CLAIMS

1. A data processing system configured to communicate with a database system and a data source external to the database system, the data processing system comprising:

at least one computer hardware processor; and

at least one non-transitory computer-readable storage medium storing processor-executable instructions that, when executed by the at least one computer hardware processor, cause the at least one computer hardware processor to perform:

obtaining a federated structured query language (SQL) query for generating output data using data to be received from the database system and data to be received from the data source external to the database system,

generating a query plan for the federated SQL query, the query plan comprising a plurality of data processing operations to be performed by the data processing system including at least a first data processing operation for processing data obtained based on the data to be received from the database system and a second data processing operation for processing data obtained based on the data to be received from the data source external to the database system;

displaying a graphical user interface (GUI) containing a plurality of GUI elements including a first GUI element representing the first data processing operation and a second GUI element representing the second data processing operation;

during execution of the federated SQL query,

gathering tracking information for the federated SQL query including: gathering first tracking information relating to execution of the first data processing operation for processing data obtained based on the data that has been received by the data processing system from the database system; and gathering second tracking information relating to execution of the second data processing operation for processing data obtained based on the data that has been received by the data processing system from the data source external to the database system; and

displaying, in the GUI, at least some of the first tracking information in association with the first GUI element and at least some of the second tracking information in association with the second GUI element,

wherein the data processing system is external to the database system and external to the data source, and

wherein the data processing system is configured to execute software that performs the obtaining the federated SQL query, the generating the query plan for the federated SQL query, and the gathering the tracking information for the federated SQL query during the execution of the federated SQL query.

2. The data processing system of claim 1, wherein the database system is a relational database system.

3. The data processing system of claim 1, wherein the data source is a flat file data source, a multi-file data source, or a Hadoop data source.

4. The data processing system of claim 1, wherein the graphical user interface is non-tabular.

5. The data processing system of claim 1, wherein displaying the GUI comprises generating a graphical representation of a graph at least in part by:

generating the first GUI element to represent a first node of the graph associated with the first data processing operation; and

generating the second GUI element to represent a second node of the graph associated with the second data processing operation.

6. The data processing system of claim 5, wherein the generating the graphical representation of a graph comprises generating a graphical representation of a dataflow graph for executing the SQL query.

7. The data processing system of claim 6, wherein generating the graphical representation of the dataflow graph further comprises generating a third GUI element representing an edge between the first node of the dataflow graph and the second node of the dataflow graph.

8. The data processing system of claim 6, wherein the processor-executable instructions further cause the at least one computer hardware processor to perform:

generating, based on the query plan, at least one data structure representing the dataflow graph, the dataflow graph comprising a plurality of nodes including the first node and the second node and a plurality of edges connecting the plurality of nodes, each of the plurality of nodes corresponding to a respective data processing operation in the plurality of data processing operations, the plurality of edges representing flows of data among nodes in the plurality of nodes.

9. The data processing system of any of claims 5, wherein gathering the first tracking information for the first data processing operation comprises tracking a number of data records processed via the first data processing operation, and wherein displaying, via the GUI at least some of the first tracking information comprises displaying, in association with the first GUI element, the number of data records processed via the first data processing operation.

10. The data processing system of claim 9, wherein displaying, in association with the first GUI element, the number of data records processed via the first data processing operation, comprises displaying the number of data records processed via the first data processing operation at an output end of the first node.

11. The data processing system of claim 10, wherein gathering the second tracking information for the second data processing operation comprises tracking a number of data records processed via the second data processing operation, and wherein displaying, via the GUI at least some of the second tracking information comprises displaying, in association with the second GUI element, the number of data records processed via the second data processing operation.

12. The data processing system of claim 1, wherein gathering first tracking information for the first data processing operation comprises gathering tracking information selected from the group consisting of: a number of data records processed via the first data processing operation, a degree of parallelism employed for performing the first data processing operation, information identifying one or more computing devices used for performing the first data processing operation, an amount of processing resources used for performing the first data processing operation, an amount of memory used for performing the first data processing operation, an amount of time used for performing the first data processing

operation, a measure of skew among processing loads on computing devices performing the first data processing operation, parameters of the first data processing operation, and information indicating whether the first data processing operation completed.

13. The data processing system of claim 1, wherein gathering first tracking information for the first data processing operation comprises identifying one or more parameters of the first data processing operation and wherein displaying at least some of the first tracking information comprises displaying at least one value of at least one of the identified one or more parameters.

14. The data processing system of claim 1, wherein the processor-executable instructions further cause the at least one computer hardware processor to perform:

displaying a second GUI displaying a list of one or more SQL queries being executed by the data processing system, the one or more SQL queries including the SQL query; and

in response to a user selection, via the second GUI, of the SQL query, generating the GUI containing the plurality of GUI elements including the first GUI element representing the first data processing operation and the second GUI element representing the second data processing operation.

