

CSP - 554
BIG DATA TECHNOLOGIES
ASSIGNMENT - 1

(5 points)

Q4) Answer each of the following questions about the article in just one to three sentences each:

➤ **What was the problem with the Google flu detection algorithm?**

Google's flu detection algorithm had a major overestimation of flu-related clinic visits, more than twice as many as actually recorded by the Centers for Disease Control and Prevention (CDC). This was the main problem in the system. The system used keyword frequency to detect flu patterns, but neglected to take into account the underlying causes of consumers' search behavior, which led to this disparity. Because the model was not updated adequately to reflect changing search behavior and pattern, it was unable to effectively anticipate flu outbreaks, which resulted in projections that were not in line with actual flu activity.

- Additionally, two key issues contributed to this error:

Big Data Hubris: Overconfidence in large datasets led to reliance on search patterns without considering user intent or context, resulting in flawed predictions.

Algorithm Dynamics: The model's dependence on outdated search data and lack of adaptability caused a misalignment with real-world flu trends.

➤ **What is big data hubris?**

The term "big data hubris" describes the arrogant belief that existing approaches to data collecting and analysis can be replaced by big data, rather than used in conjunction with them. It is still necessary to solve fundamental issues like measurement accuracy and data validation, despite the astounding sheer volume of data. Ignoring fundamental principles can result in incorrect conclusions when assuming that greater data inevitably yields better insights.

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

If the approach were changed from only counting trending flu-related terms to investigating the underlying causes of the spike in these searches, the Google flu detection algorithm would perform better. With this method, flu detection becomes more precise and useful by offering more insights into the context underlying search trends.

Furthermore, the dynamic model of the algorithm needs to be refined. It is possible to drastically minimize the probability of inaccurate data and the overall error rate by making sure the model is both accurate and limited. With these enhancements in mind, the algorithm can provide more trustworthy and insightful flu detection results.

➤ **What is “algorithm dynamics”?**

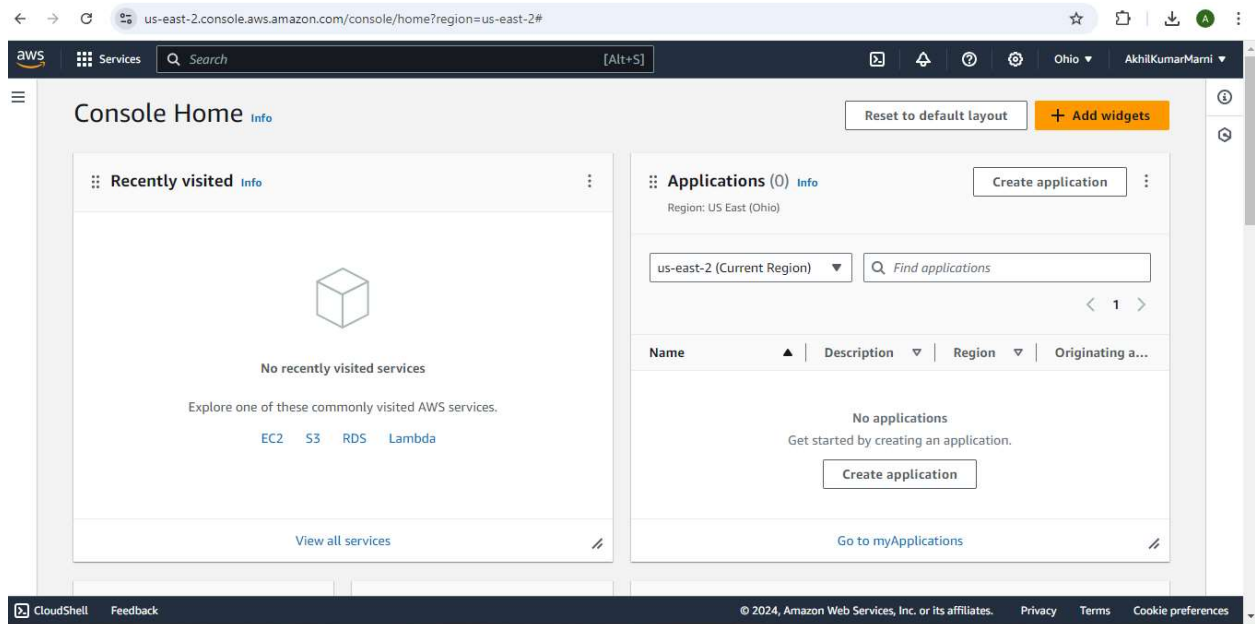
Algorithm dynamics refers to the changes engineers make when utilizing commercial services as well as the tweaks engineers make to optimize them. These dynamics are crucial because they let algorithms adjust over time to changing social trends and new innovations.

