

What is claimed is:

1. A system for coordinating processing of data by multiple networked computing resources, the system comprising at least one processor configured to:

obtain a plurality of data processing waves, each data processing wave identifying: two or more data processing segments, one or more networked corresponding computing resources to which the two or more data processing segments are to be routed; and a timing sequence in which the two or more data processing segments are to be routed;

obtain a minimum handling interval for each of the networked computing resources;

schedule an order for routing the plurality of data processing waves based on the wave timing sequences and the minimum handling intervals for the networked computing resources; and

route each of the data processing segments in the plurality of data processing waves based on the order.

2. The system of claim 1 wherein the at least one processor is configured to schedule the order for routing the plurality of data processing waves by determining an order which intersperses the data processing segments of different data processing waves without violating the minimum handling intervals of the plurality of networked computing resources or the timing sequences of the plurality of data processing waves.

3. The system of claim 1 wherein the at least one processor is configured to identify, in the timing sequences, one or more time ranges within which a first data processing segment in a particular data processing wave can be routed relative to a time at which a second data processing segment in the particular data processing wave is routed.

4. The system of claim 1 wherein the at least one processor is configured to obtain the plurality of data processing waves by obtaining M data processing waves from a

wave queue; and to schedule the order for routing the M data processing waves by: determining a total handling time for each possible arrangement of the M data processing waves, and to schedule the order as an arrangement of the possible arrangements having the shortest total handling time.

5. The system of claim 4 wherein the at least one processor is configured to select M based on a scheduling computation time and total handling times for one or more of the plurality of data processing waves.

6. The system of claim 1 wherein the at least one processor is configured to obtain the plurality of data processing waves by obtaining M data processing waves from a wave queue; and to schedule the order for routing the M data processing waves by: determining total handling times for different arrangements of the M data processing waves until a defined computation time expires, and to schedule the order as an arrangement from the different arrangements having the shortest total handling time determined before the defined computation time expired.

7. The system of claim 1, wherein the at least one processor is configured to: monitor data associated with the plurality of networked computing resources, the monitored data including data associated with data processing segments previously routed to the plurality of networked computing resources.

8. The system of claim 7, wherein the at least one processor is configured to obtain the minimum handling interval time by: determining when two or more data processing segments routed to one of the networked computing resources resulted in one or more execution latencies which are longer than a historical latency value; and to use a relative timing of the routing of the two or more data processing segments to define the minimum handling time for the one of the networked computing resources.

9. The system of claim 1 wherein the at least one processor is configured to: identify two or more data processing segments which are scheduled to be routed to one

of the plurality of networked computing resources with a relative timing that is within a threshold of the minimum handling interval of the one of the plurality of networked computing resources; and after routing the two or more data processing segments, to monitor execution latencies for the two or more data processing segments to determine whether to adjust the minimum handling interval of the one of the plurality of networked computing resources.

10. The system of claim 1, wherein the at least one processor is configured to identify a plurality of data processing segments which are scheduled to be routed to one of the plurality of networked computing resources with a relative timing that is within a threshold of the minimum handling interval of the one of the plurality of networked computing resources; and after routing the plurality of data processing segments, to monitor execution latencies for the plurality of data processing segments to determine a number of data processing segments which will trigger the minimum handling interval of the one of the plurality of networked computing resources.

11. The system of claim 1 wherein the at least one processor is configured to obtain the plurality of data processing waves from a wave queue; wherein the at least one processor is configured to fill the wave queue with at most M data processing waves from each of a plurality of session queues associated with one or more corresponding data sources.

12. The system of claim 11 wherein the at least one processor is configured to fill the wave queue by selecting data processing waves from the plurality of session queues using a priority scheme which is based at least in part on timing parameters associated with the data processing waves.

