

New York University
Computer Science Department
Courant Institute of Mathematical Sciences

Course Title: Cloud Computing
Instructor: Jean-Claude Franchitti

Course Number: csci-ga.3033-026
Session: 2

Assignment #2

I. Due

February 17, 2022 at the beginning of class.

II. Objectives

1. Understand how to leverage IaaS to manage data and compute in the Cloud in a Cloud application development context.

III. References

1. Slides and handouts posted on the course Web site
2. Textbook Part I.1 (Cloud, Big Data, and Cognitive Computing – Principle of Cloud Computing Systems)

IV. Software Required

1. Microsoft Word.
2. Win Zip as necessary.

V. Assignment

Part I: Manage data in the Cloud

1. Use the Azure Portal to create a storage account.
2. Download a CSV file and some data objects from a local store.
3. Push the data objects as blobs into the storage account.
4. Create an Azure NoSQL Table
5. For each line in the CSV file create a line in the table with a pointer (URL) to the corresponding blob.

6. Visualize your work in the Azure Storage Explorer.
7. Adapt and repeat steps 1-6 on AWS, GCP
8. Extra Credit: Adapt and repeat steps 1-6 on IBM Cloud.

Part II: Install the Linux Data Science VM

1. Create a password protected private and public key
2. Install and deploy the Linux DS VM using your public key
3. Download X2GO and use it to visit your VM
4. Run Jupyter on the VM through the jupyter hub

For Part I and Part II above:

1. Document the complete associated processes in a document with screenshots.
2. Save the file as a Word document.
3. Name the file “**firstname_lastname_hw_2.doc**” (e.g., “john_doe_hw_1.doc”).
4. Submit your assignment electronically via NYU Classes by the due date.

Use the following naming convention in the subject line of the eMail:

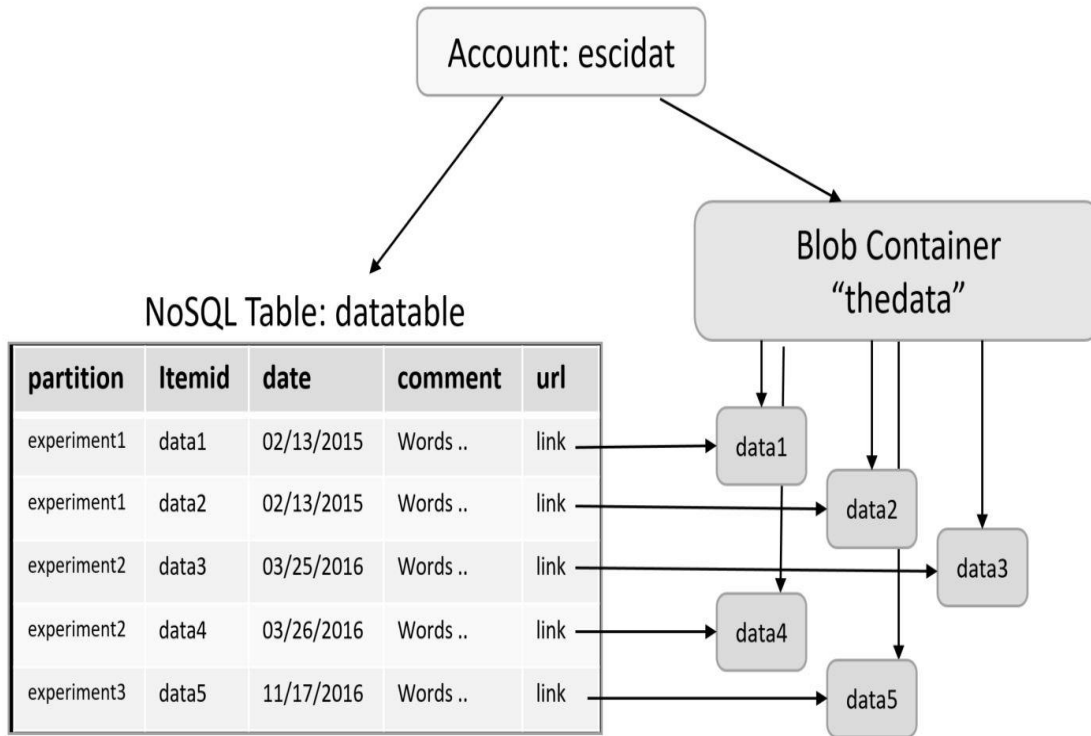
“CC - firstname lastname - homework #”
(e.g.: “CC – John Doe - homework 2”).

