



Основные моменты

- Расширенная аналитика для структурированных, частично структурированных и неструктурированных данных
 - Визуализация на профессиональном уровне, инструментарий для разработки и администрирования, способствующий повышению продуктивности
 - Ускорители приложений, помогающие ускорить процесс внедрения и сократить период окупаемости
 - Интеграция с популярными предложениями IBM, а также с решениями других поставщиков
 - Готовая для работы на предприятии аналитическая платформа на основе Apache Hadoop
 - Поддержка множества дистрибутивов Hadoop для обеспечения максимальной гибкости
-

IBM InfoSphere BigInsights Enterprise Edition

*Эффективное управление большими данными
и их анализ с целью извлечения ценных идей*

Укрощение больших данных

Многие компании наблюдают стремительный рост разнообразия, скорости передачи и объемов информации, генерируемой предприятиями, и пытаются решить задачи управления большими данными: огромным количеством традиционных структурированных данных разного рода, наряду с частично структурированными и неструктурированными данными. Большие данные предполагают огромный потенциал для глубокого анализа, позволяющего быстро, четко и тщательно принимать решения. Но для того чтобы воспользоваться всеми преимуществами этого ресурса, организациям требуются инструменты для управления данными и анализа, эффективно работающие на совершенно новом, по сравнению с прошлым, уровне.

IBM® InfoSphere® BigInsights™ Enterprise Edition позволяет организациям преобразовывать большие объемы комплексных данных в аналитические выводы по различным направлениям решения задач в бизнесе. На высшем уровне эти задачи можно разделить на три основные категории: операционная эффективность, расширенная аналитика, поиск и извлечение данных.

Операционная эффективность

Для того чтобы более эффективно управлять производительностью и экономическим эффектом растущих объемов данных, необходимо совместное использование архитектур, включающих в себя различные подходы. Например, большие объемы «холодных» данных из хранилища информации можно заархивировать в аналитической среде, а не в пассивном хранилище. Можно также интегрировать данные посредством общего репозитория на пути к хранилищу.

InfoSphere BigInsights способствует повышению операционной эффективности путем расширения (а не замены) среды хранилища данных. Его можно использовать в качестве архива с поддержкой поиска по запросам, что позволяет организациям хранить и анализировать большие объемы различным образом структурированных данных, не перегружая хранилище данных. Выступая в качестве узла предварительной обработки или «зоны посадки» данных, InfoSphere BigInsights помогает организациям изучать свои данные, определять ценные активы и извлекать данные экономически эффективным способом. Кроме того, InfoSphere BigInsights поддерживает специализированный анализ больших объемов данных, что позволяет организациям очень быстро изучать, извлекать и анализировать данные.



Расширенная аналитика

Некоторые организации стремятся не только к росту показателей операционной эффективности, но и к освоению новых, расширенных функций аналитики, хотя для этого им недостает нужных инструментов. В рамках InfoSphere BigInsights аналитика не является отдельным действием, выполняемым после этапа сохранения данных; напротив, InfoSphere BigInsights работает совместно с InfoSphere Streams, поддерживая аналитику в реальном времени, взаимосвязанную с хронологическими моделями, полученными на основе анализа неподвижных данных.

InfoSphere BigInsights включает в себя несколько предварительно скомпонованных аналитических модулей и ускорителей, с помощью которых можно распознать смысл текста в неструктурированных документах, выполнить смысловой анализ социальных данных или проанализировать данные, полученные из множества разнообразных источников.

Поиск и извлечение данных

Разнообразные источники и типы больших данных, в сочетании со стремительным ростом объемов данных, могут волной захлестнуть организацию, что окончательно затруднит процесс извлечения крупин драгоценной информации. InfoSphere BigInsights помогает создать среду, идеально подходящую для изучения и извлечения взаимосвязей и взаимозависимостей между данными в целях получения новых идей и улучшенных результатов работы бизнеса. Специалисты по работе с данными могут анализировать необработанные данные, полученные из различных источников, наряду с выборочными данными, взятыми из хранилища предприятия, в среде, напоминающей песочницу. Следовательно, они могут перемещать все только что извлеченные ценные данные в хранилище данных предприятия и сочетать их с другими надежными данными, чтобы способствовать оптимизации эксплуатационных и стратегических знаний и процесса принятия решений.

