SAS® Data Loader for Hadoop

› fact sheet

셀프 서비스로 Hadoop의 빅 데이터를 직접 간편하게 관리할 수 있는 솔루션

빅 데이터는 시장에서 점차 중요해지고 있습니다. 실제로, 분석을 비롯한 여러 고급 기술에 빅 데이터를 활용할 수 있게 되면서 많은 기업들이 빅 데이터를 Hadoop 시스템에 저장하고 있습니다. SAS Data Loader for Hadoop은 Hadoop 환경에서 데이터 프로파일링, 관리, 정리, 이동을 가능하게 해주는 간편한 인터페이스를 제공하여 일반 사용자가 데이터를 손쉽게 조작할 수 있습니다.

Hadoop 데이터 정제
데이터 품질 변환(data quality transforms)을 실행하여 Hadoop 데이터를 정제합니다.

Hadoop 데이터 변환
Hadoop 테이블의 데이터를 변환합니다.

Hadoop 데이터 조회 또는 통합
테이블을 생성하거나 복수의 테이블들을 통합합니다.

Hadoop 적용 및 실행
Hadoop 내부에서 분석 모델을 직접 적용, 데이터의 이동 감소 및 빠른 분석 제공

데이터 프로파일링
데이터에 대한 프로파일 보고서를 생성합니다.

SAS® Data Loader for Hadoop은 어떤 솔루션인가?

SAS Data Loader for Hadoop은 간편하고 쉬운 사용자 인터페이스를 통해 Hadoop 환경에서 직접 데이터를 액세스하고 관리할 수 있게 해주는 솔루션으로, 누구나 간단한 교육을 거쳐 손쉽게 셀프 서비스 데이터 준비 작업을 수행 할 수 있습니다. 또한 전문 스킬을 갖춘 고급 사용자는 Hadoop에서 직접 SAS 코드를 작성하고 실행 할 수 있으므로 업무 성과와 거버넌스가 개선됩니다.

SAS® Data Loader for Hadoop이 왜 중요한가?

대용량 데이터를 저장하기 위해 Hadoop을 선택하는 조직이 늘고 있습니다. 하지만 Hadoop은 전문적인 코딩 스킬을 요하기 때문에 데이터를 관리하는 데 여전히 어려움이 따르고 있는 실정입니다. SAS Data Loader for Hadoop은 이 같은 스킬 격차를 해소하여 전문 지식이 없는 일반 사용자도 데이터를 손쉽게 액세스할 수 있도록 해주는 솔루션입니다.

SAS® Data Loader for Hadoop은 누구를 위한 솔루션인가?

SAS Data Loader for Hadoop은 일반 비즈니스 사용자가 코드를 작성할 필요 없이 빅 데이터를 최대한 활용하고, SAS 프로그래머와 데이터 사이언티스트가 업무 성과와 생산성을 개선할 수 있도록 설계된 솔루션입니다.

SAS Data Loader for Hadoop 특징

- **최소한의 스킬로 데이터 관리 가능**
  전문 교육을 실시하거나 값비싼 인력을 채용할 필요가 없습니다. SAS Data Loader for Hadoop은 IT 부서의 도움이 없이도 일반 사용자가 직접 데이터 통합, 데이터 품질, 데이터 준비 등의 작업을 수행할 수 있게 해줍니다.
- **빅 데이터 가치 활용 극대화**
  일단 스킬 장벽이 해소되면 데이터로 할 수 있는 일이 무궁무진해집니다. SAS Data Loader for Hadoop이 바로 이런 잠재력을 실현시켜 주며, 사용자는 데이터 프로파일링, 정리, 통합, 변환을 통해 고급 분석에 필요한 양질의 정보를 산출할 수 있게 됩니다.
- **확장성과 성능 강화**
  비즈니스 사용자는 SAS Data Loader for Hadoop을 이용해서 분석 및 의사결정 업무 지원에 초점을 맞추고, 데이터 사이언티스트와 SAS 프로그래머는 속도와 효율성, 민첩성을 향상시키는 데 주력할 수 있습니다. 또한 Code Accelerator 기능이 Hadoop의 파워를 활용해 성능과 속도를 크게 높여주며, 그 밖에도 데이터 이동을 최소화하여 데이터 보안을 강화할 수 있습니다.
- **Hadoop 데이터 로딩에 특화된 설계**
  SAS Data Loader for Hadoop은 기존의 IT 전용 툴을 용도만 변경한 것이 아니라, 처음부터 Hadoop 환경에서 빅 데이터를 관리할 목적으로 설계되었습니다.

