

Lập trình bài toán Word Count

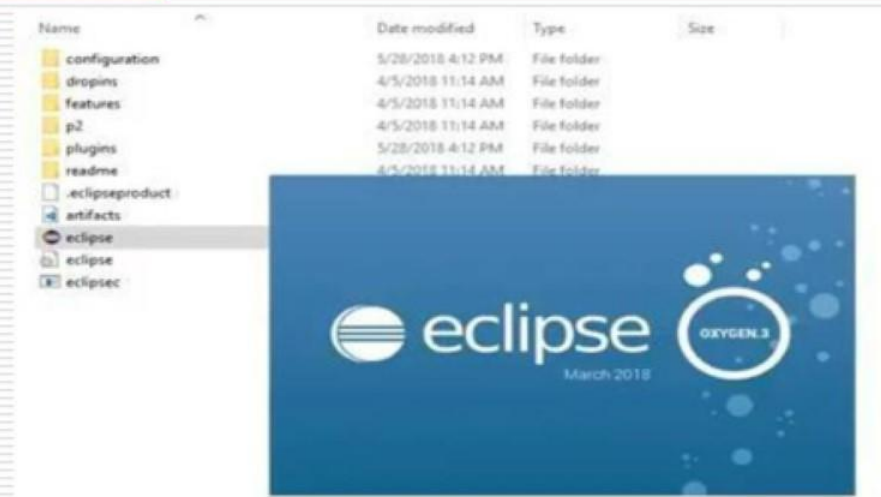
- ❖ Mô tả bài toán Word Count
- ❖ Khởi tạo project MapReduce
- ❖ Các nhiệm vụ chính
- ❖ Mã nguồn lớp WordCountDriver
- ❖ Mã nguồn lớp WordCountMapper
- ❖ Mã nguồn lớp WordCountReducer

Mô tả bài toán Word Count

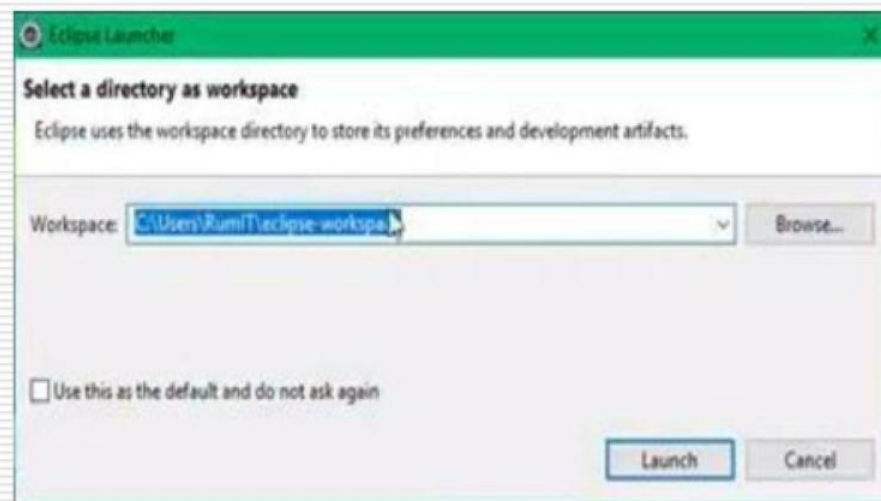
- ❖ Input: File input.txt có nội dung:
 - ❖ Bus, Car, bus, car, train, car, bus, car, train, bus, TRAIN, BUS, buS, caR, CAR, car, BUS, TRAIN
- ❖ Output: File text có nội dung:
 - ❖ BUS 2
 - ❖ Bus 1
 - ❖ CAR 1
 - ❖ Car 1
 - ❖ TRAIN 2
 - ❖ buS 1
 - ❖ bus 3
 - ❖ caR 1
 - ❖ car 4
 - ❖ train 2

Khởi tạo project MapReduce

- ❖ B1: Vào thư mục eclipse, click đúp file ứng dụng eclipse để khởi động eclipse và đợi

