CLAIMS

1. A method performed by a data processing apparatus comprising:

retrieving a data stream comprising a plurality of data records;

determining a predetermined size capacity based on a memory size of a cache memory associated with the data processing apparatus;

aggregating the plurality of data records of the data stream to form a plurality of record packets, each of the plurality of record packets having the predetermined size capacity; and

transferring respective ones of the plurality of record packets to respective ones of a plurality of threads associated with one or more processing operations of the data processing apparatus.

2. The method of claim 1, wherein the one or more processing operations are associated with a data analytics workflow executing on the data processing apparatus.

3. The method of claim 2, further comprising:

executing each of the one or more processing operations to perform a corresponding data analytics function on the plurality of record packets in a linear order, wherein the linear order is according to an operational sequence set in the data analytics workflow.

4. The method of claim 3, wherein executing each of the one or more processing operations comprises parallel processing performed by executing each thread on a processor from among a plurality of processors associated with the data processing apparatus.

5. The method of claim 1, wherein the memory size of the cache memory associated with the data processing apparatus is dynamically determined from

an operating system or a central processing unit (CPU) of the processing apparatus.

6. The method of claim 1, wherein the predetermined size capacity is an order of magnitude of the memory size of the cache memory.

7. The method of claim 1, wherein a number of data records aggregated into a record packet is a variable determined for each of the plurality of record packets and does not exceed the predetermined size capacity.

8. The method of claim 1, wherein the aggregating is performed upon retrieving the data stream in its entirety.

9. The method of claim 1, wherein the aggregating is performed in parallel with retrieving the data stream.

10. The method of claim 1, further comprising:

re-aggregating data records associated with two or more record packets of the plurality of record packets into an additional record packet, upon determining that the two or more record packets have a number of data records less than a predetermined minimum capacity.

11. The method of claim 1, wherein transferring respective ones of the plurality of record packets to respective ones of a plurality of threads associated with one or more processing operations of the data processing apparatus comprises:

transferring each of the plurality of record packets to a respective one of the plurality of threads for the record packet, wherein the respective threads for the plurality of record packets concurrently

run within a processing operation to concurrently process the plurality of record packets.

12. The method of claim 1, further comprising:

for each record packet of a subset of the plurality of record packets:

determining a size for the record packet based on the predetermined size capacity of the record packet and a threshold latency time associated with processing the record packet;

determining a number of data records to form the record packet based on the size of the record packet; and

aggregating the number of data records to form the record packet.

13. A data processing apparatus comprising:

a non-transitory memory storing executable computer program code; and

a plurality of computer processors having a cache memory and communicatively coupled to the memory, the computer processors executing the computer program code to perform operations comprising:

retrieving a data stream comprising a plurality of data records;

determining a predetermined size capacity based on a memory size of a cache memory associated with the data processing apparatus;

aggregating the plurality of data records of the data stream to form a plurality of record packets, each of the plurality of record packets having the predetermined size capacity; and

transferring respective ones of the plurality of record packets to respective ones of a plurality of threads associated with one or more processing operations of the plurality of processors.

14. The data processing apparatus of claim 13, wherein the one or more processing operations are associated with a data analytics workflow executing on the data processing apparatus.

15. The data processing apparatus of claim 14, wherein the operations further comprise:

executing each of the one or more processing operations to perform a corresponding data analytics function on the plurality of record packets in a linear order, wherein the linear order is according to an operational sequence set in the data analytics workflow.

16. The data processing apparatus of claim 15, wherein executing each of the one or more processing operations comprises parallel processing performed by executing each thread on a processor from among the plurality of processors.

17. The data processing apparatus of claim 13, wherein the predetermined size capacity is an order of magnitude of the memory size of the cache memory.

18. A non-transitory computer-readable memory storing computer program code executable to perform operations using a plurality of computer processors having a cache memory, the operations comprising:

retrieving a data stream comprising a plurality of data records;

determining a predetermined size capacity based on a memory size of a cache memory associated with the data processing apparatus;

aggregating the plurality of data records of the data stream to form a plurality of record packets, each of the plurality of record packets having the predetermined size capacity; and

transferring respective ones of the plurality of record packets to respective ones of a plurality of threads associated with one or more processing operations of the data processing apparatus.

19. The memory of claim 18, wherein the one or more processing operations are associated with a data analytics workflow executing on the plurality of processors.

20. The memory of claim 19, the operations further comprising:

executing each of the one or more processing operations to perform a corresponding data analytics function on the plurality of record packets in a linear order, wherein the linear order is according to an operational sequence set in the data analytics workflow.

21. The memory of claim 20, wherein executing each of the one or more processing operations comprises parallel processing performed by executing each thread on a processor from among the plurality of processors.

22. The memory of claim 18, wherein the predetermined size capacity is an order of magnitude of the memory size of the cache memory.