Подведем итог: предприятия наконец-то получают возможность «объять» огромные массивы нетронутых данных и извлечь из них ценный смысл и идеи с помощью эффективного, оптимизированного и масштабируемого подхода.

Внедрение Hadoop на предприятии

InfoSphere BigInsights объединяет Apache Hadoop с инновациями IBM в целях реализации масштабной обработки и анализа больших объемов данных, обладая внутренними характеристиками гибкости и отказоустойчивости. Компания IBM создала функциональные возможности на основе Hadoop, добавив функции упрощенного администрирования и управления, расширенные инструменты для разработчиков и мощные функции аналитики, при этом устранив сложности, связанные с началом работы с Hadoop. Одна из наиболее

сложных проблем, связанная с созданием приложений с помощью дистрибутивов Hadoop с открытым кодом или предоставляемых сторонними поставщиками, заключается в необходимости обладания специальными навыками на высоком уровне; с помощью InfoSphere BigInsights разработчики и другие пользователи могут создавать приложения и выполнять анализ без дополнительных усилий, опираясь только на имеющиеся у них знания.

Администраторы начинают работу с инструмента установки на основе графического пользовательского интерфейса и управляемого процесса установки, позволяющего указать, какие необязательные компоненты следует установить и как настроить платформу. Отчет о выполнении установки выдается в реальном времени, а встроенная функция контроля работоспособности предназначена для автоматической проверки успеха выполнения установки. Эти расширенные функции установки способствуют сокращению затрат времени на установку и настройку, что позволяет освободить администраторов для работы над другими важными проектами.

После установки решения надежные функции управления заданиями позволяют организации четко контролировать выполнение заданий InfoSphere BigInsights. Технические специалисты могут руководить созданием заданий, их запуском и отменой; также они получают информацию об изменении рабочей нагрузки с помощью интегрированных индикаторов состояния выполнения заданий, протоколов и счетчиков, передающих сведения о конфигурации, задачах, попытках и других важных элементах выполнения. Кроме того, InfoSphere BigInsights предоставляет функции администрирования, включая Hadoop Distributed File System (HDFS) и администрирование MapReduce, управление кластерами и серверами, представления на основе ролей и возможность просмотра содержимого файлов HDFS.

Наращивание возможностей Hadoop

InfoSphere BigInsights добавляет к открытому коду Hadoop функциональность и возможности интеграции, необходимые для удовлетворения важнейших потребностей бизнеса. Организации могут запускать крупномасштабные, распределенные аналитические задания в кластерах на базе экономически-эффективного серверного оборудования. Такая инфраструктура позволяет с максимальной пользой применять структуру Hadoop MapReduce для работы с очень большими наборами данных благодаря распределению данных среди нескольких узлов и координации обработки информации в массивно-параллельной среде. После сохранения необработанных данных в распределенном кластере, запросы и анализ данных можно выполнять более эффективно, с динамичной интерпретацией формата данных во время считывания.

Повышение гибкости, развитие возможностей потребления и управления

В версию 2.0 продукта InfoSphere BigInsights включены расширенные инструменты визуализации, мониторинга и разработки, позволяющие пользователям, обладающим различными ролями в организации, совместно работать над раскрытием ценности данных. В результате необходимые данные будут скомпонованы для анализа в обход «бункеров» данных, существующих в организации.

Специалисты по бизнес-аналитике

InfoSphere BigInsights V2.0 предлагает централизованную информационную панель на основе нового механизма создания диаграмм. С помощью этой панели специалисты по бизнес-аналитике могут глубже проникать в смысл полученных ими данных, просматривать результаты работы аналитических приложений и отслеживать параметры. Связь между приложениями позволяет бизнес-аналитикам, не обладающим специальными навыками программирования, создавать новые приложения на основе уже существующих и с помощью IBM BigSheets — инструмента визуализации по типу электронных таблиц. Бизнес-аналитики также могут воспользоваться связью между приложениями для вызова аналитических приложений из веб-консоли (см. рисунок 1).

