

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 a relationship between an effect of one node and upon another node, each edge having a single direction;
generating instances of a data structure representing the portion of metadata and storing the instances of the data structure in ~~random access~~-memory, 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 memory locations associated with respective portions of the ~~random access~~-memory, wherein at least one of the memory locations corresponds to a location of another generated instance of the data structure, each pointer representing an edge between the corresponding node and another node, and associated with a node identified by a corresponding respective identification value included in the other generated instance of the data structure;
 - one or more type values that identify a type of an edge represented by a respective pointer;receiving a query that includes an identification of at least one particular element of data; ~~and~~
in response to the query, traversing at least some of the instances of the data structure stored in memory to determine lineage of the particular data element by:
 - accessing an instance of the data structure that corresponds to the at least one particular element of data identified by the query; and
 - accessing one or more other instances of the data structure based at least in part on the one or more pointers, the one or more property values, and the one or more type values contained in the instance of the data structure; and

~~generating a response to the query, the response including data representing the determined 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 at least one 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 portion 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, 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.

10. The method of any of claims 1 to 9, 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.

11. The method of any of claims 1 to 10, 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.

12. 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 performance of the steps of any of claims 1-11.

13. A non-transitory computer readable storage device storing instructions that, when executed, carry out the steps of any of claims 1-11.

14. A software program that, when executed, carries out the steps of any of claims 1-11.

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 a relationship between~~an effect of~~ one node and upon another node,~~each edge having a single direction~~;
means for generating instances of a data structure representing the portion of metadata and storing the instances of the data structure in ~~random access~~-memory, 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 memory locations associated with respective portions of the ~~random access~~-memory, wherein at least one of the memory locations corresponds to a location of another generated instance of the data structure, each pointer representing an edge between the corresponding node and another node, and~~associated with a node identified by a corresponding respective identification value included in the other generated instance of the data structure~~;

one or more type values that identify a type of an edge represented by a respective pointer;
means for receiving a query that includes an identification of at least one particular element of data; ~~and~~
means for, in response to the query, traversing at least some of the instances of the data structure stored in memory to determine lineage of the particular data element by:
accessing an instance of the data structure that corresponds to the at least one particular element of data identified by the query; and
accessing one or more other instances of the data structure based at least in part on the one or more pointers, the one or more property values, and the one or more type values contained in the instance of the data structure; and
means for generating a response to the query, the response including data representing the determined 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 at least one 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, wherein the nodes correspond to respective instances of a data structure stored in random access memory, and the edges correspond to respective pointers to respective locations in the random access memory corresponding to respective instances of the data structure; 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 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.

18. The method of any of claims 16 or 17, wherein the walk plan corresponds to a data type of the particular element of data identified by the query.

19. The method of any of claims 16 to 18, wherein the walk plan includes a structured document.

20. 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 performance of the steps of any of claims 16-19.

21. A non-transitory computer readable storage device storing instructions that, when executed, carry out the steps of any of claims 16-19.

22. A software program that, when executed, carries out the steps of any of claims 16-19.

23. A system including:

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

means for receiving an identification of a type of lineage;

means for 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;

means for traversing, based on the walk plan, nodes and edges, wherein the nodes correspond to respective instances of a data structure stored in random access memory, and the edges correspond to respective pointers to respective locations in the random access memory corresponding to respective instances of the data structure; and

means for 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.