Homework HW4 and Sample Projects

Release

IU BDAA Course Team

December 05, 2015

Contents

First some background: You will need modest software for some of lectures and homeworks. This can be most easily done in Python but Java is also possible; no other languages are relevant. You MIGHT use software in final project. Note the final project will be individual or up to a three person team and can either be a significant **PAPER** or a **SOFTWARE PROJECT**. We can support Software in Java on clouds or Python but other software such as R allowed in final project Your team chooses. You can find sample software and paper projects. here.

Preparing Software

Java and Python are installed on our cloud as explained in Unit 11. Here you choose between Python on your laptop, Python in cloud or Java in cloud.

- **DO - Python on your laptop: Download Enthought Canopy Express (free) from**
  https://store.enthought.com/downloads/ including NumPy SciPy Matplotlib

- OR - Set up Python in cloud or Java in cloud. See Unit 11.

Tasks

Submit results to show your software is set up and running. Solve tasks following instructions here .

- Submit your Java OR Python program results to IU Canvas

Discussion D5

Create a NEW post to discuss final project you want to do and look for team members on Slack (http://bdaafall2015.slack.com) for Software Projects if that's what you want.

Contents

3.1 Homework H4: Get your software set up for later lectures and final project

In this homework, you are expected to run Python or Java programs on FutureSystems or on your local machine. A few examples for beginners will help you to understand how to write and run Java or Python programs on your environment.

3.1.1 Setup

Java and Python are installed on our cloud as explained in Unit 11. Here you choose between Python on your laptop, Python in cloud or Java in cloud.

Local Setup

Download Enthought Canopy Express (free) from https://store.enthought.com/downloads/ including NumPy SciPy Matplotlib

Cloud

Set up Python in cloud or Java in cloud. See Unit 11.

3.1.2 First Program

This code explains how to display a simple string on your screen. You can download or write your own code using your editor.

Java

Download: FirstProgram.java

```java
/**
 * Sample Program to print out a message
 *
 * Compile : javac FirstProgram.java
 *    Run   : java FirstProgram
 */
public class FirstProgram {
    public static void main(String[] args){
            System.out.println("My first program on Big Data Applications and Analytics!");
        }
}
```

This example prints out the message on your screen by `println` method in the `System` class. In Java Programming, you need to complie your code to execute.

Compiling and Execution

```
javac FirstProgram.java
```

Now, you will have FirstProgram.class file on your system. Java Compiler (javac) creates Java bytecode with a `.class` extension. We will execute the class file with `java` command.

```
java FirstProgram
My first program on Big Data Applications and Analytics!
```

Python

Let's write a same program in Python.

Download: FirstProgram.py

```python
# Run python FirstProgram.py
print 'My first program on Big Data Applications and Analytics!'
```

Python function `print` simply displays a message on your screen. Compiling is not necessary in Python. You can run your code directly with `python` command.

```
python FirstProgram.py
My first program on Big Data Applications and Analytics!
```

3.1.3 Display System Information

This is an extension of your first program. We will lean how to import functions and use them to get system information like hostname or username.

Java

We now understand how to print out a message using Python or Java. System information such as time, date, user name or hostname (machine name) can be displayed as well with built-in functions in each language.

Download: FirstProgramWithSystemInfo.java

```java
import java.util.Date;
import java.text.DateFormat;
import java.text.SimpleDateFormat;
import java.net.InetAddress;
import java.net.UnknownHostException;


/**
 *  * Sample Program with system information
 *  *
 *  * Compile : javac FirstProgramWithSystemInfo.java
 *  *   Run    : java FirstProgramWithSystemInfo
 *  */
public class FirstProgramWithSystemInfo {
        public static void main(String[] args){

                System.out.println("My first program with System Information!");

                // Print Date with Time
                DateFormat dateFormat = new SimpleDateFormat("yyyy/MM/dd HH:mm:ss");
                Date date = new Date();
                System.out.println("Today is: " + dateFormat.format(date));
                // Print Username
                System.out.println("Username is: " + System.getProperty("user.name"));
                // Print hostname
                try {
                        java.net.InetAddress localMachine = java.net.InetAddress.getLocalHost();
                        System.out.println("Hostname is: " + localMachine.getHostName());
                } catch (UnknownHostException e) {
                        e.printStackTrace();
                        System.out.println("No host name: " + e.getMessage());
                }
        }
}
```

