What is claimed is:

1. An apparatus including:

a computing system including one or more nodes, the computing system configured to execute a plurality of data processing programs that each process at least one stream of data units; and

at least one data storage system accessible to at least one of the one or more nodes, the data storage system, in use, storing shared data accessible by at least two of the plurality of data processing programs;

wherein processing at least one stream of data units using at least a first data processing program of the one or more data processing programs includes:

processing a first stream of data units to generate output for each of a plurality of subsets of contiguous data units within the first stream of data units;

initiating termination of processing within the first data processing program, between processing a first subset of contiguous data units and processing a second subset of contiguous data units adjacent to the first subset of contiguous data units within the first stream of data units;

durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units after determining that the termination of processing within the first data processing program has completed;

resuming processing within the first data processing program after the changes have been durably stored; and

releasing, from the first data processing program, first output generated for the first subset of contiguous data units after the changes have been durably stored.

2. The apparatus of claim 1, wherein releasing, from the first data processing program, the first output generated for the first subset of contiguous data units includes releasing the first output to an external program that is not included in the plurality of data processing programs executing on the computing system.

3. The apparatus of claim 1, wherein processing is resumed within the first data processing program after a first portion of changes have been durably stored but before a second portion of changes have been durably stored.

4. The apparatus of claim 1 wherein processing the at least one stream of data units includes ceasing ingestion of new data units of the at least one stream of data units and continuing to process data units of the first subset of contiguous data units after termination of processing is initiated and until the first subset of contiguous data units are substantially fully processed.

5. The apparatus of claim 4 wherein durably storing the at least some changes to the shared data commences after the first subset of contiguous data units are substantially fully processed.

6. An apparatus including:

means for executing a plurality of data processing programs that each process at least one stream of data units; and

means for storing shared data accessible by at least two of the plurality of data processing programs;

wherein processing at least one stream of data units using at least a first data processing program of the one or more data processing programs includes:

processing a first stream of data units to generate output for each of a plurality of subsets of contiguous data units within the first stream of data units;

initiating termination of processing within the first data processing program, between processing a first subset of contiguous data units and processing a second subset of contiguous data units adjacent to the first subset of contiguous data units within the first stream of data units;

durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units after determining that the termination of processing within the first data processing program has completed;

resuming processing within the first data processing program after the changes have been durably stored; and

releasing, from the first data processing program, first output generated for the first subset of contiguous data units after the changes have been durably stored.

7. A method including:

executing, on a computing system including one or more nodes, a plurality of data processing programs that each process at least one stream of data units; and

storing, on at least one data storage system accessible to at least one of the one or more nodes, shared data accessible by at least two of the plurality of data processing programs;

wherein processing at least one stream of data units using at least a first data processing program of the one or more data processing programs includes:

processing a first stream of data units to generate output for each of a plurality of subsets of contiguous data units within the first stream of data units;

initiating termination of processing within the first data processing program, between processing a first subset of contiguous data units and processing a second subset of contiguous data units adjacent to the first subset of contiguous data units within the first stream of data units;

durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units after determining that the termination of processing within the first data processing program has completed;

resuming processing within the first data processing program after the changes have been durably stored; and

releasing, from the first data processing program, first output generated for the first subset of contiguous data units after the changes have been durably stored.

8. Software stored in a non-transitory form on a computer-readable medium, the software including instructions for causing a computing system to:

execute a plurality of data processing programs that each process at least one stream of data units; and

store shared data accessible by at least two of the plurality of data processing programs;

wherein processing at least one stream of data units using at least a first data processing program of the one or more data processing programs includes:

processing a first stream of data units to generate output for each of a plurality of subsets of contiguous data units within the first stream of data units;

initiating termination of processing within the first data processing program, between processing a first subset of contiguous data units and processing a second subset of contiguous data units adjacent to the first subset of contiguous data units within the first stream of data units;

durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units after determining that the termination of processing within the first data processing program has completed;

resuming processing within the first data processing program after the changes have been durably stored; and

releasing, from the first data processing program, first output generated for the first subset of contiguous data units after the changes have been durably stored.

