

## 1. Introduction

- 2. Prerequistes
- 3. Set Environment Variables
- 4. Setup SSH daemon
- 5. <u>Download hadoop and place it</u> in the home directory.
- 6. <u>Unpack hadoop</u>
- 7. Configure Hadoop
- 8. Format the namenode
- 9. Setup hadoop plugin
- 10. Start the cluster
- 11. Setup hadoop location
- 12. Upload data
- 13. Create and run a test project.

```
🟮 add this 📲 😭 🏘 ...)
```

## **Prerequisites**

Before we begin, make sure the following components are installed on your workstation:

- Java 1.6
- Eclipse Europa 3.3.2

This tutorial has been written for and tested with Hadoop version **0.19.1**. If you are using another version, some things may not work.

Make sure you have exactly the same versions of the software as shown above. Hadoop will not work with versions of Java earlier than 1.6 and versions of Eclipse later than 3.3.2 due to plug-in API incompatibility.

## **Installing Cygwin**

After installing the prerequisite software, the next step is to install the <u>Cygwin</u> environment. Cygwin is a set of Unix packages ported to Microsoft Windows. It is needed to run the scripts supplied with Hadoop because they are all written for the Unix platform.

To install the cygwin environment follow these steps:

- 1. Download cygwin installer from <a href="http://www.cygwin.com">http://www.cygwin.com</a>.
- 2. Run the downloaded file. You will see the window shown on the screenshots below.





3. When you see the above screenshot, keep pressing the 'Next' button until you see the package selection screen shown below. Make sure you select 'openssh'. This package is required for the correct functioning of the Hadoop cluster and Eclipse plug-in.

Category	Current	New	B	S	Size	Package
Category	Conoric	A Skip	n/a	n/a	57k	libwrap-devel: (devel) provides host-based access restric
	7.6-4	Кеер	n/a		12k	libwrap0: (runtime) provides host-based access restriction
		🚯 Skip	n/a	n/a	270k	lighttpd: A light-weight and flexible webserver
		🚯 Skip	n/a	n/a	337k	naim: Console AIM, ICQ, IRC, and Lily client
		🚯 Skip	n/a	n/a	234k	noftp: An improved FTP client
		🚯 Skip	n/a	n/a	50k	netcat: A simple but powerful network tool
		🚯 Skip	n/a	n/a	105k	nfs-server: Universal NFS server.
		🚯 Skip	n/a	n/a	24k	nrss: A nourses-based RSS reader
		🚯 Skip	n/a	n/a	22k	nttcp: New test TCP program
		🚯 Skip	n/a	n/a	1,038k	openIdap: Lightweight Directory Access Protocol suite
		let Skip	n/a	n/a	704k	openIdap-devel: Lightweight Directory Access Protocol :
	5.0p1-1	€ 5.1p1-10	$\times$		580k	openssh: The OpenSSH server and client programs
1	0.9.8g-1	😯 0.9.8j-1	$\times$		983k	openssl: The OpenSSL runtime environment
		🚯 Skip	n/a	n/a	554k	openssl097: The OpenSSL 0.9.7 runtime environment
		Skip	rýs.	Ty's	Ok.	ping: A basic network tool to test IP network conectivity
		🚯 Skip	n/a	n/a	211k	planet: Flexible RDF, RSS and Atom feed aggregator
		0.011		•		0003-0003-0-1-
	ulete packages					
🐉 start	📄 🗁 cygv	vin	Cygwir	n Setu	ıp - Selec	
4 After	you seled	cted these pac	kages	pres	ss the 'N	lext' button to complete the
insta	11711111111					
insta	nacion.					

2008 - Vlad Korolev