❯ Обзор решения

Очистка данных в Hadoop
Очистка данных непосредственно в Hadoop с помощью готовых преобразований для повышения качества данных

Преобразование данных в Hadoop
Преобразование данных из таблиц Hadoop

Запрос или объединение данных в Hadoop
Создание таблицы или объединение данных из нескольких таблиц

Профилирование данных
Составление отчета о профиле данных в таблице

Возможности SAS® Data Loader для Hadoop

SAS Data Loader для Hadoop позволяет использовать и обрабатывать данные в Hadoop с помощью удобного интерфейса. Пользователь с минимальным уровнем подготовки может с легкостью самостоятельно подготовить необходимые данные. Пользователи, обладающие техническими навыками, могут создавать и выполнять код SAS, обрабатывающий данные в Hadoop, повышая производительность и управляемость кода.

Сферы применения SAS® Data Loader

Все больше организаций применяют Hadoop для хранения больших объемов данных. Для управления данными в Hadoop часто требуются специфические навыки, поэтому этот процесс зачастую сопряжен с определенными сложностями. SAS Data Loader для Hadoop позволяет избежать подобных затруднений. Это решение предоставляет пользователям удобный доступ к данным вне зависимости от уровня их технической подготовки.

Потенциальные потребители SAS® Data Loader для Hadoop

Приложение предназначено для бизнес-пользователей, сталкивающихся с необходимостью обрабатывать массивы «больших данных» без написания кода, а также для программистов на языке SAS и специалистов по обработке данных, которые благодаря ему смогут повысить эффективность и производительность своей работы.

SAS Data Loader для Hadoop

Управляйте своими большими данными самостоятельно, не прибегая к помощи специалистов ИТ-отдела

На сегодняшний день важность больших данных уже не вызывает сомнений. Организации из разных сфер бизнеса стремятся максимально эффективно реализовать возможности, которые скрывают такие массивы информации, соответственно, видят ценность в их анализе и использовании разнообразных современных технологий. Для этого большие данные собирают и хранят в таких системах, как Hadoop.

Очевидно, что хранение данных и управление ими — абсолютно разные процессы. Обращение к данным в Hadoop требует написания специализированного кода. Его разработка и поддержка — сложная задача, требующая специфических навыков. В итоге данные лежат мертвым грузом и не приносят пользы. Вариантов решения в этой ситуации немного: привлечь к решению задач ИТ-специалиста, освоить самостоятельно программирование — или найти какое-то промежуточное решение.

SAS Data Loader для Hadoop содержит интуитивно понятный интерфейс для исследования, трансформации, очистки и перемещения данных в Hadoop. Вы сможете пользоваться данными, не имея навыков программирования. Сотрудники ИТ-отдела, избавленные от необходимости заниматься рутинной работой, смогут сконцентрироваться на более сложных задачах, например, на повышении производительности обработки или на защите информации.

Преимущества

Управление данными без специальной подготовки.
Вам не придется проходить сложное обучение или привлекать высокооплачиваемого специалиста. SAS Data Loader для Hadoop позволит вам выполнять задачи по интеграции, повышению качества и трансформации данных самостоятельно, без привлечения ИТ-специалистов.

Раскройте возможности больших данных.
SAS Data Loader для Hadoop снижает планку требований к уровню технической подготовки пользователей, открывая большему кругу пользователей многообразие возможностей по работе с данными. С нашим инструментом вы сможете самостоятельно выполнять исследование, очистку и трансформацию данных для задач продвинутой аналитики.

Повышение масштабируемости и производительности.
Бизнес-пользователи смогут применять SAS Data Loader для Hadoop как вспомогательное средство подготовки данных для задач аналитики и поддержки принятия решений, а специалисты по обработке данных и программисты SAS оценят повышение скорости, производительности и гибкости своей работы. Встроенный в решение обработчик кода в Hadoop позволяет ускорить подготовку и обработку данных. Кроме того, минимизируя количество перемещений данных, вы обеспечиваете их дополнительную защиту.

sas
СИЛА ЗНАНИЯ®

Обзор продукта

SAS Data Loader для Hadoop — это пакет продуктов SAS, включающий в себя такие средства интеграции и управления качеством данных, как SAS Data Loader, SAS/ACCESS Interface для Hadoop, SAS In Database Code Accelerator для Hadoop и SAS Data Quality Accelerator для Hadoop.

