

Draw flow charts and solution diagrams online using Kontext Diagrams. It is free!

[Learn more](#)

24



# Install Hadoop 3.3.0 on Windows 10 using WSL



Raymond 👁 14,057 💬 4 ⌚ 2 years ago



Hadoop 3.3.0 was released on July 14 2020. It is the first release of Apache Hadoop 3.3 line. There are significant changes compared with Hadoop 3.2.0, such as Java 11 runtime support, protobuf upgrade to 3.7.1, scheduling of opportunistic containers, non-volatile SCM support in HDFS cache directives, etc.

This article provides step-by-step guidance to install Hadoop 3.3.0 on Windows 10 via WSL (Windows Subsystem for Linux). These instructions are also be applied to Linux systems to install Hadoop. Most of the content is based on article [Install Hadoop 3.2.0 on Windows 10 using Windows Subsystem for Linux \(WSL\)](#).

## Prerequisites

Follow the page below to enable WSL and then install one of the Linux systems from Microsoft Store.

[Windows Subsystem for Linux Installation Guide for Windows 10](#)



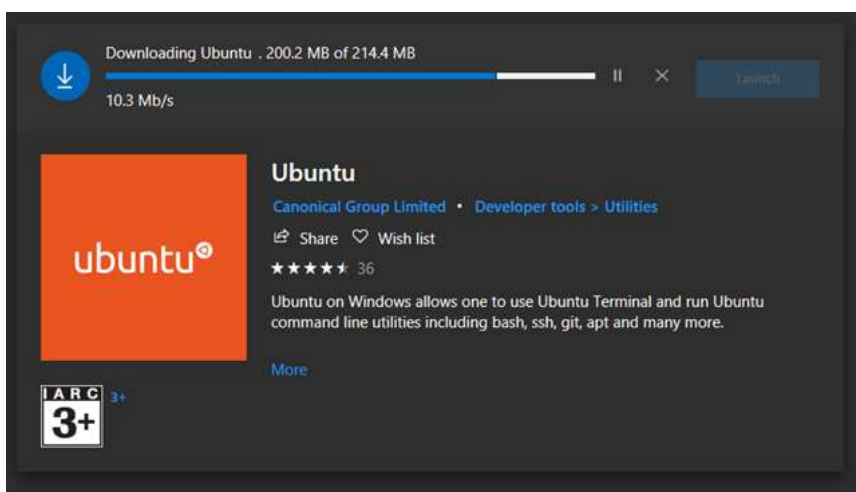
## Direct Cloud Connection

### Sign up

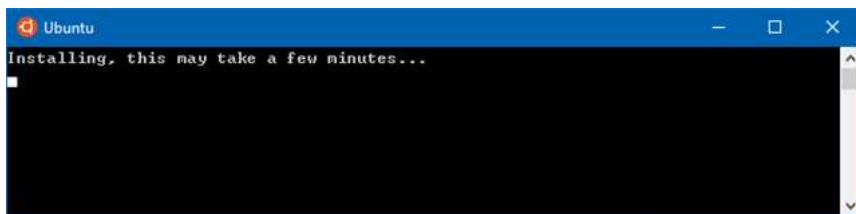
To be specific, enable WSL by running the following PowerShell code as Administrator (or enable it through Control Panel):

By using this site, you acknowledge that you have read and understand our [Cookie policy](#), [Privacy policy](#) and [Terms](#).

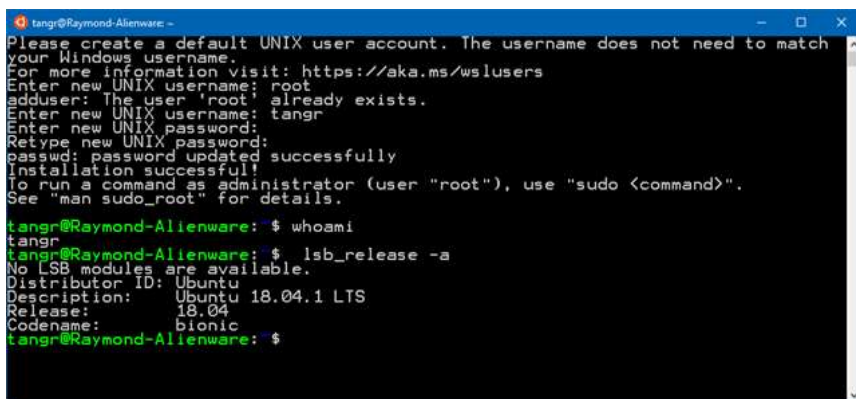
[Accept](#)



