

DSBDA_Lab No :- 13

Steps to Install Apache Hadoop 2.7.0 Single Node Cluster on Ubuntu 16.04 are as follows:

**Go to the FTP → Download the Hadoop and Word Doc file —> Paste on Home Directory
→ Hadoop folder winzip Home location → open terminal → write on terminal hadoop
version → if yes then go to Step 3.) spark- install → if no —>**

hadoop version

```
Hadoop 2.7.0
Subversion https://git-wip-us.apache.org/repos/asf/hadoop.git -r
d4c8d4d4d203c934e8074b31289a28724c0842cf
Compiled by jenkins on 2015-04-10T18:40Z
Compiled with protoc 2.5.0
From source with checksum a9e90912c37a35c3195d23951fd18f
This command was run using /usr/local/hadoop/hadoop-
2.7.0/share/hadoop/common/hadoop-common-2.7.0.jar
```

jps

```
8361 Jps
```

sudo apt-get update

```
[sudo] password for student: 123
```

```
Hit:7 https://repo.mongodb.org/apt/ubuntu xenial/mongodb-org/4.4 Release
Reading package lists... Done
```

sudo apt-get install openssh-server

sudo cp /etc/ssh/sshd_config /etc/ssh/sshd_config.factory-defaults

sudo gedit /etc/ssh/sshd_config

49 Change LogLevel INFO to LogLevel **VERBOSE**

Add the following Lines to the End:

```
AllowUsers student
PermitRootLogin no
PubkeyAuthentication yes
```

Save and Restart SSH.

sudo systemctl restart ssh

ssh-keygen -t rsa -P ""

Press Enter

Type y

Press Enter

Out Put

Generating public/private rsa key pair.

Enter file in which to save the key (/home/student/.ssh/id_rsa):

/home/student/.ssh/id_rsa already exists.

Overwrite (y/n)? y

Your identification has been saved in /home/student/.ssh/id_rsa.

Your public key has been saved in /home/student/.ssh/id_rsa.pub.

The key fingerprint is:

SHA256:MYUjT0BE1zLQN5rYNAMLzjDuGmOLt5pETZ42NgSYlvI student@student-OptiPlex-3020

The key's randomart image is:

+---[RSA 2048]-----+

|o+ .+B=.o. |

|++.= ..o@.+ |

|o..oo .BoX . |

|.E . . =o |

|oo.O S |

|+++ o |

|oo. |

|o . |

|o.. |

+----[SHA256]-----+

cat \$HOME/.ssh/id_rsa.pub >> \$HOME/.ssh/authorized_keys

ssh localhost

Step 1: Adding Jave Repository and Installing Oracle Java 8

add-apt-repository ppa:webupd8team/java

apt-get update

apt-get install oracle-java8-installer

sudo chown -R student:student /usr/local/hadoop

readlink -f /usr/bin/java

vim .bashrc

Press INSERT Button

Copy & Paste below line (end of the screen)

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

```
export HADOOP_INSTALL=/usr/local/hadoop/hadoop-2.7.0
export PATH=$PATH:$HADOOP_INSTALL/bin
export PATH=$PATH:$HADOOP_INSTALL/sbin
export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_HOME=$HADOOP_INSTALL
export HADOOP_HDFS_HOME=$HADOOP_INSTALL
export YARN_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib/native"
```

Save the file and exit

source ~/.bashrc

sudo mkdir -p /app/hadoop/tmp

[sudo] password for student: 123

sudo chown student:student /app/hadoop/tmp

sudo mkdir -p /usr/local/hadoop_store/hdfs/namenode

sudo mkdir -p /usr/local/hadoop_store/hdfs/datanode

sudo chown -R student:student /usr/local/hadoop_store

hdfs namenode -format

Write ? (Y or N) Y

Press Enter

start-dfs.sh

start-yarn.sh

jps

Out Put

```
11633 NodeManager
11497 ResourceManager
11929 Jps
11324 SecondaryNameNode
10973 NameNode
```

sudo apt-get install scala

sudo apt install scala

scala -version

Out Put

Scala code runner version 2.11.6 -- Copyright 2002-2013, LAMP/EPFL

Apache Spark Framework Installation

old -- wget <https://apachemirror.wuchna.com/spark/spark-3.2.1/spark-3.2.1-bin-hadoop2.7.tgz>

New-- wget <https://archive.apache.org/dist/spark/spark-3.2.0/spark-3.2.0-bin-hadoop3.2.tgz>

nected.