SAS Data Loader для Hadoop сочетает удобные возможности для обычных пользователей и специализированные инструменты для ИТ-профессионалов, что делает его универсальным решением для пользователей с любым уровнем подготовки.

Интуитивный пользовательский интерфейс

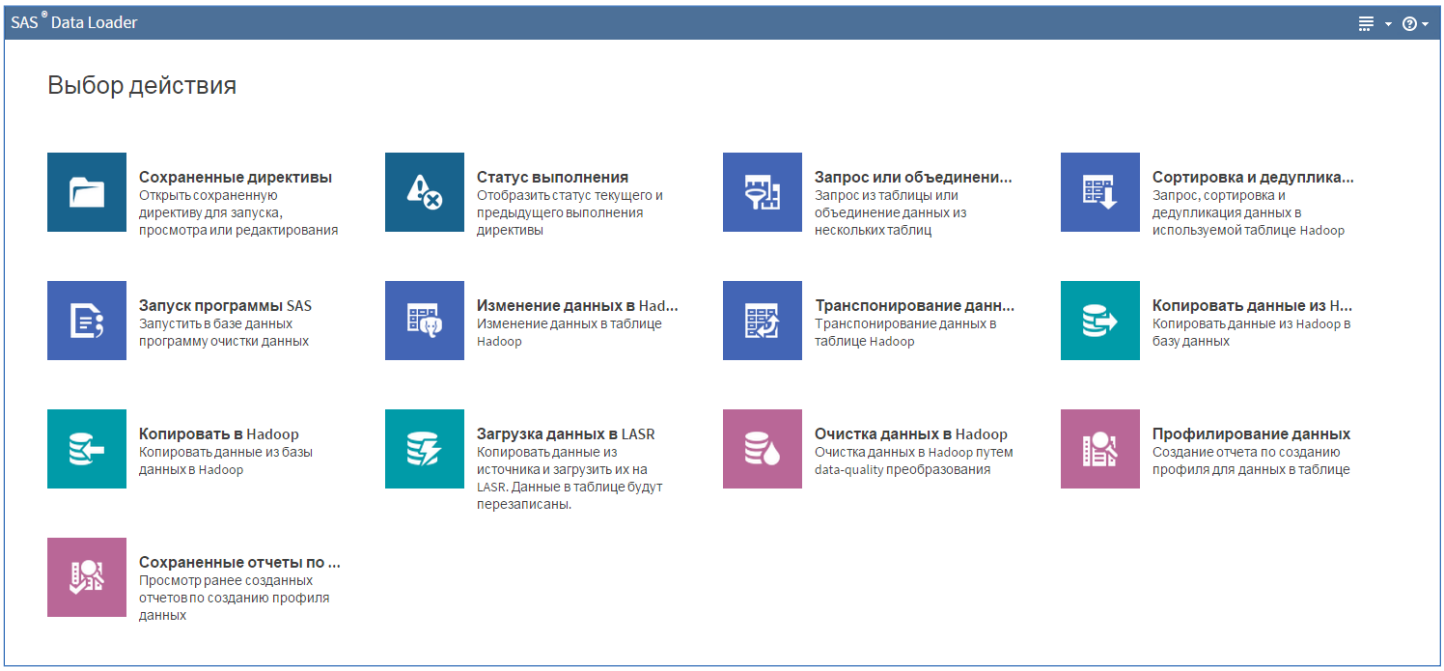
SAS Data Loader для Hadoop создавался для бизнес-пользователей. Его интуитивно понятный интерфейс и возможность решения множества задач с помощью мастеров позволяет с легкостью обращаться к данным в Hadoop и управлять ими без привлечения ИТ-специалиста с навыками работы в Hadoop

Изначально создавался для работы с данными в Hadoop.

