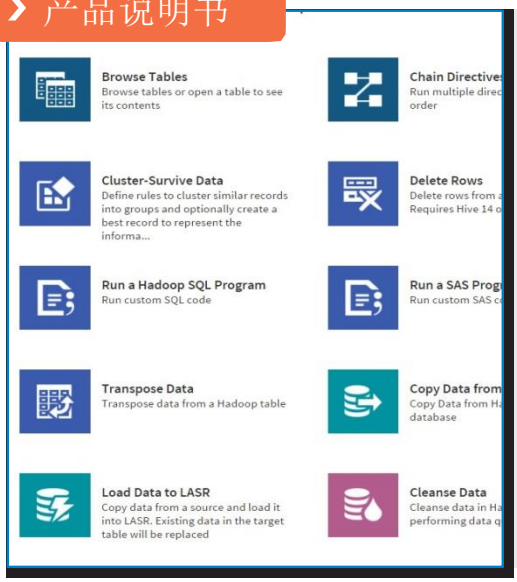
› 产品说明书



SAS® Hadoop 数据加载器可以做什么？

SAS 为 Hadoop 打造的数据加载器，通过一个直观的用户界面，帮助您访问和管理 Hadoop 平台上的数据。它让自助式数据准备变得十分简单，用户只需接受最基本的培训就可以了。具有一定技能的用户可以在 Hadoop 平台上编写和运行 SAS 代码，提高性能和治理能力。

SAS® Hadoop 数据加载器为什么重要？

随着越来越多的企业转向使用 Hadoop 来存储海量的数据，他们同时也发现，因为 Hadoop 常常要求专业的编程能力，使得管理数据变得十分困难。SAS® Hadoop 数据加载器弥补了用户技能上的差距，让用户可以轻松地访问他们的数据，无需太高的技能要求。

SAS® Hadoop 数据加载器为谁而设计？

该解决方案专为业务用户而设计，业务用户可以从他们拥有的大数据中获得价值，而且他们不需要编写代码；方案还可以为 SAS 编程人员和数据科学家所用，他们可以利用这个解决方案来改善性能和生产效率。

SAS® Hadoop 数据加载器

通过自助式的大数据整合，让您掌控您的数据，释放 IT 人员的压力

企业终于认识到了大数据的重要性。他们知道可以将大数据用于分析技术和其它高端技术，所以他们在诸如 Hadoop 之类的系统中操作和存储大数据。

然而，存储数据是一回事，能够管理数据却完全是另一回事。要访问 Hadoop 中的数据，需要难以创建和维护的代码，这造成必要的管理技能方面的差距。而且，如果您不能处理您需要的数据，那么您在第一时间收集这些数据也变得毫无意义。这意味着您只有一些有限的选择：要么依赖 IT 团队，要么自己学习编程——或者找到一个能够克服这种差距的解决方案。

SAS Hadoop 数据加载器（SAS Data Loader for Hadoop）为您提供一个直观的界面，可以在 Hadoop 中进行数据的特征刻画、管理、清洗和迁移，所以您可以对数据进行操作而无需掌握编程知识。而且，IT 部门也被解放出来，可以专注于收益更高的技术工作上，如提高处理性能和提高数据的安全性。

带来的好处

只需最低限度的技能即可管理数据

无需高级培训或招聘高薪人才。SAS Hadoop 数据加载器让您拥有自助式数据整合、数据质量管理和数据准备的能力，不需要依靠 IT 部门的帮助。

利用大数据的能力

一旦清除技能组合方面的障碍，您利用数据所能够做的事情将潜力无限。SAS Hadoop 数据加载器就是这一切的驱动力。您将能够刻画特征、清洗、连接和转换数据，创建高质量的信息，更有效地进行高级分析。

提高可扩展性和性能

业务用户注重使用 SAS Hadoop 数据加载器来支持分析和决策制定，而数据科学家和 SAS 编程人员可以使用它来提高速度、效率和灵活性。解决方案中的代码加速器利用 Hadoop 的能力，可以带来更快的性能。而且，通过数据迁移的最小化，您可以提高数据的安全性。

免费试用

SAS Hadoop 数据加载器可下载免费试用 45 天，与生产 Cloudera 或 Hortonworks 集群配套使用。试用版可转为完整版，不需要重新安装软件。

sas
THE POWER TO KNOW.