9. An apparatus including:

a computing system including one or more nodes, the computing system configured to execute a plurality of data processing programs that each process at least one stream of data units; and

at least one data storage system accessible to at least one of the one or more nodes, the data storage system, in use, storing shared data accessible by at least two of the plurality of data processing programs;

wherein processing two or more streams of data units using at least a first group of multiple data processing programs of the plurality of data processing programs includes:

processing, for each data processing program in the first group, a respective stream of data units that includes a plurality of subsets of contiguous data units;

initiating termination of processing within each data processing program in the first group, between processing a first subset of contiguous data units and processing a second subset of contiguous data units adjacent to the first subset of contiguous data units within the respective stream of data units;

durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units after determining that the termination of processing within each data processing program in the first group has completed; and

resuming processing within each data processing program in the first group after the changes have been durably stored.

10. The apparatus of claim 9, wherein processing two or more streams of data units using at least the first group of multiple data processing programs further includes:

after determining that the termination of processing within each data processing program in the first group has completed, durably storing stream state information associated with each respective stream of data units processed by any of the data processing programs in

the first group; or

releasing, from the first group of multiple data processing programs, first output generated for the first subset of contiguous data units after the changes have been durably stored.

11. The apparatus of claim 10, wherein releasing, from the first group of multiple data processing programs, the first output generated for the first subset of contiguous data units includes releasing the first output to:

one of the plurality of data processing programs executing on the computing system that is not included in the first group of multiple data processing programs; or

an external program that is not included in the plurality of data processing programs executing on the computing system.

12. The apparatus of claim 11, wherein the external program sends a request to access particular shared data that is accessible by at least one data processing program in the first group, and a result of the request is released to the external program after all changes to the particular shared data that occurred before the request was received have been durably stored.

13. The apparatus of claim 10, wherein durably stored changes to the shared data caused by processing the first subset of contiguous data units are distinguished from durably stored changes to the shared data caused by processing the second subset of contiguous data units.

14. The apparatus of claim 13, wherein at least some changes to the shared data caused by processing the first subset of contiguous data units are durably stored after at least some changes to the shared data caused by processing the second subset of contiguous data units have started, where the first subset of contiguous data units are before the second subset of contiguous data units within the first stream of data units.

15. The apparatus of claim 14, wherein the first output generated for the first subset of contiguous data units is released from the first group of multiple data processing programs after all changes caused by processing the first subset of contiguous data units have been durably stored.

16. The apparatus of claim 9, wherein processing two or more streams of data units

includes processing four or more streams of data units using at least the first group of multiple data processing programs and a second group of multiple data processing programs of the plurality of data processing programs.

17. The apparatus of claim 16, wherein each group of multiple data processing programs terminates processing of respective streams of data units periodically, and the computing system begins durably storing at least some changes to the shared data caused by processing data units while all data processing programs in that group are terminated.

18. The apparatus of claim 17, wherein the first group of data processing programs terminates and processing of respective streams of data units at a first frequency, and the second group of data processing programs terminate processing of respective streams of data units at a second frequency different from the first frequency.

19. The apparatus of claim 9, wherein processing is resumed within each data processing program in the first group after a first portion of changes have been durably stored but before a second portion of changes have been durably stored.

20. The apparatus of claim 9, wherein the first group of multiple data processing programs terminates processing the two or more streams of data units periodically, and the computing system begins durably storing at least some changes to the shared data caused by processing data units while all data processing programs in the first group are terminated.

21. An apparatus including:

means for executing a plurality of data processing programs that each process at least one stream of data units; and

means for storing shared data accessible by at least two of the plurality of data processing programs;

wherein processing two or more streams of data units using at least a first group of multiple data processing programs of the plurality of data processing programs includes:

processing, for each data processing program in the first group, a respective stream of data units that includes a plurality of subsets of contiguous data units;

initiating termination of processing within each data processing program in the first group, between processing a first subset of contiguous data units and processing a

second subset of contiguous data units adjacent to the first subset of contiguous data units within the respective stream of data units;

durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units after determining that the termination of processing within each data processing program in the first group has completed; and

resuming processing within each data processing program in the first group after the changes have been durably stored.