13. A method for coordinating processing of data by multiple networked computing resources, the method comprising:

obtaining a plurality of data processing waves, each data processing wave identifying: two or more data processing segments, one or more networked

corresponding computing resources to which the two or more data processing segments are to be routed; and a timing sequence in which the two or more data processing segments are to be routed;

obtaining a minimum handling interval for each of the networked computing resources;

scheduling an order for routing the plurality of data processing waves based on the wave timing sequences and the minimum handling intervals for the networked computing resources; and

routing each of the data processing segments in the plurality of data processing waves based on the order.

14. The method of claim 13 wherein scheduling the order for routing the plurality of data processing waves comprises determining an order which intersperses the data processing segments of different data processing waves without violating the minimum handling intervals of the plurality of networked computing resources or the timing sequences of the plurality of data processing waves.

15. The method of claim 13 wherein the timing sequences identify one or more time ranges within which a first data processing segment in a particular data processing wave can be routed relative to a time at which a second data processing segment in the particular data processing wave is routed.

16. The method of claim 13 wherein obtaining the plurality of data processing waves comprises obtaining M data processing waves from a wave queue; and wherein scheduling the order for routing the M data processing waves comprises: determining a total handling time for each possible arrangement of the M data processing waves, and scheduling the order as an arrangement of the possible arrangements having the shortest total handling time.

17. The method of claim 16 wherein M is selected based on a scheduling

computation time and total handling times for one or more of the plurality of data processing waves.

18. The method of claim 13 wherein obtaining the plurality of data processing waves comprises obtaining M data processing waves from a wave queue; and wherein scheduling the order for routing the M data processing waves comprises: determining total handling times for different arrangements of the M data processing waves until a defined computation time expires, and scheduling the order as an arrangement from the different arrangements having the shortest total handling time determined before the defined computation time expired.

19. The method of claim 13, comprising: monitoring data associated with the plurality of networked computing resources, the monitored data including data associated with data processing segments previously routed to the plurality of networked computing resources.

20. The method of claim 19, wherein obtaining the minimum handling interval time comprises: determining when two or more data processing segments routed to one of the networked computing resources resulted in one or more execution latencies which are longer than a historical latency value; and using a relative timing of the routing of the two or more data processing segments to define the minimum handling time for the one of the networked computing resources.

21. The method of claim 13 comprising: identifying two or more data processing segments which are scheduled to be routed to one of the plurality of networked computing resources with a relative timing that is within a threshold of the minimum handling interval of the one of the plurality of networked computing resources; and after routing the two or more data processing segments, monitoring execution latencies for the two or more data processing segments to determine whether to adjust the minimum handling interval of the one of the plurality of networked computing resources.

22. The method of claim 13, comprising identifying a plurality of data processing segments which are scheduled to be routed to one of the plurality of networked computing resources with a relative timing that is within a threshold of the minimum handling interval of the one of the plurality of networked computing resources; and after routing the plurality of data processing segments, monitoring execution latencies for the plurality of data processing segments to determine a number of data processing segments which will trigger the minimum handling interval of the one of the plurality of networked computing resources.

23. The method of claim 13 wherein the plurality of data processing waves is obtained from a wave queue; wherein the at least one processor is configured to fill the wave queue with at most M data processing waves from each of a plurality of session queues associated with one or more corresponding data sources.

24. The method of claim 23 wherein filling the wave queue comprises selecting data processing waves from the plurality of session queues using a priority scheme which is based at least in part on timing parameters associated with the data processing waves.

25. A non-transitory, computer-readable medium or non-transitory, computer-readable media having stored thereon instructions which when executed by at least one processor, configure the at least one processor to:

obtain a plurality of data processing waves, each data processing wave identifying: two or more data processing segments, one or more networked corresponding computing resources to which the two or more data processing segments are to be routed; and a timing sequence in which the two or more data processing segments are to be routed;

obtain a minimum handling interval for each of the networked computing resources;

schedule an order for routing the plurality of data processing waves based on the wave timing sequences and the minimum handling intervals for the networked computing resources; and

route each of the data processing segments in the plurality of data processing waves based on the order.