In the case source code is submitted, include your name as a comment at the top of each file (note: all files submitted should include your name).

VI. Sample Steps for the Microsoft Azure Cloud

Part I:

What you are doing is build a structure like the picture below:

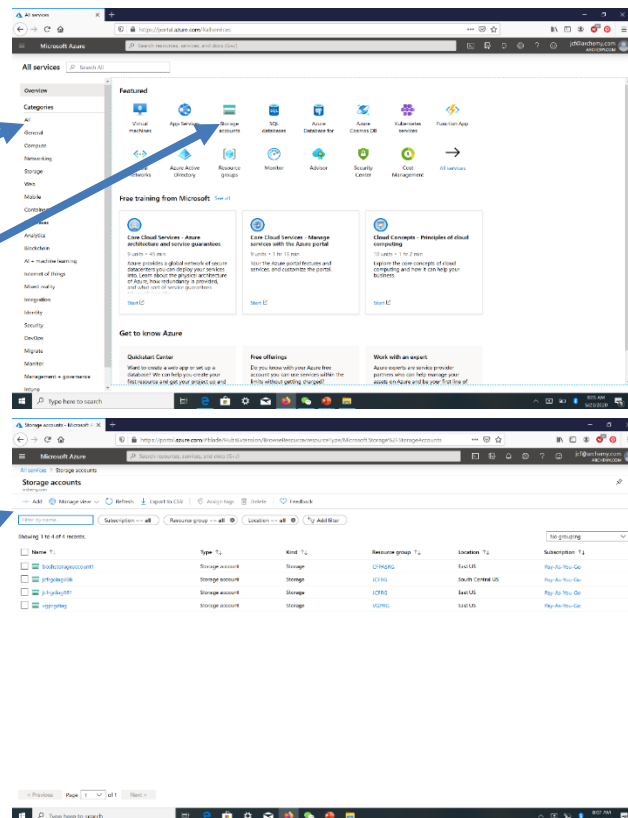


STEP 1: Got to the Azure Portal

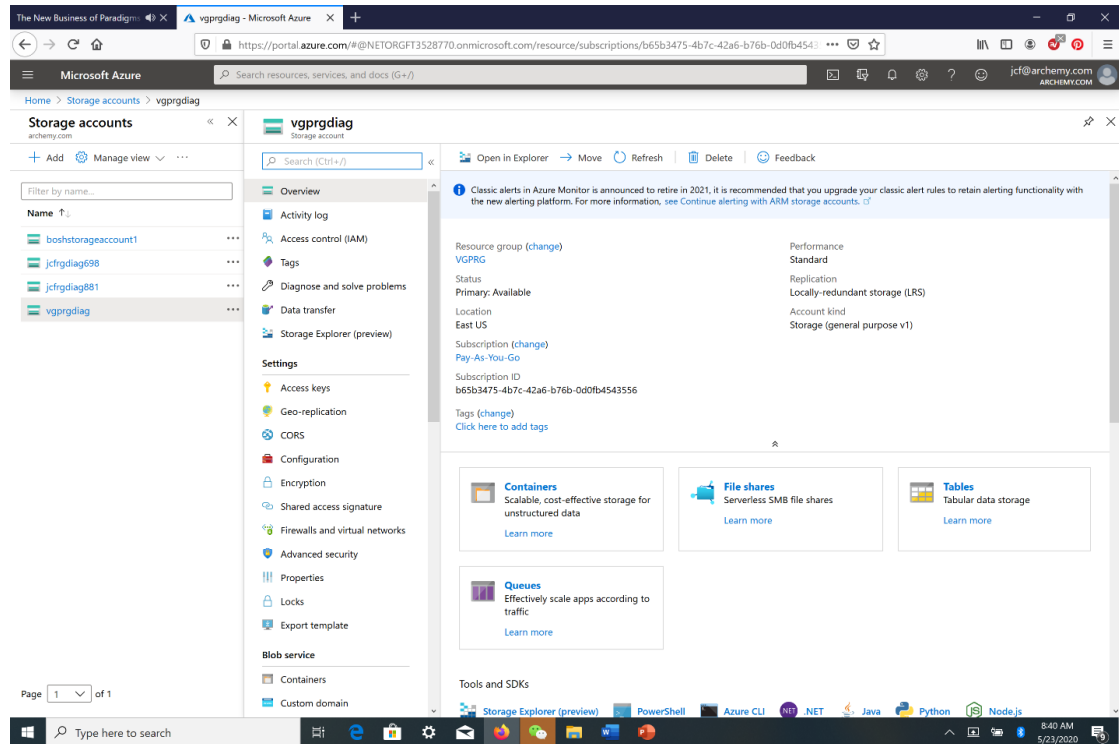
Selecting "All" gives this list of services.