22. A method including:

executing, on a computing system including one or more nodes, a plurality of data processing programs that each process at least one stream of data units; and

storing, on at least one data storage system accessible to at least one of the one or more nodes, shared data accessible by at least two of the plurality of data processing programs;

wherein processing two or more streams of data units using at least a first group of multiple data processing programs of the plurality of data processing programs includes:

processing, for each data processing program in the first group, a respective stream of data units that includes a plurality of subsets of contiguous data units;

initiating termination of processing within each data processing program in the first group, between processing a first subset of contiguous data units and processing a second subset of contiguous data units adjacent to the first subset of contiguous data units within the respective stream of data units;

durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units after determining that the termination of processing within each data processing program in the first group has completed; and

resuming processing within each data processing program in the first group after the changes have been durably stored.

23. Software stored in a non-transitory form on a computer-readable medium, the software including instructions for causing a computing system to:

execute a plurality of data processing programs that each process at least one stream of data units; and

store shared data accessible by at least two of the plurality of data processing programs;

wherein processing two or more streams of data units using at least a first group of multiple data processing programs of the plurality of data processing programs includes:

processing, for each data processing program in the first group, a respective stream of data units that includes a plurality of subsets of contiguous data units;

initiating termination of processing within each data processing program in the first group, between processing a first subset of contiguous data units and processing a second subset of contiguous data units adjacent to the first subset of contiguous data units within the respective stream of data units;

durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units after determining that the termination of processing within each data processing program in the first group has completed; and

resuming processing within each data processing program in the first group after the changes have been durably stored.

24. An apparatus including:

a computing system including one or more nodes, the computing system configured to execute a plurality of data processing programs that each process at least one stream of data units; and

at least one data storage system accessible to at least one of the one or more nodes, the data storage system, in use, storing shared data accessible by at least two of the plurality of data processing programs;

wherein processing at least one stream of data units using at least a first data processing program of the plurality of data processing programs includes:

processing a first stream of data units that includes a plurality of subsets of contiguous data units;

initiating termination of processing within the first data processing program, between processing a first subset of contiguous data units and processing a second subset of contiguous data units adjacent to the first subset of contiguous data units within the first stream of data units;

durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units after determining that the termination of processing within the first data processing program has completed; and

resuming processing within the first data processing program before all of the changes have completed being durably stored.

25. The apparatus of claim 22, wherein processing at least one stream of data units using at least the first data processing program further includes storing at least one snapshot of the shared data and storing a journal of changes to the shared data caused by processing data units after the snapshot was stored.

26. The apparatus of claim 25, wherein durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units includes storing at least a portion of the snapshot and storing at least a portion of the journal of changes.

27. The apparatus of claim 22, wherein processing at least one stream of data units using at least the first data processing program further includes, before determining that the termination of processing within the first data processing program has completed, durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units.

28. The apparatus of claim 27, wherein processing at least one stream of data units using at least the first data processing program further includes, after resuming processing within the first data processing program, durably storing at least some changes to the shared data caused by processing the second subset of contiguous data units.

29. The apparatus of claim 22, wherein processing at least one stream of data units using at least the first data processing program further includes further includes generating output for each of the plurality of subsets of contiguous data units, and releasing from the first data processing program, first output generated for the first subset of contiguous data units after the changes have completed being durably stored.

30. The apparatus of any one of claims 1 or 22, wherein processing at least one stream of data units using at least the first data processing program further includes, after determining that the termination of processing within the first data processing program has completed, durably storing stream state information associated with the first stream of data units.

31. The apparatus of any one of claims 1 or 22, wherein the first output generated for the first subset of contiguous data units is released from the first data processing program after

all changes caused by processing the first subset of contiguous data units have been durably stored.

