

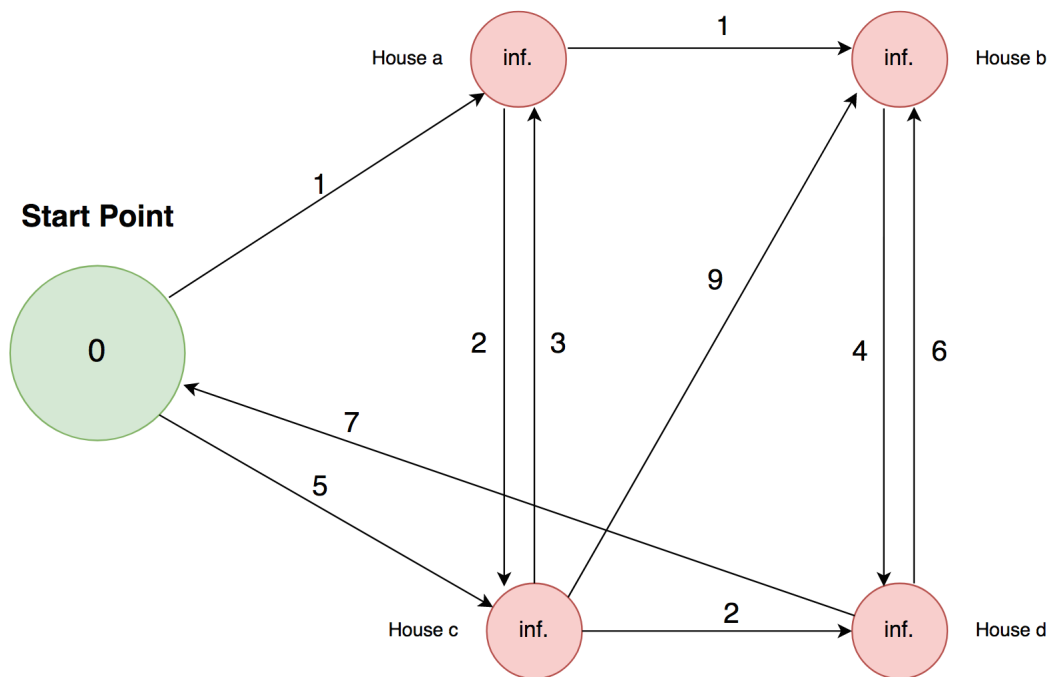
Name:

Std. Number:

Solutions, TH-Quiz 7 (Big data analysis)

Questions

1. A postman wants to move from the start point and deliver several letters to the houses, the location map is shown in the figure below. Use the Map Reduce algorithm to get the shortest path from the starting point to each house. First define exactly Map and Reduce functions and then solve this problem.



Input Data Format:

Node: $\langle id, costFromSource, prevHopFromSource, AdjacencyList \rangle$
AdjacencyList: $\{neighborNode, costToNeighborNode\}$

Initial Input Data:

Node s: $\langle s, 0, -, \{(Node\ a, 1), (Node\ c, 5)\} \rangle$

Node a: $\langle a, \infty, -, \{(Node\ b, 1), (Node\ c, 2)\} \rangle$

Node b: $\langle b, \infty, -, \{(Node\ d, 4)\} \rangle$

Node c: <c, ∞ , - , { (Node a, 3), (Node b, 9), (Node d, 2) }>

Node d: <d, ∞ , - , { (Node s, 7), (Node b, 6) }>

MapReduce Iteration Mapping:

map1 : Node.id: Node

$\rightarrow \{list(Node.neighborNode.id: (Node.id, SUM(Node.costToNeighborNode, Node.costFromSource))\}$

map2 : Node.id: Node $\rightarrow \{list(Node.NeighborNode.id, Node.NeighborNode)\}$

reduce : Node.id: {list(prevHopFromSource, costFromSource, Node)} \rightarrow Node.id: Node'

where

$Node'.costFromSource = MIN(costFromSource),$

$Node'.prevHopFromSource = prevHopFromSource_{min}$

Note: Reducer only emits value if Node structure is updated, i.e., the iteration found a new shortest path from the source.

Iteration 1:

Map Input: s: <s, 0, - , { (Node a, 1), (Node c, 5) }>

Map Output: (a: s, 1), (a: Node a), (c: s, 5), (c, Node c)

Reduce 1 Input: a: (s, 1, Node a)

Reduce 1 Output: a: <a, 1, s, { (Node b, 1), (Node c, 2) }>

Reduce 2 Input: c: (s, 5, Node c)

Reduce 2 Output: c: <c, 5, s, { (Node a, 3), (Node b, 9), (Node d, 2) }>

The reader is encouraged to continue the example and see how the solution converges in 4 iterations.

2. In which cases Map Reduce algorithm can't solve the problem efficiently? (At least two cases should be mentioned and the cause of each should be briefly explained).

[Solution.q2.a](#)

[Solution.q2.b](#)

3. In which cases Graph Lab is a better choice than Map Reduce to design and implement parallel systems? Why? Write down advantages and disadvantages of this framework.

[Solution.q3.a](#)

[Solution.q3.b](#)

4. Compare Spark and Hadoop. Write down advantages and disadvantages of each one.

[Solution.q4](#)