Q1) In a system with a high churn rate (i.e. servers are frequently being removed and added), one might think that Memcache will be very inefficient since it will take a lot of time for the caches to get repopulated (following cache misses and fetch operations from disk). How does Memcache deal with this problem? Explain. [3 marks]

Cold Cluster Warmup **[1 mark for mentioning name]**

Clients in the frontend cluster that have an empty cache (or “cold cluster”) can retrieve data from a cluster that has caches with normal hit rates (or “warmer cluster”) rather than data stored in persistent storage databases. **[1 mark for explanation]**

This scheme makes use of the data replication that happens across frontend clusters and allows cold clusters to be repopulated within a few hours rather than a few days. **[1 mark for justifying efficiency]**

Q2) In Memcached, what happens when a client times out while waiting for a GET request to be served (i.e. it receives no reply from the memcached server)? [2 marks]

The client assumes the server has failed and issues the request again to a special Gutter pool. **[1 mark]**

If this second request misses, the client will query the database **[0.5 mark]**
And insert the appropriate key-value pair into the Gutter machine after **[0.5 mark]**