

THE INVENTION CLAIMED IS:

1. A computer-implemented method executed by at least one processor for load balancing for processing large data sets based on distribution entropy using a plurality of processors, comprising:

identifying, with the at least one processor, a number of segments based on the number of the plurality of processors, and a transaction data set comprising transaction data for a plurality of payment transactions distributed over a range of transaction values, the transaction data for each payment transaction of the plurality of payment transactions comprising a transaction value representing a payment amount;

determining, with the at least one processor, a distribution entropy of the transaction data set based on the transaction value of each payment transaction of the plurality of payment transactions;

repeatedly segmenting, with the at least one processor, the transaction data set into respective pairs of segments until the number of segments is reached, each segment of the number of segments having various sizes, based on the distribution entropy of the transaction data set and balancing respective distribution entropies of each segment of the number of segments such that respective distribution entropies of each segment are balanced to within a predefined tolerance of the distribution entropy of at least one adjacent segment; and

distributing, with the at least one processor, processing tasks associated with each segment of the number of segments to the plurality of processors to process each payment transaction in each respective segment.

2. The computer-implemented method of claim 1, wherein the transaction data set comprises a plurality of transaction values sorted in a data structure stored in memory of a computer system.

3. The computer-implemented method of claim 2, wherein the data structure comprises an array, and wherein the plurality of transaction values are stored in the array as floating point values.

4. The computer-implemented method of claim 1, wherein the plurality of processors comprises a plurality of cores of at least one processor, and wherein each segment is assigned to a core of the at least one processor.

5. The computer-implemented method of claim 1, wherein balancing respective entropies of each segment of the number of segments comprises adjusting a scope of each segment based on a comparison to the at least one distribution entropy of the at least one adjacent segment.

6. A computer-implemented method executed by at least one processor for load balancing for processing large data sets based on distribution entropy, comprising:

(a) identifying, with the at least one processor, a number of segments based on the number of the plurality of processors, and a transaction data set for a plurality of

payment transactions to be processed, the transaction data set comprising data distributed over a transaction value representing a payment amount of each payment transaction and at least one other parameter;

(b) determining, with the at least one processor, a distribution entropy of the transaction data set;

(c) segmenting, with the at least one processor, the transaction data set into respective pairs of transaction data segments having various sizes based on the distribution entropy of the transaction data set and a closest match between respective distribution entropies of each individual transaction data set-segment in the pair of transaction data segments, such that respective distribution entropies of each segment are balanced to within a predefined tolerance of the distribution entropy of at least one adjacent segment;

(d) repeating step (c) for each transaction data set formed by segmentation until a total number of transaction data segments reaches the number of segments; and

(e) distributing, with the at least one processor, processing tasks associated with each segment of the number of segments to the plurality of processors to process each payment transaction in each respective segment.

7. The computer-implemented method of claim 6, wherein the transaction data set comprises a plurality of transaction values sorted in a data structure stored in memory of a computer system.

8. The computer-implemented method of claim 7, wherein the data structure comprises an array, and wherein the plurality of transaction values are stored in the array as floating point values.

9. The computer-implemented method of claim 6, wherein the plurality of processors comprises a plurality of cores of at least one processor, and wherein each segment is assigned to a core of the at least one processor.

10. The computer-implemented method of claim 6, wherein balancing respective entropies of each segment of the number of segments comprises adjusting a scope of each segment based on a comparison to the at least one distribution entropy of the at least one adjacent segment.

11. A computer program product for load balancing for processing large data sets based on distribution entropy using a plurality of processors, comprising at least one non-transitory computer-readable medium including program instructions that, when executed by at least one processor, cause the at least one processor to:

identify a number of segments based on the number of the plurality of processors, and a transaction data set comprising transaction data for a plurality of payment transactions distributed over a range of transaction values, the transaction data for each payment transaction of the plurality of payment transactions comprising a transaction value representing a payment amount;

determine a distribution entropy of the transaction data set based on the transaction value of each payment transaction of the plurality of payment transactions;

repeatedly segment the transaction data set into respective pairs of segments until the number of segments is reached, each segment of the number of segments having various sizes, based on the distribution entropy of the transaction data set and balancing respective distribution entropies of each segment of the number of segments such that respective distribution entropies of each segment are balanced to within a predefined tolerance of the distribution entropy of at least one adjacent segment; and

distribute processing tasks associated with each segment of the number of segments to the plurality of processors to process each transaction in each respective segment.

12. The computer program product of claim 11, wherein the transaction data set comprises a plurality of transaction values sorted in a data structure stored in memory of a computer system.

13. The computer program product of claim 12, wherein the data structure comprises an array, and wherein the plurality of transaction values are stored in the array as floating point values.

14. The computer program product of claim 11, wherein the plurality of processors comprises a plurality of cores of at least one processor, and wherein each segment is assigned to a core of the at least one processor.

15. The computer program product of claim 11, wherein balancing respective entropies of each segment of the number of segments comprises adjusting a scope of each segment based on a comparison to the at least one distribution entropy of the at least one adjacent segment.

16. A computer program product for load balancing for processing large data sets based on distribution entropy using a plurality of processors, comprising at least one non-transitory computer-readable medium including program instructions that, when executed by at least one processor, cause the at least one processor to:

(a) identify a number of segments based on the number of the plurality of processors, and a transaction data set for a plurality of payment transactions to be processed, the transaction data set comprising data distributed over a transaction value representing a payment amount of each payment transaction and at least one other parameter;

(b) determine a distribution entropy of the transaction data set;

(c) segment the transaction data set into respective pairs of transaction data segments having various sizes based on the distribution entropy of the transaction data set and a closest match between respective entropies of each individual transaction data segment in the pair of transaction data segments, such that respective distribution

entropies of each segment are balanced to within a predefined tolerance of the distribution entropy of at least one adjacent segment;

(d) repeat step (c) for each transaction data set formed by segmentation until a total number of transaction data segments reaches the number of segments; and

(e) distribute processing tasks associated with each segment of the number of segments to the plurality of processors to process each payment transaction in each respective segment.

17. The computer program product of claim 16, wherein the transaction data set comprises a plurality of transaction values sorted in a data structure stored in memory of a computer system.

18. The computer program product of claim 17, wherein the data structure comprises an array, and wherein the plurality of transaction values are stored in the array as floating point values.

19. The computer program product of claim 16, wherein the plurality of processors comprises a plurality of cores of at least one processor, and wherein each segment is assigned to a core of the at least one processor.

20. The computer program product of claim 16, wherein balancing respective entropies of each segment of the number of segments comprises adjusting a

scope of each segment based on a comparison to the at least one distribution entropy of the at least one adjacent segment.