15. The data processing system of claim 1, wherein the processor-executable instructions further cause the at least one computer hardware processor to perform:

displaying a second GUI displaying a list of one or more SQL queries being executed by the data processing system, the one or more SQL queries including the SQL query; and

in response to a user selection, via the second GUI, of the SQL query, stopping execution of the SQL query.

16. The data processing system of claim 1, wherein the displaying further comprises refreshing the displayed at least some of the first tracking information.

17. At least one non-transitory computer-readable storage medium storing processor-executable instructions that, when executed by at least one computer hardware processor in a data processing system configured to communicate with a database system and a data source external to the database system, cause the at least one computer hardware processor to perform:

obtaining a federated structured query language (SQL) query for generating output data using data to be received from the database system and data to be received from the data source external to the database system,

generating a query plan for the federated SQL query, the query plan comprising a plurality of data processing operations to be performed by the data processing system including at least a first data processing operation for processing data obtained based on the data to be received from the database system and a second data processing operation for processing data obtained based on the data to be received from the data source external to the database system;

displaying a graphical user interface (GUI) containing a plurality of GUI elements including a first GUI element representing the first data processing operation and a second GUI element representing the second data processing operation;

during execution of the federated SQL query,

gathering tracking information for the federated SQL query including:

gathering first tracking information relating to execution of the first data processing operation for processing data obtained based on data that has been received by the data processing system from the database system; and

gathering second tracking information relating to execution of the second data processing operation for processing data obtained based on data that has been received by the data processing system from the data source external to the database system; and

displaying, in the GUI, at least some of the first tracking information in association with the first GUI element and at least some of the second tracking information in association with the second GUI element,

wherein the data processing system is external to the database system and external to the data source, and

wherein the data processing system is configured to execute software that performs the obtaining the federated SQL query, the generating the query plan for the federated SQL query, and the gathering the tracking information for the federated SQL query during execution of the federated SQL query.

18. A method, performed by at least one computer hardware processor in a data processing system configured to communicate with a database system and a data source external to the database system, the method comprising:

obtaining a federated structured query language (SQL) query for generating output data using data to be received from the database system and data to be received from the data source external to the database system,

generating a query plan for the federated SQL query, the query plan comprising a plurality of data processing operations to be performed by the data processing system including at least a first data processing operation for processing data obtained based on the data to be received from the database system and a second data processing operation for processing data obtained based on the data to be received from the data source external to the database system;

displaying a graphical user interface (GUI) containing a plurality of GUI elements including a first GUI element representing the first data processing operation and a second GUI element representing the second data processing operation;

during execution of the federated SQL query,

gathering tracking information for the federated SQL query including:

gathering first tracking information relating to execution of the first data processing operation for processing data obtained based on data that has been received by the data processing system from the database system; and

gathering second tracking information relating to execution of the second data processing operation for processing data obtained based on data that has been received by the data processing system from the data source external to the database system; and

displaying, in the GUI, at least some of the first tracking information in association with the first GUI element and at least some of the second tracking information in association with the second GUI element,

wherein the data processing system is external to the database system and external to the data source, and

wherein the data processing system is configured to execute software that performs the obtaining the federated SQL query, the generating the query plan for the federated SQL query, and the gathering the tracking information for the federated SQL query during execution of the federated SQL query.

19. At least one non-transitory computer-readable storage medium storing processor-executable instructions for execution by at least one computer hardware processor in a data

processing system configured to communicate with a database system and a data source external to the database system, the processor-executable instructions comprising:

means for obtaining a federated structured query language (SQL) query for generating output data using data to be received from the database system and data to be received from the data source external to the database system;

means generating a query plan for the federated SQL query, the query plan comprising a plurality of data processing operations to be performed by the data processing system including at least a first data processing operation for processing data obtained based on the data to be received from the database system and a second data processing operation for processing data obtained based on the data to be received from the data source external to the database system;

means for displaying a graphical user interface (GUI) containing a plurality of GUI elements including a first GUI element representing the first data processing operation and a second GUI element representing the second data processing operation;

means for, during execution of the federated SQL query,

gathering tracking information for the federated SQL query including: gathering first tracking information relating to execution of the first data processing operation for processing data obtained based on data that has been received by the data processing system from the database system; and gathering second tracking information relating to execution of the second data processing operation for processing data obtained based on data that has been received by the data processing system from the data source external to the database system; and

means for displaying, in the GUI, at least some of the first tracking information in association with the first GUI element and at least some of the second tracking information in association with the second GUI element,

wherein the data processing system is external to the database system and external to the data source, and

wherein the data processing system is configured to execute software that performs the obtaining the federated SQL query, the generating the query plan for the federated SQL query, and the gathering the tracking information for the federated SQL query during the execution of the federated SQL query.