

# VMC Control-M для Hadoop

## Автоматизация пакетной обработки данных Hadoop и связанных промышленных деловых процессов

Сегодня многие компании стремятся достичь конкурентного преимущества за счет анализа Больших Данных при помощи Hadoop. Ускоренная разработка новых приложений Hadoop и их оперативное внедрение являются важнейшими условиями достижения этой цели. Поскольку большинство процессов обработки данных Hadoop выполняются в пакетном режиме, эффективное управление их выполнением становится важным условием получения быстрой отдачи от инвестиций в Большие Данные.

Control-M для Hadoop помогает организациям быстрее и без излишних затрат добиваться эффекта от инвестиций в Hadoop и Большие Данные. Решение помогает разработчикам приложений ускорить предоставление новых бизнес-услуг, сделать их более простыми при внедрении и эксплуатации в существующей ИТ-инфраструктуре. Это достигается путем предоставления возможностей для разработчиков по созданию, коллективной разработке и управлению цепочками заданий (как внутри Hadoop, так и интегрированных с другими системами) посредством простого, но мощного пользовательского интерфейса.

### УСКОРЕННОЕ ВНЕДРЕНИЕ ПРИЛОЖЕНИЙ HADOOP

При разработке нового приложения Hadoop сразу после написания кода (будь то Java, Pig, Hive и т.д.) необходимо разработать цепочку заданий для потоковой обработки данных при помощи нового приложения. Обычно это выполняется путем создания скриптов, что является, по сути, отдельным проектом. Кроме того, в этот момент должны быть определены критерии успеха выполнения каждого шага в процессе работы нового приложения, способы отладки и поиска точек отказа, а также действия в случае нештатных ситуаций: если некоторые задания в процессе завершились успешно, а некоторые нет и т.д. и т.п.

Control-M для Hadoop учитывает все эти требования, что позволяет за считанные секунды создавать новые цепочки заданий, добавлять связи простым перетаскиванием мышью и определять уведомления для таких событий, как позднее начало задачи, раннее ее окончание или более продолжительное, чем предполагалось, выполнение. Control-M для Hadoop сохраняет выходные данные и обеспечивает доступ к ним для анализа проблемы. При возникновении ошибок поточная обработка данных автоматически прекращается, если только не будет задан другой вариант действий. Отмена задачи осуществляется путем щелчка кнопки мыши и выбора меню «повторный запуск» (rerun). Благодаря этим возможностям приложения можно внедрять и запускать в эксплуатацию быстрее и с более высоким качеством, что позволит вместо отладки скриптов для уже готовых приложений перейти к созданию нового.

### ИСПОЛЬЗУЙТЕ HADOOP КАК ПРИВЫЧНЫЙ ИНСТРУМЕНТ

Control-M для Hadoop распространяет на Hadoop все возможности BMC Control-M Workload Automation. Это означает, что вы можете управлять поточной обработкой данных Hadoop с применением мощных графических пользовательских интерфейсов, которые предоставляют единую точку контроля всей рабочей нагрузки. Это позволяет убедиться, что по заранее составленным спискам адресатов были автоматически сформированы уведомления о неудачном выполнении задачи Hadoop — путем отправки электронного письма, отправки информации о событии на консоль предприятия или открытия инцидента в системе технической поддержки. В пользовательской консоли в графическом виде представлена информация о том, какие задачи будут выполняться, и как это может повлиять на рабочую нагрузку Hadoop. А проактивный мониторинг с использованием статистической информации и раннего предупреждения о предстоящей задержке выполнения цепочки заданий поможет убедиться, что сервисы Hadoop соответствуют SLA. Соблюдать управленческие требования и требования к аудиту предлагается посредством комплексного контроля версий приложений и цепочек заданий, сбора контрольных событий, авторизации пользователей и гибкой отчетности.

При обнаружении сбоя работы приложений Hadoop для скорейшего устранения проблемы выходные данные Hadoop JobTracker могут быть проанализированы в интерфейсе Control-M или отправлены в виде электронного письма сотруднику, отвечающему за анализ подобных проблем. Кроме того, Control-M обеспечивает возможность мониторинга сразу нескольких цепочек заданий Hadoop из единой консоли управления.