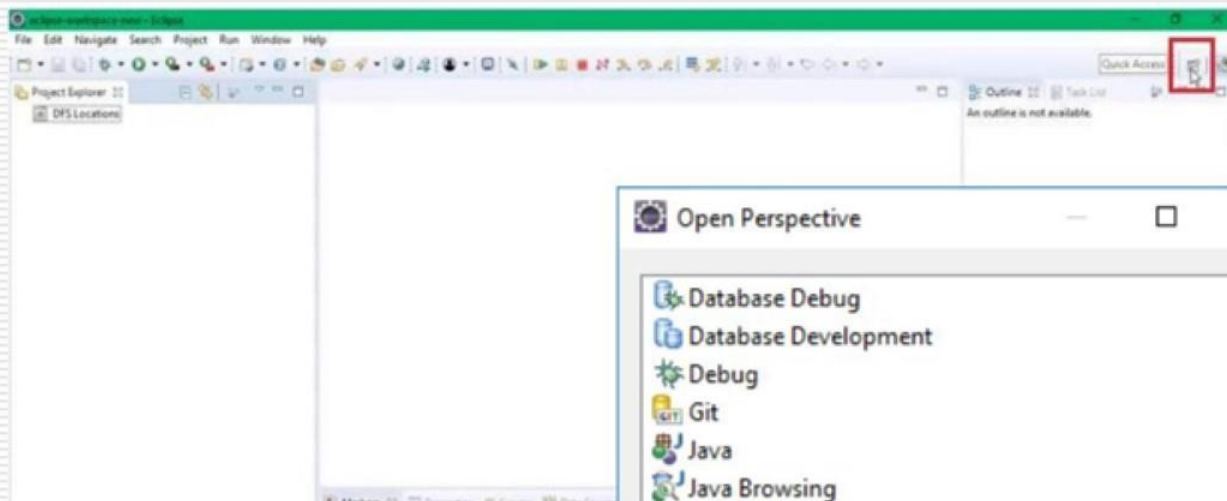


- ❖ B2: Chọn thư mục workspace và nhấn nút Launch



Khởi tạo project MapReduce

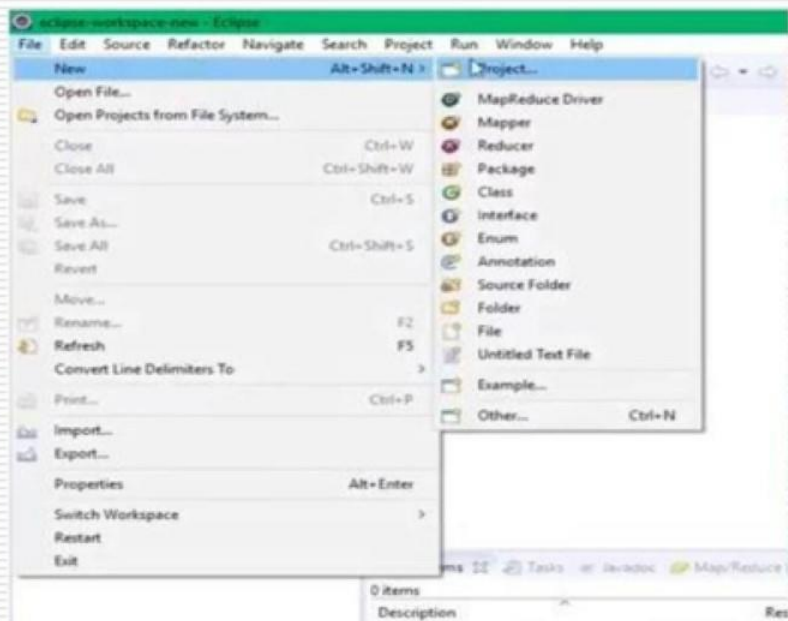
- ❖ B3: Nhấn nút “Open perspective” để mở hộp thoại “Open perspective”