32. The apparatus of any one of claims 1 or 22, wherein the first data processing program terminates processing the first stream of data units periodically, and the computing system begins durably storing at least some changes to the shared data caused by processing data units while the first data processing program is terminated.

33. The apparatus of any one of claims 1, 9 or 22, wherein the plurality of data processing programs each process at least one stream of data units with no program state information being maintained over more than two adjacent data units in the stream.

34. The apparatus of any one of claims 1, 9 or 22, wherein the data storage system:

includes a non-volatile storage medium, and durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units includes storing the changes in the non-volatile storage medium;

includes a communication medium coupled to a plurality of the nodes, and durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units includes sending the changes from a first node to at least a second node of the plurality of the nodes over the communication medium; or

also stores stream state information associated with one or more streams of data units processed by at least one of the plurality of data processing programs.

35. The apparatus of any one of claims 1 or 22, wherein durably stored changes to the shared data caused by processing the first subset of contiguous data units are distinguished from durably stored changes to the shared data caused by processing the second subset of contiguous data units.

36. The apparatus of any one of claims 1 or 22, wherein at least some changes to the shared data caused by processing the first subset of contiguous data units are durably stored after at least some changes to the shared data caused by processing the second subset of contiguous data units have started, where the first subset of contiguous data units are before the second subset of contiguous data units within the first stream of data units.

37. The apparatus of any one of claims 1, 9 or 22, wherein initiating termination of processing within the first data processing program includes inserting a stream-ending indicator between the first subset of contiguous data units the second subset of contiguous data units, and the termination of processing within the first data processing program has completed after all processes that perform tasks specified by the first data processing program have exited normally in response to the stream-ending indicator.

38. The apparatus of any one of claims 1, 9 or 22, wherein the shared data is accessible by all of the plurality of data processing programs.

39. An apparatus including:

means for executing a plurality of data processing programs that each process at least one stream of data units; and

means for storing shared data accessible by at least two of the plurality of data processing programs;

wherein processing at least one stream of data units using at least a first data processing program of the plurality of data processing programs includes:

processing a first stream of data units that includes a plurality of subsets of contiguous data units;

initiating termination of processing within the first data processing program, between processing a first subset of contiguous data units and processing a second subset of contiguous data units adjacent to the first subset of contiguous data units within the first stream of data units;

durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units after determining that the termination of processing within the first data processing program has completed; and

resuming processing within the first data processing program before all of the changes have completed being durably stored.

40. A method including:

executing, on a computing system including one or more nodes, a plurality of data processing programs that each process at least one stream of data units; and

storing, on at least one data storage system accessible to at least one of the one or more nodes, shared data accessible by at least two of the plurality of data processing programs;

wherein processing at least one stream of data units using at least a first data processing program of the plurality of data processing programs includes:

processing a first stream of data units that includes a plurality of subsets of contiguous data units;

initiating termination of processing within the first data processing program, between processing a first subset of contiguous data units and processing a second subset of contiguous data units adjacent to the first subset of contiguous data units within the first stream of data units;

durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units after determining that the termination of processing within the first data processing program has completed; and

resuming processing within the first data processing program before all of the changes have completed being durably stored.

41. Software stored in a non-transitory form on a computer-readable medium, the software including instructions for causing a computing system to:

execute a plurality of data processing programs that each process at least one stream of data units; and

store shared data accessible by at least two of the plurality of data processing programs;

wherein processing at least one stream of data units using at least a first data processing program of the plurality of data processing programs includes:

processing a first stream of data units that includes a plurality of subsets of contiguous data units;

initiating termination of processing within the first data processing program, between processing a first subset of contiguous data units and processing a second subset of contiguous data units adjacent to the first subset of contiguous data units within the first stream of data units;

durably storing at least some changes to the shared data caused by processing the first subset of contiguous data units after determining that the termination of processing within the first data processing program has completed; and

resuming processing within the first data processing program before all of the changes have completed being durably stored.