Compiling and Execution

```
javac FirstProgramWithSystemInfo.java
```

```
java FirstProgramWithSystemInfo
My first program with System Information!
Today is: 2015/01/01 18:54:10
Username is: albert
Hostname is: bigdata-host
```

Python

Download FirstProgramWithSystemInfo.py

```python
from datetime import datetime
import getpass
import socket


# Run python FirstProgramWithSystemInfo.py
print ('My first program with System Information!')
```

3.1. Homework H4: Get your software set up for later lectures and final project 9

```
print ("Today is: " + str(datetime.now()))
print ("Username is: " + getpass.getuser())
print ("Hostname is: " + socket.gethostname())
```

Execution

```
python  FirstProgramWithSystemInfo.py
My first program with System Information!
Today is: 2015-01-01 18:58:10.937227
Username is: albert
Hostname is: bigdata-host
```

3.1.4 Submission of HW4

Submit these compiled files or screenshot image files to IU Canvas

[Java]

- ******FirstProgram.class or a screenshot image of the 'FirstProgram' execution (25%) **

- **FirstProgramWithSystemInfo.class or a screenshot image of the 'FirstProgramWithSystemInfo' execution (25%)**

[Python]

- **FirstProgram.pyc or a screenshot image of the 'FirstProgram' execution (25%)**

    - run `python -m compileall FirstProgram.py` to generate `FirstProgram.pyc`

- **FirstProgramWithSystemInfo.pyc or a screenshot image of the 'FirstProgramWithSystemInfo' execution (25%)**

    - run `python -m compileall FirstProgramWithSystemInfo.py` to generate `FirstProgramWithSystemInfo.pyc`

3.1.5 Challenge tasks

- **Run any Java or Python on a FutureSystems OpenStack instance**
    - Submit screenshot images of your terminal executing Java or Python code on FutureSystems
- **Run NumPyTutorial Python on IPython Notebook**
    - Submit screentshot images of your web browser executing NumPyTutorial on FutureSystems
- **Tips: See** tutorials for Big Data Applications and Analytics Shell on FutureSystems

3.1.6 Preview Course Examples

- The Elusive Mr.Higgs [Java][Python]

- Number Theory [Python]

- Calculated Dice Roll [Java][Python]

- KNN [Java][Python]

- PageRank [Java][Python]

- KMeans [Java][Python]

3.2 Discussion D5

Create a NEW post to discuss final project you want to do and look for team members on Slack for Software Projects if that's what you want. Please use the #project channel https://bdaafall2015.slack.com/messages/project/ for that purpose. If you are not part of Slack yet, please send an email to bdaacoursehelp@googlegroups.com

3.3 Project Details

You have a choice to write a term paper or do a software project using our cloud computing test bed called FutureSystems. However, we encourage you to do a software project that can be done as a group project. You can form teams of up to 3 people. Sample project details are given later. You can use the BDAA slack project channel to form project teams and also create a private group for your team if you like.

3.3.1 Accessing Hadoop Cluster

If your project is going to use Hadoop Cluster, the next page will be useful. You will get a Hadoop Cluster on FutureSystems with your fg-475 membership.

- Hadoop Cluster Access

3.3.2 Supported Software Projects

- TeraSort Hadoop Benchmark

- Stock Analysis with MPI

- IPython Notebook

- NIST NBIS

- DFSIO Distributed I/O Benchmark

- MRBench MapReduce Benchmark

- NNBench NameNode Benchmark

3.3.3 OpenStack Kilo FutureSystems (IaaS)

If you need to use the latest OpenStack Cloud, consider to use OpenStack Kilo on FutureSystems. The software of the supported projects are available in OpenStack Kilo as virtual machine images. You may want to start your project by creating your own virtual instance with the image. The details are:

- OpenStack Kilo

3.3.4 Sample Software Projects or Term Paper

We provide a few examples of projects and papers. Please use them as your references.

- Term Paper
- Sample Software Project