

WHAT IS CLAIMED IS:

1. A method for automatically generating an executable application to transform and load data into ~~a structured dataset~~ one or more data warehouses for storing data from one or more ~~networked~~ data sources, comprising:

receiving at least one metadata file that specifies values for parameters and ~~ruleset mappings specifying that specifies~~ logical rules for transforming data feeds, wherein each data feed is to be ~~received-transformed~~ from a ~~networked~~ data source, and for loading the ~~result~~transformed data feeds into a ~~structured dataset~~ data warehouse of the one or more data warehouses, and wherein the values of the parameters and the logical rules are associated with the data warehouse and specify actions to be executed at runtime to transform the data feeds such that the data feeds are structured for maintaining data integrity in the data warehouse in which the data feeds are to be stored;

for each data feed, validating that the values of the parameters and the logical rules for transforming the data feeds are not inconsistent ~~for each data feed;~~

generating data rules ~~that~~ for transforming each data feed into transformed data for loading into the data warehouse, wherein the data rules specify one or more standards of data integrity in accordance with ~~the~~ validated values of the parameters and ~~the~~ validated logical rules ~~for transforming each data feed into a transformed record for loading into the structured dataset;~~ and

generating an executable data processing application for a runtime environment, the executable data processing application configurable to:

receive source data comprising a data feed from one or more data sources; and

transform the source data into transformed data that satisfies the one or more standards of ~~the structured dataset~~ data integrity for the data warehouse in compliance with the generated data rules.

2. The method of claim 1, wherein the executable data processing application comprises a dataflow graph, a dataflow subgraph, or a plurality of dataflow graphs.

3. The method of claim 1 or 2, wherein the ~~structured dataset comprises a database~~ data warehouse comprises a data store configured to store the transformed

data in structured data records according to the logical rules and the values for parameters of the metadata file.

4. The method of any of the preceding claims, wherein the source data comprises a data record, and wherein a parameter of the metadata file specifies a field of the data record that represents a key value for the data record of the source data.

5. The method of ~~any of claims 1-3 as dependent on claim 4~~ claim 3, wherein a parameter of the metadata file specifies a mapping between the field that represents the key value of the source data and another field of the structured ~~data records~~ dataset that represents another key value of the structured ~~data records~~ dataset.

6. The method of any of claims 1-3 as dependent on claim 4, wherein a parameter of the metadata file specifies a format of the key value, and wherein the executable data processing application is configured to transform the key value to have the format specified by the parameter.

7. The method any of the preceding claims, further comprising:
retrieving a default value for a parameter from a data storage; and
defining the data rules based on the default value of the parameter.

8. The method of ~~any of the preceding claims~~ claim 3, wherein a logical rule specifies a format for a field of the structured ~~dataset~~ data records, the field including a data history value.

9. The method ~~any of the preceding claims as dependent on~~ of claim 8, wherein the data history value includes a time stamp indicative of when the structured ~~dataset~~ data records including the data history value is updated.

10. The method of any of the preceding claims, wherein receiving the metadata file comprises parsing a header row of the metadata file to determine which parameters have specified values in the metadata file.

11. The method of ~~any of the preceding claims~~ claim 3, wherein transforming

the source data into ~~structured-transformed~~ data that satisfies the one or more standards ~~for the structured dataset~~ of data integrity for the data warehouse as defined by the data rules comprises:

determining that at least two different portions of the source data specify identical key values; and

specifying a new key value for at least one of the two different portions of the source data, the new key value being different from the identical key values and based on a key sequence of the structured ~~data records~~ dataset.

12. The method of any of the preceding claims, comprising retrieving one or more default values for one or more additional parameters that are not specified by the at least one metadata file, wherein defining the data rules is based on the default values of the one or more additional parameters.

13. The method of any of the preceding claims, wherein the metadata file include one or more semantic rules that specify a label for interpreting values of the transformed ~~record~~ data.

14. The method of any of the preceding claims, wherein the parameters of the metadata file comprise a data quality parameter that specifies acceptable data values for including in the transformed ~~record~~ data.