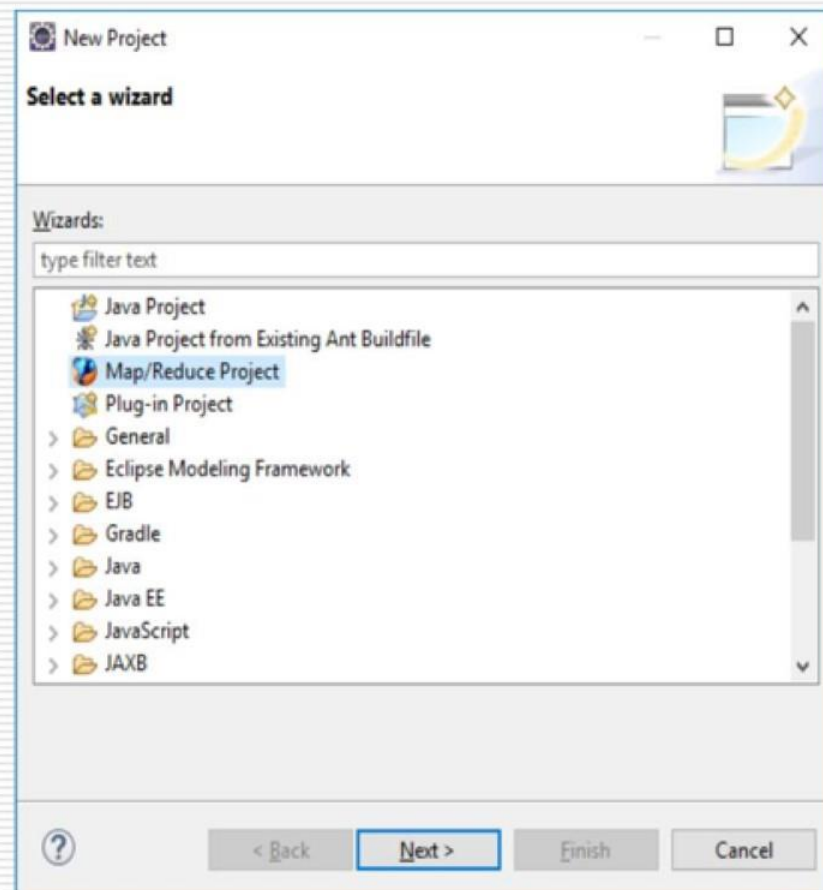
- ❖ B4: Chọn “Map/Reduce”, nhấn Open

Khởi tạo project MapReduce

- ❖ B5: Tạo project Map/Reduce: Vào File->New->Project

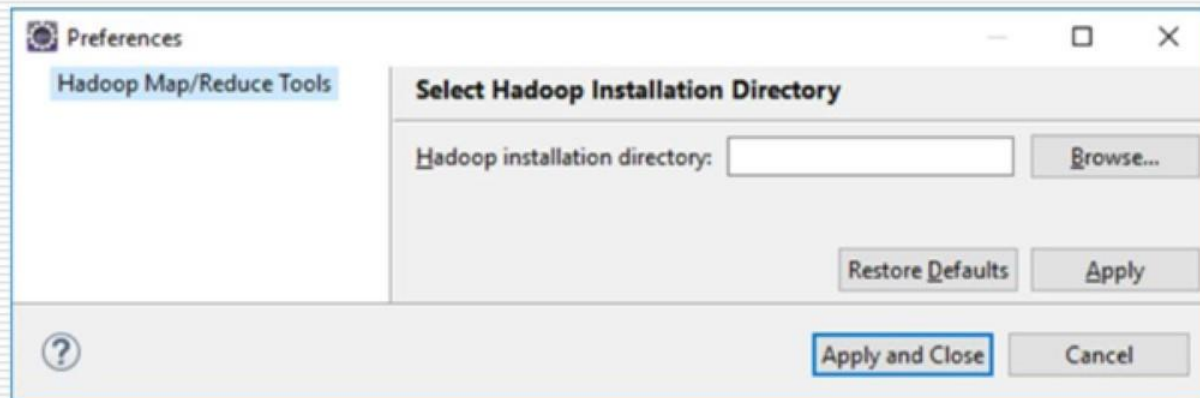
