Assignment -6: Linux and Bash by Hemanth

1.Set a cron job at host sh01to start your individual instance at the defined work period and stop it out of your work time. Turn in host, cron job line and screenshot to validate the implementation.

*Answer:*

1.Create a new bash script by running following command

nano start-stop-ec2.sh

2.Connect to the instance using SSH and create a shell script that starts and stops the instance based on a given schedule. For example, you can create a script named **start-stop-ec2.sh** with the following contents:

Text

Description automatically generated

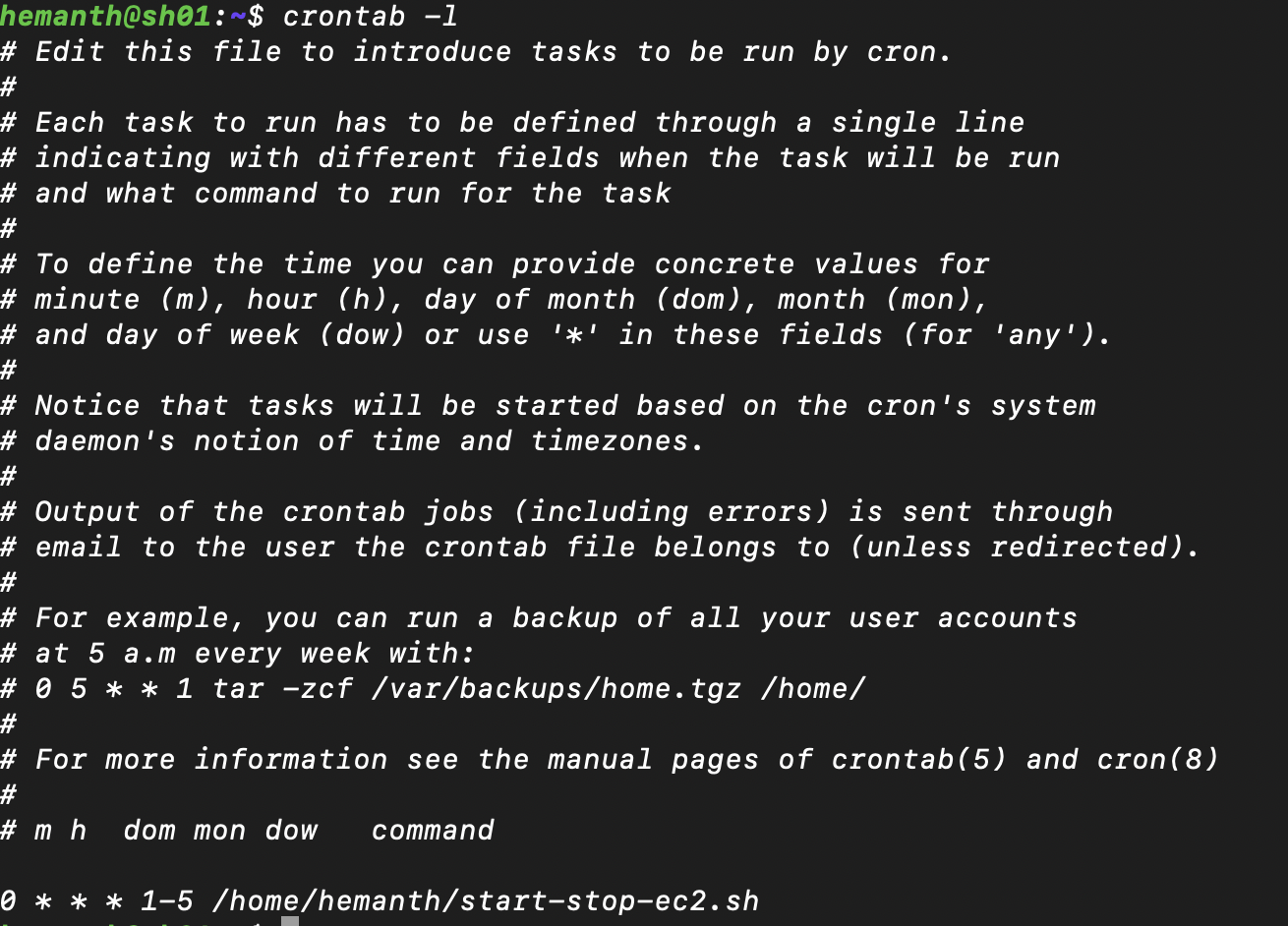
3.Make the script executable by running the following command:

chmod +x start-stop-ec2.sh

./start-stop-ec2.sh

This will start or stop the specified EC2 instance based on the current time, and print a message to the console indicating whether the instance was started or stopped.

1. In the crontab editor, add a new line at the bottom of the file with the following syntax:



This cron job will run the **start-stop-ec2.sh** script at the top of every hour on weekdays (Monday through Friday), which will start or stop the EC2 instance based on the current time.

2. The 1000 Genomes Project is a collaboration among research groups in the US, UK, China and Germany to produce an extensive catalog of human genetic variation that will support future medical research studies. write a bash script with following requirementsInput: the setofthefile names (SRR062634\_1 SRR062634\_2 SRR062635\_1 SRR062635\_2)Function and Output: oDownload the data to subdirectory at host sh01under /scratch/[user]/hg38oPrint the base name of the file and the base name with no extension from the absolute path of filei.e.downloaded data at /scratch/[user]/hg38/xyz.fastq.gz, output as xyz.fastq.gz and xyzSRR062634\_1.filt.fastq.gzSRR062634\_2.filt.fastq.gzSRR062635\_1.filt.fastq.gzSRR062635\_2.filt.fastq.gzSRR062641\_1.filt.fastq.gzhttps://ftp.ncbi.nih.gov/1000genomes/ftp/phase3/data/HG00096/sequence\_read/Turn in the location of your script, the command to run your script and screenshot of the result.

*Answer:*

*genome.sh*

#!/bin/bash

# Set the base URL for the data

URL="ftp://ftp.ncbi.nih.gov/1000genomes/ftp/phase3/data/HG00096/sequence\_read/"

# Set the destination directory for the downloaded data

DEST\_DIR="/scratch/hemanth/hg38/"

# Make sure the destination directory exists

mkdir -p "$DEST\_DIR"

# Loop over the input file names

for FILENAME in SRR062634\_1 SRR062634\_2 SRR062635\_1 SRR062635\_2; do

# Construct the URL for the file

FILE\_URL="${URL}${FILENAME}.filt.fastq.gz"

# Download the file to the destination directory

wget -P "$DEST\_DIR" "$FILE\_URL"

# Extract the base name and the base name with no extension from the absolute path of the file

BASENAME=$(basename "$DEST\_DIR/${FILENAME}.filt.fastq.gz")

BASENAME\_NO\_EXT="${BASENAME%.\*}"

# Print the base name and the base name with no extension

echo "$BASENAME"

echo "$BASENAME\_NO\_EXT"

done

Text

Description automatically generated

This will download the files to the subdirectory **/scratch/hemanth/hg38** and print the base name and base name with no extension for each file. we can verify that the files have been downloaded by checking the target directory.\

Text

Description automatically generatedText

Description automatically generated