Расширение возможностей использования позволяет бизнес-аналитикам просматривать потоки данных BigSheets между наборами данных и внутри таких наборов, что обеспечивает ускоренную навигацию и возможность создания связей между анализом и диаграммами.

Специалисты по работе с данными

Расширение функций анализа текста обеспечивает дополнительную поддержку модульных приемников. Благодаря функции многократного использования общих приемников, словарей, правил и таблиц, эти функциональные возможности способствуют расширению потенциала использования и повышению продуктивности разработки приложений для анализа текста. Специалисты по работе с данными также могут управлять жизненным циклом текстовой аналитики благодаря новым возможностям, обеспечивающим: создание выборок и подмножеств данных; разработку, тестирование и публикацию приложений для текстовой аналитики из инструментов InfoSphere BigInsights для Eclipse; а также разработку, выполнение и мониторинг приложений для анализа текста с помощью веб-консоли.

Кроме того, интеграция с языком программирования R с открытым исходным кодом позволяет специалистам по работе с данными выполнять специальные задания R непосредственно из веб-консоли InfoSphere BigInsights.

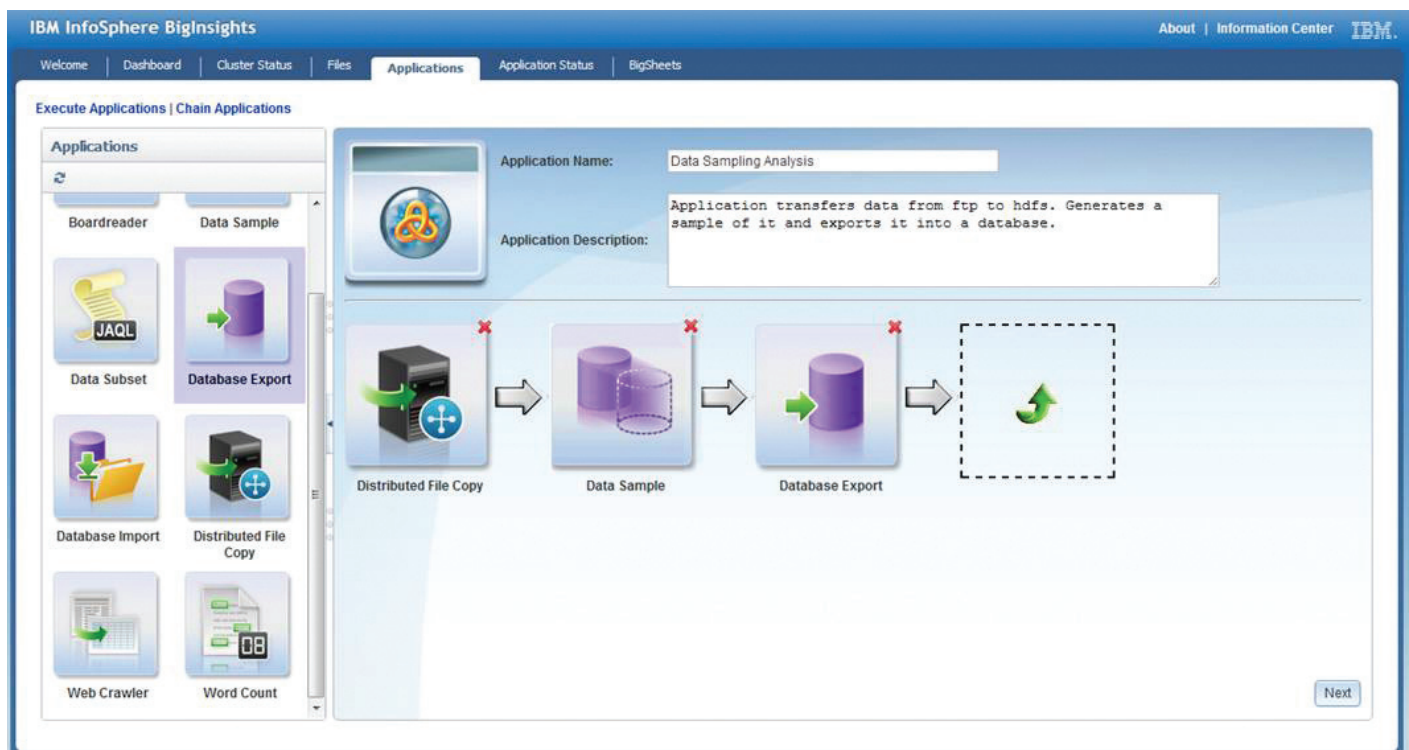


Рисунок 1. Связь между приложениями для бизнес-аналитиков.