Once download is completed, click Launch button to launch the application. It may take a few minutes to install:



During the installation, you need to input a username and password. Once it is done, you are ready to use the Ubuntu terminal:



## Install Java JDK

Run the following command to update package index:

By using this site, you acknowledge that you have read and understand our [Cookie policy](#), [Privacy policy](#) and [Terms](#). Accept

```
java -version
```

Command 'java' not found, but can be installed with:

```
sudo apt install default-jre
sudo apt install openjdk-11-jre-headless
sudo apt install openjdk-8-jre-headless
```

Install OpenJDK via the following command:

```
sudo apt-get install openjdk-8-jdk
```

Check the version installed:

```
java -version
openjdk version "1.8.0_191"
OpenJDK Runtime Environment (build 1.8.0_191-8u191-b12-2ubuntu0.18.04.1-b12)
OpenJDK 64-Bit Server VM (build 25.191-b12, mixed mode)
```

You can also use Java 11 from this version as it is now supported.

## Download Hadoop binary

Go to release page of Hadoop website to find a download URL for Hadoop 3.3.0:

### Hadoop Releases

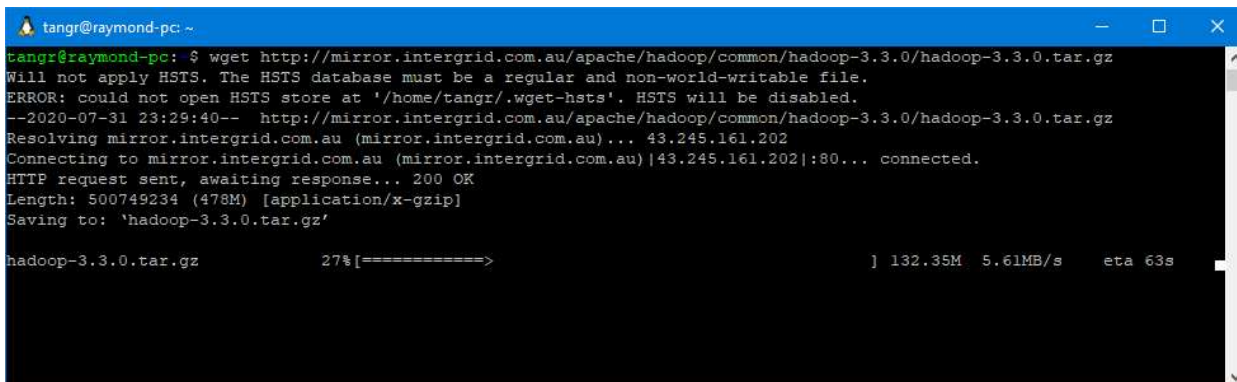
For me, the closest mirror is:

<http://mirror.intergrid.com.au/apache/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz>

Run the following command in Ubuntu terminal to download a binary from the internet:

```
wget http://mirror.intergrid.com.au/apache/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz
```

Wait until the download is completed:



```
tangr@raymond-pc: ~
tangr@raymond-pc: $ wget http://mirror.intergrid.com.au/apache/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz
Will not apply HSTS. The HSTS database must be a regular and non-world-writable file.
ERROR: could not open HSTS store at '/home/tangr/.wget-hsts'. HSTS will be disabled.
--2020-07-31 23:29:40--  http://mirror.intergrid.com.au/apache/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz
Resolving mirror.intergrid.com.au (mirror.intergrid.com.au)... 43.245.161.202
Connecting to mirror.intergrid.com.au (mirror.intergrid.com.au)|43.245.161.202|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 500749234 (478M) [application/x-gzip]
Saving to: 'hadoop-3.3.0.tar.gz'

hadoop-3.3.0.tar.gz      27%[=====>] 132.35M  5.61MB/s  eta 63s
```

