Procedure of Data Analysis on HPC

1. Make sure R file can run well locally.
2. Test R file can run well on the server side. First try
   - Log on the server from the terminal "ssh [uarkID]@hpc-portal2.hpc.uark.edu"
   - Create folders on the server side
   - Set up no_cores as 3
   - Do not use "setwd" function in R file that stored in the server, using relative file path. For example, if you data files are stored in the folder "R12002030", make sure the folder is in the same folder as your R code file and you read in data using function like read.csv(file = "R12002030/XXXXX.csv")
   - Try to run it directly, *srun –N1 –n1 –c1 –p cloud72 –q cloud –t 2:00:00 ––pty /bin/bash*
   - Rscript XXXX.R
   - Check whether the task is running, using squeue –u [uarkID]
3. Submit the job file
   - Create a bash file for task scheduler, named it **task.sh** (see the example file below)
   - Submit the bash file to the computing node, using *sbatch task.sh*.

ProjectA/task.sh
cd ProjectA
data stored in ProjectA/DataFiles/XXX.csv
In r_codes_na.R, read.csv("DataFiles/XXXX.csv")

Example bash file for task scheduler

Assume your project folder named "ProjectA", in which there are some subfolders, like OutputFiles, Rcode

```
#!/bin/bash
#SBATCH ––job–name=tres288
#SBATCH ––output=/home/jzhang/ProjectA/OutputFiles/tres288_%j.out
#SBATCH –p tres288
#SBATCH –q tres
#SBATCH ––nodes=1
#SBATCH ––ntasks–per–node=1
#SBATCH ––cpus–per–task=32
#SBATCH ––time=72:00:00

source /etc/profile.d/modules.sh
module purge
module load intel/21.2.0 mkl/21.3.0 R/4.3.0 gcc/11.2.1
cd $SLURM_SUBMIT_DIR

Rscript /home/jzhang/ProjectA/Rcode/r_codes_na.R
```

Looking for help

Email to: David J. Chaffin dchaffin@uark.edu for detailed specifications of computers.