

What is claimed is:

1. A method performed by a data processing apparatus, the method including:
 - receiving a portion of metadata from a data source, the portion of metadata describing nodes and edges, at least some of the edges each representing an effect of one node upon another node, each edge having a single direction;
 - generating instances of a data structure representing the portion of metadata, at least one instance of the data structure including:
 - an identification value that identifies a corresponding node,
 - one or more property values representing respective properties of the corresponding node, and
 - one or more pointers to respective identification values, each pointer representing an edge associated with a node identified by the corresponding respective identification value;
 - storing the instances of the data structure in random access memory;
 - receiving a query that includes an identification of at least one particular element of data; and
 - using at least one instance of the data structure to cause a display of a computer system to display a representation of lineage of the particular element of data.
2. The method of claim 1, wherein the query includes an identification of a type of lineage and a walk plan that identifies which types of nodes and edges are relevant to the identified type of lineage.
3. The method of claim 2, wherein the walk plan includes conditions for following or collecting an edge based on one or more property values representing respective properties of a corresponding node.
4. The method of any of claims 1 to 3, wherein at least one of the identification values represents a node corresponding to the particular element of data identified in the query.

5. The method of any of claims 1 to 4, wherein the data processing apparatus is configured to return a response to the query, the response including metadata describing lineage of the particular element of data.

6. The method of claim 5, wherein the metadata describing lineage of the particular element of data includes metadata describing a sequence of nodes and edges, wherein one of the nodes of the sequence represents the particular element of data.

7. The method of any of claims 1 to 6, wherein the set of metadata includes text strings associated with respective nodes, wherein the data structure omits at least one of the text strings.

8. The method of any of claims 1 to 7, wherein generating the instances of the data structures includes identifying metadata that is relevant to lineage and storing the identified metadata in a format defined by the data structure.

9. The method of any of claims 1 to 8, wherein at least one of the pointers includes a reference to a memory location associated with a portion of the random access memory, wherein the memory location corresponds to a location of a node corresponding to one of the identity values.

10. The method of any of claims 1 to 9, including
in response to receiving the query:
accessing data stored using the data structure, including accessing an instance of the data structure containing at least one identifier that corresponds to the particular element of data identified by the query; and
based on the data stored using the data structure, generating a response to the query, the response including lineage metadata, the lineage metadata describing a sequence of nodes and edges, wherein one of the nodes of the sequence represents the particular element of data.

11. The method of any of claims 1 to 10, including
accessing a first instance of the data structure, the first instance of the data structure identified by a first identifier that corresponds to the particular element of data identified by the query;
accessing a pointer associated with the first identifier, the pointer associated with the first identifier;
accessing a second instance of the data structure, the second instance of the data structure identified by a second identifier, the second identifier being stored at a memory location referenced by the pointer associated with the first identifier; and
collecting data of the second instance of the data structure;
wherein the response includes data representing a node associated with the second identifier.

12. The method of any of claims 1 to 11, including transmitting requests to the data source on scheduled intervals, wherein the portion of metadata is received from the data source in response to one of the requests.

13. A system including:
at least one non-transitory computer-readable storage medium storing executable instructions; and
one or more processors configured to execute the instructions;
wherein execution of the instructions causes:
receiving a portion of metadata from a data source, the portion of metadata describing nodes and edges, at least some of the edges each representing an effect of one node upon another node, each edge having a single direction;
generating instances of a data structure representing the portion of metadata, at least one instance of the data structure including:
an identification value that identifies a corresponding node,
one or more property values representing respective properties of the corresponding node, and

one or more pointers to respective identification values, each pointer representing an edge associated with a node identified by the corresponding respective identification value;

storing the instances of the data structure in random access memory;

receiving a query that includes an identification of at least one particular element of data; and

using at least one instance of the data structure to cause a display of a computer system to display a representation of lineage of the particular element of data.

14. A non-transitory computer readable storage device storing instructions that, when executed, carry out operations including:

receiving a portion of metadata from a data source, the portion of metadata describing nodes and edges, at least some of the edges each representing an effect of one node upon another node, each edge having a single direction;

generating instances of a data structure representing the portion of metadata, at least one instance of the data structure including:

an identification value that identifies a corresponding node,

one or more property values representing respective properties of the corresponding node, and

one or more pointers to respective identification values, each pointer representing an edge associated with a node identified by the corresponding respective identification value;

storing the instances of the data structure in random access memory;

receiving a query that includes an identification of at least one particular element of data; and

using at least one instance of the data structure to cause a display of a computer system to display a representation of lineage of the particular element of data.

15. A system including:

means for receiving a portion of metadata from a data source, the portion of metadata describing nodes and edges, at least some of the edges each representing an effect of one node upon another node, each edge having a single direction;

means for generating instances of a data structure representing the portion of metadata, at least one instance of the data structure including:

an identification value that identifies a corresponding node,

one or more property values representing respective properties of the corresponding node, and

one or more pointers to respective identification values, each pointer representing an edge associated with a node identified by the corresponding respective identification value;

means for storing the instances of the data structure in random access memory;

means for receiving a query that includes an identification of at least one particular element of data; and

means for using at least one instance of the data structure to cause a display of a computer system to display a representation of lineage of the particular element of data.

16. A method performed by a data processing apparatus, the method including:

receiving a query that includes an identification of at least one particular element of data;

receiving an identification of a type of lineage;

receiving a walk plan which identifies which types of nodes and edges are relevant to the identified type of lineage,

wherein the walk plan includes 1) at least one identifier corresponding to a type of edge, and 2) at least one identifier corresponding to a direction of traversal for the type of edge, and 3) at least one indicator corresponding to an action to be taken when traversing the edge;

traversing, based on the walk plan, nodes and edges; and

based on the traversal of the nodes and edges, causing a display of a computer system to display a representation of lineage of the particular element of data.

17. The method of claim 16, wherein the nodes correspond to respective instances of a data structure, and the edges correspond to respective pointers to a respective instance of the data structure.

18. The method of claim 17, wherein an instance of the data structure includes:
an identification value that identifies a corresponding node,
one or more property values representing respective properties of the corresponding node, and
one or more pointers to respective identification values, each pointer representing an edge associated with a node identified by the corresponding respective identification value.

19. The method of any of claims 16 to 18, wherein the walk plan is chosen based on a data type of the particular element of data identified by the query.

20. The method of any of claims 16 to 19, wherein the walk plan includes a structured document.