HTTP request sent, awaiting response... 200 OK

Length: 300965906 (287M) [application/x-gzip]

Saving to: 'spark-3.2.0-bin-hadoop3.2.tgz.2'

spark-3.2.0-bin-had 100%[=====>] 287.02M 327KB/s in 18m 20s

2024-04-03 15:18:03 (267 KB/s) - 'spark-3.2.0-bin-hadoop3.2.tgz.2' saved
[300965906/300965906]

old-- tar -xvzf spark-3.1.1-bin-hadoop2.7.tgz

New- -tar -xvzf spark-3.2.0-bin-hadoop3.2.tgz

spark-3.2.0-bin-hadoop3.2/licenses/LICENSE-vis-timeline.txt

spark-3.2.0-bin-hadoop3.2/licenses/LICENSE-blas.txt

spark-3.2.0-bin-hadoop3.2/licenses/LICENSE-re2j.txt

spark-3.2.0-bin-hadoop3.2/licenses/LICENSE-kryo.txt

spark-3.2.0-bin-hadoop3.2/licenses/LICENSE-cloudpickle.txt

sudo mv spark-3.2.0-bin-hadoop3.2 /opt/spark

[sudo] password for student: 123

mv: cannot stat 'spark-3.1.1-bin-hadoop2.7': No such file or directory

(base) student@student-OptiPlex-3020:~\$

echo "export SPARK_HOME=/opt/spark" >> ~/.profile

echo "export PATH=\$PATH:/opt/spark/bin:/opt/spark/sbin" >> ~/.profile

echo "export PYSPARK_PYTHON=/usr/bin/python3" >> ~/.profile

source ~/.profile

ls -l /opt/spark

Out Put

total 164

drwxr-xr-x 2 student student 4096 Oct 6 2021 bin

drwxr-xr-x 2 student student 4096 Oct 6 2021 conf

drwxr-xr-x 5 student student 4096 Oct 6 2021 data

```
drwxr-xr-x 4 student student 4096 Oct 6 2021 examples
drwxr-xr-x 2 student student 16384 Oct 6 2021 jars
drwxr-xr-x 4 student student 4096 Oct 6 2021 kubernetes
-rw-r--r-- 1 student student 22878 Oct 6 2021 LICENSE
drwxr-xr-x 2 student student 4096 Oct 6 2021 licenses
drwxrwxr-x 2 student student 4096 Mar 13 16:37 logs
-rw-r--r-- 1 student student 57677 Oct 6 2021 NOTICE
drwxr-xr-x 9 student student 4096 Oct 6 2021 python
drwxr-xr-x 3 student student 4096 Oct 6 2021 R
-rw-r--r-- 1 student student 4512 Oct 6 2021 README.md
-rw-r--r-- 1 student student 167 Oct 6 2021 RELEASE
drwxr-xr-x 2 student student 4096 Oct 6 2021 sbin
drwxr-xr-x 13 student student 4096 Oct 6 2021 spark-3.2.0-bin-hadoop3.2
drwxrwxr-x 2 student student 4096 Mar 13 16:37 work
drwxr-xr-x 2 student student 4096 Oct 6 2021 yarn
```

Start Apache Spark in Ubuntu

start-master.sh

spark-shell

Out Put:-

Spark session available as 'spark'.
Welcome to

```

  ____      ____
 /  _ \    /  _ \  /  _ \
 \  __/    \  __/  \  __/
  \_/      \_/    \_/

 /  _ \   .  _ \  /  _ \   version 3.2.0
 /  _ \

```

Using Scala version 2.12.15 (OpenJDK 64-Bit Server VM, Java 1.8.0_292)
Type in expressions to have them evaluated.
Type :help for more information.

scala> (base) student@student-OptiPlex-3020:~\$ s

Note:- Perss Control C

start-workers.sh spark://localhost:7077

Note:

(if workers not starting then remove and install openssh:

sudo apt-get remove openssh-client openssh-server

sudo apt-get install openssh-client openssh-server)

http://localhost:8080/

URL: spark://student-OptiPlex-3020:7077

Alive Workers: 1

Cores in use: 4 Total, 0 Used

Memory in use: 2.8 GiB Total, 0.0 B Used

Resources in use:

Applications: 0 [Running](#), 0 [Completed](#)

Drivers: 0 Running, 0 Completed

Status: ALIVE .