Пользователи могут управлять собственной рабочей нагрузкой на приложение Hadoop с помощью инструментов самообслуживания Control-M Self Service, не обладая специальными знаниями об особенностях работы с Hadoop или Control-M. Для этого на экран монитора выводятся названия и описания бизнес-задач, и они могут выбрать то, что им необходимо, либо из простого приложения на основе интернет-технологий, либо из мобильного приложения для устройств на платформах iOS и Android.

VMC Control-M ускоряет разработку приложений Hadoop, предоставляет надежное оперативное управление и облегчает интеграцию в существующую ИТ-инфраструктуру, что позволяет получить наибольшую выгоду от инвестиций в Большие Данные

### ОСОБЕННОСТИ BMC CONTROL-M

- Управление задачами Hadoop с помощью единого, мощного и простого в эксплуатации решения
- Использование графических форм, объединенных при помощи программы-мастера, которые требуют минимум обучения
- Концентрация на действительно значимых вопросах для бизнеса с помощью автоматического управления SLA
- Интеграция передачи файлов, бизнес-аналитики и аналитических приложений с заданиями Hadoop для повышения прозрачности выполнения бизнес-процессов
- Снижение операционных расходов за счет использования виртуализации и облачных технологий для динамического использования серверных ресурсов
- Простота соблюдения корпоративных норм регулирования и соблюдение нормативных требований за счет встроенных средств аудита и отчетности о выполненных задачах
- Предоставление пользователям полномочий по самостоятельному управлению цепочками заданий из интернет-браузера или мобильного устройства

## ОКУПАЕМОСТЬ

### Международная сеть супермаркетов

- В более короткие сроки создаются конкурентные цены и маркетинговые мероприятия
- Управление всеми бизнес-процессами, включая Hadoop, из единой консоли управления

### Страхование и финансовые услуги

- Многомиллионная выгода благодаря повышению эффективности обработки Больших Данных, что привело к реинжинирингу процесса работы со страховыми претензиями
- Интернет-оператор розничной торговли
- Повышение эффективности обработки обращений в интернет-магазин путем отслеживания поведения заказчиков
- BMC Control-M управляет обработкой аналитических данных на основе корпоративного хранилища данных и журналов транзакций пользователей интернет-магазина.

### Международная телекоммуникационная и ИТ-корпорация

- Собирает данные, введенные с клавиатуры
- Оптимизирует пользовательский интерфейс на основании анализа операций и поведения пользователей
- Обрабатывает сотни терабайтов данных ежедневно
- Фильтрует данные и помещает в различные разделы в Hadoop с последующей финальной аналитической обработкой

## УПРАВЛЕНИЕ БИЗНЕС-УСЛУГАМИ НА ОСНОВЕ БОЛЬШИХ ДАННЫХ В МАСШТАБЕ ПРЕДПРИЯТИЯ

В процессе работы Hadoop на основании входных данных генерирует на выходе результат в зависимости от используемой цепочки заданий. Входные данные могут поступать из разных источников, включая корпоративное хранилище данных (Enterprise Data Warehouse, EDW), реляционные базы данных и традиционные бизнес-приложения. Выходные данные могут часто использоваться приложениями, не относящимися к Hadoop. Вы должны не только соединить все эти приложения, но и управлять передачей файлов и данных данных по цепочке заданий. Корпоративные архитекторы теперь могут сконцентрироваться на проектировании таких потоков файлов и данных, которые необходимы для достижения требуемых бизнес-результатов, не беспокоясь о том, как технически создавать связи среди разнородных приложений, технологий и инструментов. Control-M позволяет осуществлять сбор данных и управлять их потоками, которые включают в себя широкий набор технологий:

**Hadoop:** Java MapReduce, Pig, Hive, Sqoop, HDFS

**ОС:** Windows, Unix, Linux, z/OS, iSeries, Tandem, OpenVMS

**Технологии:** FTP, sFTP, интернет-сервисы, EJB и Java-приложения

**ETL, BA и BI:** Informatica, IBM Cognos, SAP Business Objects, Oracle Business Intelligence, SQL Server SSIS

**RDBMS:** Oracle, Sybase, DB2/UDB, SQL Server и PostgreSQL

**Виртуализация и облачные технологии:** BMC BladeLogic, VMware, Amazon EC2

**Резервное копирование:** Symantec NetBackup, IBM Tivoli Storage Manager

## ГОТОВНОСТЬ ДЛЯ ВИРТУАЛЬНЫХ ЦЕНТРОВ ОБРАБОТКИ ДАННЫХ (ЦОД)

Современным решениям необходима гибкость для внедрения в физические, виртуальные, облачные или гибридные ЦОД. BMC Control-M использует виртуализованные и облачные среды, а также расширяет их функциональность.

