Final Project Tips and Guidance

by
Diane Howard, Nishava Inc.

Deep Azure @McKesson

Overview

- **Final Project Breakdown**
  - How to approach the Final Project
  - Helpful Tips
  - Useful links for data sets & editors
  - Skeleton Report, One Page Summary, Slides
  - Creating YouTube Account, Video tools, Upload
  - Certificates & Grading Criteria
  - 2 Sample One Page Summaries
  - Your Questions

Final Project Approach

1. Research your technology. Try the demos.
2. Define a problem statement to demonstrate your AZ technology.
3. Select a data source (There are many data sets available online).
4. Develop/implement/code a simple demonstration to solve your problem using a programming language of your choice: .NET, Java, Python, C.
5. Produce a visualization of your data/results using D3, R, Python, Matlab, JS Framework or graphical tool (Tableau).
6. Document your work (slides, one page summary, report). Report is very detailed showing all steps: install, config, runs, results (to include graphical results).
7. Create 2 YouTube videos (2 minute, 15 minute). Show your demo in your 2 minute PLEASE! Creating a Youtube video can take up quite a bit of time!

Helpful Tips

- Demonstrate AZ technology similar to a homework example that solves your defined problem.

- On-line demos / examples are EXCLUDED. You may use AZ on-line code but you must modify it. Also do NOT use Hello World type of examples.

- AZ limitations …
  1. AZ SDKs are LIMITED so may need to also use AZ CLI, Power Shell or Portal.
  2. AZ CLI is not considered software code to meet s/w requirements. Some code & portal/power shell set up = Excellent work!
  3. Generally AZ technologies are dependent on other AZ services for set up so get these dependencies working ASAP.
  4. Use Windows 10 (avoid Windows 7 as SDKs are NOT supported!)

  Other…
  1. There may be firewall issues if using laptops @ work.
  2. Avoid using 32-bit laptops for VMs. Use AZ VMs if you have a 32-bit laptop.

More Helpful Tips

- Give reference to any documentation & source code found on the web.
- Put all code in your report too. DO NOT REMOVE licenses within code (such as APACHE). This is plagiarism.
- Every step must be documented in your report. Ensure your project demo is reproducible.
- Do not include your AZ security identifiers/keys in your reports.

- Zip files:
  - Please keep zip file sizes < 20 MB please. We are limited on our web site space usage.
  - VMs or Docker files should not be included in your zip file.
  - Do NOT include binaries in your zip file.

- SHOW YOUR DEMO in your 2 minute YouTube presentation. You can use your slides to introduce topic but do show your demo to include your output results and graphics.

Repositories for public data sets

- Microsoft Azure: https://docs.microsoft.com/en-us/azure/sql-database/sql-database-public-data-sets
- Amazon: https://aws.amazon.com/public-data-sets/
- Mobile Analytics Data from Microsoft Research collection:

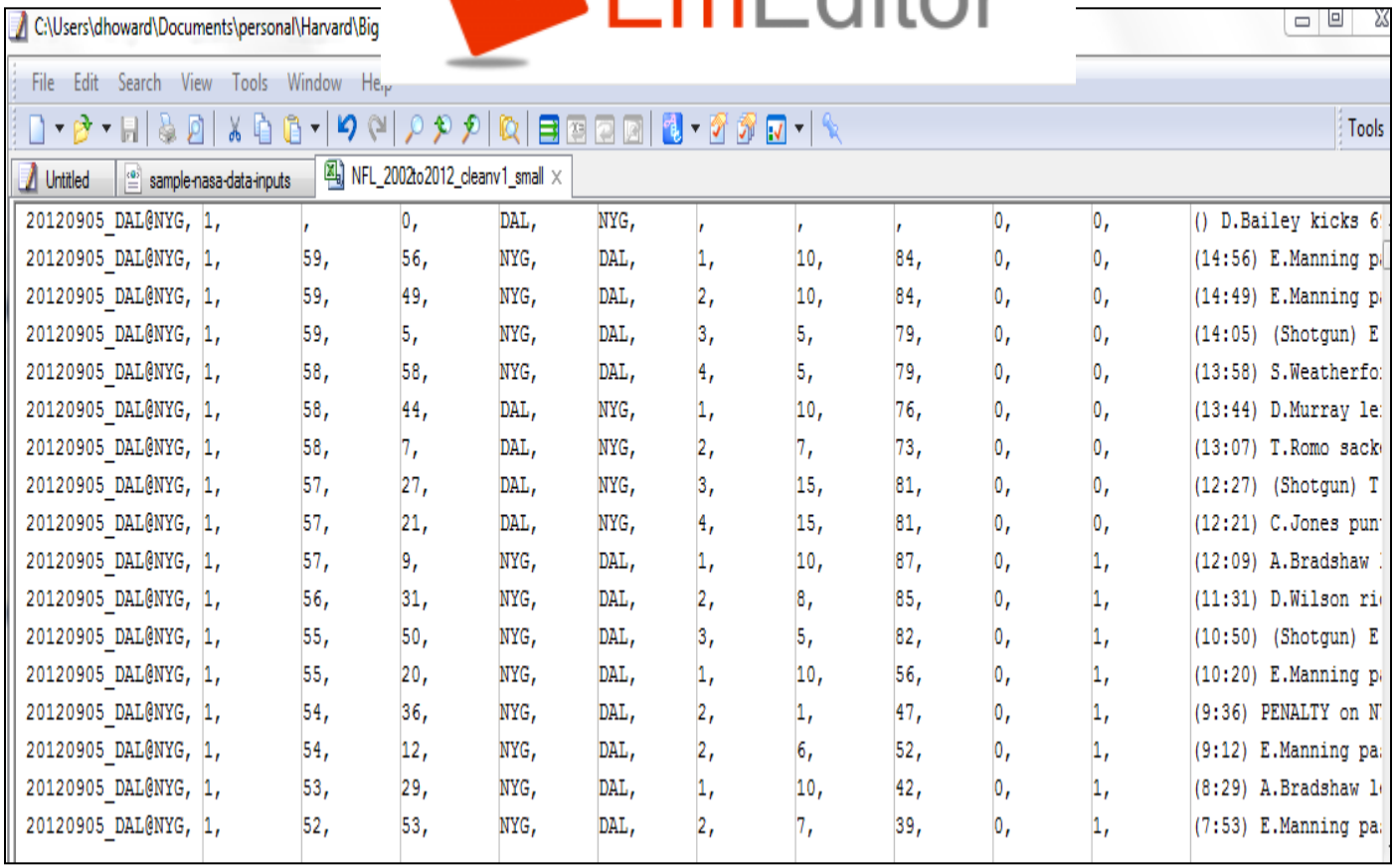
https://www.microsoft.com/en-us/research/project/mobile-data-collection/

- U.S. government: https://www.data.gov/
- Kaggle: https://www.kaggle.com/datasets

Give yourself enough time to understand and clean data!

Large Data File Editor for Windows

EmEditor http://www.emeditor.com/



Large Data File Editor for Mac, Linux

vim http://www.vim.org/

Example Problem Statements

Problem Statement:

Build a job posting application that utilizes Azure's Service where you can search a person's resume by first or last name or key words related to the job skill set.

Below is missing what will be solved. Describe a problem to solve using the technology.

An introduction to Microsoft Azure's Search service with a walkthrough of how to use it.

Please do not miss the boat on defining your problem statement properly.

More Examples of Problem Statements

Problem: Predict future health care costs using the Spark Machine learning libraries by analyzing data from the Centers for Medicare & Medicaid Services in 2008-2010 to predict cost savings based on the number of times a patients has visited doctors and their ailments.

Problem

There are billions of electronic devices. Homes are getting more and more smart – with the evolution of IPv6 and smart gadgets, all the electronics at home (for example, washer, toaster, fridge, electrical switches, electrical outlets, dryer etc) are getting connected to the internet via a hub device located at the residence.
In this project we use a combination of technologies (sqlite3, python, Cassandra, High Charts, High Maps etc) to visualize appliance devices across the state of Pennsylvania from data provided at pennstate.gov to determine what and where are the most frequently used internet based appliances used in homes.

