WHAT IS CLAIMED IS:

1. A method implemented by a data processing system for providing a development environment and storage that stores datasets having one or more attributes, and with the development environment providing access to the one or more attributes of the datasets, including:

accessing a schema that specifies relationships among datasets represented in the schema, one or more computations on one or more of the datasets, or one or more transformations of one or more of the datasets;

identifying, from among the datasets, a plurality of the datasets in storage, by:

selecting a dataset from among the datasets; and

identifying, from the schema, one or more other datasets that are related to the selected dataset;

identifying attributes of the plurality of the datasets;

generating logical data representing identified attributes of the plurality of the datasets and further representing one or more relationships among the attributes;

providing, to a development environment, the logical data;

providing, by the development environment, access to one or more portions of the logical data representing the identified attributes of the plurality of the datasets;

receiving, from the development environment, a specification that specifies at least one of the identified attributes in performing an operation; and

based on the specification and on the one or more relationships among the identified attributes represented by the logical data, generating a computer program that is configured to perform the operation by accessing, from storage, at least one dataset from the plurality, with the at least one dataset accessed having the at least one of the attributes specified in the specification.

2. The method of claim 1, wherein the development environment provides access to the one or more portions of the logical data without accessing the plurality of datasets from storage.

3. The method of claim 1, comprising:

identifying a dataset from the plurality of datasets including the at least one of the attributes specified in the specification; and

accessing, from storage, the identified dataset.

4. The method of claim 1, comprising executing the computer program using the at least one dataset accessed from storage.

5. The method of claim 1, comprising optimizing the computer program to produce an optimized computer program that is configured to perform the operation by accessing, from storage, only those datasets in the plurality of datasets having the at least one of the attributes specified in the specification.

6. The method of claim 1, wherein the one or more attributes include field names of the plurality of the datasets.

7. The method of claim 1, wherein the one or more attributes include information for accessing the plurality of the datasets in storage.

8. The method of claim 1, comprising identifying, from the schema, one or more parameters for joining the selected dataset and the one or more other datasets.

9. The method of claim 8, wherein the one or more parameters include a key for joining the selected dataset and at least one of the one or more other datasets.

10. The method of claim 1, comprising receiving, from a client device, selection data specifying the selected dataset.

11. The method of claim 1, wherein the selected dataset comprises a root node of the logical data, and wherein at least one of the one or more other datasets are joined to the selected dataset.

12. The method of claim 1, wherein the one or more computations on one or more of the datasets or one or more transformations of one or more of the datasets define a virtual field for at least one of the plurality of the datasets.

13. The method of claim 1, comprising generating, based on the specification and on the one or more relationships among the identified attributes represented by the logical data, an executable dataflow graph that is configured to perform the operation, wherein the executable dataflow graph includes at least one of the one or more attributes as an input.

14. The method of claim 1, comprising removing from the computer program an operation to access, from storage, at least one dataset in the plurality of datasets that does not include the at least one of the attributes specified in the specification.

15. The method of claim 1, wherein the computer program is configured to access, from storage, at least some data from the plurality by a select statement, wherein the select statement is minimized to select only the at least one of the attributes specified in the specification.

16. The method of claim 1, wherein the development environment reads the logical data as a data source.

17. A system for providing a development environment and storage that stores datasets having one or more attributes, and with the development environment providing access to the one or more attributes of the datasets, including:

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

accessing a schema that specifies relationships among datasets represented in the schema, one or more computations on one or more of the datasets, or one or more transformations of one or more of the datasets;

identifying, from among the datasets, a plurality of the datasets in storage, by:

selecting a dataset from among the datasets; and

identifying, from the schema, one or more other datasets that are related to the selected dataset;

identifying attributes of the plurality of the datasets;

generating logical data representing identified attributes of the plurality of the datasets and further representing one or more relationships among the attributes;

providing, to a development environment, the logical data;

providing, by the development environment, access to one or more portions of the logical data representing the identified attributes the plurality of the datasets;

receiving, from the development environment, a specification that specifies at least one of the identified attributes in performing an operation; and

based on the specification and on the one or more relationships among the identified attributes represented by the logical data, generating a computer program that is configured to perform the operation by accessing, from storage, at least one dataset from the plurality, with the at least one dataset accessed having the at least one of the attributes specified in the specification.

18. The system of claim 17, wherein the computer program is configured to access, from storage, only those datasets having the at least one of the attributes specified in the specification.