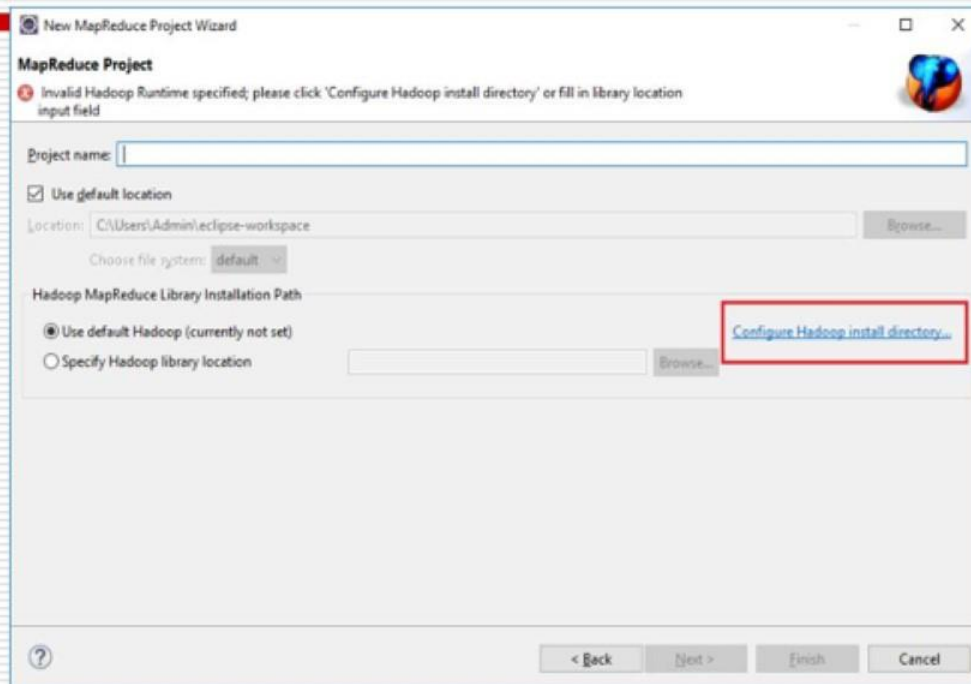


- ❖ B6: Trong cửa sổ “New Project”, chọn Map/ReduceProject và nhấn Next



Khởi tạo project MapReduce

- ❖ B7: Nhấn vào link “Configure Hadoop install directory”