Describe:
What you are solving & why + data you are using + technology
EQUALS
A Well Defined Problem Statement

Skeleton for Full Report

Problem Statement

Description of Data & provide URL where you got your Data set

Description of Hardware

Description of Software (Provide URLs of downloads)

Describe **installation** & **configuration** (if needed) steps. Capture all screens!

Show **example data file**.

Describe **cleansing of data** you had to do. Show a **sample of your data**.

Show **upload of data & any transformations.**

Put **all your code in your document**. Describe it. Show packages/libraries used.

Show how to compile if needed. Run code.

Show results and visualization.

Summarize: Lessons Learned: Describe Issues / Benefits (Pros & Cons), what did you like/not like. What would you do next?

References: Cite your URLs from code or technical info from online sources.

Skeleton for One Page Summary

Define your Problem Statement:

Overview of Technology:

High Level Overview of steps:

1. Installed (list URLs)
2. Configured
3. Cleaned data
4. Ingested data
5. Ran JAVA code

Describe Data Set:

URL: http://ita.ee.lbl.gov/html/contrib/NASA-HTTP.html

Size: 205.2 MB, sample size: 1MB

Format of data file: apache-access log

Hardware:

Mac OS x running VMWare with Hortonworks HDP 2.2 Sandbox

Software:

VMWare Fusion 7.1.1

Logstash 1.4.2 release (https://download.elasticsearch.org/logstash/logstash/logstash-1.4.2.tar.gz)

ElasticSearch 1.4.2 release (https://download.elastic.co/elasticsearch/elasticsearch/elasticsearch-1.4.2.tar.gz)

Fluentd 2.1.4 release (http://packages.treasuredata.com/2/macosx/td-agent-2.1.4-0.dmg)

Hortonworks HDP 2.2 Sandbox for Kafka 0.8.0 environment

References:

Lessons Learned & Pros/Cons

YouTube URLs:

Skeleton for PowerPoint Presentation

- 10-30 slides
- Use template provided. Only white background slides please.

Skeleton

- Problem statement
- Description of AZ Technology and s/w Description of Data
- Overview of Install/Config/Set up
- Demonstration (Results and Visualization)
- Summarize: Lessons Learned & Pros/Cons
- YouTube URLs (last slide)

YouTube Video Editing Software

You will need video editing software: Camtasia (Windows) Note: For demo version a logo will appear and can't be removed!

https://www.iskysoft.com/video-editing/free-video-editing-software-for-youtube.html

Windows 10:

- Open the app you want to record. ...
- Press the Windows key and the letter G at the same time to open the Game Bar dialog.
- Check the "Yes, this is a game" checkbox to load the Game Bar. ...
- Click on the Start Recording button (or Win + Alt + R) to begin capturing video.

Mac:

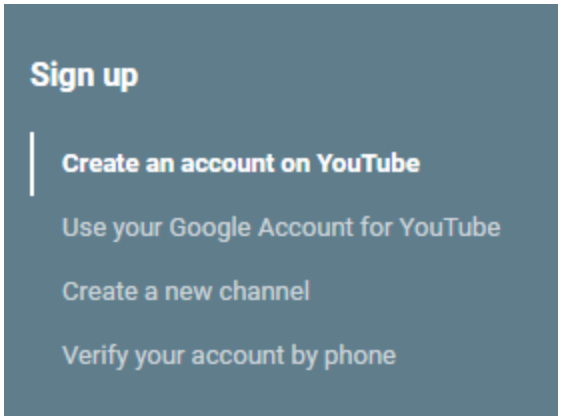
The easiest way to **record video** on your **Mac** is with the built-in QuickTime app.

- Open your Applications folder to find QuickTime.
- Once it's open, go to File > New Screen Recording and then click the **Record** button. You can choose between recording a portion of your screen or the entire screen.

How to create a YouTube Account

https://support.google.com/youtube/answer/161805?co=GENIE.Platform%3DDesktop&hl=en

1. Need a gmail account
2. Go to youtube.com.
3. In the top right, click **Sign in**.
4. Click **More options > Create Account**.



