

Hadoop HDFS/MapReduce Overview

2015/12/21

Scott Miao

@takeshi.miao

A HDFS/MapReduce overview

- Here I will refer to a **MRv1** tutorial made by my colleague MengSung Wu.
 - `hadoop_training_3.pdf`

Come on

Let's do some exercise !!

Prerequisite

- It is better to use Unix-like OS for developing Hadoop related softwares
- Use Java in this exercise, use any IDE you like
 - Or follow the instruction to download Eclipse IDE
- No need for a running Hadoop cluster
 - But you can still deploy your code to run it !

1. Install Java7 & Eclipse IDE

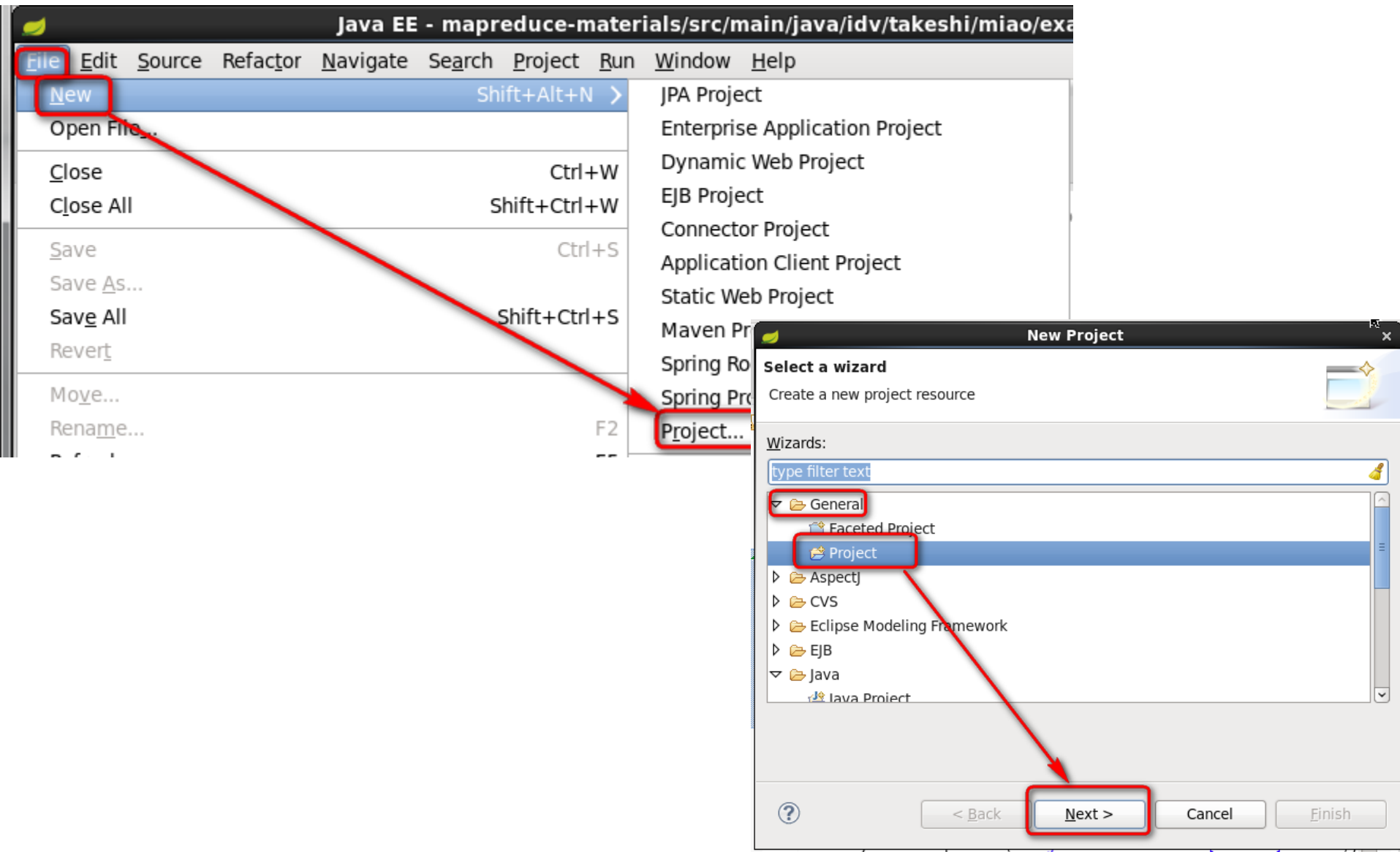
- Download and install Java7
 - <http://www.oracle.com/technetwork/java/javase/downloads/jdk7-downloads-1880260.html>
 - Why we choose Java7 ?
 - Due to till current Hadoop-2.6.x, they still officially support to Java7 only
- Download and unzip/untar your Eclipse IDE
 - <http://www.eclipse.org/downloads/packages/eclipse-ide-java-ee-developers/mars1>