产品概述

SAS Hadoop 数据加载器（SAS Data Loader for Hadoop）是一个 SAS 产品包，包含 SAS 数据加载器、SAS 的 Hadoop 访问接口，SAS 的 Hadoop 库内代码加速器和 SAS 的 Hadoop 数据质量加速器。它提供了数据整合和数据质量管理等技术。

SAS Hadoop 数据加载器以其友好的用户界面和高级的技术特性，成为一种让企业双方受益的解决方案。

直观的用户界面

SAS Hadoop 数据加载器专为业务用户而设计。直观的向导式界面使得访问和管理存储在 Hadoop 中的数据变得十分容易，减少了对 IT 的依赖或招聘 Hadoop 专业人才的需要。

专为将数据载入或导出 Hadoop 而设计

SAS Hadoop 数据加载器是专为管理 Hadoop 上的大数据而设计开发，并非借用现有的 IT 工具。

大数据质量

掌控您的 Hadoop 环境中的数据。SAS Hadoop 数据加载器让您能够对数据进行特征刻画，了解其总体质量。然后，您可以使用轻量级 SAS 执行引擎 SAS 嵌入过程进行标准化、解析、匹配和执行 Hadoop 内的其他核心数据质量功能。

其他数据质量指令包括封装、性别分析、模式分析和字段提取等。从而，用户可以将情况变化应用到数据中，根据相关数值猜测性别以改进客户细分，并根据字段值猜测可接受的数据模式。字段提取可从字段内的非结构化或任意文本中获取有用的符号，例如姓名、组织、地址、电子邮箱和电话号码等。

特征刻画在 Hadoop 集群上并行运行，提高性能，添加发展趋势图表，可随着时间的推移跟踪数据变化。识别分析确定一列中显示的数据类型。例如，一列数据内的“NC”、“North Carolina”和“N Carolina”可划分到“州”的类别中，辅助进行数据探查。

