

CSP – 554 Big Data Technologies

Assignment #1

Rutul Mehta

A20476293

Question 3

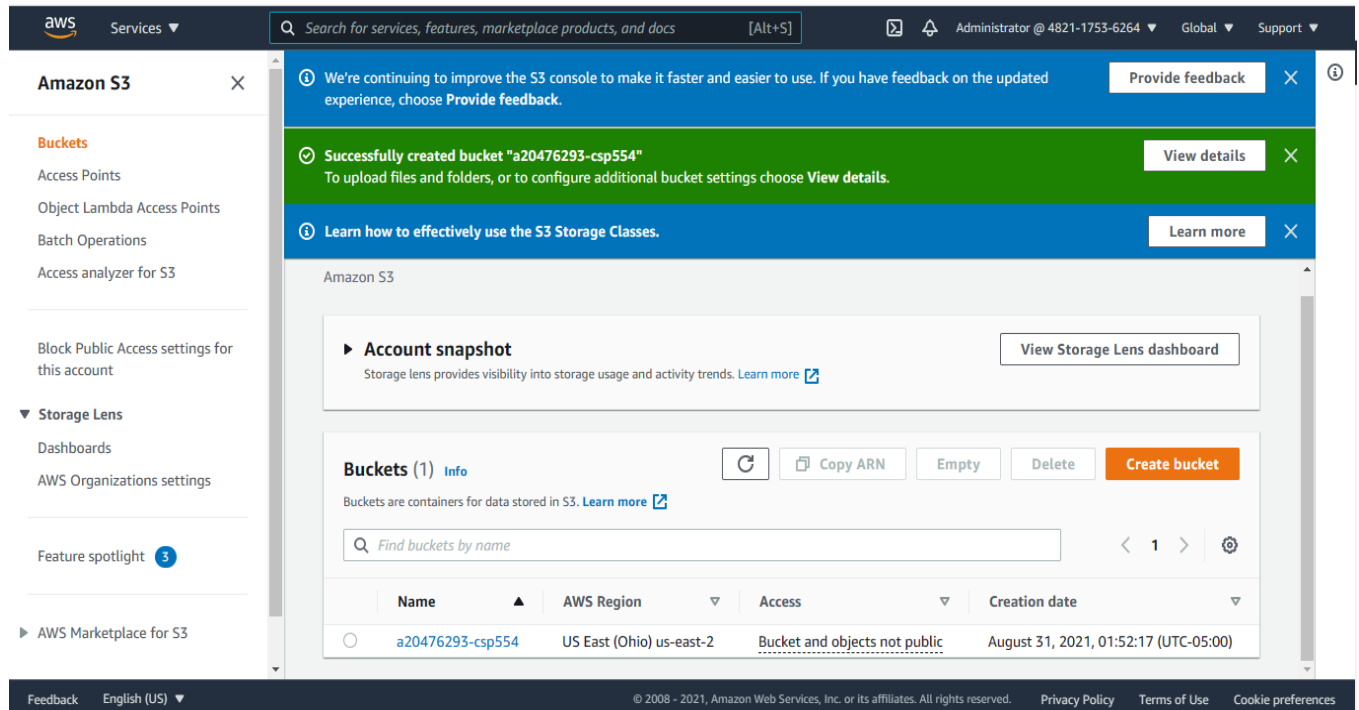
1. I am attending an in-person class on this subject.
2. To take exposure in the Data Science field, I took several courses in my previous semesters like Data mining, Introduction of artificial intelligence.
 - I would like to share my experience when I was doing a project on data mining, I used several classification and clustering algorithms to get better accuracy, but my accuracy didn't increase above 30%.
 - After analyzing the confusion matrix, I have figured out that my data was inclined towards only 1 label (biased data). So, I performed down-sampling / up-sampling on my training dataset and I improved my accuracy to 50%.
3. I have an experience in SQL (Structured Query Language) which is performed on structured data but nowadays our most of the data are unstructured. So, getting an exposure in NoSQL or MongoDB would be better for future prospects.
4. The course structure is very comprehensive I don't think anything remains in the course structure.
5. I don't have any personal issues.