Selecting "Storage Account" gives the secondary menu of types of storage apps.

To create a storage account select Add. Then give it a name, location, etc.



- While you are in the Azure portal, you should grab a copy of the access key.
- Clicking on the blue “Microsoft Azure” in the upper left corner takes you to a top level view of your resources.
- Look for your new storage account and click on that name to display a view similar to the one below.
- Click on the “Access key” button and copy the 2nd access key and save it somewhere like in notepad.



- Make sure you have docker installed on your machine. You can download Docker for you PC or Mac from <https://docs.docker.com/engine/installation/>
- Then run the following Docker command, which will take a while:

```
docker run -i -t -p 8888:8888 dbgannon/tutorial
```
- Once the above command completes, go to <https://localhost:8888> (you will need to add security exceptions in the browser, it is safe). The password is “tutorial”
- Open azure.ipynb in Jupyter (or, if using a different jupyter, use azure.ipynb from the tutorial tarball under “notebooks”).

STEPS 2-5:

- These steps are in the Jupyter notebook. The first cell is a comment. You probably should remove the # symbol and turn it into a command and run it to make sure that the azure python modules are installed.
- The first steps in the notebook create a “container” in your blob storage account.
- The next step creates a table called “DataTable” in the same storage account.
- In the next step you will load the CSV file. If you are using the tutorial container, it is local. If you are not running the tutorial container, you will need to load it. The notebook has two commented commands. Remove the # mark and execute that box and it should download the needed file.
- When you are done working through the notebook look at the table in the Data Explorer.

Part II:

STEP 1: Obtain a public-private key pair

- If you have a mac or linux do this:

```
>ssh-keygen
```

And follow the instruction. This will generate two files. One is your private key and the other has the extension .pub. This is the public key you will upload to your VM.

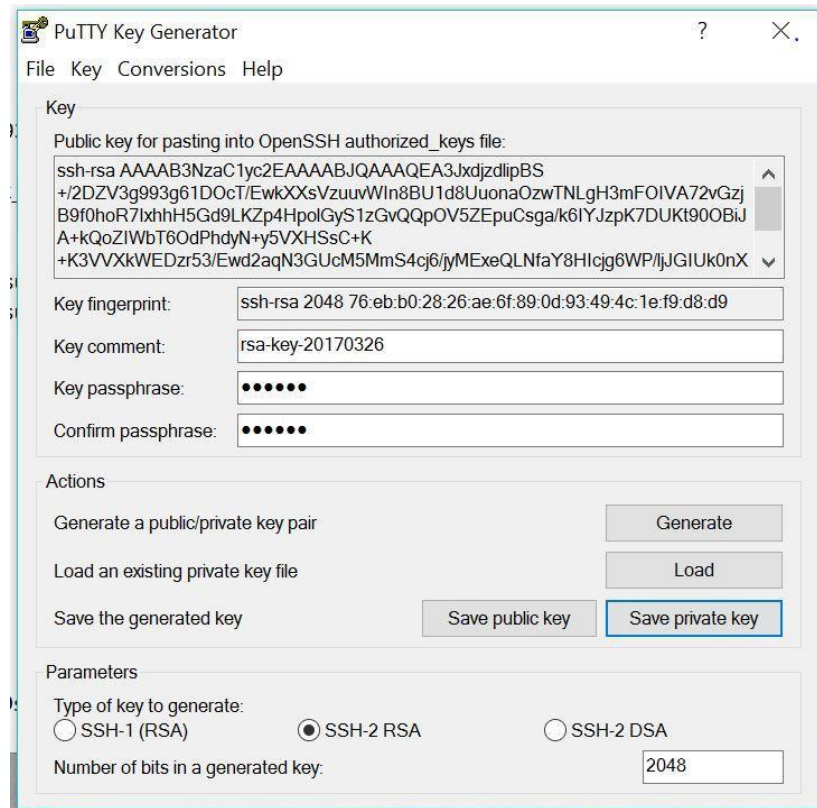
- If you are on a windows machine do this:

Download and Install PuTTY. From www.putty.org

This should contain two programs. PuTTY and PuTTYgen.

Run PuTTYgen.