How to upload your video to YouTube

https://www.youtube.com/upload

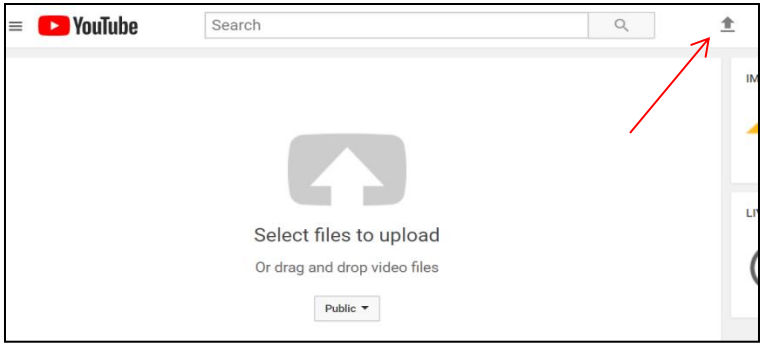
Upload videos

You can upload videos to YouTube in a few easy steps. Use the instructions below to upload your videos from a computer or from a mobile device.

COMPUTER ANDROID IPHONE & IPAD

1. Sign in to YouTube.
2. Click on **Upload** at the top of the page.
3. Before you start uploading the video, you can choose the video privacy settings.
4. Select the video you'd like to upload from your computer. You can also import a video from Google Photos.
5. As the video is uploading, you can edit both the basic information and the advanced settings of the video and decide if you want to notify subscribers (if you uncheck this option, no communication will be shared with your subscribers). Partners will also be able to adjust their Monetization settings.
6. Click **Publish** to finish uploading a public video to YouTube. If you set the video privacy setting to Private or Unlisted, just click **Done** to finish the upload or click **Share** to privately share your video.
7. If you haven't clicked **Publish**, your video won't be viewable by other people. You can always publish your video at a later time in your Video Manager.

Once the upload is completed we will send you an email to notify you that your video is done uploading and processing. You can then forward that email to friends or family for easy sharing. If you prefer not to receive notification, you can opt out by visiting your email settings.



YouTube Hints

- 2 minute (short & sweet)
  - 1-6 slides on Problem Statement & quick overview of technology, data set.
  - **Show your demo, show your demo , show your demo – show output results and visualization.**
  - No need to let us know that you are doing this for McKesson. (i.e., leave out extraneous info PLEASE!)
  - Do NOT point us to your 15 minute YouTube to see your demo. ☹
  - Add a quick one-liner: what did you learn, like, didn't like?
  - Be aware of the cadence of your delivery – *not slow!*
- 15 minute (of course this has more details)
  - **Describe your technology & code! Show your demo! Show your demo!**
  - **Add your thoughts with Pros/Cons/Lessons Learned, what would you do next?**

