CLAIMS

1. A method for generating an updated dataflow graph based on a base graph, in which each of the base graph and the updated dataflow graph is configured to process data in one or more fields of data records having a record format, the method including:

identifying differences between a first dataflow graph and each of a second dataflow graph and a third dataflow graph, wherein the first dataflow graph includes the base graph, the second dataflow graph includes a source graph, and the third dataflow graph includes a target graph,

wherein the specification of each of the dataflow graphs defines nodes, at least one of the nodes representing a data processing component defining an operation to be performed to process data in one or more fields of data records having a record format, the data records being provided to the data processing component, and one or more links connecting the nodes and each representing a flow of data records, and

wherein the identifying includes:

comparing a specification of the base graph and a specification of the source graph to identify a first set of differences between the base graph and the source graph, the first set of differences including information indicative of one or more of a first deleted node, a first deleted link, a first added node, a first added link, a first modified node, or a first modified link, and

comparing a specification of the base graph and a specification of the target graph to identify a second set of differences between the base graph and the target graph, the second set of differences including information indicative of one or more of a second deleted node, a second deleted link, a second added node, a second added link, a second modified node, or a second modified link;

characterizing a correspondence between the differences in the first set of differences and the differences in the second set of differences, including identifying one or more of (i) a node or link identified in the first set of differences that corresponds to a node or link identified in the second set of differences, (ii) a node or link identified in the first set of differences that does not correspond to any node or link identified in the second set of differences, or (iii) a node or link identified in the second set of differences that does not correspond to any node or link identified in the first set of differences;

determining a set of combined differences based on the characterized correspondence, the first set of differences between the base graph and the source graph, and the second set of differences between the base graph and the target graph;

generating the updated dataflow graph based on the base graph and the set of combined differences, including generating a specification of the updated dataflow graph that defines nodes, at least one of the nodes representing a data processing component defining an operation to be performed to process data in one or more fields of data records having a record format, the data records being provided to the data processing component, and that defines one or more links connecting the nodes and each representing a flow of data records; and

executing the updated dataflow graph to process data in one or more fields of input data records received from a data source external to the updated dataflow graph, the executing including:

receiving input data records into an input port of a first node of the updated dataflow graph, wherein the first node represents a first data processing component defining a first operation;

processing the data in the one or more fields of the input data records according to the first operation defined by the first data processing component; and

outputting output data records based on the processing, the output data records including one or more fields containing output data, including outputting the output data records onto a link connected to an output port of the first node, wherein the link is connected either to an input port of a second node of the updated dataflow graph or to a data destination external to the updated dataflow graph.

2. The method of claim 1, in which the source graph, the base graph, the target graph, and the updated dataflow graph are executable dataflow graphs.

3. The method of claim 1, in which the source graph is a first modified version of the base graph, and in which the target graph is a second modified version of the base graph.

4. The method of claim 1, including generating the first set of differences and the second set of differences, each set of differences including, for each difference in the set:

information identifying a type of the difference, and

one or more of (i) information indicative of one or more nodes associated with the difference or (ii) information indicative of one or more links associated with the difference.

5. The method of claim 4, in which the type of the difference includes one or more of a deletion, an addition, or a modification.

6. The method of claim 4, in which the information indicative of one or more nodes associated with the difference includes (i) information identifying a deleted, added, or modified node and one or more of (ii) information identifying a node upstream of the deleted, added, or modified node and (iii) information identifying a node downstream of the deleted, added, or modified node.

7. The method of claim 4, in which the information indicative of one or more links associated with the difference includes one or more of (i) information identifying an added ´link and (ii) information identifying a deleted link.

8. The method of claim 1, in which generating the set of combined differences includes including, in the set of combined differences, information indicative of the node or link identified in the first set of differences that does not correspond to any node or link identified in the second set of differences.

9. The method of claim 1, in which generating the set of combined differences includes including, in the set of combined differences, information indicative the node or link identified in the second set of differences that does not correspond to any node or link identified in the first set of differences.

10. The method of claim 1, including determining a compatibility between the node or link identified in the first set of differences and the corresponding node or link identified in the second set of differences.