Old----Spark-shell

New-- spark-shell

sudo apt install snapd

Out put:-

Reading package lists... Done

Building dependency tree

Reading state information... Done

snapd is already the newest version (2.48.3).

0 upgraded, 0 newly installed, 0 to remove and 11 not upgraded.

Old--snap find "intellij"

New-- snap find intellij

Reading package lists... Done

Building dependency tree

Reading state information... Done

snapd is already the newest version (2.48.3).

0 upgraded, 0 newly installed, 0 to remove and 11 not upgraded.

(base) student@student-OptiPlex-3020:~\$ snap find intellij

Name	Version	Publisher	Notes	Summary
intellij-idea-community	2023.3.6	jetbrains**	classic	Capable & Ergonomic Java IDE
intellij-idea-ultimate	2023.3.6	jetbrains**	classic	Capable & Ergonomic Java IDE for Enterprise, Web & Mobile Development
intellij-idea-educational	2022.2.2	jetbrains**	classic	IntelliJ IDEA Educational Edition
datagrip	2023.3.4	jetbrains**	classic	DataGrip
rider	2023.3.4	jetbrains**	classic	A fast & powerful cross-platform .NET IDE
rustrover	2023.3-EAP	jetbrains**	classic	RustRover
goland	2023.3.6	jetbrains**	classic	GoLand
gradle	7.2	snapcrafters*	classic	An open-source build automation tool

old---sudo snap install intellij-idea-community - - classic

new--sudo snap install intellij-idea-community – classic

Out Put:

snap "intellij-idea-community" is already installed, see 'snap help refresh'



```
1 # Package generated configuration file
2 # See the sshd_config(5) manpage for details
3
4 # What ports, IPs and protocols we listen for
5 Port 22
6 # Use these options to restrict which interfaces/protocols sshd
   will bind to
7 #ListenAddress ::
8 #ListenAddress 0.0.0.0
9 Protocol 2
10 # HostKeys for protocol version 2
11 HostKey /etc/ssh/ssh_host_rsa_key
12 HostKey /etc/ssh/ssh_host_dsa_key
13 HostKey /etc/ssh/ssh_host_ecdsa_key
14 HostKey /etc/ssh/ssh_host_ed25519_key
15 #Privilege Separation is turned on for security
16 UsePrivilegeSeparation yes
17
18 # Lifetime and size of ephemeral version 1 server key
19 KeyRegenerationInterval 3600
20 ServerKeyBits 1024
21
22 # Logging
23 SyslogFacility AUTH
24 LogLevel VERBOSE
25
26 # Authentication:
27 LoginGraceTime 120
28 PermitRootLogin prohibit-password
29 StrictModes yes
30
31 RSAAuthentication yes
32 PubkeyAuthentication yes
33 #AuthorizedKeysFile      %h/.ssh/authorized_keys
34
35 # Don't read the user's ~/.rhosts and ~/.shosts files
36 IgnoreRhosts yes
```



```
# ~/.bashrc: executed by bash(1) for non-login shells.
# see /usr/share/doc/bash/examples/startup-files (in the package bash-doc)
# for examples

# If not running interactively, don't do anything
case $- in
    *i*) ;;
    *) return;;
esac

# don't put duplicate lines or lines starting with space in the history.
# See bash(1) for more options
HISTCONTROL=ignoreboth

# append to the history file, don't overwrite it
shopt -s histappend

# for setting history length see HISTSIZE and HISTFILESIZE in bash(1)
HISTSIZE=1000
HISTFILESIZE=2000

# check the window size after each command and, if necessary,
# update the values of LINES and COLUMNS.
".bashrc" 1451 4802C
```



3.2.0

Spark Master at spark://student-OptiPlex-3020:7077

URL: spark://student-OptiPlex-3020:7077

Alive Workers: 1

Cores in use: 4 Total, 0 Used

Memory in use: 2.8 GiB Total, 0.0 B Used

Resources in use:

Applications: 0 Running, 0 Completed

Drivers: 0 Running, 0 Completed

Status: ALIVE

Workers (1)

Worker Id	Address	State	Cores	Memory	Resources
worker-20240403152928-172.16.107.242-37616	172.16.107.242:37616	ALIVE	4 (0 Used)	2.8 GiB (0.0 B Used)	

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

Completed Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

hadoop 2.7.0_updated

File Edit View Insert

hadoop 2.7.0_updated

Welcome to IntelliJ IDEA

IntelliJ IDEA 2023.3.5

Projects

Customize

Plugins

Learn

Marketplace

Installed

Type / to see options

Staff Picks

Show all

Scala

Restart IDE

20.2M 4.40 JetBrains ...

aws AWS Toolkit - ...