Администраторы

Благодаря новым возможностям мониторинга, версия 2.0 продукта InfoSphere BigInsights поддерживает гибкие функции администрирования. Централизованное представление сводной панели позволяет администраторам визуализировать ключевые индикаторы производительности, включая показатели использования центрального процессора, диска, памяти

и сети в кластере, а также показатели служб данных Apache, таких как HDFS, HBase, Zookeeper и Flume, и служб приложений Apache, включая MapReduce, Hive и Oozie (см. рисунок 2). Также администраторы получают расширенные сведения о состоянии и могут контролировать основные функции приложений и кластеров.

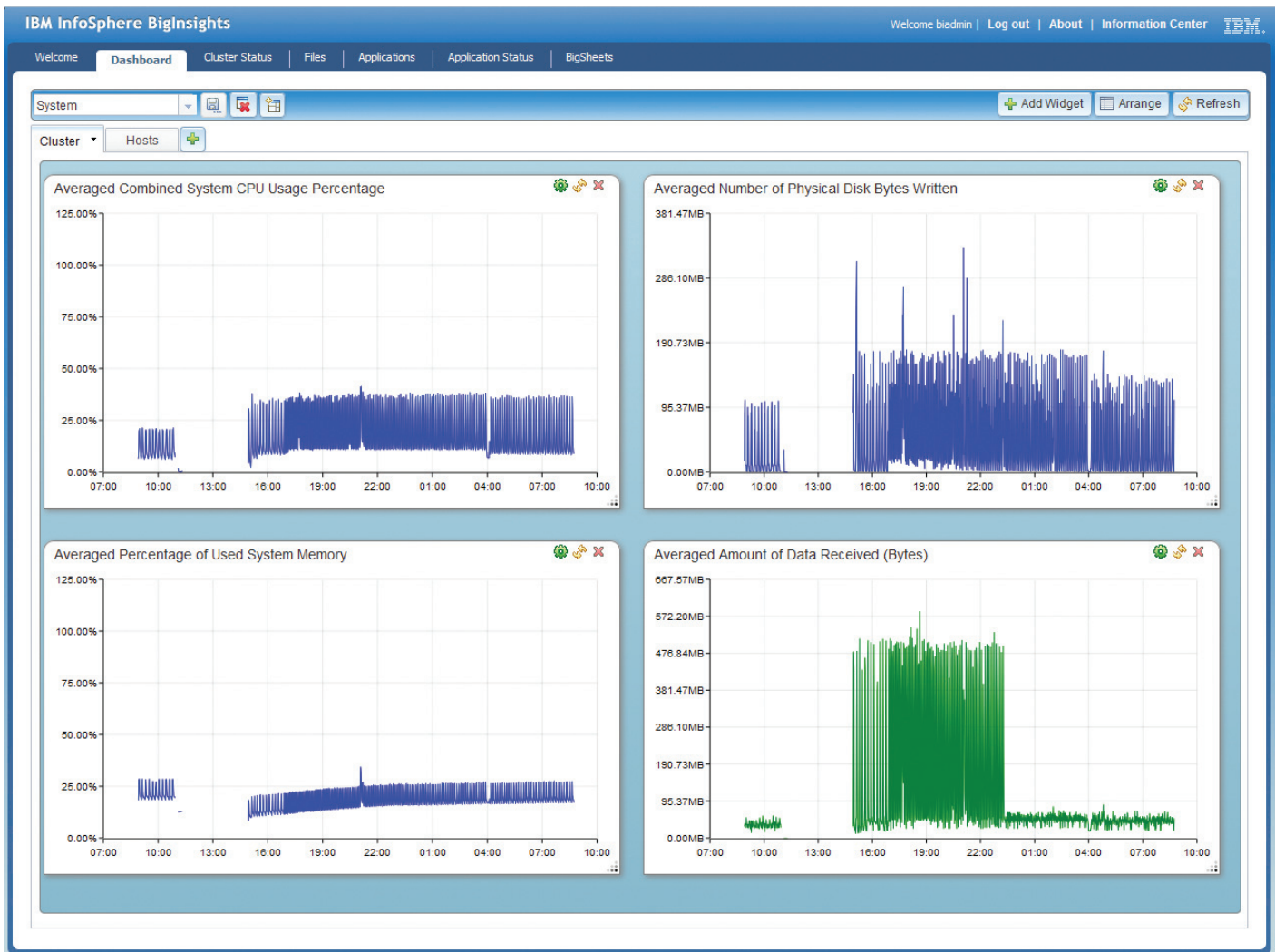


Рисунок 2. Расширенные возможности мониторинга для администраторов.

Разработчики

Разработчики обладают доступом к инструментарию разработки приложений на профессиональном уровне. Этот инструментарий позволяет создавать приложения, не обладая навыками эксперта Nadoor. Благодаря унифицированному инструментарию, разработчики создают выборки данных и определяют, тестируют и развертывают аналитические приложения из инструментов Eclipse, а также администрируют, запускают и отслеживают развернутые приложения из веб-консоли. В состав инструментария входят следующие компоненты:

- Новые и расширенные редакторы, например, редактор рабочих потоков, существенно упрощающий создание сложных рабочих потоков Oozie с удобным интерфейсом; редактор Pig с функцией помощника по содержанию и выделения синтаксиса; улучшенный редактор Jaql с расширенной поддержкой синтаксиса Jaql, расширенным помощником по содержанию и улучшенной функцией обратной связи выполнения.
- Поддержка новых типов приложений с возможностью разработки макросов и программ чтения BigSheets.
- Расширенная разработка текстовой аналитики, включая поддержку модульных наборов правил.
- Расширенный ассортимент артефактов для разработки на этапе развертывания, включая артефакты для анализа текста; сценарии для Jaql, Hive SQL и Pig; макросы и программы чтения BigSheets.

Глубокое погружение в большие данные с помощью функций аналитики и ускорителей приложений