11. The method of claim 10, in which the node or link identified in the first set of differences includes a first modified node or link and the node or link identified in the second set of differences includes a second modified node or link, and

in which determining a compatibility includes:

determining that a modification associated with the first modified node or link is compatible with a modification associated with the second modified node or link ; and

including, in the set of combined differences, information indicative of both the node or link identified in the first set of differences and the node or link identified in the second set of differences.

12. The method of claim 10, in which the node or link identified in the first set of differences includes a first modified node or link and the node or link identified in the second set of differences includes a second modified node or link, and

in which determining a compatibility includes:

determining that a modification associated with the first modified node or link is compatible with a modification associated with the second modified node or link;

selecting the first modified node or link or the second modified node or link; and

including information indicative of the selected node or link in the set of combined differences.

13. The method of claim 1, in which determining the set of combined differences includes:

for a pair of differences determined to be corresponding, determining that the differences are compatible; and

including information indicative of both differences of the pair in the set of combined differences.

14. The method of claim 1, in which determining the set of combined differences includes:

for a pair of differences determined to be corresponding, determining that the differences are conflicting;

selecting one of the differences of the pair; and

including information indicative of the selected difference in the set of combined differences.

15. The method of claim 14, including selecting one of the differences based on a selection criterion.

16. The method of claim 14, including:

presenting, on a user interface, information indicative of the both differences of the pair; and

selecting the one of the differences based on user input.

17. The method of claim 1, in which generating the updated dataflow graph includes:

generating a copy of the base graph; and

applying each difference in the set of combined differences to the copy.

18. The method of claim 1, in which comparing a specification of the base graph and a specification of the source graph includes determining a correspondence between the base graph and the source graph.

19. The method of claim 18, in which determining a correspondence between the base graph and the source graph includes one or more of (1) identifying a first node or link of the base graph that corresponds to a node or link of the source graph, (2) identifying a second node or link of the base graph that does not correspond to any node of the source graph, and (3) identifying a third node or link of the source graph that does not correspond to any node or link of the base graph.

20. The method of claim 19, in which identifying a first set of differences between the base graph and the source graph includes identifying a difference between the first node or link of the base graph and the corresponding node or link of the source graph.

21. The method of claim 20, in which identifying a difference between the first node or link and the corresponding node or link of the source graph includes identifying a difference between one or more of a parameter and an expression for a parameter of the first node or link and a parameter or an expression for a parameter of the corresponding first node or link.

22. The method of claim 1, in which the base graph contains a base subgraph and in which the source graph contains a first subgraph, and including:

comparing a specification of the base subgraph and a specification of the first subgraph to identify one or more differences between the base subgraph and the first subgraph.

23. The method of claim 1, in which the first dataflow graph is the base graph, the second dataflow graph is the source graph, and the third dataflow graph is the target graph.

24. The method of claim 1, wherein the operation defined by each data processing component is defined at least in part by one or more parameters that are referenced by the data processing component.

25. The method of claim 1, wherein the updated dataflow graph is based on a copy of the base graph.

26. The method of claim 1, wherein the base graph, the source graph, and the target graph include data processing graphs.

27. The method of claim 1, wherein the base graph, the source graph, and the target graph include control graphs.

28. The method of claim 1, wherein the base graph, the source graph, and the target graph include plans.

29. The method of claim 1, wherein the determining of the correspondence between the differences includes:

identifying (1) differences that involve nodes or links that are identified in only one of the first and second sets of differences and (2) differences that involve nodes or links that are identified in both of the first and second sets of differences.

30. The method of claim 1, wherein the determining of the set of combined differences includes:

for each difference of the first and second sets of differences that involves a node or a link that is identified in only one of the first and second sets of differences, information indicative of that difference is added to the set of combined differences, and

for each difference of the first and second sets of differences that involves a node or a link that is identified in both of the first and second sets of differences, a compatibility analysis is performed.

31. The method of claim 30, wherein the performing of the compatibility analysis includes:

determining whether or not a first difference is conflicting with a second difference, wherein a pair of conflicting differences are conflicting if application of one of the pair of conflicting differences to the base executable graph would prevent application of the other one of the pair of conflicting differences to the base executable graph.

32. The method of claim 31, including:

in case of any pair of conflicting differences, resolving the conflict of the pair of conflicting differences by selecting one of the pair of conflicting differences according to a selection criterion, and

adding information indicative of the selected difference the set of combined differences.

33. The method of claim 31, including:

providing a graphical user interface; and

displaying, via the graphical user interface, information visually representing the pair of conflicting differences.