19. A non-transitory computer-readable storage medium storing instructions for causing a computing system to:

access a schema that specifies relationships among datasets represented in the schema, one or more computations on one or more of the datasets, or one or more transformations of one or more of the datasets;

identify, from among the datasets, a plurality of the datasets in storage, by:

selecting a dataset from among the datasets; and

identifying, from the schema, one or more other datasets that are related to the selected dataset;

identify attributes of the plurality of the datasets;

generate logical data representing identified attributes of the plurality of the datasets and further representing one or more relationships among the attributes;

provide, to a development environment, the logical data;

provide, by the development environment, access to one or more portions of the logical data representing the identified attributes the plurality of the datasets;

receive, from the development environment, a specification that specifies at least one of the identified attributes in performing an operation; and

based on the specification and on the one or more relationships among the identified attributes represented by the logical data, generate a computer program that is configured to perform the operation by accessing, from storage, at least one dataset from the plurality, with the at least one dataset accessed having the at least one of the attributes specified in the specification.

20. The non-transitory computer-readable storage medium of claim 19, wherein the computer program is configured to access, from storage, only those datasets having the at least one of the attributes specified in the specification.

WHAT IS CLAIMED IS:

1. A method implemented by a data processing system for providing one or more interfaces and storage that stores datasets having one or more attributes, and with the one or more interfaces providing access to the one or more attributes of the datasets, including:

accessing a schema that specifies relationships among datasets represented in the schema, one or more computations on one or more of the datasets, or one or more transformations of one or more of the datasets;

identifying, from among the datasets, a plurality of the datasets in storage, by:

selecting a dataset from among the datasets; and

identifying, from the schema, one or more other datasets that are related to the selected dataset;

identifying attributes of the plurality of the datasets;

generating logical data representing identified attributes of the plurality of the datasets and further representing one or more relationships among the attributes;

providing, through one or more interfaces, access to one or more portions of the logical data representing the identified attributes of the plurality of the datasets;

receiving, from the one or more interfaces, a specification that specifies at least one of the identified attributes in performing an operation; and

based on the specification and on the one or more relationships among the identified attributes represented by the logical data, generating a computer program that is configured to perform the operation by accessing, from storage, at least one dataset from the plurality, with the at least one dataset accessed having the at least one of the attributes specified in the specification.

2. The method of claim 1, wherein the one or more interfaces provides access to the one or more portions of the logical data without accessing the plurality of datasets from storage.

3. The method of claim 1, comprising:

identifying a dataset from the plurality of datasets including the at least one of the attributes specified in the specification; and

accessing, from storage, the identified dataset.

4. The method of claim 1, comprising executing the computer program using the at least one dataset accessed from storage.

5. The method of claim 1, comprising optimizing the computer program to produce an optimized computer program that is configured to perform the operation by accessing, from storage, only those datasets in the plurality of datasets having the at least one of the attributes specified in the specification.

6. The method of claim 1, wherein the one or more attributes include field names of the plurality of the datasets.

7. The method of claim 1, wherein the one or more attributes include information for accessing the plurality of the datasets in storage.

8. The method of claim 1, comprising identifying, from the schema, one or more parameters for joining the selected dataset and the one or more other datasets.

9. The method of claim 8, wherein the one or more parameters include a key for joining the selected dataset and at least one of the one or more other datasets.

10. The method of claim 1, comprising receiving, from a client device, selection data specifying the selected dataset.

11. The method of claim 1, wherein the selected dataset comprises a root node of the logical data, and wherein at least one of the one or more other datasets are joined to the selected dataset.

12. The method of claim 1, wherein the one or more computations on one or more of the datasets or one or more transformations of one or more of the datasets define a virtual field for at least one of the plurality of the datasets.

13. The method of claim 1, comprising generating, based on the specification and on the one or more relationships among the identified attributes represented by the logical data, an

executable dataflow graph that is configured to perform the operation, wherein the executable dataflow graph includes at least one of the one or more attributes as an input.

14. The method of claim 1, comprising removing from the computer program an operation to access, from storage, at least one dataset in the plurality of datasets that does not include the at least one of the attributes specified in the specification.

15. The method of claim 1, wherein the computer program is configured to access, from storage, at least some data from the plurality by a select statement, wherein the select statement is minimized to select only the at least one of the attributes specified in the specification.

16. The method of claim 1, wherein the logical data is read as a data source.

17. The method of claim 1, further including:

providing a development environment that includes one or more graphical user interfaces for viewing the one or more portions of the logical data representing the identified attributes of the plurality of the datasets and further for defining the specification that specifies at least one of the identified attributes in performing the operation.

18. A system for providing one or more interfaces and storage that stores datasets having one or more attributes, and with the one or more interfaces providing access to the one or more attributes of the datasets, including:

one or more processors and one or more storage devices storing instructions that are operable, when executed by the one or more processors, to cause the one or more processors to perform operations according to the method of any one of claims 1 to 17.

19. A non-transitory computer-readable storage medium storing instructions for causing a computing system to perform operations according to the method of any one of claims 1 to 17.