Running Hadoop on Windows

[1 Executive Summary 2](#_Toc504061727)

[1.1 Overview 2](#_Toc504061728)

[1.2 Target Customers/Areas 2](#_Toc504061729)

[1.3 Reusable Characteristics 2](#_Toc504061730)

[2 High Level System Use Cases 3](#_Toc504061731)

[3 Solution Architecture 3](#_Toc504061732)

[3.1 Core Solution Components and its functions 3](#_Toc504061733)

[3.2 Deployment Architecture 3](#_Toc504061734)

[4 Installation & Configuration Procedures 3](#_Toc504061735)

# Executive Summary

This document details the solution outline of running Hadoop on Windows machines. The solution is described in such a manner and to a level of detail that any project stakeholders interested in this solution will be able to understand the solution characteristics and its reusability in other applicable areas. This document will explain the system behavior of the Component, its reusable characteristics, processing functions and outputs.

## Overview

Hadoop and the components of its ecosystem are designed to run on Linux. It is seen that most Hadoop components fail to run on windows environment as they require a running Hadoop environment which is difficult to run in windows. A vendor Horton works provides windows versions but they are for enterprise use and come with a cost. Also the usage is quite difficult. Most Ericsson users have Windows installed. This component overcomes this challenge and all big data developers can successfully run Hadoop and all components of its ecosystem by this solution.

Scope

All Hadoop eco system projects need Hadoop to run on windows for developers to debug and write code. The component solves this problem.

## Target Customers/Areas

Developers working in all big data projects for all customer projects will be benefitted from this Hadoop. This reduces a lot of time and cost. They need to set this in their environment variable and can run Flume, Spark, Hadoop and other related projects on their Windows machine.

## Reusable Characteristics

The jar can just be downloaded and unpackaged and used.

# High Level System Use Cases

1.Running Spark on Windows.

2.Running Flume on Windows.

3. Running Kafka on Windows.

4.Running Flink on Windows.

5. Running Hadoop Map-Reduce on Windows.

# Solution Architecture

## Core Solution Components and its functions

The main component is a rar file namely **hadoop-2.6.0.rar** This just needs to be unzipped and the path upto bin needed to be added in environment variable. This will enable the Windows user to debug and run Hadoop in Windows and also run programs in eclipse.

For Eg: If this is in C:\Hadoop then :

Path to Environment variable is: **C:\Hadoop\hadoop-2.6.0\bin**

Also any version from apache distribution can be downloaded and built for windows use.

## Deployment Architecture

Download the rar and set the path upto bin in environment variable for the already built package.

# Installation & Configuration Procedures

**5.1 Installation Procedure**

1. Download the **hadoop-2.6.0.rar**
2. Unzip it in a folder in the local machine.
3. Right click on My Computer Icon and click on Properties.
4. Click on Advanced system settings.
5. Click on Environment variables.
6. Select the PATH variable under System variables.
7. Add the path of the unzipped **hadoop-2.6.0.rar** upto bin folder in the PATH variable.
8. Apply and Save.

**5.2 Building from an apache distribution**

**Prerequisites:**

1. Java Version 1.7 or higher need to be set.

2) Download the latest Hadoop release from:

[ASF Hadoop download page](http://mirrors.ibiblio.org/apache/hadoop/common/)

1. Maven 3.0 or later
2. ProtocolBuffer 2.5.0
3. CMake 2.6 or newer
4. Windows SDK ([**http://www.microsoft.com/en-us/download/details.aspx?id=8279**](http://www.microsoft.com/en-us/download/details.aspx?id=8279)) download link.

**Building Procedure:**

Keep the source code tree in a short path to avoid running into problems related

to Windows maximum path length limitation. (For example, C:\hdc).

Run builds from a Windows SDK Command Prompt. (Start, All Programs,

Microsoft Windows SDK v7.1, Windows SDK 7.1 Command Prompt.)

The Platform environment variable must be set to x64.

It must be "**Platform**", not "PLATFORM" or "platform".

Environment variables on Windows are usually case-insensitive, but Maven treats them as case-sensitive. Failure to set this environment variable correctly will cause msbuild to fail while building the native code in hadoop-common.

**set Platform=x64**

All Maven goals are the same as described above with the exception that

native code is built by enabling the 'native-win' Maven profile. -Pnative-win

is enabled by default when building on Windows since the native components

are required (not optional) on Windows.

To build a binary distribution run the following command from the root of the source tree.

**mvn package -Pdist,native-win -DskipTests -Dtar**

(Note that this command must be run from a Windows SDK command prompt as documented in BUILDING.txt. A successful build generates a binary hadoop .tar.gz package in hadoop-dist\target\. )