34. The method of claim 33, including:

displaying, via the graphical user interface, one or more visual representations of one or more selections of a difference among the pair of conflicting differences and a visual representation of the respective updated dataflow graph that would result from the respective selections.

35. The method of claim 34, including:

receiving, from a user and via the graphical user interface, an input including a selection of one of the visual representations, wherein the selection criterion is defined by the input.

36. A non-transitory computer readable medium storing instructions for causing a computing system to generate an updated dataflow graph based on a base graph, in which each of the base graph and the updated dataflow graph is configured to process data in one or more fields of data records having a record format, the instructions including instructions for causing the computing system to:

identify differences between a first dataflow graph and each of a second dataflow graph and a third dataflow graph, wherein the first dataflow graph includes the base graph, the second dataflow graph includes a source graph, and the third dataflow graph includes a target graph,

wherein the specification of each of the dataflow graphs defines nodes, at least one of the nodes representing a data processing component defining an operation to be performed to process data in one or more fields of data records having a record format, the data records being provided to the data processing component, and one or more links connecting the nodes and each representing a flow of data records, and

wherein the identifying includes:

comparing a specification of the base graph and a specification of the source graph to identify a first set of differences between the base graph and the source graph, the first set of differences including information indicative of one or more of a first deleted node, a first deleted link, a first added node, a first added link, a first modified node, or a first modified link, and

comparing a specification of the base graph and a specification of the target graph to identify a second set of differences between the base graph and the target graph, the second set of differences including information indicative of one or more of a second deleted node, a second deleted link, a second added node, a second added link, a second modified node, or a second modified link;

characterize a correspondence between the differences in the first set of differences and the differences in the second set of differences, including identifying one or more of (i) a node or link identified in the first set of differences that corresponds to a node or link identified in the second set of differences, (ii) a node or link identified in the first set of differences that does not correspond to any node or link identified in the second set of differences, or (iii) a node or link identified in the second set of differences that does not correspond to any node or link identified in the first set of differences;

determine a set of combined differences based on the characterized correspondence, the first set of differences between the base graph and the source graph, and the second set of differences between the base graph and the target graph;

generate the updated dataflow graph based on the base graph and the set of combined differences, including generating a specification of the updated dataflow graph that defines nodes, at least one of the nodes representing a data processing component defining an operation to be performed to process data in one or more fields of data records having a record format, the data records being provided to the data processing component, and that defines one or more links connecting the nodes and each representing a flow of data records; and

execute the updated dataflow graph to process data in one or more fields of input data records received from a data source external to the updated dataflow graph, the executing including:

receiving input data records into an input port of a first node of the updated dataflow graph, wherein the first node represents a first data processing component defining a first operation;

processing the data in the one or more fields of the input data records according to the first operation defined by the first data processing component; and

outputting output data records based on the processing, the output data records including one or more fields containing output data, including outputting the output data records onto a link connected to an output port of the first node, wherein the link is connected either to an input port of a second node of the updated dataflow graph or to a data destination external to the updated dataflow graph.

37. A computing system for generating an updated dataflow graph based on a base graph, in which each of the base graph and the updated dataflow graph is configured to process data in one or more fields of data records having a record format, the computing system including:

one or more processors coupled to a memory, the one or more processors and memory configured to:

identify differences between a first dataflow graph and each of a second dataflow graph and a third dataflow graph, wherein the first dataflow graph includes the base graph, the second dataflow graph includes a source graph, and the third dataflow graph includes a target graph,

wherein the specification of each of the dataflow graphs defines nodes, at least one of the nodes representing a data processing component defining an operation to be performed to process data in one or more fields of data records having a record format, the data records being provided to the data processing component, and one or more links connecting the nodes and each representing a flow of data records, and

wherein the identifying includes:

comparing a specification of the base graph and a specification of the source graph to identify a first set of differences between the base graph and the source graph, the first set of differences including information indicative of one or more of a first deleted node, a first deleted link, a first added node, a first added link, a first modified node, or a first modified link, and

comparing a specification of the base graph and a specification of the target graph to identify a second set of differences between the base graph and the target graph, the second set of differences including information indicative of one or more of a second deleted node, a second deleted link, a second added node, a second added link, a second modified node, or a second modified link;