15. The method of any of the preceding claims, wherein the parameters of the metadata file comprise a data integrity parameter that specifies a key mapping scheme for the transformed ~~record~~ data.

16. The method of any of the preceding claims, wherein the parameters of the metadata file comprise a data reporting parameter that specifies whether the transformed ~~data record of the structured dataset~~ is configured to be read-optimized or write-optimized.

17. The method of any of the preceding claims, wherein validating that the values of the parameters and the logical rules for transforming the plurality of the data feeds are not inconsistent for each data feed comprises performing a check on feed-

specific metadata specifying a key surrogation rule and load-specific metadata specifying a data historization rule.

18. The method of any of the preceding claims, wherein the executable data processing application is further configured to load the transformed data into ~~the a~~ structured dataset.

19. The method of any of the preceding claims, further comprising:
executing, by the runtime environment, the executable application, the executing including:
receiving the source data comprising a data feed from one or more data sources; and
transforming the source data into the transformed data that satisfies the one or more standards of data integrity for the data warehouse;
loading the transformed data in compliance with the data rules into the ~~structured-database~~data warehouse.

20. A system for generating an executable data processing application to transform and load data into ~~a structured dataset~~ one or more data warehouses for storing data from one or more ~~networked~~ data sources, the system comprising:
an interface configured to receive at least one metadata file that specifies values for parameters and that specifies logical rules for transforming data feeds, wherein each data feed is to be ~~received-transformed~~ from a ~~networked~~ data source, and for loading the ~~results-transformed data feeds~~ into ~~a structured dataset~~ data warehouse of the one or more data warehouses, and wherein the values of the parameters and the logical rules are associated with the data warehouse and specify actions to be executed at runtime to transform the data feeds such that the data feeds are structured for maintaining data integrity in the data warehouse in which the data feeds are to be stored;
a configurator configured to:
for each data feed, validate that the values of the parameters and the logical rules for transforming the plurality of the data feeds are not inconsistent ~~for each data feed~~; and
generate data rules ~~that~~ for transforming each data feed into

transformed data for loading into the data warehouse, wherein the data rules specify one or more standards of data integrity in accordance with the validated values of the parameters and the validated logical rules for transforming each data feed of the data feeds into a transformed record for loading into the structured dataset; and

an application generation engine configured to generate an executable data processing application for a runtime environment, the executable data processing application configurable to:

receive source data comprising a data feed from one or more data sources; and

transform the source data into transformed data that satisfies the one or more standards for the structured dataset of data integrity for the data warehouse in compliance with the generated data rules.

21. The system of claim 20, wherein the executable data processing application comprises a dataflow graph, a dataflow subgraph, or a plurality of dataflow graphs.

22. The system of claims 20 or 21, wherein the ~~structured dataset comprises a database~~ data warehouse comprises a data store configured to store the transformed data in structured data records according to the logical rules and the values for parameters of the metadata file.

23. The system of any of claims 20-22 wherein the source data comprises a data record, and wherein a parameter of the metadata file specifies a field of the data record that represents a key value for the data record of the source data.

24. The system of ~~any of claims 20-22 as dependent on claim 23~~ claim 22, wherein a parameter of the parameters of the metadata file specifies a mapping between the field of the source data that represents the key value and another field of the structured ~~data records dataset~~ data records dataset that represents another key value of the structured ~~data records dataset~~.

25. The system of ~~any of claims 20-22 as dependent on claim 23~~ claim 22,

wherein a parameter of the parameters of the metadata file specifies a format of the key value for a structured dataset of the structured ~~data records~~dataset, and wherein the executable data processing application is configured to transform the key value to have the format specified by the parameter.

26. The system of any of claims 20-25, wherein the configurator is further configured to:

retrieve a default value for a parameter from a data storage; and
define the data rules based on the default value of the parameter.

27. The system of ~~any of claims 20-26~~claim 22, wherein a logical rule specifies a format for a field of the structured ~~data records~~dataset, the field including history data.

28. The system of any of claims 20-27, wherein receiving the metadata file comprises parsing a header row the metadata file to determine which parameters have specified values in the metadata file.