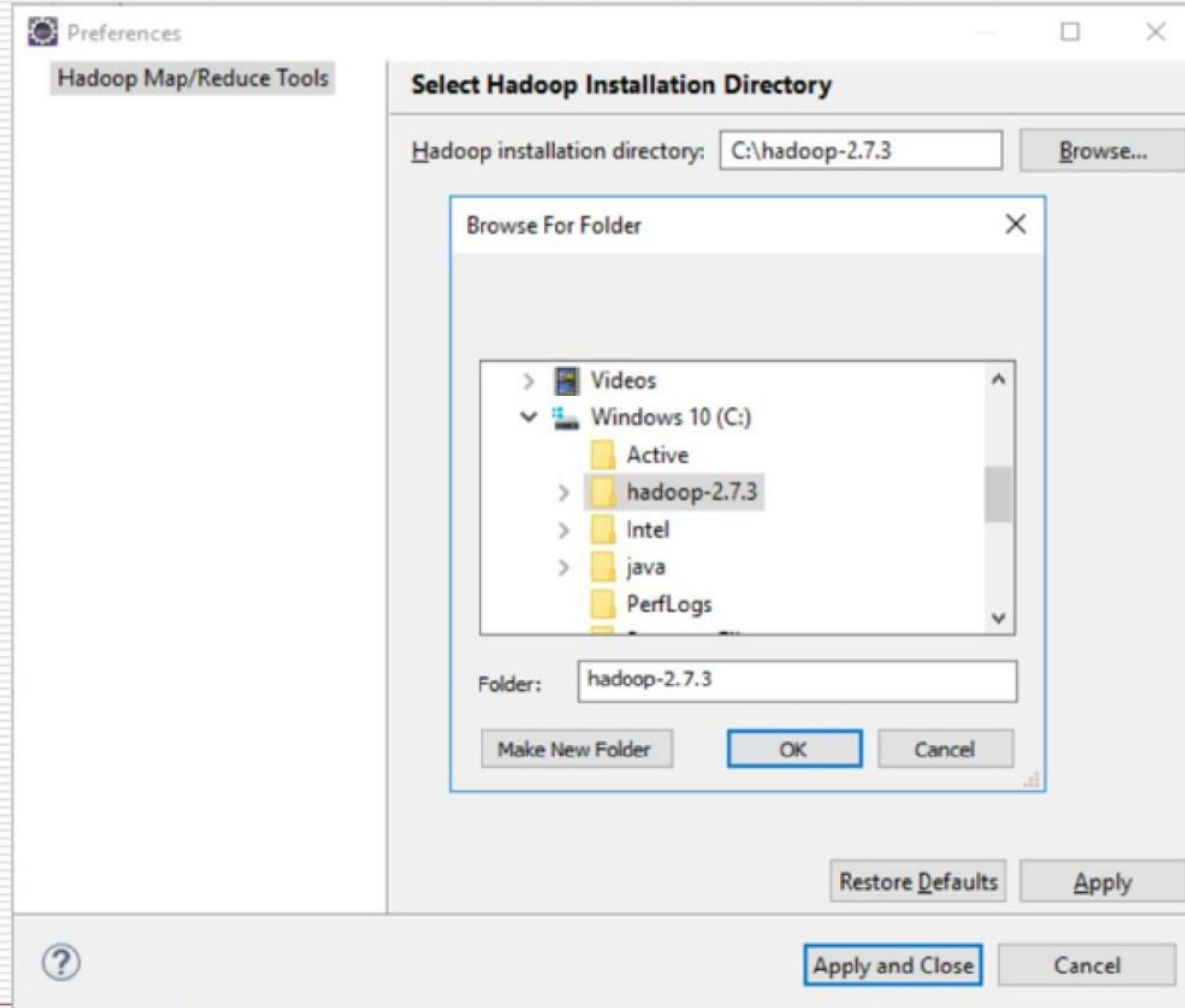


Khởi tạo project MapReduce

❖ B8: Nhấn nút Browser để duyệt đến thư mục cài Hadoop

❖ B9: Nhấn nút Browser để duyệt đến thư mục cài Hadoop

❖ B10: Nhấn OK, nhấn “Apply and Close”



Khởi tạo project MapReduce

- ❖ B11: Gõ tên project vào hộp Textbox “Project name” và nhấn Finish

New MapReduce Project Wizard

MapReduce Project
Create a MapReduce project.

Project name:

☒ Use default location
Location: [Browse...](#)
Choose file system: [default](#)