After a few mouse moves you will see

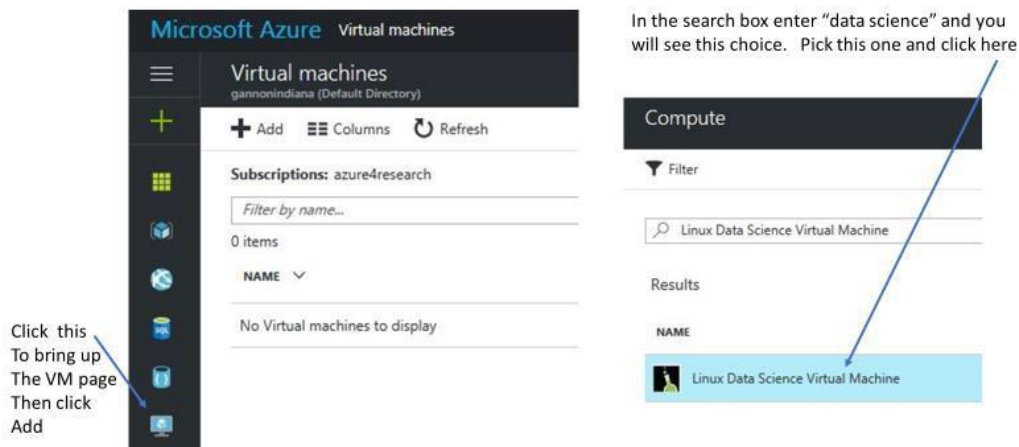


Give it a password (key passphrase) and save the public and private keys.

STEP 2: Install the Linux Data Science VM

- Using your Azure account connect to the portal and sign in.

Using the Azure portal to create a VM



Create virtual machine

Basics

1 Basics
Configure basic settings

2 Size
Choose virtual machine size

3 Settings
Configure optional features

4 Summary
Linux Data Science Virtual Machine

5 Buy

* Name
myDataScienceVM

VM disk type
SSD

* User name
dbgannon

* Authentication type
SSH public key Password

* SSH public key
----- BEGIN SSH2 PUBLIC KEY -----
Comment: "rsa-key-20170210"
AAAAB3NzaC1yc2EAAAABJQAAAQEAi+S
oqE+zhRcAt8wsF31YDgpwTQSnVMwQ5c

Subscription
azure4research

* Resource group
☐ Create new ☒ Use existing
bookRG

Location
South Central US

OK

- Next do the basic configuration as shown below and give your VM a name, give yourself a user ID, copy and paste your public key, create a resource group (“create new” if you do not have one), and select a location (South Central US should work or you will be able to tell which regions you are allowed to use).

- If everything has a green check mark then click OK.
- You next must select a server type for this VM. You will get three choices

DS2_V2 Standard	DS3_V2 Standard	DS14_V2 Standard
2 Cores	4 Cores	16 Cores
7 GB	14 GB	112 GB
4 Data disks	8 Data disks	32 Data disks
6400 Max IOPS	12800 Max IOPS	50000 Max IOPS
14 GB Local SSD	28 GB Local SSD	224 GB Local SSD
Load balancing	Load balancing	Load balancing
Premium disk support	Premium disk support	Premium disk support
94.49 USD/MONTH (ESTIMATED)	189.72 USD/MONTH (ESTIMATED)	989.52 USD/MONTH (ESTIMATED)

Pick the one you can “afford”. Then go to step 3. Just click OK to accept all the defaults. And finally click “Buy”.

- When the VM starts you should be able to see it in the portal on you dashboard (which you get to by clicking on the big [Microsoft Azure](#) in the upper left corner. You should see:

Connect Start Restart Stop Capture Delete

Essentials

Resource group (change)	Computer name
bookRG	myDataScienceVM
Status	Operating system
Running	Linux
Location	Size
South Central US	Standard DS14 v2 (16 cores, 112 GB mem...
Subscription name (change)	Public IP address/DNS name label
azure4research	13.84.55.19/<none>
Subscription ID	Virtual network/subnet
f518fe6b-5262-4e5a-80cb-05b7a39f9298	bookRGvnet771/default