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

The Global Flu Trends (GFT) system shows notable irregularities as a result of algorithmic fluctuations. As a result, it starts to portray the incidence of the flu in an increasingly erratic and untrustworthy way. Based on the relative frequency of search phrases, GFT estimations have and these errors were due to biases introduced by changes in Google’s algorithm. A skewed data collection procedure has resulted from these changes and modifications, which have caused significant data distortion. Flu prevalence estimations have a high percentage of inaccuracy due to the system reporting twice as many data points as there actually are as a result of this distortion.

(3 points)

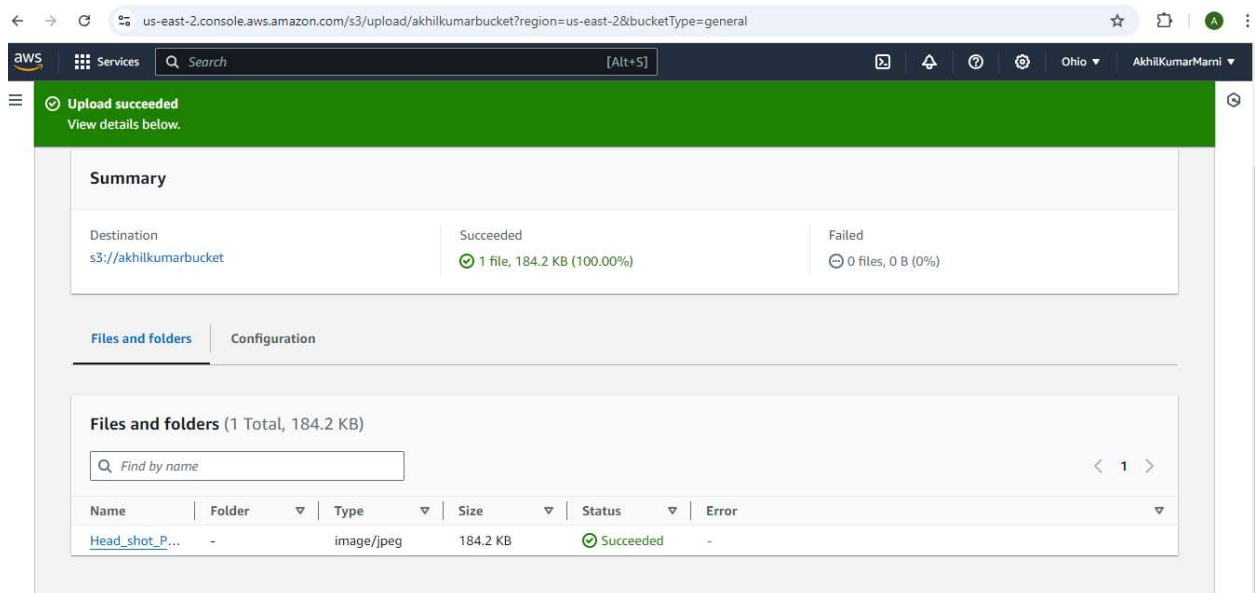
Q5) C. Below screen shot shows the S3 (Simple Storage Service) dashboard of the AWS cloud.



(2 points)

6Q) C. A bucket is created with name “akhilkumarbucket”

D. Uploaded the sample object successfully in newly created S3 bucket



A20554334
AKHIL KUMAR MARNI

6Q) E. Deleting the newly created bucket to avoid extra costing. To delete a bucket, it is mandatory that it should be empty and there is no object inside it.