Software Development

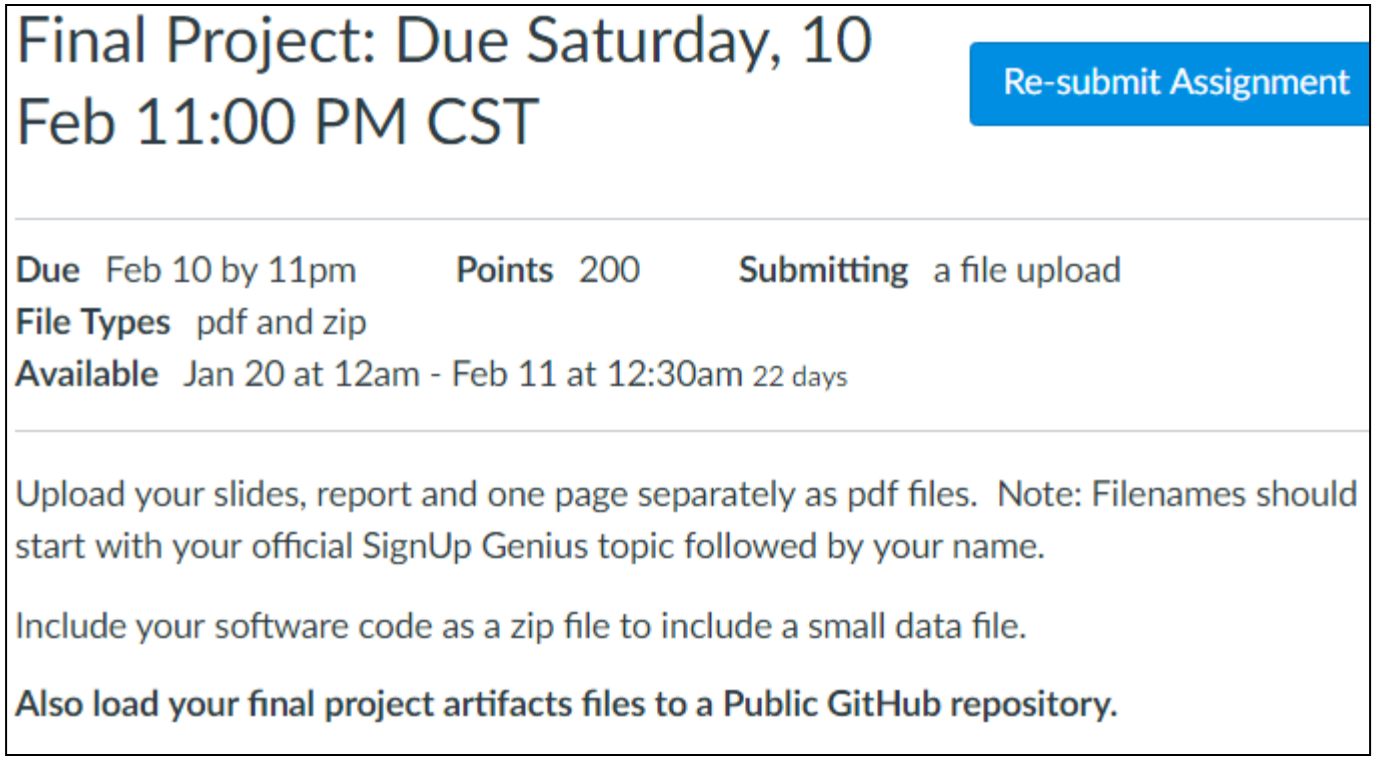
- Use a programming language of your choice: **.NET , Java, Python, C.**
  - Other choices are fine too! Examples: R, GO, Matlab.

- If you have a tool that is mostly a GUI – you must include a software programming aspect of your demo. For example: ingest, parse, transform data.

- Produce a visualization of your data/results using **D3, R, Python, html, CSS or other JS Framework (or R or Matlab!)**

Submission Upload Requirements

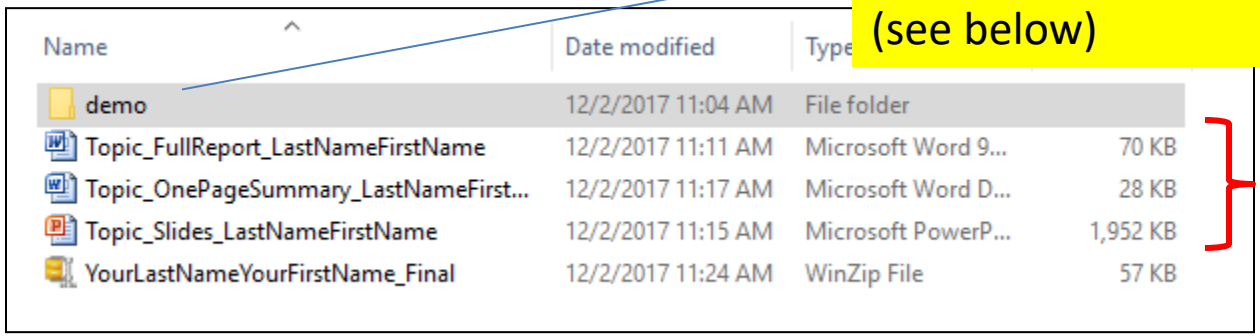
- Canvas submission (pdf of report, summary and slides, zip of code and data).
- Also post at URL of the public GitHub repository .