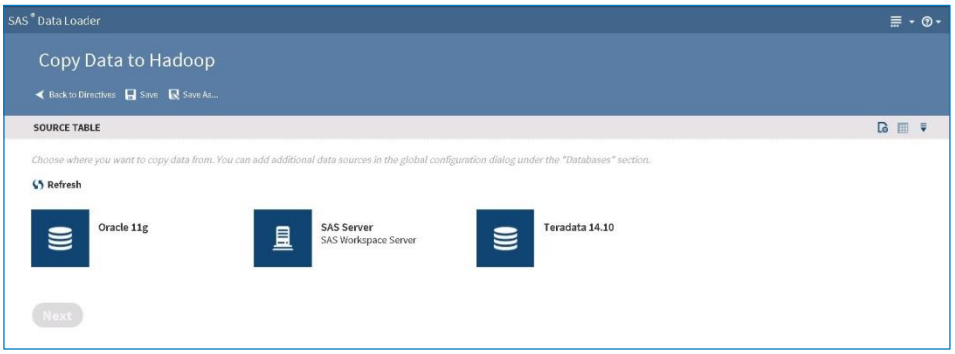


图 1：自助式功能让您可以将数据复制到 Hadoop 中

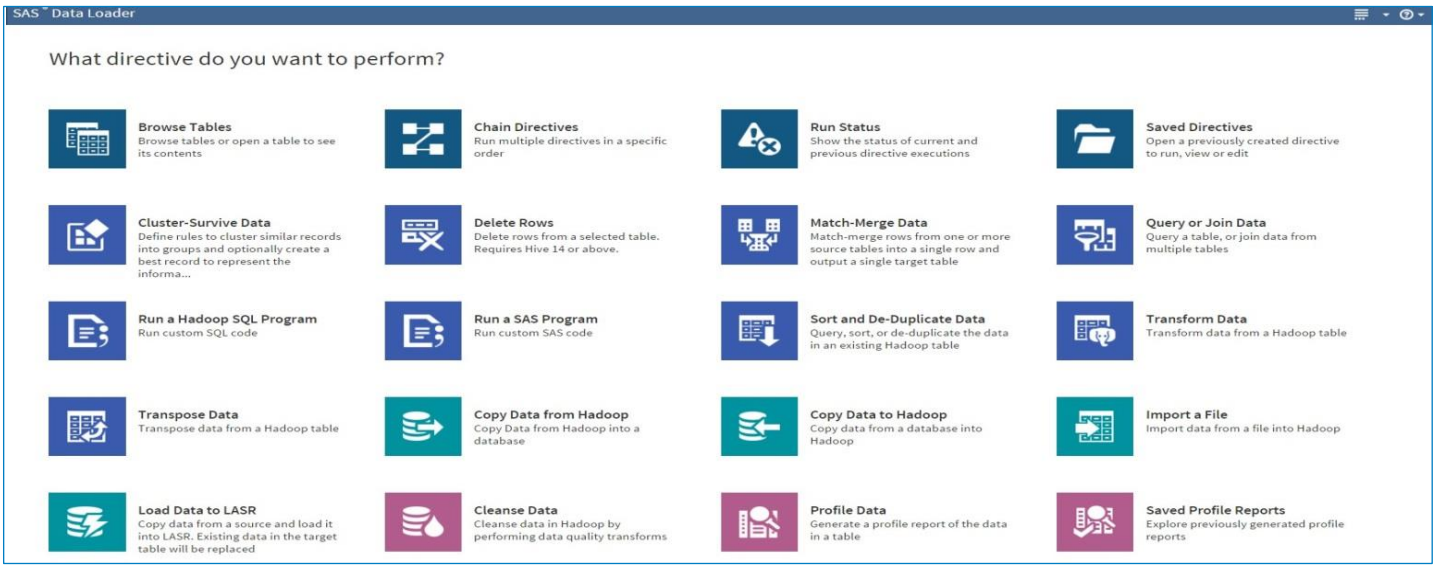


图 2：SAS 数据加载器使用嵌入式指令，最大限度减少培训需求。这是“符合归并”数据指令的一个实例，它让您可以将两个表归并到一个表中，并在 Hadoop 内运行，提高性能。

主要特点

内存分析服务器

不需要等待 IT 获取您需要的数据进行报告、可视化或分析。SAS Hadoop 数据加载器使得业务用户可以将数据载入到 SAS® LASR™ 分析服务器，在内存中为 SAS 可视化分析准备数据。

安全性

SAS Hadoop 数据加载器支持安全访问启用 Hortonworks 和 Cloudera KerberosHadoop 的集群。您还可以与活动目录或 LDAP 集成，以支持用户授权。

集群内 SAS®代码执行

使用 SAS Hadoop 数据加载器，您可以在 Hadoop 生态圈内执行分析处理，以比传统解决方案更低的成本，更快速地获取结果。由于减少了数据迁移，同时实现并行处理，带来可扩展性和性能上的提升，将让您受益匪浅。

在 Hadoop 中转换和转置数据

- 通过并行化的批量数据迁移，将关系数据库和 SAS 数据集中的数据载入或导出 Hadoop。
- 将数据从 CSV 及其他限定文件输入到 Hadoop 中，删除 Hadoop 表中的行。
- 通过过滤和汇总行以及管理列,来转换数据。
- 转置和分组选定的列。

安全访问大数据

- 安全访问启用 Kerberos 的 Hadoop 集群（Hortonworks 和 Cloudera）。
- 支持活动目录和基于 LDAP 的用户验证。

在 Hadoop 中清洗数据

- 在 Hadoop 中对数据进行标准化、去重、匹配和解析。
- 智能过滤功能，支持将数据特征刻画中的数值导入过滤器和转换指令。
- 在现有的 Hadoop 数据表中查询、排序或数据去重。
- 基于数值确定一列中的数据类型，加快数据探查。
- 通过使用其他数据质量功能，可应用封装，确定性别，进行模式分析，从非结构化文本字段提取符号。

在 Hadoop 中查询或连接数据

- 无需掌握 SQL 技术，即可查询数据表或进行多表连接。
- 对选定的列和过滤源数据进行汇总。
- 高级用户可以生成或编辑 HiveQL 查询、或粘贴现有 HiveQL 查询语句。