THE POWER TO KNOW.

주요 기능

SAS Data Loader for Hadoop은 SAS 소프트웨어 번들 패키지로, 여기에는 데이터 통합 및 데이터 품질 작업을 위한 SAS Data Loader, SAS/ACCESS® Interface to Hadoop, SAS In-Database Code Accelerator for Hadoop, SAS Data Quality Accelerator for Hadoop 등이 포함되어 있습니다.

SAS Data Loader for Hadoop 솔루션을 통해 사용자 친화적 기능과 최첨단 기능의 이점을 동시에 누릴 수 있습니다.

Hadoop 환경에서의 데이터 정제: 빅 데이터 품질 관리

Hadoop 환경 내에서 데이터를 관리할 수 있습니다. SAS Data Loader for Hadoop이 데이터를 프로파일링하여 데이터 품질의 전반적인 상태를 파악할 수 있게 해주면, 사용자는 간단한 SAS 실행 엔진인 SAS Embedded Process를 사용하여 Hadoop 환경에서 데이터의 표준화, 파싱, 매칭 등 핵심적인 데이터 품질 기능을 수행할 수 있습니다.

- Hadoop 환경에서 데이터 표준화, 중복 제거, 매칭, 파싱 등 다양한 데이터 품질 기능 수행
- 인텔리전트 필터링 기능으로 Profile 값을 필터 및 변환
- 기존 Hadoop 테이블에서 데이터 조회·정렬 및 데이터 중복 제거

Hadoop 환경에서의 데이터 변환 및 전환

필요한 데이터를 선택하여 조작하고 사용하고자 하는 데이터 형식으로 재구성합니다.

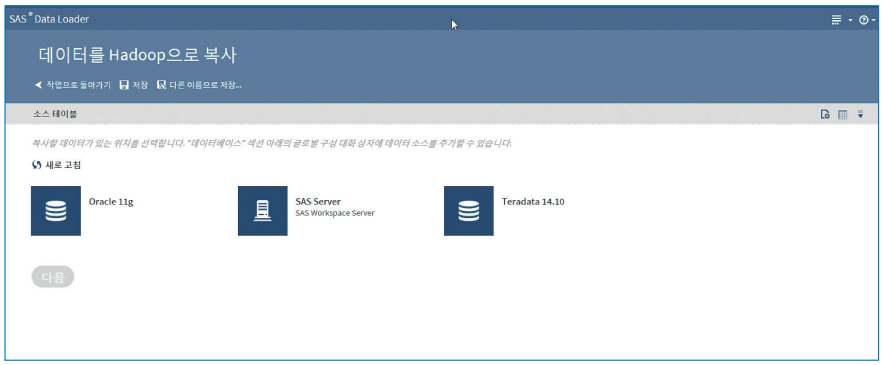
- 대용량 데이터의 병렬 이동이 가능하므로 Hadoop에 RDBMS 및 SAS 데이터 셋을 복사
- 행 필터링, 열 관리, 행 요약을 통한 데이터 변환
- 선택한 열의 전환 및 분류

Hadoop 환경에서의 데이터 조회 / 통합

공통 Key 가 없는 데이터 포함, 다양한 데이터 셋을 합치거나 중복된 데이터를 제거하고, 집단 내 새로운 데이터를 생성합니다.

- SQL없이 테이블 조회 및 복수 테이블 통합
- 선택한 열 집계(aggregations) 및 소스 데이터 필터링
- 고급 사용자를 위한 HiveQL 쿼리 생성·편집 기능과 기존 HiveQL 쿼리 붙여넣기 기능

마법사 기능을 지원하는 웹 기반 인터페이스를 통해 Hadoop에 저장된 데이터를 손쉽게 액세스하고 관리할 수 있게 해줍니다.



셀프 서비스 기능을 이용해서 Hadoop에 데이터를 복사할 수 있습니다.

