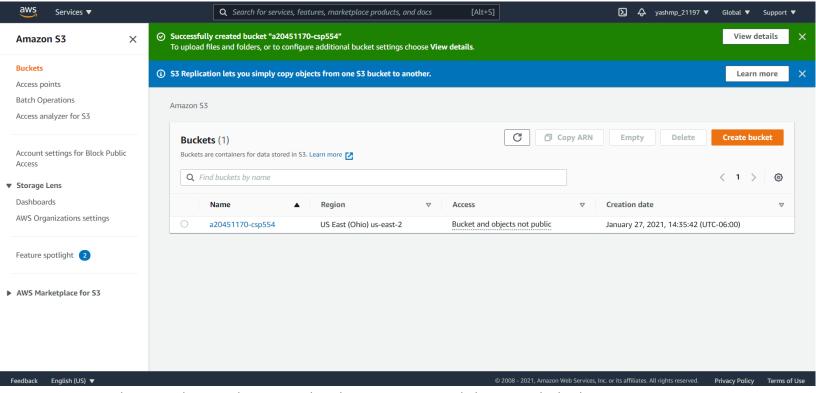# CSP 554—Big Data Technologies

Assignment #1 (Modules 01a & 01b, 12 points)

1. (2 points) Submit very brief answers (or bullet points) to the following questions:
   - What location or time zone are you in when you attend the course?
     - When I attend the course, I am in Chicago, USA.
   - Describe any prior experience you might have with use of public cloud, data mining, machine learning, statistics, data science and big data.
     - I have taken machine learning, introduction to artificial intelligence, natural language processing, probabilistic graphical models, computer vision, online social network analysis courses.
     - Hands-on experience in firebase.
   - Share any big data interests and personal learning goals for the course.
     - Want to learn technologies like Kafka, spark, hive, etc.
     - Want to learn about architectures like the Hadoop file distribution systems.
   - Indicate if there are additional topics in the scope of the course of special interest to you.
     - I have an interest to learn presto.
   - Do you have any anticipated personal issues such as expected absences or other necessary accommodations with course impact? (Of course, these will be held in strictest confidence.)
     - No, I don't have any as of now.

2. (5 points) Answer each of the following questions about the article in just one to three sentences each:
   o What was the problem with the Google flu detection algorithm?
     - In contrast to the centers for disease control and prevention, Google flu detection algorithm predicted about twice as many doctor visits for influenza-like diseases as the Centers for Disease. Algorithm Dynamics and Big Data Hubris are the two problems that add up.
   o What is big data hubris?
     - Big data hubris is the tacit belief that big data is a replacement for conventional data collection and analysis, rather than a complement.
   o What approach could have been used to improve the Google flu detection algorithm?
     - By integrating real-time health information and dynamically recalibrating Google flu detection algorithm, the Google flu detection algorithm could have been improved.
   o What is "algorithm dynamics"?
     - Algorithm dynamics refers to the modifications made by engineers with a view to enhancing the commercial service and the use of the service by customers.
   o What aspect of algorithm dynamics impacted the Google flu detection algorithm?
     - Media stock panic is a factor that has disrupted monitoring by the Google flu detection algorithm. Another one is a search algorithm of Google. Modifications

added to provide users with valuable knowledge quickly and this suggested search affects the estimate of the Google flu detection algorithm.

3. (5 points) Set up an Amazon Web Services (AWS) cloud account, if you don't already have one (see below for details), and then follow the tutorial about how to work with a storage service called S3. Since we will do most of our assignments using AWS, this will get you started. In a while we will come to understand S3 as one critical element of a big data processing architecture know as the "data lake."
   a. To receive credit for this question, provide a screen shot showing the S3 bucket you have created. The bucket name should be named something like "YourIITId-CSP554", for example: "A1234567_CSP554"

   b. Now also provide a screen shot showing some named object is in the bucket.

⊘ Upload succeeded
View details below.

Upload: status

Exit

ⓘ The information below will no longer be available after you navigate away from this page.

Summary

Destination	Succeeded	Failed
s3://a20451170-csp554	⊘ 1 file, 199.9 KB (100.00%)	⊝ 0 files, 0 B (0%)

Files and folders Configuration

Files and folders (1 Total, 199.9 KB)

Find by name

‹ 1 ›

Name ▲	Folder ▽	Type ▽	Size ▽	Status ▽	Error ▽
AWS01 v2.pdf	-	application/pdf	199.9 KB	⊘ Succeeded	-

Amazon S3 ✕

Buckets
Access points
Batch Operations
Access analyzer for S3

Account settings for Block Public Access

▼ Storage Lens
Dashboards
AWS Organizations settings

Feature spotlight 2

▶ AWS Marketplace for S3

Amazon S3 > a20451170-csp554

a20451170-csp554

Objects Properties Permissions Metrics Management Access points

Objects (1)

Objects are the fundamental entities stored in Amazon S3. For others to access your objects, you'll need to explicitly grant them permissions. Learn more ↗

⟳   Delete   Actions ▼   Create folder   **Upload**

Find objects by prefix

‹ 1 › ⚙

	Name ▲	Type ▽	Last modified ▽	Size ▽	Storage class ▽
☐	📄 AWS01 v2.pdf	pdf	January 27, 2021, 14:40:40 (UTC-06:00)	199.9 KB	Standard