## Unzip Hadoop binary

Run the following command to create a **hadoop** folder under user home folder:

```
mkdir ~/hadoop
```

And then run the following command to unzip the binary package:

```
tar -xvzf hadoop-3.3.0.tar.gz -C ~/hadoop
```

Once it is unpacked, change the current directory to the Hadoop folder:

```
cd ~/hadoop/hadoop-3.3.0/
```

## Configure hadoop-env.sh

By using this site, you acknowledge that you have read and understand our [Cookie policy](#), [Privacy policy](#) and [Terms](#).

Accept

ssh localhost

If you cannot ssh to localhost without a passphrase, run the following command to initialize your private and public keys:

```
ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
chmod 0600 ~/.ssh/authorized_keys
```

If you encounter errors like 'ssh: connect to host localhost port 22: Connection refused', run the following commands:

```
sudo apt-get install ssh
And then restart the service:
sudo service ssh restart
```

If the above commands still don't work, try the solution in [this comment](#).

\*The comment link will redirect you to another article for a different version of Hadoop installation.

## Configure the pseudo-distributed mode (Single-node mode)

Now, we can follow the official guide to configure a single node:

### Pseudo-Distributed Operation

#### 1) Setup environment variables (optional)

Setup environment variables by editing file `~/.bashrc`.

```
vi ~/.bashrc
```

Add the following environment variables:

```
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64
export HADOOP_HOME=~/.hadoop/hadoop-3.3.0
export PATH=$PATH:$HADOOP_HOME/bin
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
```

Run the following command to source the latest variables:

```
source ~/.bashrc
```

#### 2) Edit `etc/hadoop/hadoop-env.sh` file:

```
vi etc/hadoop/hadoop-env.sh
```

Set a JAVA\_HOME environment variable:

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

#### 3) Edit `etc/hadoop/core-site.xml`:

```
vi etc/hadoop/core-site.xml
```

Add the following configuration:

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

#### 4) Edit `etc/hadoop/hdfs-site.xml`:

```
vi etc/hadoop/hdfs-site.xml
```

By using this site, you acknowledge that you have read and understand our [Cookie policy](#), [Privacy policy](#) and [Terms](#).

Accept

```
<name>dfs.replication</name>
<value>1</value>
</property>
</configuration>
```

5) Edit file **etc/hadoop/mapred-site.xml**:

```
vi etc/hadoop/mapred-site.xml
```

Add the following configuration:

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>mapreduce.application.classpath</name>
    <value>${HADOOP_MAPRED_HOME}/share/hadoop/mapreduce/*:${HADOOP_MAPRED_HOME}/share/hadoop/mapreduce/lib/*</value>
  </property>
</configuration>
```

6) Edit file **etc/hadoop/yarn-site.xml**:

```
vi etc/hadoop/yarn-site.xml
```

Add the following configuration:

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.env-whitelist</name>

<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_M
APRED_HOME</value>
  </property>
</configuration>
```

## Format namenode

Run the following command to format the name node:

```
bin/hdfs namenode -format
```

## Run DFS daemons

1) Run the following commands to start NameNode and DataNode daemons:

```
tangr@raymond-pc:~/hadoop/hadoop-3.3.0$ sbin/start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [raymond-pc]
```

2) Check status via **jps** command:

```
tangr@raymond-pc:~/hadoop/hadoop-3.3.0$ jps
2212 NameNode
2423 DataNode
2682 SecondaryNameNode
2829 Jps
```

When the services are initiated successfully, you should be able to see these four processes.

By using this site, you acknowledge that you have read and understand our [Cookie policy](#), [Privacy policy](#) and [Terms](#) .