Hadoop 인-클러스터 SAS® 코드 적용 및 실행

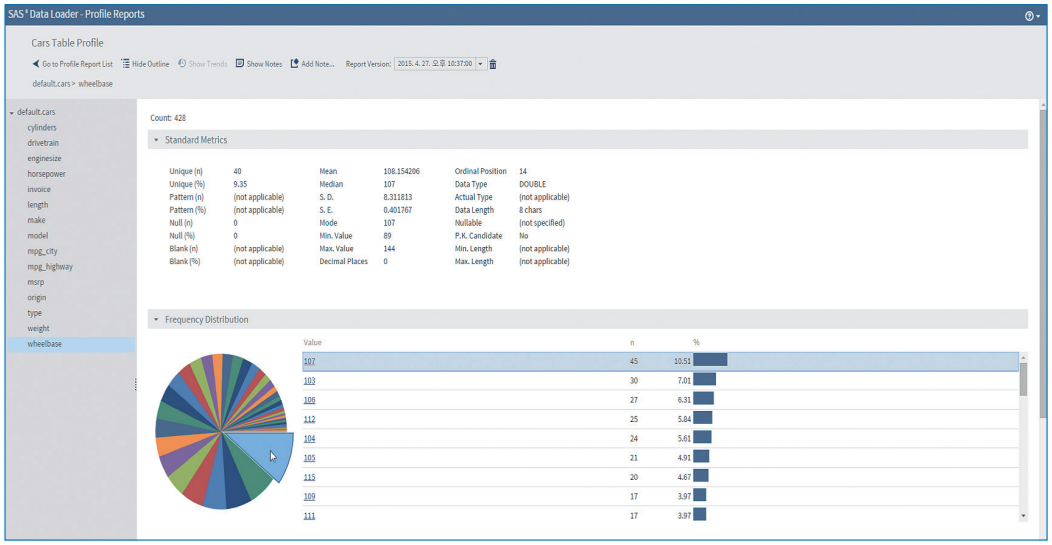
SAS Data Loader for Hadoop로 Hadoop 에코시스템 내에서 분석 프로세싱을 실행할 수 있으므로 기존의 솔루션보다 더 경제적이고 신속하게 작업 결과물을 얻을 수 있습니다. 또한 데이터 이동을 최소화하고 병렬 처리 방식을 지원하므로 사용자는 확장성과 성능 향상에 따른 이점을 동시에 누릴 수 있습니다.

- 간이 SAS 실행 엔진인 SAS Embedded Process로 Hadoop에서 DS2 언어를 사용하는 SAS 프로그램 실행

데이터 프로파일링 및 프로파일 보고서 저장

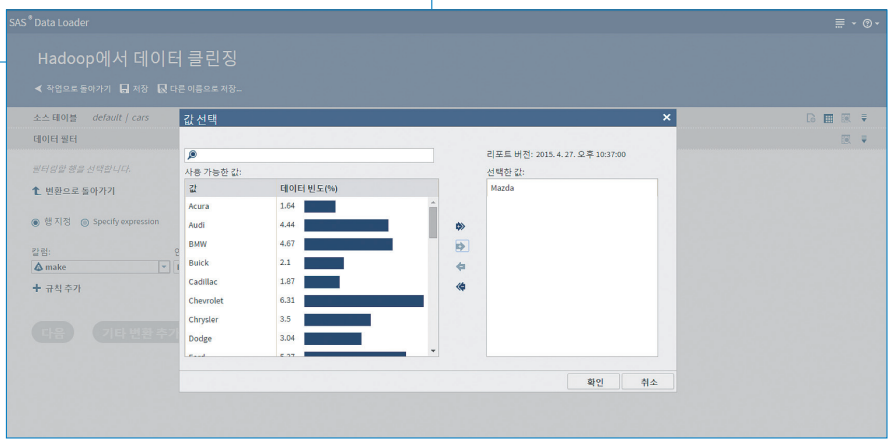
데이터의 구조나 컨텐츠에 대한 프로파일 보고서를 생성합니다.

- 하나 이상의 테이블에서 소스 열 선택 후 데이터의 고유 특성, 불완전성 및 패턴 파악
- 프로파일 데이터 기능으로 생성된 보고서의 목록 표시 및 특정 보고서 확인
- 노트 생성 및 저장



프로파일링 프로세스가 Hadoop 클러스터 내부에서 처리되어 성능을 향상시킵니다.

Hadoop 환경에서 데이터 파싱 및 표준화 등 데이터 품질 기능을 사용할 수 있습니다.



데이터 프로파일링 피드백을 통해 데이터 품질 루틴의 효율을 높일 수 있습니다.

