Hi Nathan, hope you are doing well. I received your profile from Hans Hesse who works with us at Amazon AWS in Cape Town. I will call you a bit later this week also.

As a first step in determining a possible fit for the Software Development Engineer position at our Cape Town Development Centre, please complete the following task and return the results to Raldo (cc'd in this email) once completed (preferably in the next 2 days).

The task consists of calculating data about a running cloud infrastructure. The cloud is made up of a fleet of hosts on which instances are launched. You are given as input a list of hosts and a list of instances. Instances are the virtual computers that run on the hosts. The HostState.txt input lists the different hosts and their location. The InstanceState.txt input lists the running instances and on which host they are running.

HostState.txt input format:

```
<hostID>,<numberOfSlots>,<datacentreID>
...
```

- **host ID is the unique ID identifying the host**
- **numberOfSlots is the number of instance slots that the host has to launch instances into**
- **datacentre ID is the unique ID of the datacentre in which the host resides**

InstanceState.txt input format:

```
<instanceID>,<customerID>,<hostID>
...
```

- **instance ID is the unique ID identifying the instance**
- **customer ID is the unique ID of the customer who owns this instance**
- **hostID is the unique ID of the host on which this instance is running**

Write a program that loads the the data contained in "HostState.txt" and "InstanceState.txt" then computes and writes out the following summary statistics to the output file "Statistics.txt":

- **Output the customer with the largest fraction of their total fleet of instances on a single host and output the value of that fraction.**
- **Output the customer with the largest fraction of their total fleet of instances in a single datacentre and output the value of that fraction.**
- **Output a list of all the hosts which have at least one empty slot.**

The output file must have the following format:

```
HostClustering:<customerID>,<fractionOfFleetOnHost>
DatacentreClustering:<customerID>,<fractionOfFleetInDataCentre>
AvailableHosts:<hostID_1>,<hostID_2>,...
```

Your program should handle errors, including malformed and invalid input, appropriately.

You can choose the programming language for your program. It should preferably be one of the more mainstream object-orientated languages such as Java, C#, C++, Python or Ruby. Please motivate your choice of programming language.

Please provide the source code and automated tests and attach any additional data or information (e.g. assumptions you've made, known bugs, etc.) that you think we would need to fairly judge your submission. Also provide the output file "Statistics.txt" as generated by your program along with the input files used to generate it. Aim for code that is at the level where it could be run on a production system.

Your program will be judged on the quality and robustness of the code as well as the correctness of the output.

Please don't send executable code, as this will be blocked by our mail-servers.

Please **zip your code and provide a download location (e.g. dropbox url, S3 bucket URL, ftp/http url) where we can download your code from.** Your code should be of a sufficient quality that it can run on a live business-critical system.

Please confirm receipt of this email by signing and returning the attached NDA.

Kind regards,