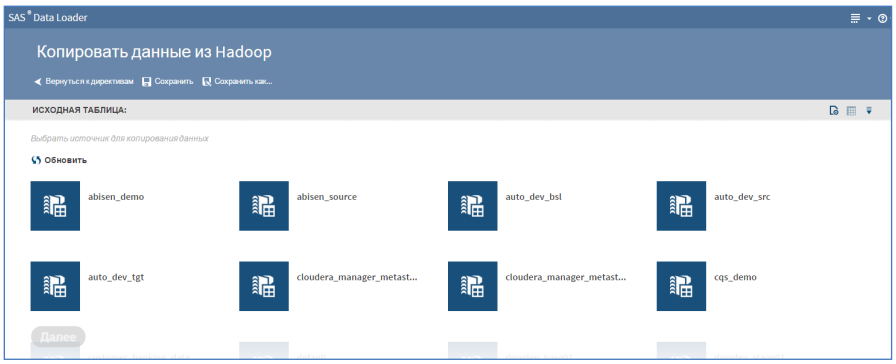
SAS Data Loader для Hadoop — это не адаптация имеющихся решений для платформы Hadoop. Данное средство изначально разрабатывалось для управления большими данными в Hadoop.

Качество больших данных

Контролируйте данные, размещенные в Hadoop. SAS Data Loader для Hadoop позволяет исследовать данные и составить представление об их общем качестве. После этого вы сможете осуществлять стандартизацию, статистический разбор, сопоставление и другие базовые задачи обеспечения качества данных внутри Hadoop с помощью нетребовательного к ресурсам механизма выполнения программ SAS — SAS Embedded Process.



Пользовательский веб-интерфейс на основе мастеров существенно упрощает доступ к данным в Hadoop и управление ими с помощью директив.



Информационные службы для самостоятельного копирования данных в Hadoop.

Аналитический сервер с обработкой данных в оперативной памяти

Чтобы сформировать необходимый набор данных для отчета, визуализации или анализа, больше не придется привлекать ИТ-специалиста. SAS Data Loader для Hadoop позволяет бизнес-пользователям загружать подготовленные данные на сервер аналитики SAS LASR Analytic Server, чтобы подготовить их к обработке в SAS Visual Analytics.

Выполнение кода SAS в кластере

SAS Data Loader для Hadoop позволяет осуществлять аналитическую обработку в экосистеме Hadoop, получая результаты быстрее и с меньшими затратами, чем при использовании более традиционных решений. Такой подход позволяет сократить перемещения данных и обеспечить параллельную обработку, а значит, повысить масштабируемость и производительность системы.

SAS Data Loader для Hadoop поддерживает следующие дистрибутивы Hadoop: Cloudera и Hortonworks.

Основные функции

Преобразование и перемещение данных в Hadoop

- Импортируйте и экспортируйте данные из реляционных баз данных и наборов данных SAS в Hadoop с помощью средств параллельной загрузки больших массивов данных.
- Преобразуйте данные: поддерживается фильтрация строк, управление столбцами, суммирование строк.
- Перемещайте и объединяйте выбранные столбцы.

Защищенный доступ к большим данным

- Используйте защищенный доступ к кластерам Hadoop с поддержкой Kerberos.

