

CSP554 BIG DATA TECHNOLOGIES ASSIGNMENT-1

1. What was the problem with the Google flu detection algorithm?
2. What is Big Data Hubris?
3. What approach could have been used to improve the Google flu detection algorithm?
4. What is “algorithm dynamics?”
5. What aspect of algorithm dynamics impacted the Google flu detection algorithm?

What was the problem with the Google flu detection algorithm?

- Despite GFT was built to predict CDC reports, data based on surveillance reports from labs across US. But, because of algorithm dynamics, it was predicting more than double the proportion of doctor visits for influenza like illness (ILL) than CDC.

What is Big Data Hubris?

- It is an assumption that big data is substitute for, rather than a supplement, traditional data collection and analysis.

What approach could have been used to improve the Google flu detection algorithm?

- **It is not always about the size of the data**, GFT has heavily dependent on its google search algorithm and the enormous **BIG DATA**, though it provides a broader view of things sometimes traditional “small data” often provides information that is not contained in the big data. So, using data from both the traditional and new sources could improve the flu detection algorithm.

What is “algorithm dynamics?”

- Algorithm Dynamics are the changes made by the engineers to improve the commercial service and by consumers in using that service.

What aspect of algorithm dynamics impacted the Google flu detection algorithm?

- In the process of improving its service to consumers, Google is continuously changing the data-generation process so as to support its business model, by providing useful information to the users quickly along with increasing the ad revenue by promoting more related advertisements.

AWS ACCOUNT :-

STEP 1:- CREATING A S3 BUCKET (a20526972-csp554).

The screenshot displays the AWS S3 console interface. The left sidebar shows the navigation menu with options like Buckets, Access Points, and Storage Lens. The main content area is titled 'Account snapshot' and shows the 'Buckets (1)' section. A table lists the bucket 'a20526972-csp554' with details: AWS Region (US East (N. Virginia) us-east-1), Access (Objects can be public), and Creation date (January 14, 2023, 14:24:27 (UTC-06:00)).

Account snapshot

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

Buckets (1) [Info](#)

Buckets are containers for data stored in S3. [Learn more](#)

[Refresh](#) [Copy ARN](#) [Empty](#) [Delete](#) [Create bucket](#)

	Name	AWS Region	Access	Creation date
<input type="radio"/>	a20526972-csp554	US East (N. Virginia) us-east-1	Objects can be public	January 14, 2023, 14:24:27 (UTC-06:00)

Feedback Looking for language selection? Find it in the new [Unified Settings](#)

© 2023, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

30°F Mostly sunny 2:25 PM 1/14/2023

STEP 2:- UPLOADING AN OBJECT TO THE S3 BUCKET TITLED(a20526972-csp554).

The screenshot displays the AWS S3 console interface for the bucket 'a20526972-csp554'. The breadcrumb navigation shows 'Amazon S3 > Buckets > a20526972-csp554'. The bucket name 'a20526972-csp554' is prominently displayed with an 'Info' link. Below this, a tabbed interface includes 'Objects' (selected), 'Properties', 'Permissions', 'Metrics', 'Management', and 'Access Points'. The 'Objects' section shows a list of objects with the heading 'Objects (1)'. A descriptive text explains that objects are fundamental entities in Amazon S3 and provides a link to 'Amazon S3 inventory'. A toolbar offers actions like 'Copy S3 URI', 'Copy URL', 'Download', 'Open', 'Delete', 'Actions', 'Create folder', and 'Upload'. A search bar is labeled 'Find objects by prefix' and a 'Show versions' toggle is present. The object list table contains one entry: 'Tom White Hadoop_ The Definitive Guide, 4th Edition, 2015 - PDF Room.pdf', which is a 'pdf' file, last modified on 'January 14, 2023, 14:26:49 (UTC-06:00)', with a size of '9.6 MB' and 'Standard' storage class. The bottom of the image shows a Windows taskbar with a search bar, various application icons, and a system tray indicating '30°F Mostly sunny' and the time '2:27 PM 1/14/2023'. A large, light-gray watermark 'A20526972' is oriented diagonally across the lower half of the image.

Amazon S3 > Buckets > a20526972-csp554

a20526972-csp554 [Info](#)

Objects | Properties | Permissions | Metrics | Management | Access Points

Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

[Refresh](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions](#) [Create folder](#) [Upload](#)

[Show versions](#) [Previous](#) **1** [Next](#) [Settings](#)

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	Tom White Hadoop_ The Definitive Guide, 4th Edition, 2015 - PDF Room.pdf	pdf	January 14, 2023, 14:26:49 (UTC-06:00)	9.6 MB	Standard

Feedback Looking for language selection? Find it in the new [Unified Settings](#)

© 2023, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

30°F Mostly sunny Search [Taskbar icons] 2:27 PM 1/14/2023