Organization of Zip file

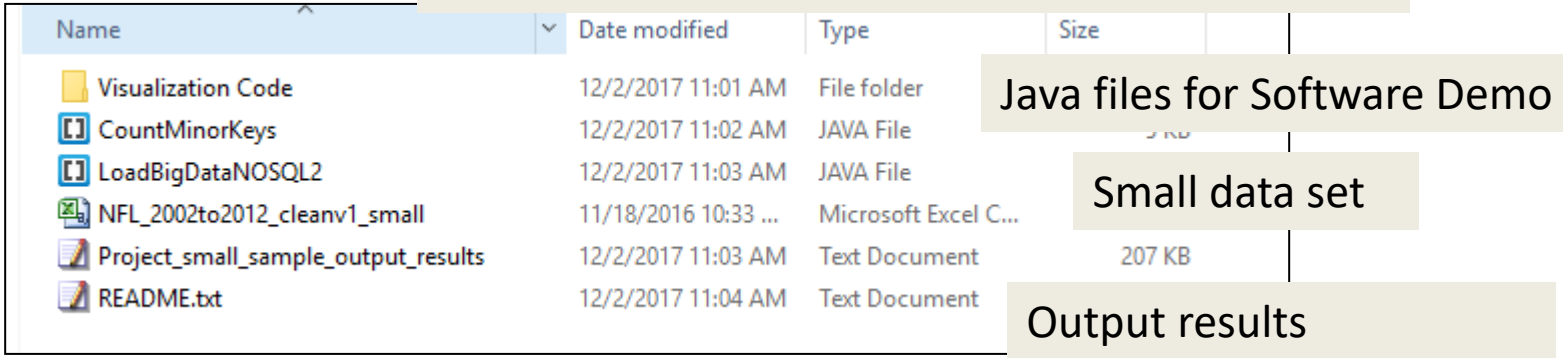
ZIP file contains:

Separate zip file of code (see below)