characterizing a correspondence between the differences in the first set of differences and the differences in the second set of differences, including identifying one or more of (i) a node or link identified in the first set of differences that corresponds to a node or link identified in the second set of differences, (ii) a node or link identified in the first set of differences that does not correspond to any node or link identified in the second set of differences, or (iii) a node or link identified in the second set of differences that does not correspond to any node or link identified in the first set of differences;

determining a set of combined differences based on the characterized correspondence, the first set of differences between the base graph and the source graph, and the second set of differences between the base graph and the target graph;

generating the updated dataflow graph based on the base graph and the set of combined differences, including generating a specification of the updated dataflow graph that defines nodes, at least one of the nodes representing a data processing component defining an operation to be performed to process data in one or more fields of data records having a record format, the data records being provided to the data processing component, and that defines one or more links connecting the nodes and each representing a flow of data records; and

executing the updated dataflow graph to process data in one or more fields of input data records received from a data source external to the updated dataflow graph, the executing including:

receiving input data records into an input port of a first node of the updated dataflow graph, wherein the first node represents a first data processing component defining a first operation;

processing the data in the one or more fields of the input data records according to the first operation defined by the first data processing component; and

outputting output data records based on the processing, the output data records including one or more fields containing output data, including outputting the output data records onto a link connected to an output port of the first node, wherein the link is connected either to an input port of a second node of the updated dataflow graph or to a data destination external to the updated dataflow graph.

38. A computing system for generating an updated dataflow graph based on a base graph, in which each of the base graph and the updated dataflow graph is configured to process data in one or more fields of data records having a record format, the computing system including:

means for identifying differences between a first dataflow graph and each of a second dataflow graph and a third dataflow graph, wherein the first dataflow graph includes a base graph, the second dataflow graph includes a source graph, and the third dataflow graph includes a target graph,

wherein the specification of each of the dataflow graphs defines nodes, at least one of the nodes representing a data processing component defining an operation to be performed to process data in one or more fields of data records having a record format, the data records being provided to the data processing component, and one or more links connecting the nodes and each representing a flow of data records, and

wherein the identifying includes:

comparing a specification of the base graph and a specification of the source graph to identify a first set of differences between the base graph and the source graph, the first set of differences including information indicative of one or more of a first deleted node, a first deleted link, a first added node, a first added link, a first modified node, or a first modified link, and

comparing a specification of the base graph and a specification of the target graph to identify a second set of differences between the base graph and the target graph, the second set of differences including information indicative of one or more of a second deleted node, a second deleted link, a second added node, a second added link, a second modified node, or a second modified link;

means for characterizing a correspondence between the differences in the first set of differences and the differences in the second set of differences, including identifying one or more of (i) a node or link identified in the first set of differences that corresponds to a node or link identified in the second set of differences, (ii) a node or link identified in the first set of differences that does not correspond to any node or link identified in the second set of differences, or (iii) a node or link identified in the second set of differences that does not correspond to any node or link identified in the first set of differences;

means for determining a set of combined differences based on the characterized correspondence, the first set of differences between the base graph and the source graph, and the second set of differences between the base graph and the target graph; and

means for generating the updated dataflow graph based on the base graph and the set of combined differences, including generating a specification of the updated dataflow graph that defines nodes, at least one of the nodes representing a data processing component defining an operation to be performed to process data in one or more fields of data records having a record format, the data records being provided to the data processing component, and that defines one or more links connecting the nodes and each representing a flow of data records; and

means for executing the updated dataflow graph to process data in one or more fields of input data records received from a data source external to the updated dataflow graph, the executing including:

receiving input data records into an input port of a first node of the updated dataflow graph, wherein the first node represents a first data processing component defining a first operation;

processing the data in the one or more fields of the input data records according to the first operation defined by the first data processing component; and

outputting output data records based on the processing, the output data records including one or more fields containing output data, including outputting the output data records onto a link connected to an output port of the first node, wherein the link is connected either to an input port of a second node of the updated dataflow graph or to a data destination external to the updated dataflow graph.

39. The method of claim 1, in which identifying differences includes identifying differences based on an analysis of a record format file referenced by the base graph and a record format file referenced by the target graph.

40. The method of claim 39, in which identifying differences includes identifying differences between the record format file referenced by the base graph and the record format file referenced by the target graph.