

CLAIMS

1. A method for processing data in a data processing system, wherein a first dataflow graph is executed, the method including:
 - receiving multiple units of work that each include one or more records;
 - determining a characteristic of a first unit of work based on data included in the one or more records of the first unit of work;
 - identifying, by a component of the first dataflow graph, a second dataflow graph from multiple available dataflow graphs based on the determined characteristic, the multiple available dataflow graphs being stored in a data storage system;
 - processing the first unit of work using the second dataflow graph based on loading the second dataflow graph into the first dataflow graph;
 - receiving, by the first dataflow graph, one or more values of one or more performance metrics associated with the processing of the first unit of work using the loaded second dataflow graph;
 - outputting, by the first dataflow graph to a performance monitoring module, the one or more received values of the one or more performance metrics associated with the processing of the first unit of work using the loaded second dataflow graph, wherein the performance monitoring module receives a plurality of values of one or more performance metrics output by a plurality of dataflow graphs;
 - selecting, from among the plurality of values of the one or more performance metrics output by the plurality of dataflow graphs, values of one or more performance metrics associated with an identifier for the second dataflow graph; and
 - aggregating the values selected for the one or more performance metrics associated with the identifier for the second dataflow graph.
2. The method of claim 1, wherein the second dataflow graph is compiled independent of the first dataflow graph.
3. The method of claim 1, wherein the processing further includes storing the one or more values of the one or more performance metrics associated with the processing of the first unit of work using the loaded second dataflow graph in an in-memory data store.

4. The method of claim 3, wherein the processing further includes transferring the stored one or more values of the one or more performance metrics to a persistent data store.
5. The method of claim 1, wherein the values selected include first values and wherein the method further includes aggregating second values of the one or more performance metrics output by the plurality of dataflow graphs based on an identifier associated with the first dataflow graph.
6. The method of claim 1, further including displaying the one or more performance metrics to a user or displaying the aggregated values of the one or more performance metrics to the user.
7. A non-transitory computer-readable storage medium storing a computer program for processing data in a data processing system, wherein a first dataflow graph is executed, the computer program including instructions for causing a computing system to:
 - receive multiple units of work that each include one or more records;
 - determine a characteristic of a first unit of work based on data included in the one or more records of the first unit of work;
 - identify, by a component of the first dataflow graph, a second dataflow graph from multiple available dataflow graphs based on the determined characteristic, the multiple available dataflow graphs being stored in a data storage system;
 - process the first unit of work using the second dataflow graph based on loading the second dataflow graph into the first dataflow graph;
 - receive, by the first dataflow graph, one or more values of one or more performance metrics associated with the processing of the first unit of work using the loaded second dataflow graph;
 - output, by the first dataflow graph to a performance monitoring module, the one or more received values of the one or more performance metrics associated with the processing of the first unit of work using the loaded second dataflow graph, wherein the performance monitoring module receives a plurality of values of one or more performance metrics output by a plurality of dataflow graphs;

select, from among the plurality of values of the one or more performance metrics output by the plurality of dataflow graphs, values of one or more performance metrics associated with an identifier for the second dataflow graph; and

aggregate the values selected for the one or more performance metrics associated with the identifier for the second dataflow graph.

8. The medium of claim 7, wherein the second dataflow graph is compiled independent of the first dataflow graph.

9. The medium of claim 7, wherein the computer program further includes instructions for causing a computing system to store the one or more values of the one or more performance metrics associated with the processing of the first unit of work using the loaded second dataflow graph in an in-memory data store.

10. The medium of claim 9, wherein the computer program further includes instructions for causing a computing system to transfer the stored one or more values of the one or more performance metrics to a persistent data store.

11. The medium of claim 7, wherein the values selected include first values and wherein the instructions to aggregate include instructions to aggregate second values of the one or more performance metrics output by the plurality of dataflow graphs based on an identifier associated with the first dataflow graph.