InfoSphere BigInsights включает в себя расширенный набор аналитических инструментов и функций, предоставляемых без дополнительной оплаты. Готовый установочный комплект позволяет организациям немедленно приступить к выявлению закономерностей в данных и создавать мощные, нестандартные аналитические приложения, выдающие результаты и достоверные значения в соответствии с определенными потребностями конкретного бизнеса.

Расширенный анализ текстов

InfoSphere BigInsights содержит в себе многофункциональный механизм для анализа текстов, разработанный компанией IBM. Используя универсальную библиотеку правил или собственные пользовательские правила, разработчики могут быстро запрашивать и идентифицировать интересующие их элементы в документах и сообщениях, включая имена, адреса электронной почты, почтовые адреса, номера телефонов, URL-адреса, названия совместных предприятий, партнерств и т.д.

Ускоритель аналитики социальных данных

Новый компонент InfoSphere BigInsights V2.0, ускоритель аналитики социальных данных, может использоваться для эффективного анализа информации о клиентах, извлекаемой из различных типов данных социальных средств связи. Такой анализ поддерживает глубокое проникновение в суть данных для эффективного проведения целевых маркетинговых кампаний и своевременного принятия маркетинговых решений по продуктам, что помогает организациям выполнять анализ действий конкурентов, удерживать клиентов на постоянной основе и реализовывать программы привлечения новых клиентов.

Ускоритель аналитики технических данных

Еще один новый компонент InfoSphere BigInsights V2.0 — ускоритель аналитики технических данных. Он позволяет получать, анализировать и извлекать разнообразные технические данные из таких источников, как файлы протоколов, интеллектуальные устройства и телеметрические датчики, и обеспечивает обработку этих данных за считанные минуты, а не дни и недели. Благодаря ускорителю аналитики технических данных, организации выполняют анализ операций, взаимодействия с клиентами, транзакций и поведения. Полученная в результате информация может быть использована для упреждающей оптимизации операционной эффективности; выявления и устранения неполадок и изучения инцидентов, связанных с безопасностью; а также для мониторинга сквозной инфраструктуры во избежание простоев и снижения качества обслуживания.