Question 5

1. The Problem was Google Flu Trend algorithm could only measure what people search for; it didn't analyze why they were searching for those words.
2. "Big data hubris" the assumption that big data can substitute for (rather than supplement) traditional data collection and analysis.
3. They can improve the Google Flu trend algorithm by removing human input, and letting the raw data do the work, the model had to make its predictions using only search queries from the previous years.
4. Algorithm dynamics are the modifications made by engineers to improve the commercial service and by consumers in using that service.
5. Several changes in Google's search algorithm and user behavior likely affected Google Flu Trend algorithm's tracking.

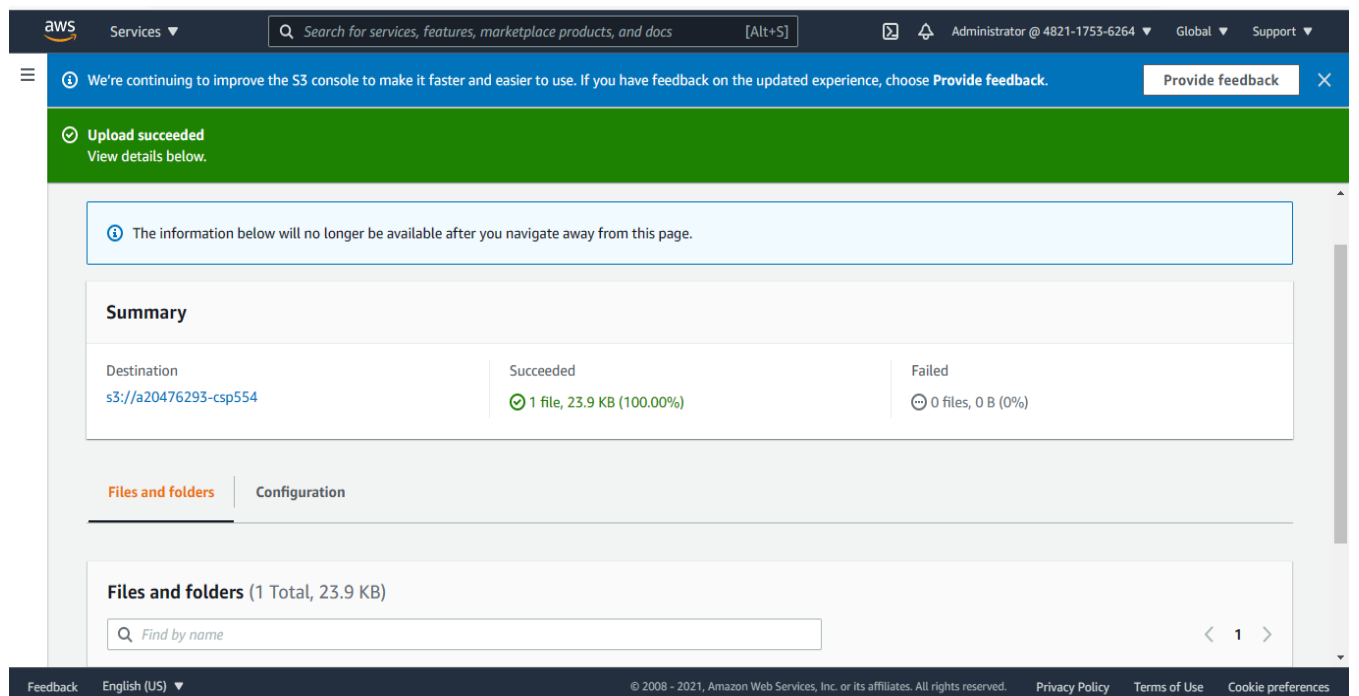
Question 6

e. Screenshot of the created bucket



f. Upload a text file as an object to the created S3 Bucket

g. Screenshot of the uploaded bucket



h. Screenshots of deleting bucket

- Emptied bucket

