ABSTRACT

DATA QUALITY ANALYSIS

A method includes receiving information indicative of an output dataset generated by a data processing system; identifying, based on data lineage information relating to the output dataset, one or more upstream datasets on which the output dataset depends; analyzing one or more of the identified one or more upstream datasets on which the output dataset depends. The analyzing includes, for each particular upstream dataset of the one or more upstream datasets, applying one or more of: (i) a first rule indicative of an allowable deviation between a profile of the particular upstream dataset and a reference profile for the particular upstream dataset, and (ii) a second rule indicative of one or more allowable values or prohibited values for each of one or more data elements in the particular upstream dataset, and based on the results of applying the one or more rules, selecting one or more of the upstream datasets. The method includes outputting information associated with the selected one or more upstream datasets.

(Fig. 1)