29. The system ~~any of claims 20-28~~claim 22, wherein transforming the source data into transformed data that satisfies the one or more standards of data integrity for the ~~structured-dataset~~data warehouse as defined by the data rules comprises:

determining that at least two different portions of the source data specify identical key values; and

specifying a new key value for at least one of the two different portions of the source data, the new key value being different from the identical key values and based on a key sequence of the structured ~~dataset~~data records.

30. The system of any of claims 20-29, comprising retrieving one or more default values for one or more additional parameters that are not specified by the at least one metadata file, wherein defining the data rules is based on the default values of the one or more additional parameters.

31. The system of ~~any of claims 20-30~~claim 22, wherein the metadata file include one or more semantic rules that specify an interpretation of the transformed

data for the structured ~~dataset~~data records.

32. The system of ~~any of claims 20-31~~claim 22, wherein the parameters of the metadata file comprise a data quality parameter that specifies acceptable data values for including in the structured ~~dataset~~data records.

33. The system of ~~any of claims 20-32~~claim 22, wherein the parameters of the metadata file comprise a data integrity parameter that specifies a key mapping scheme for the structured ~~dataset~~data records.

34. The system of ~~any of claims 20-33~~claim 22, wherein the parameters of the metadata file comprise a data reporting parameter that specifies whether ~~the a~~ structured dataset of the structured ~~dataset~~data records is configured to be read-optimized or write-optimized.

35. The system of any of claims 20-34, wherein validating that the values of the parameters and the logical rules for transforming the plurality of the data feeds are not inconsistent for each data feed comprises performing a check on feed-specific metadata specifying a key surrogation rule and load-specific metadata specifying a data historization rule.

36. The system of any of claims 20-35, wherein the executable data processing application is further configured to load the transformed data into ~~the a~~ structured dataset.

37. The system of any of claims 20-36, further comprising a runtime system configurable for:

executing, by the runtime environment, the executable application, the executing including:

receiving the source data comprising a data feed from one or more data sources; and

transforming the source data into the transformed data that satisfies the one or more standards of data integrity for the data warehouse;

loading the transformed data in compliance with the data rules into the

~~structured-database~~data warehouse.

38. One or more non-transitory computer readable media storing instructions that are executable by one or more processors configured to perform operations for generating an executable data processing application to transform and load data into ~~a structured-dataset~~one or more data warehouses for storing data from one or more ~~networked~~ data sources, the operations comprising:

receiving at least one metadata file that specifies values for parameters and that specifies logical rules for transforming data feeds, wherein each data feed is to be ~~received-transformed~~ from a ~~networked~~ data source, and for loading the ~~results transformed data feeds~~ into ~~a structured-dataset~~ a data warehouse of the one or more data warehouses, and wherein the values of the parameters and the logical rules are associated with the data warehouse and specify actions to be executed at runtime to transform the data feeds such that the data feeds are structured for maintaining data integrity in the data warehouse in which the data feeds are to be stored;

for each data feed, validating that the values of the parameters and the logical rules for transforming the plurality of the data feeds are not inconsistent ~~for each data feed;~~

generating data rules ~~that for transforming each data feed into transformed data for loading into the data warehouse, wherein the data rules~~ specify one or more standards of data integrity in accordance with ~~the~~ validated values of the parameters and ~~the~~ validated logical rules ~~for transforming each data feed of the data feeds into a structured-dataset;~~ and

generating an executable data processing application for a runtime environment, the executable data processing application configurable to:

receive source data comprising a data feed from one or more data sources; and

transform the source data into transformed data that satisfies the one or more standards of data integrity for the ~~structured-dataset~~ data warehouse in compliance with the generated data rules.

39. The method of any one of claims 1-19, further comprising:

verifying that key values of the source data, which identify the records of the source data, have referential integrity within the data warehouse or replacing the key

values with surrogate key values that are not inconsistent within the system of references in the data warehouse,

wherein the surrogate keys are replacement keys that replace the keys used in the source data and ensure that each record in the data warehouse has a unique key value and/or a consistent key value with the rest of the data in the data warehouse.