

- 39 -

CLAIMS

What is claimed is:

1. A 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,
 - 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 and a second data processing operation;
 - 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 for the first data processing operation and second tracking information for the second data processing operation; 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.
2. The data processing system of claim 1, wherein executing the federated SQL query comprises reading data records from and/or writing data records to a first data source of a first type and reading data records from and/or writing data records to a second data source of a second type different from the first type.
3. The data processing system of claim 1 or 2, wherein the plurality of operations comprises at least one data processing operation for accessing data records in a first data

- 40 -

source of a first type and at least another data processing operation for accessing data records in a second data source of a second type different from the first type.

4. The data processing system of claim 3, wherein the first data source of the first type is a flat file data source, a multi-file data source, a Hadoop data source, an Oracle data source, a Teradata data source, a DB2 data source, a SQL Server data source, an Informix data source, a MongoDB data source, an SAP data source, a SAS data source, or a metadata data source.

5. The data processing system of any preceding claim, wherein the graphical user interface is non-tabular.

6. The data processing system of any preceding claim, 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.

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

8. The data processing system of claim 7, 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.

9. The data processing system of claim 7 or 8, 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

- 41 -

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.

10. The data processing system of any of claims 6 to 9, wherein gathering the first tracking information for the first 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.

11. The data processing system of claim 10, 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.

12. The data processing system of claim 11, wherein gathering the second tracking information for the first 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.

13. The data processing system of any preceding claim, wherein gathering first tracking information for the first 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, the total amount of time taken to perform the first data processing operation, a measure of skew among processing loads on

- 42 -

computing devices performing the first data processing operation, parameters of the first data processing operation, information indicating whether the first data processing operation completed.

14. The data processing system of any preceding claim, 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 tracking information comprises displaying at least one value of at least one of the identified one or more parameters.

15. The data processing system of any preceding claim, 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.

16. The data processing system of any preceding claim, 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.

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

- 43 -

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

- obtaining a federated structured query language (SQL) query,

- 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 and a second data processing operation;

- 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 for the first data processing operation and second tracking information for the second data processing operation; 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.

19. A method, performed by at least one computer hardware processor, the method comprising:

- obtaining a federated structured query language (SQL) query,

- 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 and a second data processing operation;

- 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

- 44 -

gathering first tracking information for the first data processing operation and second tracking information for the second data processing operation; 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.

20. At least one non-transitory computer-readable storage medium storing processor-executable instructions comprising:

means for obtaining a federated structured query language (SQL) query;

means generating a query plan for the federated SQL query, the query plan comprising a plurality of operations to be performed by the data processing system including at least a first data processing operation and a second data processing operation;

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 for the first data processing operation and second tracking information for the second data processing operation; 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.