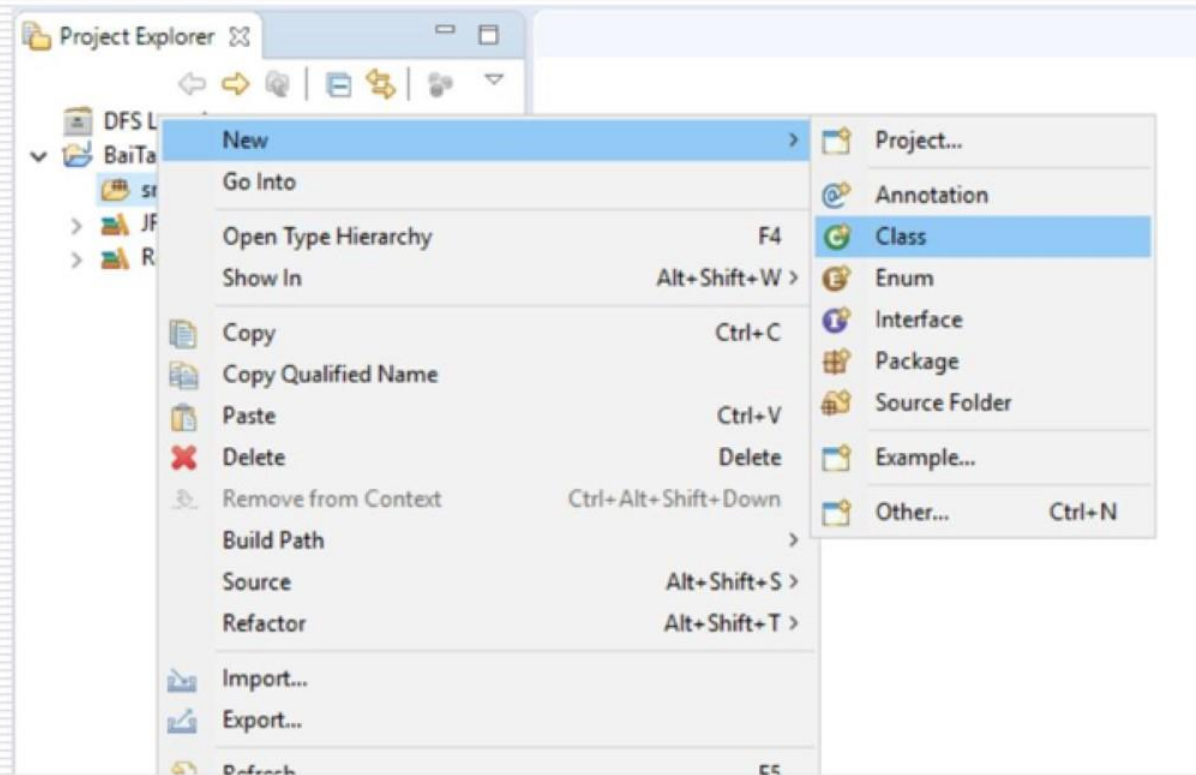
Hadoop MapReduce Library Installation Path

☒ Use default Hadoop [Configure Hadoop install directory...](#)
☐ Specify Hadoop library location [Browse...](#)

[?](#) [< Back](#) [Next >](#) [Finish](#) [Cancel](#)

Tạo file java trong project MapReduce

- ❖ B1: Chuột phải vào mục src trong mục project, chọn New->Class để mở hộp thoại “New Java Class”



Tạo file java trong project MapReduce

- ❖ B2: Gõ tên file vào hộp Textbox Name và Enter

New Java Class

Java Class

⚠ The use of the default package is discouraged.

Source folder: BaiTap1/src [Browse...](#)

Package: (default) [Browse...](#)

☐ Enclosing type: [Browse...](#)

Name: **WordCount**

Modifiers: ☒ public ☐ package ☐ private ☐ protected
☐ abstract ☐ final ☐ static

Superclass: java.lang.Object [Browse...](#)

Interfaces: [Add...](#) [Remove](#)

Which method stubs would you like to create?

☐ public static void main(String[] args)
☐ Constructors from superclass
☒ Inherited abstract methods

Do you want to add comments? (Configure templates and default value [here](#))
☐ Generate comments

[Finish](#) [Cancel](#)

Các nhiệm vụ chính

- ❖ Định nghĩa lớp Driver: bao gồm hàm main
 - ❖ Thông tin vào ra
 - ❖ Thông tin cấu hình Job
 - ❖ Khởi tạo Job
- ❖ Định nghĩa lớp Mapper: chứa hàm map
- ❖ Định nghĩa lớp Reducer: chứa hàm reduce

Mã nguồn lớp Driver

- ❖ Xem file WordCount1Driver.java
- ❖ Tạo file

Mã nguồn lớp Mapper

- ❖ Xem file WordCount1Mapper.java

Mã nguồn lớp Reducer

- ❖ Xem file WordCount1Reducer.java