2. Download the sample codes

- Download sample codes from my github
 - <https://github.com/takeshimiao/workshop-materials/archive/master.zip>
- Unzip it to where you like

3 Put the sample codes to Eclipse

3.1 Create a new project

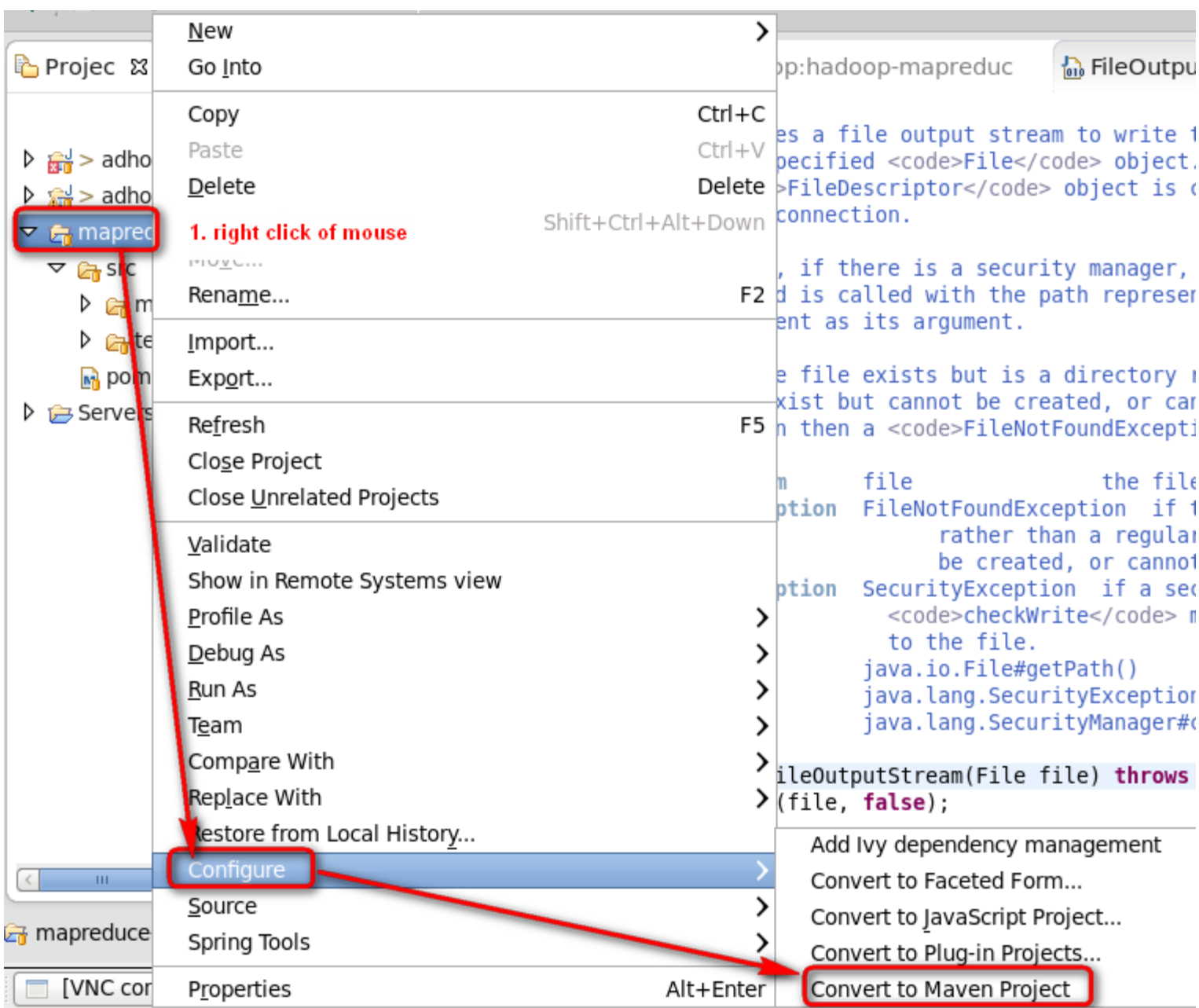


3.2 point the new project to where the sample code you put

The screenshot shows the 'New Project' dialog box with the following elements and annotations:

- Project name:** The text 'mapreduce-materials' is entered in the text box. A red box highlights the text, and a callout bubble labeled '1.' points to it. Another callout bubble labeled '3.' points to the same text, stating: '3. name your project, can be same as 'mapreduce-materials''.
- Use default location:** The checkbox is unchecked. A red box highlights the checkbox and its label, with a callout bubble labeled '1.' pointing to it.
- Location:** The text 'cuments/projects/workshop-materials/mapreduce-materials' is entered in the text box. A red box highlights the text, and a callout bubble labeled '2.' points to it, stating: '2. point to where sample codes you put, to the sub directory 'mapreduce-materials''.
- Choose file system:** The dropdown menu is set to 'default'.
- Working sets:** The checkbox 'Add project to working sets' is unchecked. Below it, the 'Working sets:' text is followed by a dropdown menu and a 'Select...' button.
- Buttons:** At the bottom, there are four buttons: '< Back', 'Next >', 'Cancel', and 'Finish'. A red box highlights the 'Finish' button, and a callout bubble labeled '4.' points to it.