간편한 사용자 인터페이스: 마법사 기능을 통한 작업 관리 및 재사용

SAS Data Loader for Hadoop은 비즈니스 사용자를 위해 특별히 설계된 솔루션입니다. 마법사 기능을 지원하는 손쉬운 인터페이스가 Hadoop에 저장된 데이터를 손쉽게 액세스·관리할 수 있게 해주므로, 사용자는 IT 부서에 의존할 필요가 없고 Hadoop 전문 인력을 고용해야 하는 부담도 덜 수 있습니다.

- 작업 로그의 목록 및 상태 확인
- 작업 시작/중지, 해당 로그 및 생성 코드 파일 확인
- 재사용을 위해 저장된 작업의 실행, 보기 또는 편집

인-메모리 분석 서버

리포팅, 시각화 또는 분석에 필요한 데이터를 얻기 위해 IT 부서에 도움을 요청하고 기다릴 필요가 없습니다. SAS Data Loader for Hadoop이 SAS® LASR™ Analytic Server 메모리에 데이터를 로드하여 SAS Visual Analytics에 사용할 데이터를 준비할 수 있게 해줍니다.

- SAS Visual Analytics 또는 SAS Visual Statistics로 메모리에 저장되어 있는 특정 Hadoop 칼럼을 SAS LASR Analytic Server에 로드 및 분석 수행

안전한 빅 데이터 액세스

- Kerberos 프로토콜이 지원되는 Hadoop 클러스터에 안전하게 액세스

cloudera CERTIFIED

SAS Data Loader for Hadoop은 Cloudera 및 Hortonworks Hadoop 배포판을 지원합니다.

SAS Data Loader for Hadoop에 관한 더 자세한 내용은 www.sas.com/korea/data-loader-hadoop 페이지를 참조하십시오.

시스템 사양

호스트 플랫폼 및 서버 환경

- HP/UX on Itanium : 11iv3 (11.31)
- IBM AIX R64 on POWER architecture 7.1
- IBM z/OS : V1R11 and higher
- Linux x64 (64- bit) : Novell SuSE 11 SP1; Red Hat Enterprise Linux 6.1; Oracle Linux 6.1
- Microsoft Windows on x 64 (64-bit):
  - Desktop: Windows 7*x 64 SP1; Windows 8**x 64
  - Server : Windows Server 2008 x 64 SP2 Family; Windows Server 2008 R2 SP1 Family; Windows Server 2012 Family
- Solaris on SPARC : Version 10 Update 9
- Solaris on x 64 (x 64-86) : Version 10 Update 9; Version 11

클라이언트 환경

- Microsoft Windows (32- bit) : Windows 7*x 86-64; Windows 8**x 86-64)
- Microsoft Windows (64- bit) : Windows 7*x 64 SP1; Windows 8**x 64)

미드-티어

- HP/UX on Itanium
- IBM AIX on POWER
- Linux x 64 (x 86-64)
- Microsoft Windows x 64 (x 86-64)
- Solaris (SPARC and x 64)

지원 브라우저

- Internet Explorer 9 : Windows 7 (32- bit and 64- bit web browsers)
- Internet Explorer 10 : Windows 7 and Windows 8 (32- bit and 64-bit web browsers)
- Internet Explorer 11 : Windows 7 and Windows 8 (32- bit and 64-bit web browsers)
- Firefox 6 and up : Windows 7 and Windows 8 (32- bit and 64-bit web browsers); Linux x 64 : RHEL 6 and SLES 11 (32- bit web browsers)
- Chrome 15 and up: Windows 7 and Windows 8 (32- bit and 64-bit web browsers); Linux x 64 : RHEL 6.1 and SLES 11 SP 1 (32- bit web browsers)
- You can also review the Third- Party Support page for details about requirements of third- party software for use with SAS 9.4)

* NOTE : Windows 7 supported editions are : Professional, Ultimate and Enterprise.

**NOTE : Supported editions include : Windows 8, Windows 8 Pro, Windows 8 Enterprise.

한국새스소프트웨어(유) 서울 강남구 테헤란로 408 (대치동, 대치빌딩 8~10층) (우 135-839)
SAS 및 기타 모든 SAS Institute Inc.의 제품 또는 서비스명은 미국 및 다른 국가에 있는 SAS Institute Inc.의 등록 상표 또는 상표입니다. ®은 미국에 등록되어 있음을 나타냅니다. 그 밖의 상표 및 제품명은 해당 기업의 등록 상표입니다. Copyright©2014, SAS Institute Inc. All rights reserved. 107223_S118780.0714

www.sas.com/korea

THE POWER TO KNOW.