CLAIMS

1. A computer-implemented method for selecting data records to cause execution of a processing rule during testing of a data processing application, the method including:

obtaining a first set of data records;

processing the first set of data records using a data processing application that includes a processing rule, wherein a processing rule operates on at least one input value and generates at least one output value, and wherein whether the processing rule is executed by the data processing application during processing of a particular data record depends directly or indirectly on a value in each of one or more data fields of the particular data record;

receiving execution information indicative of a number of times the processing rule was executed in connection with processing of the first set of data records;

obtaining a second set of data records;

analyzing values in one or more data fields of each of the data records in the second set, the analyzing including generating a profile of each of one or more of the data fields for the second set of data records, the profile of a data field characterizing the values in the data field;

obtaining a subsetting rule based on (i) the generated profile and (ii) the execution information indicative of the number of times the processing rule was executed in connection with processing the first set of data records, the subsetting rule including an identification of a particular one of the data fields of the data records in the second set as a target data field;

selecting a subset of data records from the second set of data records according to the subsetting rule, the selecting of the subset of data records being based on values in the target data field; and

processing the selected subset of data records using the data processing application.

2. The method of claim 1, wherein obtaining the subsetting rule includes formulating the subsetting rule, including identifying the one of the data fields as the target data field based on a cardinality of the identified one of the data fields.

3. The method of claim 2, wherein the target data field has a set of distinct values in the plurality of data records, and wherein selecting a subset of data records includes selecting data records such that there is at least one data record in the selected subset that has each of the distinct values for the target data field.

4. The method of claim 1, wherein generating a profile includes classifying values for a first data field of the data records in the second set of data records; and

wherein obtaining the subsetting rule includes formulating the subsetting rule, including identifying the first data field as the target data field based on the classifying.

5. The method of claim 4, wherein the target data field has a set of distinct values of the data records in the second set of data records, and wherein selecting a subset of data records includes selecting data records such that there is at least one data record in the selected subset that has each of the distinct values for the target data field.

6. The method of claim 1, wherein the subsetting rule identifies a first data field as a first target data field and a second data field as a second target data field, and wherein selecting a subset of data records includes selecting the subset of data records based on combinations of a first set of distinct values for the first target data field and a second set of distinct values for the second target data field.

7. The method of claim 1, wherein generating a profile includes identifying a relationship between data records of the second set of data records related via values of a first data field; and

wherein the at least one subsetting rule includes an identification of the relationship.

8. The method of claim 7, wherein selecting a subset of data records includes:

selecting a first data record; and

selecting one or more second data records related to the first data record via the relationship identified in the subsetting rule.

9. The method of claim 7, wherein the relationship between data records includes a relationship between data records in the second set of data records and data records in a third set of data records.

10. The method of claim 1, wherein generating a profile includes:

generating a pseudofield for at least some of the data records in the second set of data records; and

populating the pseudofield for each corresponding data record with an accumulated value, wherein the accumulated value for a first data record is determined based on the first data record and at least one other data record related to the first data record,

wherein the first data record and the at least one other data record are related via values of a first data field.

11. The method of claim 10, including determining the accumulated value based on a sum of a value for a second data field of the first data record and values for the second data field for each other related data record.

12. The method of claim 1, including:

formulating a second subsetting rule based on results of processing the selected subset of data records by the data processing application; and

selecting a second subset of data records based on the second subsetting rule.

13. A non-transitory computer-readable medium storing instructions for causing a computing system to execute the method of any one of claims 1 to 12.

14. A computing system for selecting data records to cause execution of a processing rule during testing of a data processing application, the system including:

at least one processor configured to:

obtain a first set of data records;

process the first set of data records using a data processing application that includes a processing rule, wherein a processing rule operates on at least one input value and generates at least one output value, and wherein whether the processing rule is executed by the data processing application during processing of a particular data record depends directly or indirectly on a value in each of one or more data fields of the particular data record;

receive execution information indicative of a number of times the processing rule was executed in connection with processing of the first set of data records;

obtain a second set of data records;

analyze values in one or more data fields of each of the data records in the second set, the analyzing including generating a profile of each of one or more of the data fields for the second set of data records, the profile of a data field characterizing the values in the data field;

obtain a subsetting rule based on (i) the generated profiles and (ii) the execution information indicative of the number of times the processing rule was executed in connection with processing the first set of data records, the subsetting rule including an identification of a particular one of the data fields of the data records in the second set as a target data field;

select a subset of data records from the second set of data records according to the subsetting rule, the selecting of the subset of data records being based on values in the target data field; and

processing the selected subset of data records using the data processing application.

15. The computing system of claim 14, wherein obtaining the subsetting rule includes formulating the subsetting rule, including identifying the one of the data fields as the target data field based on a cardinality of the identified one of the data fields.