3.3 convert this project to Maven project



Then click 'OK'

3.4 build the sample code first

Project Explorer

- adho
- adho
- mapreduce**
- src
- src/
- src/t
- src/t
- JRE
- Mav
- src
- target
- pc
- Servers

Show In

Copy

Copy Qualified Name

Paste

Delete

Remove from Context

Build Path

Refresh

Close Project

Close Unrelated Projects

Validate

Show in Remote Systems view

Profile As

Debug As

Run As

Team

Compare With

Replace With

Restore from Local History...

Maven

Configure

Source

Spring Tools

Shift+Alt+V

Ctrl

Ctrl

De

Shift+Ctrl+Alt+D

Shift+Alt+

1. Again, right click your mouse on the project

import...

Export...

Build Project

Refresh

Close Project

Close Unrelated Projects

Validate

Show in Remote Systems view

Profile As

Debug As

Run As

Team

Compare With

Replace With

Restore from Local History...

Maven

Configure

Source

Spring Tools

Build Path

Refresh

Close Project

Close Unrelated Projects

Validate

Show in Remote Systems view

Profile As

Debug As

Run As

Team

Compare With

Replace With

Restore from Local History...

Maven

Configure

Source

Spring Tools

Ma

Pr

Se

Da

Sni

Co

Bo

Pr

Se

<terminated> /usr/java/jdk1.7.0_71/bin/java (Dec 21, 2015, 2:30:34 AM)

[INFO] Installing /home/scott_miao/Documents/projects/worksho

[INFO] Installing /home/scott_miao/Documents/projects/worksho

[INFO] BUILD SUCCESS

[INFO] Total

[INFO] Finish

[INFO] Final

[INFO] -----

NOTICE:

You should see the 'BUILD SUCCESS'

After every time you click

'Maven clean' or 'Maven install'

F5

ugins>

Dependencies

Dependency Hierarchy

Effective POI

D

S

C

B

P

S

C

R

T

> java/jdk1.7.0_71/bin/java (Dec 21, 2015, 2:30:34 AM)

> 49,999 INFO [main] hbase.HBaseCommonTest

> 1 Java Applet Shift+Alt+X A

> 2 Java Application Shift+Alt+X J

> 3 JUnit Test Shift+Alt+X T

> 4 Maven click 'Maven clean' first, wait it run till completed -Alt+X M

> 5 Maven

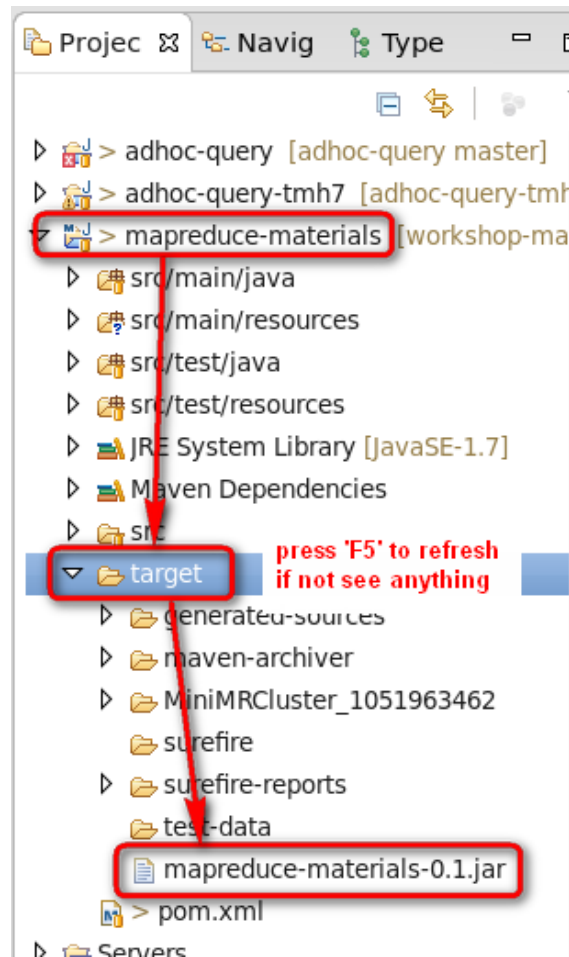
> **6 Maven clean**

> 7 Maven generate-sources

> **8 Maven install** then click 'Maven install'

> 9 Maven test

3.5 Congrats ! You finally get your first build !!



MRv2 on YARN !!

- This is a main feature brought to Hadoop-2.x
- Let's refer to my another good friend Jeff Markham's slides here
 - 6.Hadoop_2_Overview_andNew_features_Jeff.pdf