Install

6.3M 2.44 Amazon We...

Grazie Pro

Install

606.1K 4.71 JetBrains ...

IdeaVim

Install

14.8M 4.44 JetBrains ...

Solarized The...

Install

344.3K 4.42 Tyler Thr...

JetBrains Aca...

Install

3.2M 4.13 JetBrains s...

Key Promoter X

Install

5.6M 4.95 Hal's Corner

Programming Language

Scala

JetBrains s.r.o. Plugin homepage

Restart IDE

2023.3.27

Reviews

Additional Info

Overview

What's New

Adds support for the Scala language. The following features are available for free with IntelliJ IDEA Community Edition:

- Coding assistance (highlighting, completion, formatting, refactorings, etc.)

hadoop 2.7.0_updated

File Edit View Insert

hadoop 2.7.0_updated

Welcome to IntelliJ IDEA

IntelliJ IDEA 2023.3.5

Projects

Customize

Plugins

Learn

Marketplace

Installed

Type / to see options

Staff Picks

Show all

Scala

Restart IDE

20.2M 4.40 JetBrains ...

aws AWS Toolkit - ...

Install

6.3M 2.44 Amazon We...

Programming Language

Scala

JetBrains s.r.o. Plugin homepage

Restart IDE

2023.3.27

Reviews

Additional Info

What's New

IntelliJ IDEA and Plugin Updates

Restart IntelliJ IDEA to apply changes in plugins?

Restart

Not Now

Solarized The...

Install

344.3K 4.42 Tyler Thr...

JetBrains Aca...

Install

3.2M 4.13 JetBrains s....

Key Promoter X

Install

5.6M 4.95 Hal's Corner

Adds support for the Scala language. The following features are available for free with IntelliJ IDEA Community Edition:

- Coding assistance (highlighting, completion, formatting, refactorings, etc.)

hadoop 2.7.0_updated

File Edit View Insert

100

2

020.out

spark-shell

...

hadoop 2.7.0_updated

File Edit View Insert

100

2

020.out

spark-shell

...

IntelliJ IDEA
2023.3.5

Projects

Customize

Plugins

Learn

Welcome to IntelliJ IDEA


Create a new project to start from scratch.
Open existing project from disk or version control.