16. The computing system of claim 15, wherein the target data field has a set of distinct values in the plurality of data records, and wherein selecting a subset of data records includes selecting data records such that there is at least one data record in the selected subset that has each of the distinct values for the target data field.

17. The computing system of claim 14, wherein generating a profile includes classifying values for a first data field of the data records in the second set of data records; and

wherein obtaining the subsetting rule includes formulating the subsetting rule, including identifying the first data field as the target data field based on the classifying.

18. The computing system of claim 17, wherein the target data field has a set of distinct values of the data records in the second set of data records, and wherein selecting a subset of data records includes selecting data records such that there is at least one data record in the selected subset that has each of the distinct values for the target data field.

19. The computing system of claim 14, wherein the subsetting rule identifies a first data field as a first target data field and a second data field as a second target data field, and wherein selecting a subset of data records includes selecting the subset of data records based on combinations of a first set of distinct values for the first target data field and a second set of distinct values for the second target data field.

20. The computing system of claim 14, wherein generating a profile includes identifying a relationship between data records of the second set of data records related via values of a first data field; and

wherein the at least one subsetting rule includes an identification of the relationship.

21. The computing system of claim 20, wherein selecting a subset of data records includes:

selecting a first data record; and

selecting one or more second data records related to the first data record via the relationship identified in the subsetting rule.

22. The computing system of claim 14, wherein generating a profile includes:

generating a pseudofield for at least some of the data records in the second set of data records; and

populating the pseudofield for each corresponding data record with an accumulated value, wherein the accumulated value for a first data record is determined based on the first data record and at least one other data record related to the first data record,

wherein the first data record and the at least one other data record are related via values of a first data field.

23. The computing system of claim 22, including determining the accumulated value based on a sum of a value for a second data field of the first data record and values for the second data field for each other related data record.

24. The computing system of claim 14, including:

formulating a second subsetting rule based on results of processing the selected subset of data records by the data processing application; and

selecting a second subset of data records based on the second subsetting rule.

25. A computing system for selecting data records to cause execution of a processing rule during testing of a data processing application, the computing system including:

means for obtaining a first set of data records;

means for processing the first set of data records using a data processing application that includes a processing rule, wherein a processing rule operates on at least one input value and generates at least one output value, and wherein whether the processing rule is executed by the data processing application during processing of a particular data record depends directly or indirectly on a value in each of one or more data fields of the particular data record;

means for receiving execution information indicative of a number of times the processing rule was executed in connection with processing of the first set of data records;

means for obtaining a second set of data records;

means for analyzing values in one or more data fields of each of the data records in the second set, the analyzing including generating a profile of each of one or more of the data fields for the second set of data records, the profile of a data field characterizing the values in the data field;

means for obtaining a subsetting rule based on (i) the generated profiles and (ii) the execution information indicative of the number of times the processing rule was executed in connection with processing the first set of data records, the subsetting rule including an identification of a particular one of the data fields of the data records in the second set as a target data field;

means for selecting a subset of data records from the second set of data records according to the subsetting rule, the selecting of the subset of data records being based on values in the target data field; and

means for processing the selected subset of data records using the data processing application.

26. A computer-implemented method for selecting data records to cause execution of a processing rule during testing of a data processing application, the method including:

obtaining a first set of data records;

processing in a data processing system the first set of data records using a data processing application that includes a processing rule that operates on at least one input value and generates at least one output value, and wherein whether the processing rule is executed by the data processing application during processing of a particular data record depends directly or indirectly on a value in each of one or more data fields of the particular data record;

receiving execution information indicative of a number of times the processing rule was executed in connection with processing of the first set of data records;

obtaining a second set of data records;

characterizing values in one or more data fields of each of the data records in the second set;

determining an identification criterion that identifies a particular one of the data fields of the data records in the second set as a target data field, the identification criterion being based on (i) the characterization of the values in the one or more data fields of each of the data records in the second set and (ii) the execution information indicative of the number of times the processing rule was executed in connection with processing the first set of data records;

applying the identification criterion to select one or more data records from the second set of data records based on values in the target data field; and

processing the selected one or more data records using the data processing application.

27. The method of claim 26, including identifying a particular one of the data fields of the data records in the second set as a target data field based on a cardinality of each of one or more of the data fields of the data records in the second set.

28. The method of claim 27, wherein there are multiple distinct values in the target data field across the data records in the second set, and wherein selecting one or more data records from the second set of data records includes selecting at least one data record having each of the distinct values in the target data field.

29. The method of claim 26, wherein the target data field is a first target data field, and including identifying a second particular one of the data fields of the data records in the second set as a second target data field.

30. The method of claim 29, wherein selecting one or more data records from the second set of data records includes selecting the data records based on a combination of values in the first target data field and values in the second target data field.

31. The method of claim 30, wherein there are multiple distinct values in the first target data field

across the data records in the second set and multiple distinct values in the second target data field across the data records in the second set, and wherein selecting one or more data records from the second set of data records includes selecting the data records based on combinations of the distinct values in the first target data field and the distinct values in the second target data field.