The web UI looks like the following:

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

## Overview 'localhost:9000' (active)

|                |  |
|----------------|--|
| Started:       | Fri Jul 31 23:53:37 +1000 2020                             |
| Version:       | 3.3.0, raa96f1871bfd858f9bac59cf2a81ec470da649af           |
| Compiled:      | Tue Jul 07 04:44:00 +1000 2020 by brahma from branch-3.3.0 |
| Cluster ID:    | CID-705c34e6-0a2a-40a2-8ebd-aaa78759a77f                   |
| Block Pool ID: | BP-1339971671-127.0.0.1-1596203598353                      |

## Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 131.25 MB of 446 MB Heap Memory. Max Heap Memory is 3.53 GB.

Non Heap Memory used 48.43 MB of 49.94 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

You can also view the data nodes information through menu link **Datanodes**:

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

## Datanode Information

In service

Down

Decommissioning

Decommissioned

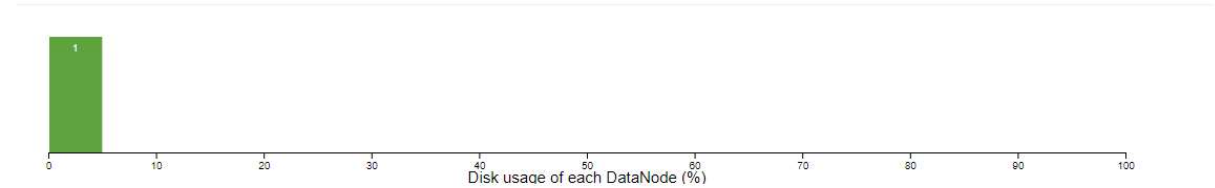
Decommissioned & dead

Entering Maintenance

In Maintenance

In Maintenance & dead

### Datanode usage histogram



### In operation

DataNode State

All

Show 25 entries

Search:

| Node  | Http Address          | Last contact | Last Block Report | Used | Non DFS Used | Capacity  | Blocks | Block pool used | Version |
|---|-----------------------|--------------|-------------------|------|--------------|-----------|--------|-----------------|---------|
| <div><div></div>localhost:9866<br/>(127.0.0.1:9866)</div> | http://localhost:9864 | 1s           | 3m                | 0 B  | 158.04 GB    | 331.39 GB | 0      | 0 B (0%)        | 3.3.0   |

Showing 1 to 1 of 1 entries

Previous1Next

## Run YARN daemon

1) Run the following command to start YARN daemon:

```
sbin/start-yarn.sh
```

By using this site, you acknowledge that you have read and understand our [Cookie policy](#), [Privacy policy](#) and [Terms](#).

Accept

WARNING: YARN\_CONF\_DIR has been replaced by HADOOP\_CONF\_DIR. Using value of YARN\_CONF\_DIR.

## 2) Check status via `jps` command

```
tangr@raymond-pc:~/hadoop/hadoop-3.3.0$ jps
2212 NameNode
5189 NodeManager
2423 DataNode
5560 Jps
5001 ResourceManager
2682 SecondaryNameNode
```

Once the services are started, you can see two more processes for **NodeManager** and **ResourceManager**.

## 3) View YARN web portal

You can view the YARN resource manager web UI through the following URL:

<http://localhost:8088/cluster>

The web UI looks like the following:

**Nodes of the cluster**

Cluster Metrics

| Apps Submitted | Apps Pending | Apps Running | Apps Completed | Containers Running | Memory Used | Memory Total | Memory Reserved | VCoers Used | VCoers Total | VCoers Reserved |
|----------------|--------------|--------------|----------------|--------------------|-------------|--------------|-----------------|-------------|--------------|-----------------|
| 0              | 0            | 0            | 0              | 0                  | 0 B         | 8 GB         | 0 B             | 0           | 8            | 0               |

Cluster Nodes Metrics

| Active Nodes | Decommissioning Nodes | Decommissioned Nodes | Lost Nodes | Unhealthy Nodes | Rebooted Nodes | Shutdown Nodes |
|--------------|-----------------------|----------------------|------------|-----------------|----------------|----------------|
| 1            | 0                     | 0                    | 0          | 0               | 0              | 0              |