+

New Project

Open

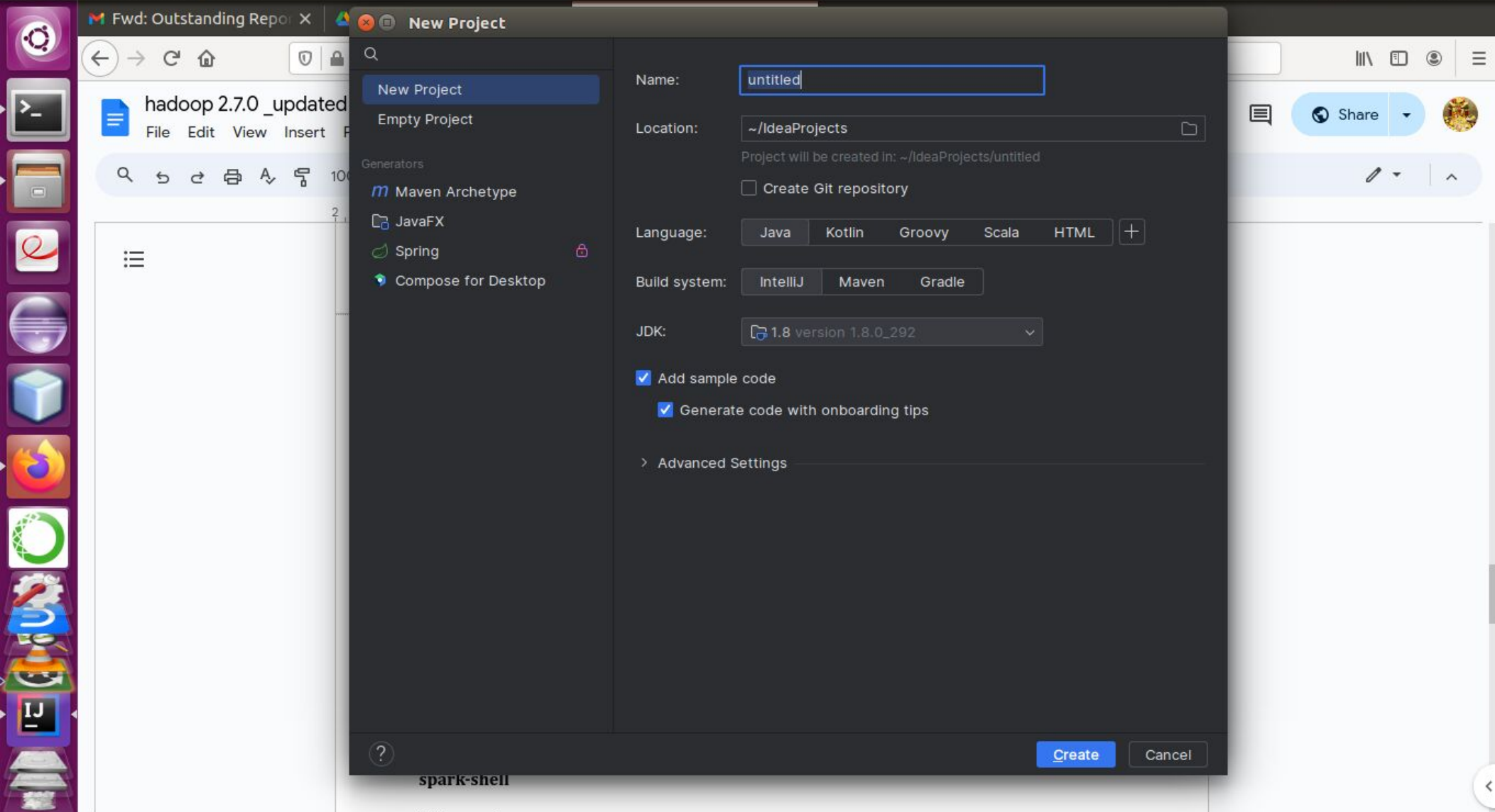
Get from VCS



Take a quick onboarding tour

New to IntelliJ IDEA? Get the most out of your IDE. Become acquainted with its tools and basic workflows.

Start Tour in Java



hadoop 2.7.0_updated

File Edit View Insert

hadoop 2.7.0_updated

2

1 2 3 4 5 6 7 8 9 10 11

IntelliJ IDEA 2023.3.5

Projects

Customize

Plugins

Learn

Welcome to IntelliJ IDEA

Create a new project to start from scratch.
Open existing project from disk or version control.

New Project Open Get from VCS

Take a quick onboarding tour

New to IntelliJ IDEA? Get the most out of your IDE. Become acquainted with its tools and basic workflows.

Start Tour in Java

hadoop 2.7.0_updated

File Edit View Insert

hadoop 2.7.0_updated

Search

Projects

Customize

Plugins

Learn

Welcome to IntelliJ IDEA

IntelliJ IDEA 2023.3.5

Marketplace Installed

Type / to see options

Staff Picks

Show all

Scala

20.2M 4.40 JetBrains ...

Install

aws AWS Toolkit - ...

6.3M 2.44 Amazon We...

Install

Grazie Pro

606.1K 4.71 JetBrains ...

Install

IdeaVim

14.8M 4.44 JetBrains ...

Install

Solarized The...

344.3K 4.42 Tyler Thr...

Install

JetBrains Aca...

3.2M 4.13 JetBrains s...

Install

Key Promoter X

5.6M 4.95 Hal's Corner

Install

Programming Language

Scala

JetBrains s.r.o. Plugin homepage

Install 2023.3.27

Reviews Additional Info

Overview What's New

Adds support for the Scala language. The following features are available for free with IntelliJ IDEA Community Edition:

- Coding assistance (highlighting, completion, formatting, refactorings, etc.)

hadoop 2.7.0_updated

Welcome to IntelliJ IDEA

IntelliJ IDEA 2023.3.5

Projects

Customize

Plugins

Learn

Marketplace

Installed

Type / to see options

Staff Picks

Show all

Scala

20.2M 4.40 2023.3.27

aws AWS Toolkit - ...