32. The method of claim 26, wherein characterizing values in one or more data fields includes identifying a first data field having a value that indicates a relationship among two or more data records in the second set of data records.

33. The method of claim 32, wherein selecting one or more data records from the second set of data records includes: selecting a first data record; identifying one or more second data records related to the first data record based on the value in the first data field of the first data record; and selecting the one or more second data records.

34. The method of claim 32, including determining an accumulated value for the first data record based on a value in a second data field of the first data record and a value in the second data field for each of the two or more data records that are related to the first data record; generating a new field for the first data record; and populating the new field with the accumulated value.

35. The method of claim 34, wherein determining the accumulated value includes determining the accumulated value based on a sum of the value in the second data field of the first data record and the value in the second data field for each of the two or more data records that are related to the first data record.

36. The method of claim 34, wherein selecting one or more data records from the second set of data records includes selecting the one or more data records based on the accumulated value in the new field for each of one or more data records in the second set of data records.

37. A system for selecting data records to cause execution of a processing rule during testing of a data processing application, the system including:

at least one processor configured to:

obtain a first set of data records;

process in a data processing system the first set of data records using a data processing application that includes a processing rule that operates on at least one input value and generates at least one output value, and wherein whether the processing rule is executed by the data processing application during processing of a particular data record depends

directly or indirectly on a value in each of one or more data fields of the particular data record;

receive execution information indicative of a number of times the processing rule was executed in connection with processing of the first set of data records;

obtain a second set of data records;

characterize values in one or more data fields of each of the data records in the second set;

determine an identification criterion that identifies a particular one of the data fields of the data records in the second set as a target data field, the identification criterion being based on (i) the characterization of the values in the one or more data fields of each of the data records in the second set and (ii) the execution information indicative of the number of times the processing rule was executed in connection with processing the first set of data records;

applying the identification criterion to select one or more data records from the second set of data records based on values in the target data field; and

process the selected one or more data records using the data processing application.

38. The system of claim 37, wherein the at least one processor is configured to identify a particular one of the data fields of the data records in the second set as a target data field based on a cardinality of each of one or more of the data fields of the data records in the second set.

39. The system of claim 38, wherein there are multiple distinct values in the target data field across the data records in the second set, and wherein selecting one or more data records from the second set of data records includes selecting at least one data record having each of the distinct values in the target data field.

40. The system of claim 37, wherein the target data field is a first target data field, and including identifying a second particular one of the data fields of the data records in the second set as a second target data field.

41. The system of claim 40, wherein selecting one or more data records from the second set of data records includes selecting the data records based on a combination of values in the first target data field and values in the second target data field.

42. The system of claim 41, wherein there are multiple distinct values in the first target data field across the data records in the second set and multiple distinct values in the second target data field across the data records in the second set, and wherein selecting one or more data records from the

second set of data records includes selecting the data records based on combinations of the distinct values in the first target data field and the distinct values in the second target data field.

43. The system of claim 37, wherein characterizing values in one or more data fields includes identifying a first data field having a value that indicates a relationship among two or more data records in the second set of data records.

44. The system of claim 43, wherein selecting one or more data records from the second set of data records includes: selecting a first data record; identifying one or more second data records related to the first data record based on the value in the first data field of the first data record; and selecting the one or more second data records.

45. The system of claim 43, wherein the at least one processor is configured to determine an accumulated value for the first data record based on a value in a second data field of the first data record and a value in the second data field for each of the two or more data records that are related to the first data record; generating a new field for the first data record; populating the new field with the accumulated value.

46. The system of claim 45, wherein selecting one or more data records from the second set of data records includes selecting the one or more data records based on the accumulated value in the new field for each of one or more data records in the second set of data records.

47. A non-transitory computer readable medium storing instructions for causing a computer system to execute the method of any one of claims 26-36.

48. A system for selecting data records to cause execution of a processing rule during testing of a data processing application, the system including:

means for obtaining a first set of data records;

means for processing in a data processing system the first set of data records using a data processing application that includes a processing rule that operates on at least one input value and generates at least one output value, and wherein whether the processing rule is executed by the data processing application during processing of a particular data record depends directly or indirectly on a value in each of one or more data fields of the particular data record;

means for receiving execution information indicative of a number of times the processing rule was executed in connection with processing of the first set of data records;

means for obtaining a second set of data records;

means for characterizing values in one or more data fields of each of the data records in the second set;

means for determining an identification criterion that identifies a particular one of the data fields of the data records in the second set as a target data field, the identification criterion being based on (i) the characterization of the values in the one or more data fields of each of the data records in the second set and (ii) the execution information indicative of the number of times the processing rule was executed in connection with processing the first set of data records;

means for applying the identification criterion to select one or more data records from the second set of data records based on values in the target data field; and

means for processing the selected one or more data records using the data processing application.