CLAIMS

What is claimed is:

1. A method for identifying a cause of data disparity among monitored data, the method comprising:

monitoring, using a server computer, parameters associated with data collected in connection with a processing computer;

detecting, using the server computer, a data disparity among the data;

identifying, using the server computer, a first event associated with the data disparity;

determining, using the server computer, a set of events associated with the first event, wherein an event among the set of events represents an occurrence that impacts at least one of an amount of data processed by, or a processing speed of, the processing computer, wherein the set of events include one or more internal events internal to an organization comprising the processing computer and one or more external events external to the organization;

calculating, using the server computer, a score for each event among the set of events as a function of a weight assigned to each event among the set of events and occurrence score determined for each event among the set of events, wherein the weight assigned to each event among the set of events is a measure of probability of that event being a root cause of the data disparity;

identifying, using a machine learning model executing on the server computer, a second event among the set of events as the cause of the data disparity, wherein the second event has the highest score among the set of events;

when a third event, different than the second event, is an actual cause of the data disparity, training the machine learning model by:

adding the third event to the set of events if the third event is not included in the set of events;

adjusting the weight of each event among the set of events by a predetermined amount, wherein adjusting includes increasing the weight of the third event; and

recalculating the score for each event among the set of events; and

taking, using the server computer, remedial or preventive actions to address the data disparity in view of the identified cause of the data disparity,

wherein the occurrence score of a selected event is determined based on runtime characteristics of all child events and parent events of the selected event, wherein the selected event occurred prior to the all child events of the selected event, and all parent events of the selected event occurred prior to the selected event.

2. The method of claim 1, further comprising two or more iterations of adjusting and recalculating, wherein the score of the third event increases at each iteration such that the third event has the highest score among the set of events at conclusion of all iterations.

3. The method of claim 1, wherein the weight of a given event is stored along with a history of the given event being the actual cause of the data disparity.

4. The method of claim 1, further comprising:

receiving an alert associated with the data disparity; and

in response to the alert, identifying the first event associated with the data disparity.

5. The method of claim 1, wherein the set of events associated with the first event includes one or more parent events of the first event, wherein the one or more parent events occurred prior to the first event.

6. The method of claim 1, wherein the first event and the set of events form a dependency graph.

7. The method of claim 6, further comprising:

adding a new event to the dependency graph, wherein the new event is associated with a fourth event and a fifth event in the set of events, wherein the fifth event is a descendent of the fourth event; and

associating the new event with the fourth event without associating with the fifth event.

8. A computer comprising:

a processor; and

a computer readable medium, the computer readable medium comprising code that, when executed by the processor, cause the processor to:

monitor parameters associated with data collected in connection with a processing computer;

detect a data disparity among the data;

identify a first event associated with the data disparity;

determine a set of events associated with the first event, wherein an event among the set of events represents an occurrence that impacts at least one of an amount of data processed by, or a processing speed of, the processing computer, wherein the set of events include one or more internal events internal to an organization comprising the processing computer and one or more external events external to the organization;

calculate a score for each event among the set of events as a function of a weight assigned to each event among the set of events and occurrence score determined for each event among the set of events, wherein the weight assigned to each event among the set of events is a measure of probability of that event being a root cause of the data disparity;

identify, using a machine learning model executing on the processor, a second event among the set of events as the cause of the data disparity, wherein the second event has the highest score among the set of events;

when a third event, different than the second event, is an actual cause of the data disparity, training the machine learning model by:

adding the third event to the set of events if the third event is not included in the set of events;

adjusting the weight of each event among the set of events by a predetermined amount, wherein adjusting includes increasing the weight of the third event; and

recalculating the score for each event among the set of events; and

take preventive or remedial actions to address the data disparity in view of the identified cause of the data disparity,

wherein the occurrence score of a selected event is determined based on runtime characteristics of all child events and parent events of the selected event, wherein the selected event occurred prior to the all child events of the selected event, and all parent events of the selected event occurred prior to the selected event.

9. The computer of claim 8, wherein the weight of a given event is stored along with a history of the given event being the actual cause of the data disparity.

10. The computer of claim 8, wherein the code, when executed by the processor, further causes the processor to:

receive an alert associated with the data disparity; and

in response to the alert, identify the first event associated with the data disparity.

11. The computer of claim 8, wherein the set of events associated with the first event includes one or more parent events of the first event, wherein the one or more parent events occurred prior to the first event.

12. The computer of claim 8, wherein the first event and the set of events form a dependency graph.

13. The computer of claim 12, wherein the code, when executed by the processor, further causes the processor to:

add a new event to the dependency graph, wherein the new event is associated with a fourth event and a fifth event in the set of events, wherein the fifth event is a descendent of the fourth event; and

associate the new event with the fourth event without associating with the fifth event.