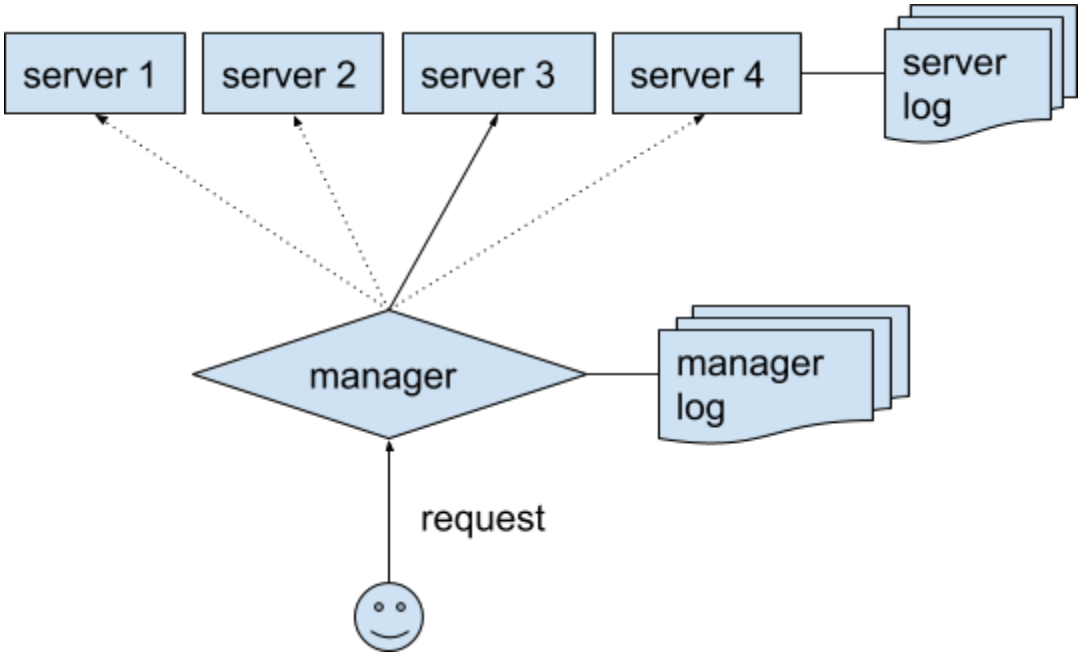
DS Home Assignment: Service Uptime

A website host promises 99.9% monthly service uptime to its customers, i.e., the service downtime for each customer's website should be less than 0.1% of the time in each monthly billing cycle. This task is to verify this claim via logs.

Architecture and logging

The backend architecture of the website host is described in the following chart:



The manager node takes care of customer accounts and load balancing. When a customer create an account, the manager creates a log message in JSON format:

```
{"event": "createAccount", "accountId": 1, "accountName": "Adam Brown", "time": "2012-11-02 14:34:02"}
```

where the time format is "yyyy-MM-dd HH:mm:ss". At the same time of account creation, the manager also decides which server node should answer requests for this customer's website. It logs another message:

```
{"event": "route", "accountId": 1, "serverId": 3, "time": "2012-11-02 14:34:02"}
```

There are 4 server nodes and they are identical to each other. So the manager can route user requests to any of the server nodes to balance the server workloads. For

example, if it re-routes the traffic for Adam's website to server 2, it will log a message like the following:

```
{"event": "route", "accountId": 1, "serverId": 2, "time": "2012-11-02 14:35:10"}
```

Then all future requests to Adam's website will be answered by server 2 until the manager re-routes it again. If Adam deletes his account, the manage logs a "deleteAccount" message:

```
{"event": "deleteAccount", "accountId": 1, "time": "2012-11-03 10:12:30"}
```

Each server node logs its workload at every exact minute, called "heartbeat":

```
{"event": "heartbeat", "serverId": 2, "workload": 0.5, "time": "2012-11-02 08:24:00"}
```

If the workload is above 1.0 or the server node stops logging heartbeats, all websites assigned to this node are considered as non-operational, i.e., service down. For simplicity, we assume that the workload during the following minute is the same as the workload logged in the heartbeat.

Tasks

Over a monthly billing cycle, all manager and server logs are collected and stored in a JSON file "logs.json" (download), with each line corresponding to a log message. Note that the logs are not necessarily ordered by time. We want to verify service uptime for customers who created their accounts during this billing cycle and explore this dataset.

Please write a report that contains the following:

1. Compute service downtimes for all customers and list customers who experienced more than 0.1% downtime, i.e., 29 days x 0.1% = 41.76 minutes. (As a sanity check, the downtime for Issac should be 34 minutes.)
2. Design a statistical hypothesis test and report whether manager's routing strategy depends on the workload.

Please send us your report along with the source code you wrote. The goal of this task is to deliver production quality work. So please take some time to organize the report, document the code, and test it.