Очистка данных в Hadoop

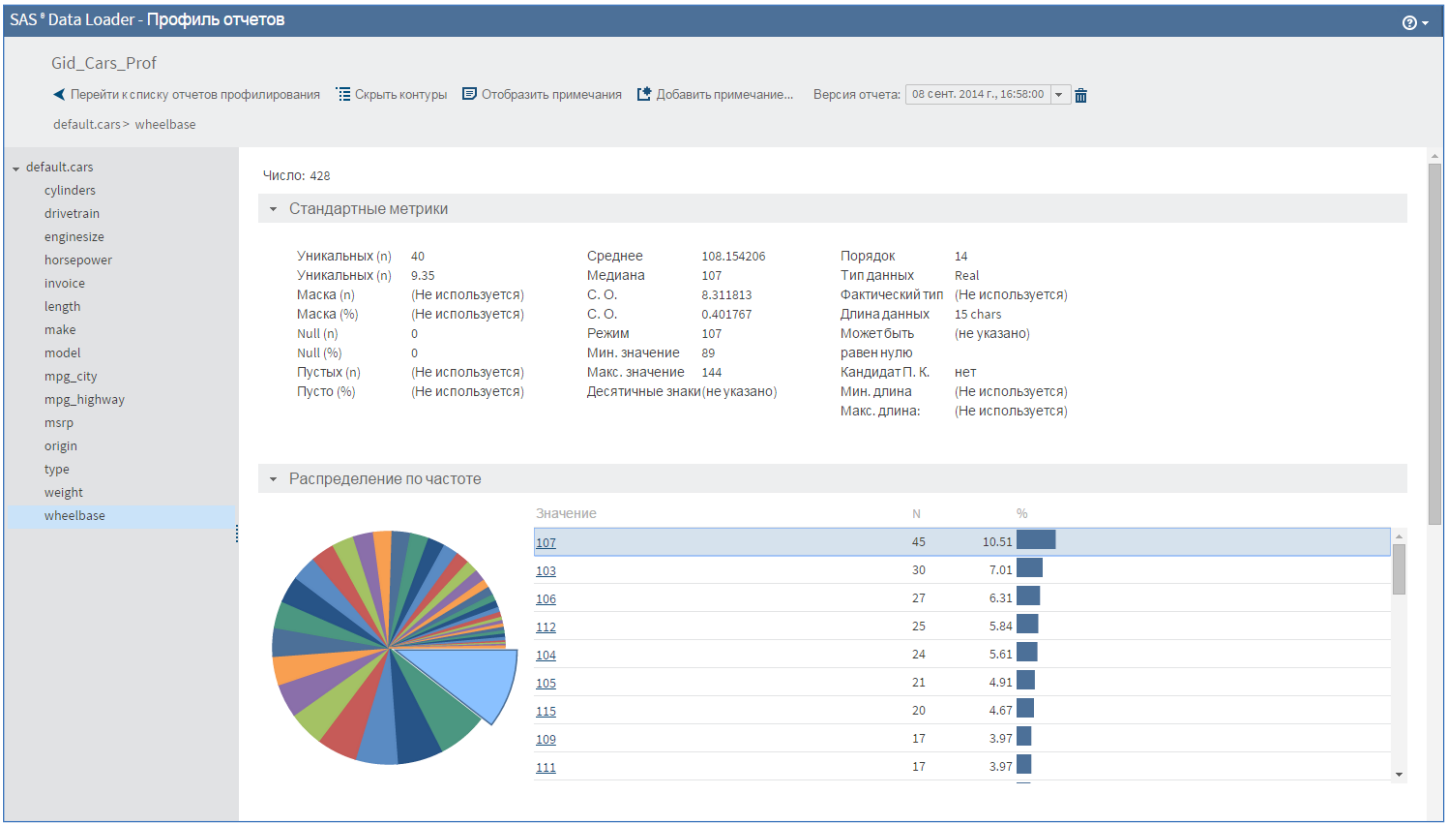
- Стандартизируйте, дуплицируйте, сопоставляйте, используйте синтаксический разбор и множество других операций по обеспечению качества данных внутри Hadoop.
- Используйте углубленную фильтрацию, которая позволяет импортировать результаты директив Профилирования в директивы Фильтрации и Трансформации.
- Запрашивайте, сортируйте, дедуплицируйте данные внутри существующих таблиц Hadoop.

Запрос или объединение данных в Hadoop

- Выполняйте запросы к таблицам или объединяйте таблицы без знания SQL.
- Агрегируйте выбранные столбцы и фильтруйте исходные данные.
- Опытные пользователи могут создавать и изменять запросы HiveQL, а также вставлять имеющиеся запросы HiveQL.

Профилирование данных и сохранение отчетов о профилях

- Определяйте уникальность, неполноту и закономерности для выбранных столбцов одной или нескольких таблиц.
- Создавайте списки отчетов и открывайте отчеты, созданные с помощью директивы профилирования данных.
- Создавайте и сохраняйте примечания.



В ходе профилирования вычисления выполняются на кластере Hadoop для повышения производительности.