Report

Visualization code



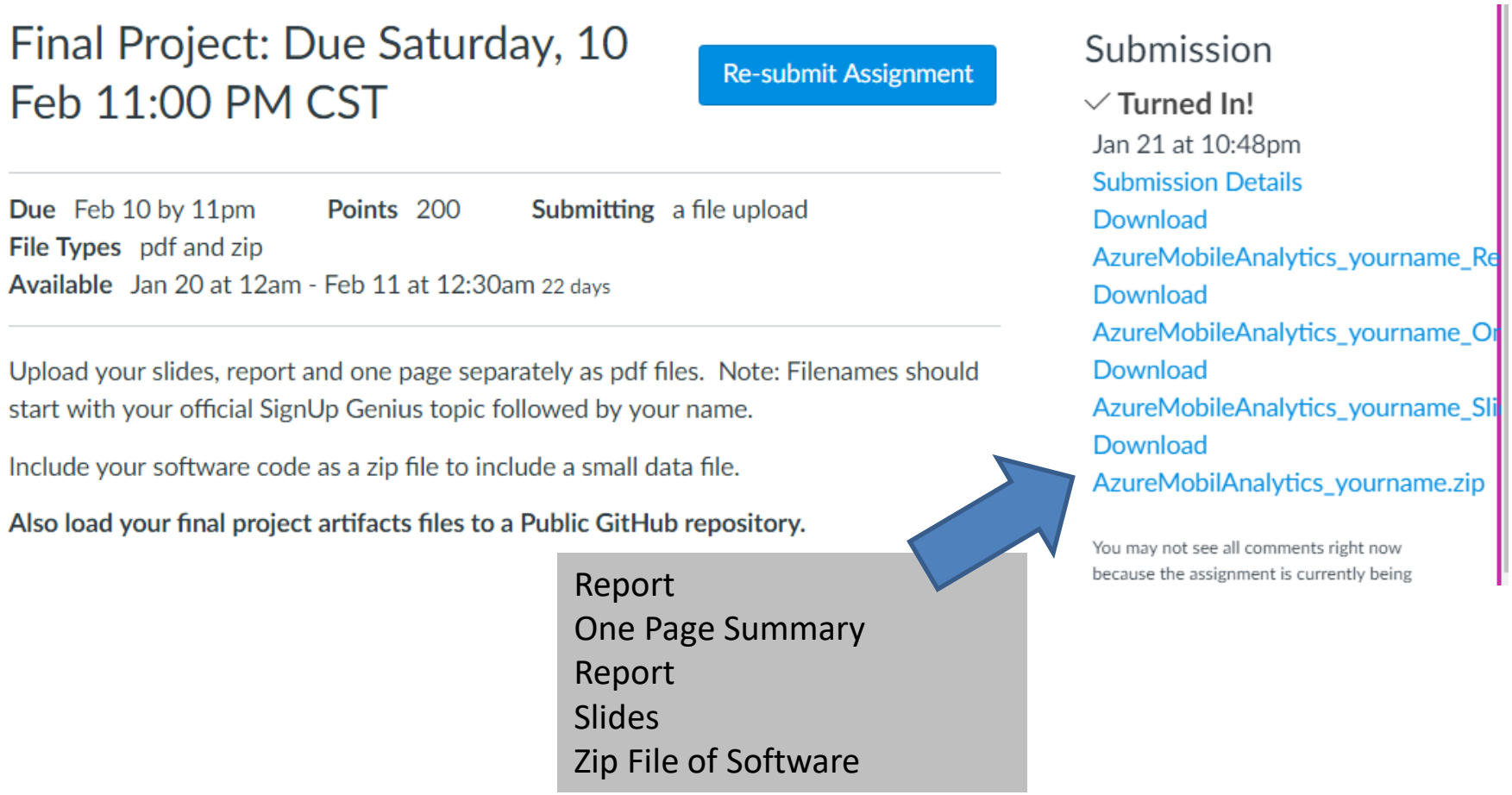
Java files for Software Demo

Small data set

Output results

Submission in Canvas

- This is how your Submission Files should be organized in Canvas.
- Deductions will be taken if it is not organized this way.



Grading Criteria

Grading criteria:

Project Report and practical software code example	50%
PowerPoint Slides	20%
15 minute YouTube video	15%
2 minute YouTube video	10%
One page summary	5%

If you fail to provide practical software code example, you will lose at least 30% of the final project grade.

If you fail to provide practical software code example, you will lose at least 30% of the final project grade.

Deductions are applied…

- Not a well defined problem statement.
- Missing steps in your report.
- Code is only config files/bash shells. S/W demo is trivial: Hello World type example.
- Used a GUI for visualization and no code involved.
- Discretionary deductions – did not submit files as requested in submission, zip file is too large, TA had to request additional info, run code, difficult to follow, missing code & results in report, depth and breadth (how many technologies did you use).

Presentation Selections

- Two minute Youtube presentations will start 6:00 PM CST on 13 Feb. End at 8:00 PM.
- We will review YouTube presentations to determine who will be presenting. A list of presenters will be sent to all students day before presentation. Please be present if you are selected.
- Zoran/TAs introduce you & your Topic via Zoom and we will start the YouTube video. Just be online in Zoom for our introduction and questions from the audience.
- We present the YouTube Videos by Topics.
- If you are done early with your project you may present your YouTube on Feb 6. Please notify Zoran
- Invite your management to watch!

Certificates

- If all homework is caught up and B average Zoran will recommend a list of students for the Azure Certificate.

- If you would like to still catch up on home works there is plenty of time… Contact Diane via Canvas Inbox.

Words of Wisdom

- *Wrap up Homework 11 &12 immediately .* Homework 13 and 14 are optional. Move to your Final Project!
- Scope your effort properly.
- Give yourself plenty of time to clean your data set from the web. This may take up to 3 days!
- Put emphasis on solving the highest percentage of points first (s/w demo & report = 50%). Then do your summary page, slides, YouTubes.
- YouTube videos take much longer than you think so plan accordingly and practice. Please speak up on your videos. We prefer to see your demo and not you.
- **HAVE FUN!**

BREAK POINT TO SAMPLE ONE PAGERS

QUESTIONS?