Дополнительная защита анализа больших данных

Строгие требования к безопасности, существующие на предприятии, должны распространяться и на большие данные, точно так же, как они применяются и для всех остальных информационных ресурсов предприятия. InfoSphere BigInsights предоставляет несколько опций повышенной сложности, позволяющих гарантировать безопасность данных и конфиденциальность.

Идентификация

Администраторы, на свое усмотрение, могут выбрать для веб-консоли InfoSphere BigInsights простой файл, Lightweight Directory Access Protocol (LDAP) или отказ от идентификации. Если используется идентификация LDAP, программа установки InfoSphere BigInsights запрашивает данные для идентификации из хранилища идентификационных данных LDAP. Затем администраторы могут предоставить доступ к консоли InfoSphere BigInsights в соответствии с ролями, что упрощает предоставление прав доступа группам пользователей.

Роли

В InfoSphere BigInsights предусмотрено четыре уровня ролей пользователей: системный администратор, администратор данных, администратор приложений и пользователь без прав администрирования. Доступ к данным и функциям зависит от роли, присвоенной пользователю.

Аудит

Задания MapReduce можно запускать под выделенными ИД учетных записей, что способствует укреплению функций защиты, контроля над доступом и аудита. Кроме того, новое хранилище ключей укрепляет безопасность паролей. А интеграция InfoSphere BigInsights с программным обеспечением IBM InfoSphere Guardium® для защиты данных позволяет организациям управлять безопасностью данных на основе Hadoop и традиционных структурированных данных.

Максимальное повышение производительности, оптимизация обработки заданий

InfoSphere BigInsights предоставляет несколько функций, способствующих повышению производительности, а также оптимизации возможностей адаптации и совместимости со средой предприятия.

InfoSphere BigInsights Scheduler для адаптируемого распределения рабочих потоков

Не все рабочие нагрузки обладают одинаковым приоритетом. Компонент InfoSphere BigInsights Scheduler предоставляет схему адаптируемого распределения рабочих потоков для заданий MapReduce, что оптимизирует обработку на основе выбранной пользователем стратегии. Этот планировщик является расширением средства Hadoop Fair Scheduler, позволяющего гарантировать, что, с течением времени, всем заданиям предоставляется пропорциональная доля ресурсов кластера.

BigIndex для крупномасштабного индексирования

BigIndex упрощает индексирование на базе Hadoop благодаря включению этой функции как встроенной в InfoSphere BigInsights. BigIndex реализует функции полнотекстового поиска в больших данных с минимальным временем ожидания. Модуль BigIndex обеспечивает компоновку, сканирование и запрос индексов в рамках потока операций. Кроме того, BigIndex поддерживает дополнительные сложные функции, такие как распределенное индексирование и многоаспектный поиск, что, в свою очередь, обеспечивает высокую степень гибкости при разработке нестандартных приложений и выборе технологий поиска.

Технология Adaptive MapReduce для ускорения выполнения заданий

Задания, запускаемые в BigInsights, в итоге нередко делятся на несколько небольших заданий, потребляющих непропорционально большую часть системных ресурсов. Для того чтобы решить эту проблему, компанией IBM была изобретена технология Adaptive MapReduce, предназначенная для ускорения обработки небольших заданий путем изменения способа обработки задач MapReduce, но без изменения способа создания заданий. Технология Adaptive MapReduce является прозрачной для операций MapReduce и операций API Hadoop.

Jaql для языка запросов больших данных

Мощный высокоуровневый язык декларативных запросов, разработанный компанией IBM и переданный в сообщество ПО с открытым исходным кодом, Jaql предоставляет возможность обработки как структурированных, так и неструктурированных данных. Его интерфейс похож на интерфейс SQL, что облегчает задачу для разработчиков, уже знакомых с языками SQL, и помогает упростить интеграцию с реляционными базами данных.