6.3M 2.44 Amazon We...

Grazie Pro

606.1K 4.71 JetBrains ...

IdeaVim

14.8M 4.44 JetBrains ...

Solarized The...

344.3K 4.42 Tyler Thr...

JetBrains Aca...

3.2M 4.13 JetBrains s...

Key Promoter X

5.6M 4.95 Hal's Corner

Nord

Programming Language

Scala

JetBrains s.r.o. Plugin homepage

Disable 2023.3.27

Reviews

Additional Info

Overview

What's New

Adds support for the Scala language. The following features are available for free with IntelliJ IDEA Community Edition:

- Coding assistance (highlighting, completion, formatting, refactorings, etc.)

020.out

spark-shell

hadoop 2.7.0_updated

hadoop 2.7.0_updated

File Edit View Insert

hadoop 2.7.0_updated

Projects

Customize

Plugins

Learn

Marketplace Installed

Type / to see options

Manage Plugin Certificates

Certificates trusted when validating plugins:

No certificates

No certificate selected

OK Cancel

Key Promoter X

5.6M 4.95 Hal's Corner

Install

Programming Language

Plugin homepage

2023.3.27

Additional Info

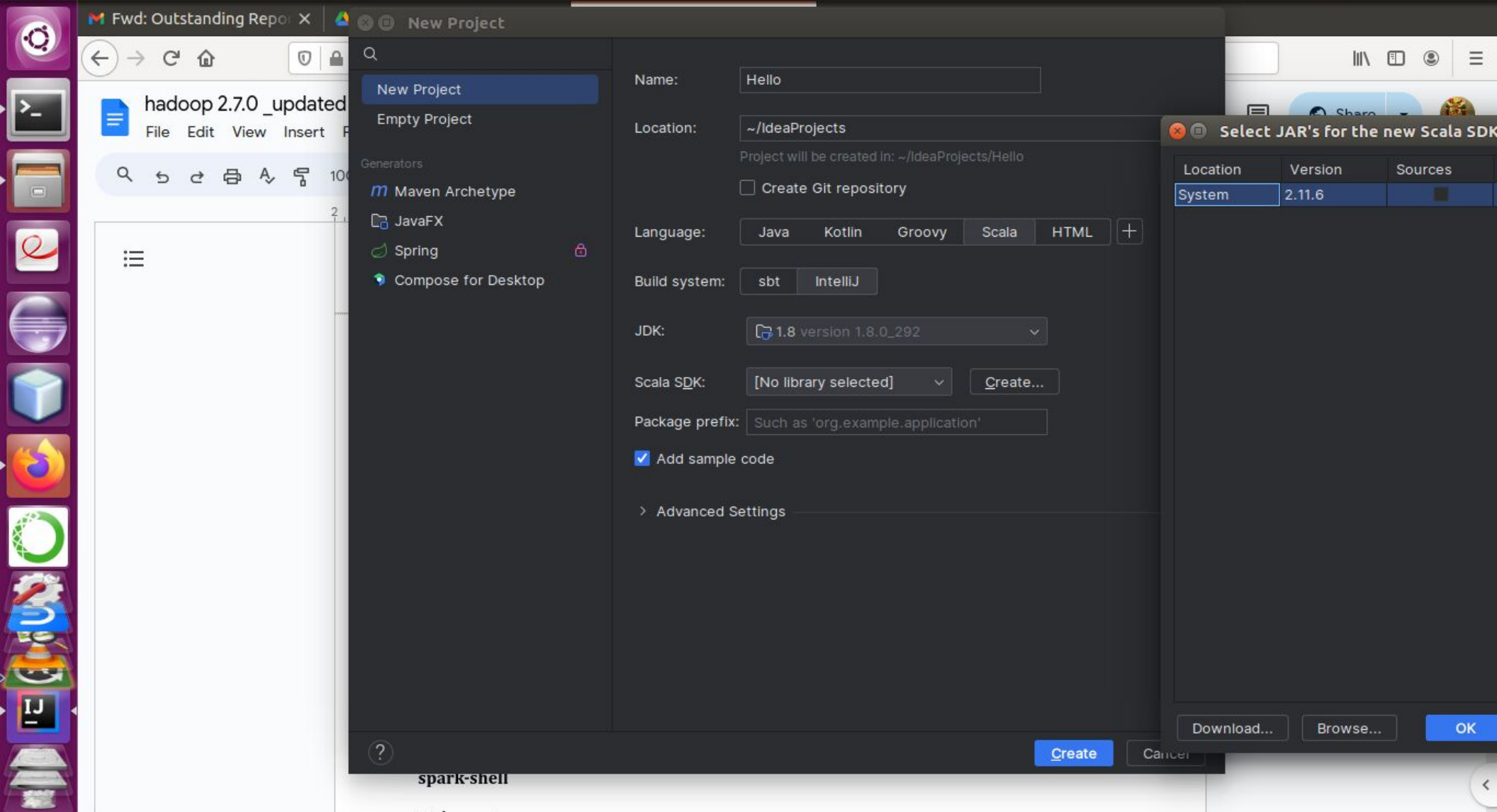
What's New

Support for the Scala language. The following features are available for free with IntelliJ IDEA Community Edition:

- Coding assistance (highlighting, completion, formatting, refactorings, etc.)

020.out

spark-shell



The screenshot displays the IntelliJ IDEA Community Edition interface. The top menu bar includes options like Code, Refactor, Build, Run, Tools, VCS, Window, and Help. The left sidebar contains a vertical stack of icons for various IDE features. The main workspace is divided into three panels:

- Project View:** Shows the project structure for 'Hello' located at '~/IdeaProjects/Hello'. It includes a folder named '.idea', a source folder 'src' containing a 'Main' class and a 'Hello.iml' file, 'External Libraries', and 'Scratches and Consoles'.
- Code Editor:** Displays the 'Main.scala' file with the following code:

```
1 object Main {  
2     def main(args: Array[String]): Unit = {  
3         println("Hello world!")  
4     }  
5 }
```
- Run View:** Shows the 'Main' class being executed.

The bottom status bar indicates the project is indexing JDK '1.8' and shows settings for 'Show all (2)', '1:1', 'LF', 'UTF-8', and '2 spaces'.