12. The medium of claim 7, wherein the computer program further includes instructions for causing a computing system to display the one or more performance metrics to a user or to display the aggregated values of the one or more performance metrics to the user.

13. A computing system for processing data, the computing system including:
an input device or port configured to receive multiple units of work that each include one or more records; and
at least one processor configured to process data, the processing including:
receiving multiple units of work that each include one or more records;

determining a characteristic of a first unit of work based on data included in the one or more records of the first unit of work;

identifying, by a component of a first dataflow graph, a second dataflow graph from multiple available dataflow graphs based on the determined characteristic, the multiple available dataflow graphs being stored in a data storage system;

processing the first unit of work using the second dataflow graph based on loading the second dataflow graph into the first dataflow graph;

receiving, by the first dataflow graph, one or more values of one or more performance metrics associated with the processing of the first unit of work using the loaded second dataflow graph;

outputting, by the first dataflow graph to a performance monitoring module, the one or more received values of the one or more performance metrics associated with the processing of the first unit of work using the loaded second dataflow graph, wherein the performance monitoring module receives a plurality of values of one or more performance metrics output by a plurality of dataflow graphs;

selecting, from among the plurality of values of the one or more performance metrics output by the plurality of dataflow graphs, values of one or more performance metrics associated with an identifier for the second dataflow graph; and

aggregating the values selected for the one or more performance metrics associated with the identifier for the second dataflow graph.

14. The system of claim 13, wherein the second dataflow graph is compiled independent of the first dataflow graph.

15. The system of claim 13, wherein the processing further includes storing the one or more values of the one or more performance metrics associated with the processing of the first unit of work using the loaded second dataflow graph in an in-memory data store.

16. The system of claim 15, wherein the processing further includes transferring the stored one or more values of the one or more performance metrics to a persistent data store.

17. The system of claim 13, wherein the values selected include first values and wherein the processing further includes aggregating second values of the one or more performance

metrics output by the plurality of dataflow graphs based on an identifier associated with the first dataflow graph.

18. The system of claim 13, wherein the processing further includes displaying the one or more performance metrics to a user or displaying the aggregated values of the one or more performance metrics to the user.

19. A computing system for processing data, the computing system including:
means for receiving multiple units of work that each include one or more records;
determining a characteristic of a first unit of work based on data included in the one or more records of the first unit of work;
means for identifying, by a component of a first dataflow graph, a second dataflow graph from multiple available dataflow graphs based on the determined characteristic, the multiple available dataflow graphs being stored in a data storage system;
means for processing the first unit of work using the second dataflow graph based on loading the second dataflow graph into the first dataflow graph;
means for receiving, by the first dataflow graph, one or more values of one or more performance metrics associated with the processing of the first unit of work using the loaded second dataflow graph;
means for outputting, by the first dataflow graph to a performance monitoring module, the one or more received values of the one or more performance metrics associated with the processing of the first unit of work using the loaded second dataflow graph, wherein the performance monitoring module receives a plurality of values of one or more performance metrics output by a plurality of dataflow graphs;
means for selecting, from among the plurality of values of the one or more performance metrics output by the plurality of dataflow graphs, values of one or more performance metrics associated with an identifier for the second dataflow graph; and
means for aggregating the values selected for the one or more performance metrics associated with the identifier for the second dataflow graph.

20. The system of claim 19, wherein the second dataflow graph is compiled independent of the first dataflow graph.

21. The system of claim 19, further including means for storing the one or more values of the one or more performance metrics associated with the processing of the first unit of work using the loaded second dataflow graph in an in-memory data store.

22. The system of claim 21, further including means for transferring the stored one or more values of the one or more performance metrics to a persistent data store.

23. The system of claim 19, wherein the values selected include first values, the system further including means for aggregating second values of the one or more performance metrics output by the plurality of dataflow graphs based on an identifier associated with the first dataflow graph.

24. The system of claim 19, further including means for displaying the one or more performance metrics to a user or displaying the aggregated values of the one or more performance metrics to the user.