

CLAIMS

1. A method, performed by one or more data processing apparatus including:
receiving user input selecting a first node in a first graph that includes nodes;
receiving a first data lineage of a first object associated with the first node, the first object having a type, the first data lineage describing relationships between the first object and one or more datasets or transforms;
receiving user input selecting a second node from a second graph that includes nodes;
receiving a second data lineage of a second object associated with the second node, the second object having the same type as the first object, the second data lineage describing relationships between the second object and one or more other datasets or transforms;
performing a comparison of the first node selected from the first graph and the first data lineage to the second node selected from the second graph and the second data lineage; and
generating a report based on the comparison.
2. The method of claim 1, wherein the type of the first node is one of a group consisting of a field of a dataset, a column of a dataset, a dataset, or a transformation.
3. The method of claim 1, wherein a graph includes nodes representing components of a data processing system and directed edges representing data flows between components of the data processing system.
4. The method of claim 1, in which the report specifies structural differences in between the first object and the second object.
5. The method of claim 1, wherein the first object is a first version of an object and the second object is a second version of the object.
6. The method of claim 5, further including:
accessing the first version of the object using a unique tag for the first version of object and a look-up table; and
accessing the second version of the object using a unique tag for the second version of the object and a look-up table.
7. The method of claim 5, in which the first version of the object and the second version of the

object have an ancestor-descendant relationship in a revision history.

8. The method of claim 5, in which the first version of the object and the second version of the object have a common ancestor in a revision history, in which the first version of the object is not a descendent in the revision history of the second version of the object, and in which the second version of the object is not a descendent in the revision history of the first version of the object.

9. The method of claim 5, in which the first version of the object and the second version of the object describe respective first and second implementations of a data processing system that are implemented concurrently.

10. A non-transitory computer storage medium encoded with computer program instructions that when executed by one or more computers cause the one or more computers to perform operations comprising:

- receiving user input selecting a first node in a first graph that includes nodes;

- receiving a first data lineage of a first object associated with the first node, the first object having a type, the first data lineage describing relationships between the first object and one or more datasets or transforms;

- receiving user input selecting a second node from a second graph that includes nodes;

- receiving a second data lineage of a second object associated with the second node, the second object having the same type as the first object, the second data lineage describing relationships between the second object and one or more other datasets or transforms;

- performing a comparison of the first node selected from the first graph and the first data lineage to the second node selected from the second graph and the second data lineage; and

- generating a report based on the comparison.

11. The medium of claim 10, wherein the type of the first node is one of a group consisting of a field of a dataset, a column of a dataset, a dataset, or a transformation.

12. The medium of claim 10, wherein a graph includes nodes representing components of a data processing system and directed edges representing data flows between components of the data processing system.

13. The medium of claim 10, in which the report specifies structural differences in between the first object and the second object.

14. The medium of claim 10, wherein the first object is a first version of an object and the second object is a second version of the object.
15. The medium of claim 14, further including:
accessing the first version of the object using a unique tag for the first version of object and a look-up table; and
accessing the second version of the object using a unique tag for the second version of the object and a look-up table.
16. The medium of claim 14, in which the first version of the object and the second version of the object have an ancestor-descendant relationship in a revision history.
17. The medium of claim 14, in which the first version of the object and the second version of the object have a common ancestor in a revision history, in which the first version of the object is not a descendent in the revision history of the second version of the object, and in which the second version of the object is not a descendent in the revision history of the first version of the object.
18. The medium of claim 14, in which the first version of the object and the second version of the object describe respective first and second implementations of a data processing system that are implemented concurrently.
19. A system comprising:
one or more computers and one or more storage devices storing instructions that are operable, when executed by the one or more computers, to cause the one or more computers to perform operations comprising:
receiving user input selecting a first node in a first graph that includes nodes;
receiving a first data lineage of a first object associated with the first node, the first object having a type, the first data lineage describing relationships between the first object and one or more datasets or transforms;
receiving user input selecting a second node from a second graph that includes nodes;
receiving a second data lineage of a second object associated with the second node, the second object having the same type as the first object, the second data lineage describing relationships between the second object and one or more other datasets or transforms;

performing a comparison of the first node selected from the first graph and the first data lineage to the second node selected from the second graph and the second data lineage;
and
generating a report based on the comparison.

20. The system of claim 19, wherein the type of the first node is one of a group consisting of a field of a dataset, a column of a dataset, a dataset, or a transformation.

21. The system of claim 19, wherein a graph includes nodes representing components of a data processing system and directed edges representing data flows between components of the data processing system.

22. The system of claim 19, in which the report specifies structural differences in between the first object and the second object.

23. The system of claim 19, wherein the first object is a first version of an object and the second object is a second version of the object.

24. The system of claim 23, further including:
accessing the first version of the object using a unique tag for the first version of object and a look-up table; and
accessing the second version of the object using a unique tag for the second version of the object and a look-up table.

25. The system of claim 23, in which the first version of the object and the second version of the object have an ancestor-descendant relationship in a revision history.

26. The system of claim 23, in which the first version of the object and the second version of the object have a common ancestor in a revision history, in which the first version of the object is not a descendent in the revision history of the second version of the object, and in which the second version of the object is not a descendent in the revision history of the first version of the object.

27. The system of claim 23, in which the first version of the object and the second version of the object describe respective first and second implementations of a data processing system that are implemented concurrently.

28. A system comprising:

one or more computers and one or more storage devices storing instructions that are operable, when executed by the one or more computers, to cause the one or more computers to perform operations comprising:

means for receiving user input selecting a first node in a first graph that includes nodes;

means for receiving a first data lineage of a first object associated with the first node, the first object having a type, the first data lineage describing relationships between the first object and one or more datasets or transforms;

means for receiving user input selecting a second node from a second graph that includes nodes;

means for receiving a second data lineage of a second object associated with the second node, the second object having the same type as the first object, the second data lineage describing relationships between the second object and one or more other datasets or transforms;

means for performing a comparison of the first node selected from the first graph and the first data lineage to the second node selected from the second graph and the second data lineage; and

means for generating a report based on the comparison.