Condo Daily Report 8-20-21

Friday

* Arrived at BC
  + Inspected the auto upload feature
  + Still timed out and overrides
  + Inspected the lambda function
    - Realized the timeout period was set to 3 seconds
      * This wouldn’t give the function enough time to execute so changed the time to 3 minutes
    - The function executed and works correctly
  + Hopped in a meeting with mike and kayce
    - Talked about soc2
      * Looked at resources in clickup
      * Started to read through the powerpoint
      * Understand the basis, and general level needed to get to type 1 soc2 compliance
  + Started to conquer the loading of s3 log info into athena
  + Attempted to populate athena table with the log info
    - Data would not populate
    - Checked out file format
      * Data was zipped in a .gz format
    - Attempted to load info in table using customized ddl
      * The ddl did not work went back to the drawing board
    - Looked up online the basic layout of data being populated from a flow log
    - Attempted to make a personal ddl that would populate the data
      * The data populated but it was offset, and certain fields were missing
    - Attempted to input data in numerous ways
      * None of the formats worked correctly
    - Drew out the process on the whiteboard to try to find a pattern of failure
      * Did not find a pattern that made sense
      * Looked into special character splitting
        + This did not come with any luck
    - Decided to download a .gz file onto my computer
      * Installed winzip to unzip file
      * The file came back as an unknown format
      * Decided to transform the file into a .txt format
      * I was able to see the layout of the data that was flowing into the database
        + Noticed that in the fields of data represented online, there was a field missing
    - Adjusted my ddl statement to account for new column
      * The data populated correctly
      * Noticed that you have to create a separate table for each different log directory tracking the separate EC2 instances
      * The problem now is the date, and time data is linked together by special chars
        + I am making a python to take read a .gz file

Ive already implemented a way to find the .gz files

The function is able to unzip the .gz file

Read the contents of it

And sort through the file contents