Claims

1. A communications server apparatus for a transportation service, comprising:

a communications server comprising a plurality of shard replicas stored on a plurality of server instances;

at least one driver communications device;

at least one user communications device; and

a communications network equipment operable for the communications server, the at least one driver communications device, and the at least one user communications device, to establish communication with each other therethrough;

wherein the at least one driver communications device comprises a first processor and a first memory, and the at least one user communications device comprises a second processor and a second memory;

the at least one driver communications device being configured, under control of the first processor, to execute first instructions stored in the first memory to:

send data regarding the driver location to the plurality of shard replicas;

the at least one user communications device being configured, under control of the second processor, to execute second instructions stored in the second memory to:

request data from the plurality of shard replicas regarding an optimal driver for a proposed path of travel, based on a user location and destination; and

the communications server comprises a third processor and a third memory, the communications server being configured, under control of the third processor, to execute third instructions stored in the third memory to:

determine if rebalancing of the plurality of server instances is required based on a plurality of data requests by the at least one user communications device and two or more quantitative metrics, wherein the

two or more quantitative metrics criteria include a number of actions to implement each candidate, and a CPU utilisation balance, and

rebalance the server instances based on selecting a plurality of candidate rebalancing plans, ranking the candidates, and executing a selected rebalancing plan from the ranking if two or more quantitative metrics criteria are met for the selection.

2. The communications server apparatus of claim 1 wherein the two or more quantitative metrics criteria further include memory safety, and a level of redundancy across different availability zones.

3. The communications server apparatus of claim 1 wherein the CPU utilisation balance includes comparing the CPU utilisation on each of the plurality of server instances.

4. The communications server apparatus of claim 1 wherein the selection of candidates is based on each potential candidate satisfying a memory limitation test, and/or each potential candidate improving CPU utilisation balance for the plurality of server instances.

5. The communications server apparatus of claim 1 wherein the ranking of candidates includes grouping the candidates according to the number of actions to implement each candidate.

6. The communications server apparatus of claim 5 wherein each group of candidates is ranked based on a selection from the group consisting of: CPU utilisation balance, memory safety, level of redundancy across different availability zones, and any combination thereof.

7. The communications server apparatus of claim 6 wherein the winners from each group of candidates are ordered based on the number of actions to implement each candidate.

8. The communications server apparatus of claim 7 wherein the top ordered winner is subsequently compared to each of the lower winners in order, based on a selection from the group consisting of: CPU utilisation balance, memory safety, level of redundancy across different availability zones, and any combination thereof.

9. The communications server apparatus of claim 1 wherein the two or more quantitative metrics criteria include a minimum CPU utilisation balance improvement.

10. A method, performed in a distributed sharding database related to a transportation service, the method comprising, under control of processors for a plurality of server instances storing a plurality of shard replicas:

determining whether rebalancing is required based on CPU utilisation balance for the plurality of server instances;

selecting candidate replacing plans based on each potential candidate satisfying a memory limitation test, and/or each potential candidate improving CPU utilisation balance;

grouping the candidates according to the number of actions to implement each candidate;

ranking each group of candidates based on a selection from the group consisting of: CPU utilisation balance, memory safety, level of redundancy across different availability zones, and any combination thereof;

ordering winners from each group of candidates based on the number of actions to implement each candidate; and

comparing the top ordered winner to each of the lower winners in order, based on a selection from the group consisting of: CPU utilisation balance, memory

safety, level of redundancy across different availability zones, and any combination thereof.

11. A computer program or computer program product comprising instructions for implementing the method of claim 10.

12. A non-transitory storage medium storing instructions, which when executed by a processor, cause the processor to perform the method of claim 10.