Основные функции (продолжение)

Управление директивами и их повторное использование с помощью пользовательского интерфейса на основе мастеров

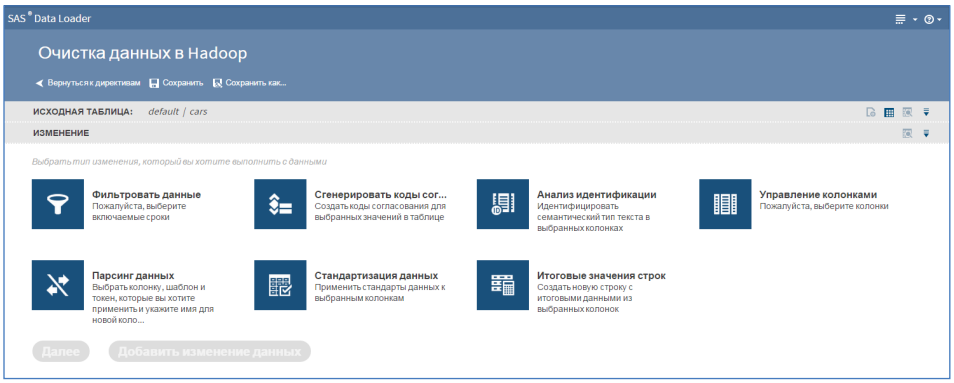
- Просматривайте списки и состояния директив и журналов задач.
- Запускайте и останавливайте директивы, просматривайте их журналы и созданные файлы кода.
- Запускайте, просматривайте и изменяйте сохраненные директивы для повторного использования.

Загрузка данных на сервер аналитики SAS LASR Analytic Server

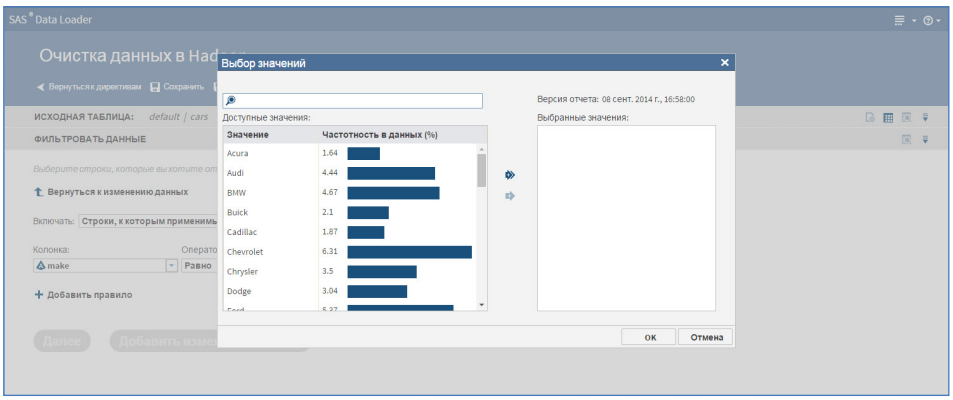
- Загружайте выбранные столбцы в таблицах Hadoop на сервер аналитики SAS LASR Analytic Server для анализа с помощью средств SAS Visual Analytics или SAS Visual Statistics (лицензии приобретаются отдельно).

Выполнение программ SAS

- Выполняйте программы SAS на языке DS2 в Hadoop с помощью нетребовательного к ресурсам механизма выполнения программ SAS — SAS Embedded Process.



Все функции управления качеством данных, такие как синтаксический анализ и стандартизация данных, поддерживаются на стороне сервера Hadoop.



Передача результатов профилирования данных в директивы управления качеством позволит быстро перейти от исследования к решению выявленных проблем.

Общество с ограниченной ответственностью «САС ИНСТИТУТ»
Россия, 109004, г. Москва, ул. Станиславского, дом 21, строение 1, • Тел.: +7 495 227 4151 • Факс: +7 495 227 4155 • www.sas.com/russia

ТОО «САС Інстітьют Ел.Ел.Сі.»
Україна, 01601, Київ, вул. Шовковична 42-44 • Тел.: +38 (044) 459 0355 • Факс: +38 (044) 490 1200

sas
СИЛА ЗНАНИЯ®