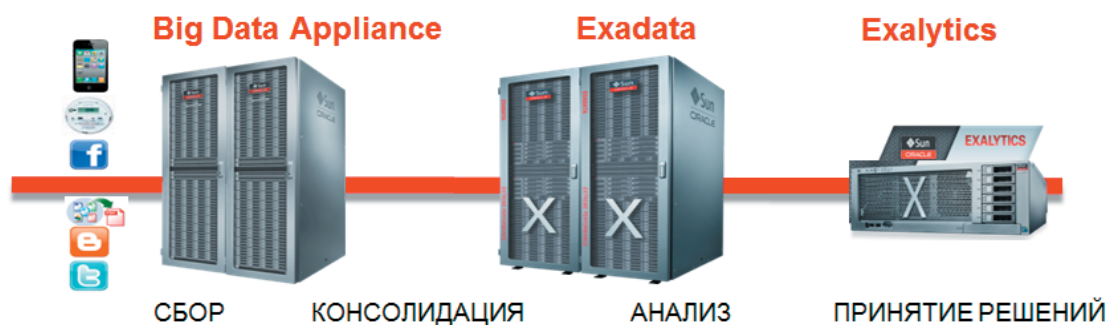
Примеры больших данных:

- данные Интернет-сайтов - тексты, фотографии, видео
- информация, содержащаяся в социальных медиа, блогах, форумах
- архивы текстовых документов
- данные медицинских обследований
- автоматически генерируемые данные, непрерывно поступающие с измерительных устройств - от приборов радиочастотной идентификации, аудио- видео- регистрации, метеорологического оборудования
- информация, накапливаемая в результате логирования систем
- данные о местонахождении абонентов сетей сотовой связи

Технологии и инструменты

Для того чтобы полностью использовать возможности, открываемые благодаря вовлечению огромных массивов разнообразных данных, предприятия должны задуматься о модернизации своей ИТ-инфраструктуры. Необходимо, чтобы она позволяла не только быстро обрабатывать растущие потоки внешних данных разных типов, но и интегрировать их с собственными корпоративными данными для последующего всестороннего анализа. В отличие от средств бизнес-анализа, где самое главное – вычисления, здесь не менее важны эффективность поисковых операций, возможность интуитивного исследования данных. Для компаний открывается широчайшее поле деятельности для выработки бизнес-стратегий и поиска нестандартных подходов к решению сложных задач. Требования к инфраструктуре, позволяющей обрабатывать Большие данные, охватывают области, связанные со сбором, консолидацией и анализом данных. Основными технологиями для работы с большими данными являются:

- **Hadoop** и **MapReduce** для долговременного распределенного хранения и обработки данных
- **NoSQL** базы данных для эффективного хранения и обработки огромных объемов нестандартных данных
- Средства статистического анализа и **data mining**, поддерживающие углубленный анализ данных большого объема и разнообразия
- Инструменты класса **Data Discovery** для интуитивного анализа с расширенными поисковыми возможностями



Oracle для Больших данных

Корпорация Oracle предлагает комплексное интегрированное решение, отвечающее всей совокупности требований к промышленной инфраструктуре уровня Big Data. В основе стратегии Oracle лежит идея эволюционного развития инфраструктуры предприятия, когда постепенно встраиваются новые функциональные возможности – в частности, для обработки Больших данных. Таким образом, достигается необходимый рост производительности систем, их надежность и гибкость, что и позволяет использовать в своей работе Большие данные, извлекая из них реальную пользу для бизнеса.

Уникальная экспертиза Oracle состоит в соединении в одном продукте всех требований, как к программному, так и аппаратному обеспечению. Разработанные корпорацией интегрированные программно-аппаратные комплексы - это отдельный класс решений, где используются инновационные технологии, ориентированные на экстремальную производительность при работе с огромными объемами данных. Базовой платформой для работы с Большими данными является программно-аппаратный комплекс Oracle Big Data Appliance, который объединяет в единое целое специально оптимизированное оборудование и сверхпроизводительное программное обеспечение для быстрого и легкого развертывания мощнейшего решения для сбора, консолидации и обработки больших данных.

Big Data Appliance содержит следующий набор продуктов, готовых для решения задач сбора и анализа больших данных:

- Apache Hadoop (Cloudera's Distribution)
- Cloudera Manager для администрирования всех компонент Cloudera
- Среда статистических исследований R (версия свободного распространения)
- Oracle NoSQL Database Community Edition
- Набор коннекторов и адаптеров для интеграции с другими продуктами Oracle
- Инструмент исследования больших данных Oracle Big Data Discovery

Полное решение, объединяющее Big Data Appliance, Oracle Exadata, и Oracle Exalytics обеспечивает сверхвысокий уровень производительности, безопасности и надежности при обработке Больших данных и позволяет эффективно решать все задачи по сбору, консолидации, хранению и анализу Больших данных.

Услуги ФОРС по направлению Big Data

Компания ФОРС активно участвует в формировании рынка Больших данных – развивает экспертизу, подготавливает специалистов, выполняет проекты, знакомит бизнес с теми преимуществами, которые может дать применение новых технологий на практике. Компания предлагает готовые технологии и решения, современные методики внедрения. При этом основные усилия сосредоточены на продвижении технологий Big Data для решения конкретных задач в различных индустриях, госсекторе, банках и ЖКХ.

В ФОРС действует собственный инженеринговый демо-центр Fors Solution Center с представленным там практически полным стеком программно-аппаратных комплексов корпорации Oracle, включая Exalytics, Exalogic, Exadata и Database Appliance. Здесь партнеры и заказчики могут получить полный комплекс консалтинговых услуг.