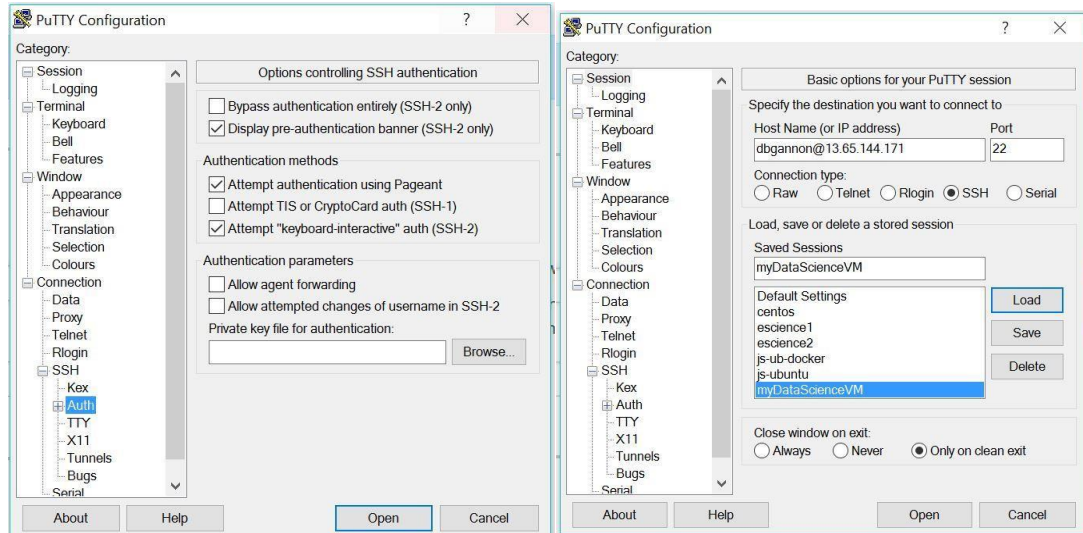
Monitoring

- Make note of the IP address. You should be able to log into your VM with your private key as follows:

For the Mac or Linux type

```
>ssh -i privatekey youruserid@ipaddress
```

For Windows run Putty. You will need to upload your private key into the putty client:



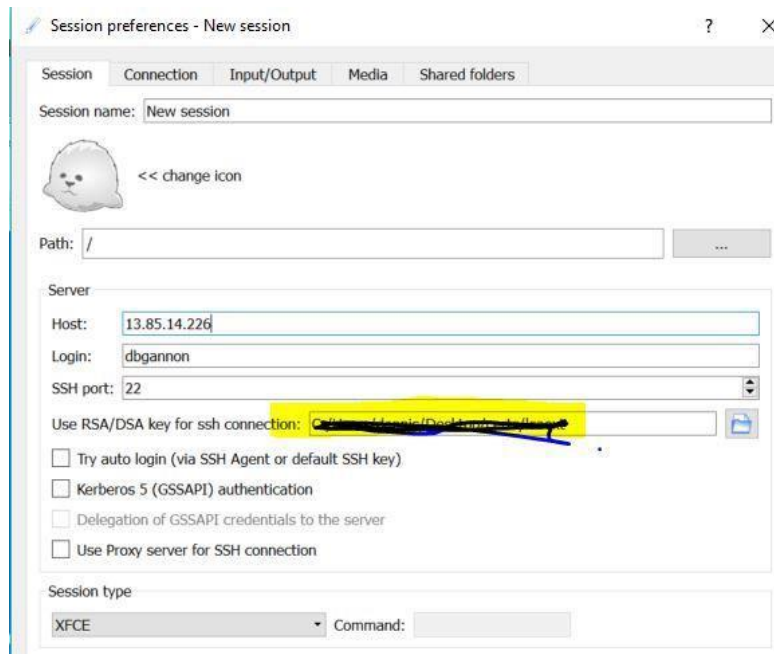
- Then return to the session tab and enter your userid@yourVM-IP-address. You can give it a name and select “save”. This will save the configuration for next time.
- Next select “open”.
- While you are logged into the VM, you should also set your local password. Type:

```
> sudo passwd userid
> enter your password twice
```