Control-M for Hadoop позволяет в динамическом режиме запросить дополнительные аппаратно-программные ресурсы, чтобы компенсировать нехватку ресурсов при обработке большой цепочки заданий или, наоборот, для максимального использования общих ресурсов ЦОД.

Развертывание Control-M для Hadoop может быть включено в различные комплексные конфигурации, обеспечиваемые решениями BMC и сторонних поставщиков, например, решением по управлению облачной инфраструктурой BMC Cloud Lifecycle Management (CLM) «Service Blueprints». Service Blueprints – компонент BMC CLM, позволяющий описать инфраструктуру ИТ-сервиса, который впоследствии может быть заказан бизнес-пользователем для автоматического развертывания. Например, разработчик Hadoop может запросить среду Hadoop для использования при разработке нового приложения. Соответствующий Service Blueprint даст команду CLM предоставить виртуальную машину Linux, а затем запустить установку Hadoop вместе с агентом Control-M и Control-M для Hadoop. Эта новая машина автоматически подключается к «Тестовой» инфраструктуре Control-M и немедленно предоставляет разработчику Hadoop кластер Hadoop с Control-M для Hadoop. Подобные подходы могут быть применены и для расширения кластера: если работа, которая запускается на существующем кластере, перегружает его, то добавляются дополнительные ресурсы, которые должны обеспечить оптимальную нагрузку.

В тех организациях, где облако еще не используется или среда Hadoop работает на физических серверах, тот же самый подход может быть применен вместе с пакетом программ BMC BladeLogic Server Automation (BSA).

## БУДЬ ГОТОВ. АВТОМАТИЗАЦИЯ УПРАВЛЕНИЯ ПАКЕТНОЙ ОБРАБОТКОЙ ДАННЫХ ПРИ ПОМОЩИ BMC CONTROL-M

Использование Hadoop ведет к повышению ценности ИТ для бизнеса. Но имейте в виду, что тенденции и технологии постоянно меняются, а объем обрабатываемых данных растет. Будьте готовы решать будущие задачи вместе с BMC Control-M Workload Automation

Единая точка контроля, позволяющая уменьшить количество ошибок и снизить потребность в обучении персонала:

- Улучшенное качество обслуживания за счет концентрации на самых важных вопросах бизнеса и автоматический контроль уровней обслуживания для пакетной рабочей нагрузки в зависимости от заранее сконфигурированных приоритетов и метрик
- Встроенные возможности по передаче файловых данных по цепочкам заданий, объединяющим разнородные системы в инфраструктуре заказчика, обеспечивают гарантированную доставку и полную видимость перемещения
- Мощное прогнозирование и анализ влияния изменений посредством сценариев «что если»
- Снижение операционных расходов за счет обеспечения динамической инфраструктуры с использованием виртуализации и облачных технологий
- Соответствие правилам аудита и гибкая отчетность
- Ориентация на бизнес, интуитивное самообслуживание и мобильность, когда бизнес-пользователям предоставляется право управлять своей рабочей нагрузкой без обременения перегруженного ИТ-персонала и нарушения корпоративных правил

### Дополнительная информация

Более подробную информацию о решениях BMC можно найти по адресу: [www.bmcsoftware.ru](http://www.bmcsoftware.ru)

BMC помогает ведущим компаниям по всему миру повысить уровень предоставления и потребления цифровых услуг и эффективно использовать ИТ-технологии для трансформации бизнеса. Компания предоставляет инновационные решения по управлению ИТ, от мейнфрейма до облачных сред и мобильных устройств, при помощи которых клиенты добиваются колоссальной производительности и делают свой бизнес более динамичным, при этом достигнутые результаты превышают их ожидания. Мы верим, что:

- Технологии – это двигатель любого бизнеса
- ИТ вводят бизнес в цифровую эру

**BMC – Bring IT to Life**