接下页

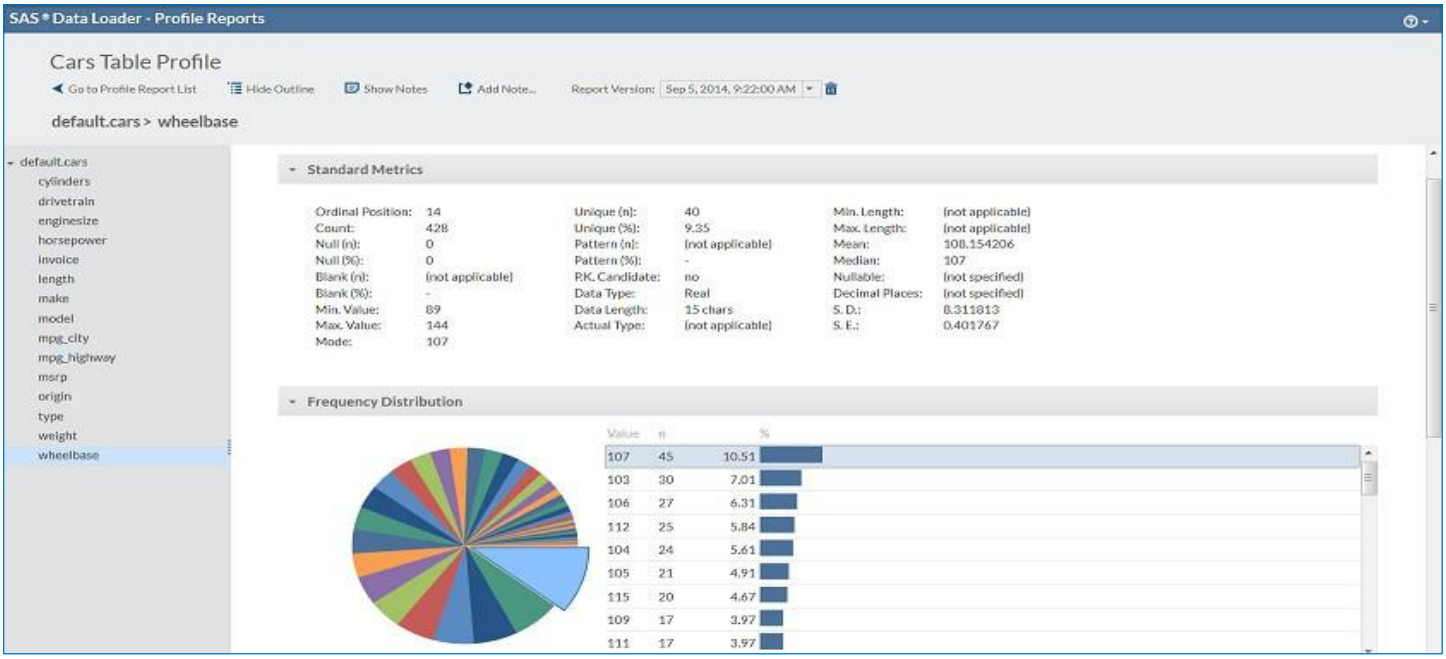


图 3：特征刻画处理下移至 Hadoop 集群，以提高性能。

主要特点（续）

通过 Spark 加快数据管理流程

- 数据质量功能在 Spark 上的内存中运行，以提高性能。
- Spark 匹配和最佳纪录创建支持大数据的主数据管理。
- 如必要，读取并写入 Spark 数据集。

将数据专业人员的工作效率提高到一个新的水平

- 加快创建 Impala 查询——更快访问 Hadoop 中的数据。
- 将多条指令组合到一起，整体运行。
- 允许使用披露的公共 API 进行对外作业调度。

在数据驻留位置管理数据

- 为 Pivotal HD 和 IBM Big Insights 以及 Hortonworks、Cloudera 和 MapR 提供 Hadoop 支持。
- 符合归并指令支持在数据库内归并多个数据源。
- 加强运行时执行日志。

刻画数据特征并保存特征刻画报告

- 从一个或多个表格中选择源数据列，确定独特性、不完整性和模式。
- 列举和打开特征刻画数据指令生成的报告。
- 创建并保存备注。
- 在 Hadoop 集群上并行运行特征刻画，以提高性能。

通过向导式用户界面，管理和重用指令

- 查看指令和作业日志的列表和状态。
- 停止和启动指令，打开日志和生成的代码文件。
- 运行、查看或编辑已保存的指令，以便重用。

将数据加载到 SAS®LASR™ 分析服务器

- 将内存中特定的 Hadoop 列加载到 SAS LASR 分析服务器，然后可以使用 SAS 可视化分析或 SAS 可视化统计（单独授予许可）进行分析。

运行 SAS® 程序

- 通过 SAS 嵌入过程（一个轻量级的 SAS 执行引擎），在 Hadoop 上运行包含 DS2 语句的 SAS 程序。

SAS Hadoop 数据加载器支持以下 Hadoop 发行版：

- Cloudera 5.2 和 5.3
- Hortonworks 2.1 和 2.2
- MapR 4.02

（不包括 Kerberos）

要了解关于 SAS Hadoop 数据加载器的更多信息，可下载白皮书，查看屏幕截图及其他相关资料，请访问：
sas.com/dataloader

THE POWER TO KNOW®