Изучение и визуализация данных

InfoSphere BigInsights включает в себя расширенный набор аналитических инструментов и функций, позволяющих организациям немедленно приступить к выявлению закономерностей в данных и создавать мощные, нестандартные аналитические приложения, выдающие результаты и достоверные значения в соответствии с определенными потребностями конкретного бизнеса.

IBM BigSheets

BigSheets — это революционный прорыв в мире инструментов для анализа данных. Он работает на основе браузера, по стилю напоминает электронные таблицы и позволяет специалистам по работе с данными и бизнес-пользователям исследовать данные, хранящиеся в приложениях InfoSphere BigInsights, и создавать аналитические запросы без написания специального кода. Встроенный аналитический макрос отвечает всем требованиям работы с данными и способствует максимальному повышению доступности данных.

В версии 2.0 ПО InfoSphere BigInsights продукт BigSheets был модернизирован, и в него были включены функции настройки диаграмм и просмотра, значительно упрощающие, по сравнению с прошлыми версиями, процессы исследования, обработки и анализа данных. Новое приложение для HBase позволяет бизнес-пользователям обращаться к HBase непосредственно из веб-консоли InfoSphere BigInsights и просматривать результаты в BigSheets.

Программное обеспечение IBM

Спецификация

С помощью BigSheets бизнес-пользователи могут выполнять следующие задачи:

- Интегрировать большие объемы неструктурированных данных из веб-хранилищ.
- Получать неструктурированные данные разных типов, полученных с определенных пользователями URL-адресов
- Извлекать и обогащать данные, используя функции анализа текста.
- Исследовать и визуализировать данные в специфических, определенных пользователями контекстах.

Интегрировать большие данные в существующую информационную инфраструктуру

Программное обеспечение InfoSphere BigInsights создано на основе Hadoop с открытым исходным кодом и расширяет его функциональные возможности, необходимые для того, чтобы превратить данное ПО в платформу уровня предприятия. Открытая гибкая архитектура этой платформы дает организациям возможность выбора между поставляемыми IBM дистрибутивами Apache Hadoop и другими дистрибутивами Hadoop, такими как Cloudera, включающее в себя Apache Hadoop (CDH). Существующие установки CDH могут использовать InfoSphere BigInsights для таких функций класса предприятия, как анализ текста, инструментарий разработчиков, удобные для пользователей функции обработки и исследования данных.

Требования к аппаратному обеспечению и поддержка операционных систем

- Серверы Intel x86, 64-разрядные, объем памяти не менее 4 Гб, емкость диска 40 Гб.
- Red Hat Enterprise Linux 5.3 и выше, Red Hat Enterprise Linux 6.2 и выше, SUSE Linux Enterprise 11 и выше, SP2 и Power Linux.

По своей природе, программное обеспечение с открытым исходным кодом не включает гарантии технической поддержки, и может поставляться на таких законных условиях, которые не устраивают большинство организаций. В противоположность этому, ПО InfoSphere BigInsights Enterprise Edition поставляется со стандартными соглашениями IBM на лицензирование и поддержку программного обеспечения. Организации могут выполнять его развертывание в соответствии с уже известными им условиями лицензирования, которые предполагают минимальную степень неопределенности и риска, а также дают уверенность в предоставлении круглосуточной поддержки в течение 7 дней в неделю, услуг обучения и организации профессионального обслуживания по всему миру.

Технологиям для работы с большими данными отводится важная роль в цепочке снабжения предприятия информацией, но только при условии глубокой и прочной интеграции с существующими системами. Компания IBM понимает это, и потому в ПО InfoSphere BigInsights включены высокоскоростные коннекторы для программного обеспечения баз данных IBM DB2®, семейства устройств хранилищ данных IBM PureData™ System, решений IBM Netezza®, IBM InfoSphere Warehouse и IBM Smart Analytics System. Эти высокоскоростные коннекторы способствуют упрощению и ускорению выполнения задач работы с данными. Более того, инструмент IBM InfoSphere DataStage® включает в себя коннектор, позволяющий использовать данные BigInsights в задании DataStage ETL.