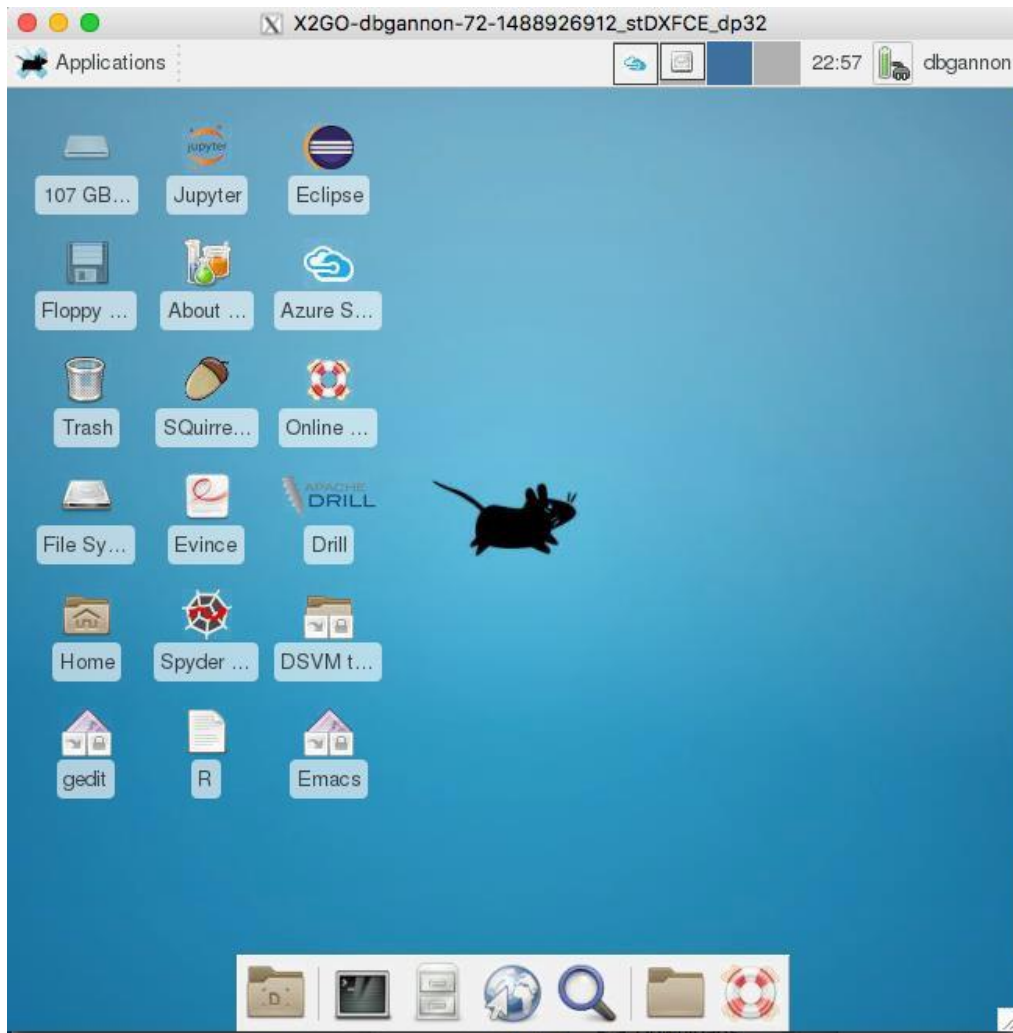
We will use this later.

STEP 3: X2Go

- Download X2Go (<http://wiki.x2go.org>) and install it.
- Create a new session.
- You will need your IP address and login id and the path to your secret key. And set the session type XFCE



- You will need to give it the password to unlock your secret key. If you are running windows it may protest but just wait. It should come up with the desktop.



- From here you can launch a terminal window, manage files, and use a web browser by looking in the bottom row of icons. There are many other tools.

STEP 3: Run Jupyter

- The system is already running a tool called jupyter hub that will allow you to log in and start a jupyter session in your browser. Go to <https://youvm-IP:8000>. Login with your userid and password. There are some great examples here.

VII. Deliverables

1. Electronic:

Your assignment file must be submitted via NYU Classes. The file must be created and sent by the beginning of class. After the class period, the homework is late. The email clock is the official clock.

2. Cover page and other formatting requirements:

The cover page supplied on the next page must be the first page of your assignment file.

Fill in the blank area for each field.

NOTE:

The sequence of the electronic submission is:

- 1. Cover sheet**
- 2. Assignment Answer Sheet(s)**

3. Grading guidelines:

Assignment Layout (15%)

- o Assignment is neatly assembled on 8 1/2 by 11 layout.
- o Cover page with your name (last name first followed by a comma then first name), username and section number with a signed statement of independent effort is included.
- o File name is correct.

Answers to Individual Questions (85%):

- o Answers to questions in V. Part I and Part II are correct.
- o Assumptions provided as required.

(100 points total, all questions weighted equally)

VIII. Sample Cover Sheet:

Name _____ Date: _____
(last name, first name)
Section: _____

Assignment 2

Total in points (100 points total): _____

Professor's Comments:

Affirmation of my Independent Effort: _____
(Sign here)