The screenshot displays the IntelliJ IDEA Community Edition interface. The top toolbar includes icons for file operations, search, and settings. The left sidebar shows the Project view with the following structure:

- Project ▾
 - ▾ Hello ~/IdeaProjects/Hello
 - > .idea
 - > out
 - ▾ src
 - ⦿ Main
 - Hello.iml
 - > External Libraries
 - ≡ Scratches and Consoles

The main editor window displays the file `Main.scala` with the following code:

```
1 object Main {  
2   def main(args: Array[String]): Unit = {  
3     println("Hello world!")  
4   }  
5 }
```

The Run tool window at the bottom shows the execution of the `Main` class. The command executed is `/usr/lib/jvm/java-1.8.0-openjdk-amd64/bin/java ...`, and the output is `Hello world!`. The process finished with exit code 0.

The status bar at the bottom indicates the current file is `Hello`, the encoding is UTF-8, and the line length is 1:1 LF.



Baliram Deshmukh <baliramdeshmukh.sae@sinhgad.edu>

Spark link

2 messages

Deepika Lalit Falak <dlFalak.sae@sinhgad.edu>

12 March 2024 at 15:35

To: Baliram Deshmukh <baliramdeshmukh.sae@sinhgad.edu>

wget <https://archive.apache.org/dist/spark/spark-3.2.0/spark-3.2.0-bin-hadoop3.2.tgz>
tar -xzf spark-3.2.0-bin-hadoop3.2.tgz

Sinhgad Technical Education Society | www.sinhgad.edu

Confidential Information. The information contained in or attached to this e-mail may be confidential information subject to protection by law or terms of applicable confidentiality agreements, and is intended only for the use of the individual or entity named above. If the reader of this message is not the intended recipient, or the employee or agent responsible to deliver it to the intended recipient, you are hereby notified that any dissemination, distribution, or copying of this communication is strictly prohibited. If you are not the addressee indicated in this message (or responsible for delivery of the message to such person), you should destroy this message and notify the sender by reply email.

Baliram Deshmukh <baliramdeshmukh.sae@sinhgad.edu>

12 March 2024 at 15:35

To: dlFalak.sae@sinhgad.edu

--

Best regards,

Prof. Baliram Deshmukh

Assistant Professor

Department of Computer Engineering.

Sinhgad Academy of Engineering Pune-48.

Email: baliramdeshmukh.sae@sinhgad.edu

Mobile : 9545947001,9850205734.