В комплект поставки InfoSphere BigInsights входит стандартный коннектор JDBC, позволяющий организациям быстро осуществлять интеграцию с разнообразными системами данных и информационными системами, включая Oracle, Microsoft SQL Server, MySQL и Teradata.

IBM InfoSphere Data Explorer

InfoSphere BigInsights V2.0 включает в себя лицензию ограниченного пользования на комплект ПО InfoSphere Data Explorer, позволяющий находить, просматривать и визуализировать огромные объемы структурированной и неструктурированной информации в InfoSphere BigInsights.

InfoSphere Streams

InfoSphere BigInsights включает в себя лицензию ограниченного пользования на продукт InfoSphere Streams, обеспечивающий последовательный анализ данных в движении в режиме реального времени. InfoSphere Streams — это система потоковой обработки класса предприятия, с помощью которой можно извлекать достоверную информацию, служащую основой для выполнения действий, из данных по мере их поступления на предприятие, при этом преобразуя данные и помещая их в InfoSphere BigInsights с высокой скоростью. InfoSphere Streams позволяет организациям фиксировать бизнес-данные и предпринимать действия на их основе в режиме реального времени — быстро получать, анализировать и сопоставлять информацию по мере ее поступления, что значительно оптимизирует функциональный компонент обработки. В состав ПО входит коннектор, позволяющий конечным пользователям выполнять чтение и запись в файловой системе InfoSphere BigInsights.

Cognos Business Intelligence

InfoSphere BigInsights включает в себя лицензию ограниченного пользования на продукт IBM Cognos® Business Intelligence, позволяющий бизнес-пользователям обращаться к необходимой им информации и анализировать ее в целях оптимизации принятия решений, более глубокого проникновения в суть данных и управления производительностью. Cognos Business Intelligence включает в себя программное обеспечение для создания запросов, отчетности, анализа и сводных панелей, а также программное обеспечение для сбора и организации информации, поступающей из различных источников.

Дополнительная информация

Для получения дополнительной информации о BigInsights Enterprise Edition обратитесь к торговому представителю компании IBM или посетите веб-сайт: ibm.com/software/data/infosphere/biginsights/enterprise.html



© Copyright IBM Corporation 2013

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Произведено в США
Январь 2013 г.

IBM, логотип IBM, ibm.com, Cognos, BigInsights, DataStage, DB2, InfoSphere, Guardium и PureData являются товарными знаками International Business Machines Corp., зарегистрированными во многих странах мира. Названия других продуктов и услуг могут являться товарными знаками IBM и прочих компаний. Актуальный список товарных знаков IBM опубликован в разделе «Сведения об авторских правах и товарных знаках» по адресу: ibm.com/legal/copytrade.shtml

Netezza является товарным знаком или зарегистрированным товарным знаком IBM International Group B.V., в составе компании IBM.

Intel является зарегистрированным товарным знаком Intel Corporation или ее дочерних фирм в США и других странах.

Linux является зарегистрированным товарным знаком Linus Torvalds в США и/или других странах.

Microsoft является зарегистрированным товарным знаком компании Microsoft Corporation в США и/или других странах.

Сведения, приведенные в настоящем документе, актуальны на момент его публикации и могут быть изменены IBM в любой момент. Перечень доступных предложений IBM зависит от региона.

ИНФОРМАЦИЯ, ПРЕДСТАВЛЕННАЯ В НАСТОЯЩЕМ ДОКУМЕНТЕ, ПРЕДОСТАВЛЯЕТСЯ «КАК ЕСТЬ», БЕЗ КАКИХ-ЛИБО ЯВНЫХ ИЛИ ПОДРАЗУМЕВАЕМЫХ ГАРАНТИЙ, В ТОМ ЧИСЛЕ ГАРАНТИЙ КОММЕРЧЕСКОЙ ЦЕННОСТИ, ПРИГОДНОСТИ ДЛЯ КАКОЙ-ЛИБО ЦЕЛИ И НЕНАРУШЕНИЯ ПРАВ. Продукты IBM предоставляются с гарантией в соответствии с положениями и условиями соответствующих соглашений.



Подлежит повторной переработке