Scheduler Metrics

| Scheduler Type     | Scheduling Resource Type      | Minimum Allocation      | Maximum Allocation      | Maximum Cluster Application Priority |
|--------------------|-------------------------------|-------------------------|-------------------------|--------------------------------------|
| Capacity Scheduler | [memory-mb (unit=Mi), vcores] | <memory:1024, vCores:1> | <memory:8192, vCores:4> | 0                                    |

Showing 1 to 1 of 1 entries

| Node Labels   | Rack    | Node State      | Node Address   | Node HTTP Address              | Last health-update | Health-report | Containers | Allocation Tags | Mem Used | Mem Avail | VCoers Used | VCoers Avail | Version |
|---------------|---------|-----------------|----------------|--------------------------------|--------------------|---------------|------------|-----------------|----------|-----------|-------------|--------------|---------|
| /default-rack | RUNNING | localhost:11507 | localhost:8042 | Sat Aug 01 00:02:34 +1000 2020 | 0                  | 0 B           | 8 GB       | 0               | 8        | 3.3.0     |             |              |         |

Showing 1 to 1 of 1 entries

You can view all the applications through this web portal.

## Shutdown services

Once you've completed explorations, you can use the following command to shutdown those daemons:

```
sbin/stop-yarn.sh
sbin/stop-dfs.sh
```

You can verify through `jps` command which will only show one process now:

```
tangr@raymond-pc:~/hadoop/hadoop-3.3.0$ jps
6593 Jps
```

## Summary

Congratulations! Now you have successfully installed a single node Hadoop 3.3.0 cluster in your Ubuntu subsystem of Windows 10. It's relatively easier compared with native Windows installation as we don't need to download or build native Hadoop HDFS libraries.

Have fun with Hadoop 3.3.0.

If you encounter any issues, please post a comment and I will try my best to help.

Last modified by Raymond 2 years ago

© This page is subject to Site terms.

Related series

# Hadoop on WSL

Big Data Tools on WSL

|   |        |
|---|--------|
| Sqoop Installation on Windows 10 using Windows Subsystem for Linux              | 3,141  |
| Apache Spark 2.4.3 Installation on Windows 10 using Windows Subsystem for Linux | 13,905 |
| Install Zeppelin 0.7.3 on Windows 10 using Windows Subsystem for Linux (WSL)    | 6,960  |
| Apache Hive 3.1.1 Installation on Windows 10 using Windows Subsystem for Linux  | 9,177  |
| Install Windows Subsystem for Linux on a Non-System Drive                       | 39,112 |
| Install Hadoop 3.2.0 on Windows 10 using Windows Subsystem for Linux (WSL)      | 29,319 |



Please log in or register to comment.

Log in

Register

Log in with external accounts

Log in with Microsoft account

Log in with Google account

4 comments

Stats

Page index 25.54

Daily views



linux

hadoop

WSL

big-data-on-wsl

Follow Kontext on LinkedIn



# Innovative Launch Support

Learn More

## More from Kontext

### Install Hadoop 3.3.0 on macOS

4,364 views 1 like 2 years ago

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

hadoop

macos

Hadoop

### How to upgrade from WSL 1 to WSL 2

12 days ago

WSL

linux

Tools

### Apache Spark 3.0.1 Installation on WSL

3,393 views 0 likes 2 years ago

spark

linux

WSL

Spark

Kontext beta

English

Settings

© Kontext 2022 v1.0.0-beta - build-20220213.1

Powered by .NET 6.0.1 on Azure.  
Made with ❤️ in Melbourne.

Features

Resources

By using this site, you acknowledge that you have read and understand our [Cookie policy](#), [Privacy policy](#) and [Terms](#).

Accept

https://kontext.tech/column/hadoop/445/install-hadoop-330-on-windows-10-using-wsl

9/10

About

- Cookie
- Privacy
- Terms
- Contact us

Featured columns

- Spark
- Hadoop
- Code Snippets & Tips
- .NET
- ASP.NET Core
- Tools
- Python Programming
- Java Programming
- Teradata
- Azure

Subscription

Subscribe to Kontext newsletter to get updates about data analytics, programming and cloud related articles.

Email

Subscribe