# SAS® Data Loader 2.4 for Hadoop

SAS Data Loader 2.4 for Hadoop consists of three delivery options:

- SAS Data Loader 2.4 for Hadoop
- SAS Data Loader 2.4 for Hadoop, Cloudera
- SAS Data Loader 2.4 for Hadoop, Hortonworks

SAS Data Loader 2.4 for Hadoop consists of two main areas for system requirements:

- SAS In-Database Technologies for Hadoop
- SAS Data Loader for Hadoop vApp

*Note:* *SAS In-Database Technologies for Hadoop must be installed before the installation of the SAS Data Loader for Hadoop vApp in order for the vApp to communicate successfully with the Hadoop cluster.*

## Hadoop Environment

System requirements for the Hadoop environment are as follows:

- Hadoop distribution:
  - Cloudera - Kerberos and LDAP are supported.
    **Note:** *SAS Data Loader for Hadoop, Cloudera requires the Cloudera manager for installation.*
    - Cloudera CDH 5.2 or 5.3 or 5.4
    - Cloudera 5.3 with optional Apache Spark 1.2 and optional Impala 2.1
    - Cloudera 5.4 with optional Apache Spark 1.3 and optional Impala 2.2

  - Hortonworks  - Kerberos is supported.
    **Note:** *SAS Data Loader for Hadoop, Hortonworks requires Ambari for installation.*
    - Hortonworks HDP 2.1 or 2.2 or 2.3
    - Hortonworks 2.2 with optional Apache Spark 1.2
    - Hortonworks 2.3 with optional Apache Spark 1.3
  - MapR  - Kerberos is not supported.
    - MapR 4.0.2  or 4.1 or 5.0
    - MapR 4.1 with optional Apache Spark 1.2 and optional Impala 1.4

- MapR 5.0 with optional Apache Spark 1.3 and optional Impala 1.4
  - o Pivotal HD - Kerberos is supported.
    - Pivotal HD 3.0
    - Pivotal HD 3.0 with optional Apache Spark 1.2
  - o IBM BigInsights - Kerberos is not supported.
    - IBM BigInsights 4.0 or 4.1
    - IBM BigInsights 4.0 with optional Apache Spark 1.2
    - IBM BigInsights 4.1 with optional Apache Spark 1.2

For information about Hadoop distribution requirements, see the "SAS 9.4 Support for Hadoop" web page at:
http://support.sas.com/resources/thirdpartysupport/v94/hadoop/

- MapReduce2. (MapReduce 1 is not supported.)
- Hive 13 (0.13.0) or Hive 14 (0.14.0) or higher
- SQOOP and OOZIE must be enabled in your Hadoop cluster in order to communicate with the DBMS to which SAS Data Loader for Hadoop users need access.

  *Note:Ensure that Oozie 4.0 or higher is installed.*
- For SAS Data Loader 2.4 for Hadoop, Cloudera and SAS Data Loader 2.4, Hortonworks: JRE Version 1.7 or higher.
- JDBC drivers required by the DBMS must be installed on the Hadoop cluster. The JDBC drivers are for the DBMS to which SAS Data Loader for Hadoop users need access.

  *Note:    The version of JDBC driver must be supported by Sqoop.*

*Note:    If the Hadoop cluster is secured using Kerberos, then the vApp must be configured for Kerberos.*

For more information about configuring Hadoop cluster components, refer to the *SAS 9.4 In-Database Products: Administrator's Guide.*

## SAS 9.4 M3 Install for SAS In-Database Technologies for Hadoop

### Operating System

AIX, HP IPF, Linux for x64,Oracle Solaris SPARC, Oracle Solaris for x64, Windows 32-bit server or workstation, and Windows for x64 server or workstation.

# SAS Data Loader for Hadoop vApp

## vApp Client Host Environment

The client host for SAS Data Loader 2.4 for Hadoop must meet the following hardware and software requirements:

*Note:    If the Hadoop cluster is secured using Kerberos, then the vApp must be configured for Kerberos. For details, see the* **SAS Data Loader 2.4 for Hadoop: vAPP Deployment Guide**.

29 October 2015

## Software

SAS Data Loader 2.4 for Hadoop.

## Hardware

- System memory (RAM): Minimum 8 GB, but 16 gigabytes or more is recommended. The SAS Data Loader virtual machine requires at least 4 GB of available memory.
- Disk space: At least 30 GB of free hard-drive space.
- BIOS: Must be enabled for virtualization technology.
- Processors: At least 2 cores (4 logical processors), but 4 core (8 local processors) or more is recommended.

## Operating System

64-bit Windows 7 or later, or Windows Server 2008 R2 or later.

## Hypervisor

- For Windows hosts only:
  - VMware Player Pro 7
  - VMware Workstation 12 andVMware Workstation 12 Pro
  - Oracle VM VirtualBox 5

  *Note:* *Oracle VM VirtualBox 5 is not supported for SAS Data Loader for Hadoop, Cloudera and SAS Data Loader for Hadoop, Hortonworks.*

- Virtual Network Configuration Network Address Translation (NAT)

## Web Browsers

- Web browsers, without Kerberos security: Windows Internet Explorer 10 or later, Mozilla Firefox 14 or later, or Google Chrome 21 or later.
- Web browsers, with Kerberos security:
  - MIT client:
    - Mozilla Firefox 14 or later
  - Active Directory with MIT:
    - Google Chrome 21 or later
    - Mozilla Firefox 14 or later