CLAIMS

1. A system for segmenting an input data stream, comprising:

a processor comprising a plurality of processing pipelines, the processor configured to:

split an input data stream to a plurality of data sub-streams such that each of the plurality of data sub-streams has an overlapping portion with a consecutive data sub-stream of the plurality of data sub-streams;

create a plurality of segmented data sub-streams by concurrently segmenting each of the plurality of data sub-streams, whereby each of the plurality of data sub-streams are segmented independently using a corresponding processing pipeline of the plurality of processing pipelines of the processor; and

join the plurality of segmented data sub-streams to create a segmented data stream having a same order as the input data stream by synchronizing a sequencing of each of the plurality of segmented data sub-streams according to at least one overlapping segment in the overlapping portion of each two consecutive data sub-streams of the plurality of data sub-streams.

2. The system according to claim 1, wherein the processor is a single-instruction-multiple-data, SIMD, processor.

3. The system according to claim 1 or 2, wherein the processor is adapted to segment each of the plurality of data sub-streams by calculating a rolling hash value for a rolling sequence applied over each of a plurality of sequences of consecutive data items of each of the plurality of data sub-streams and designating each of the plurality of sequences as a variable size segment when the rolling hash value complies with at least one segmentation criterion, wherein the rolling sequence comprises a subset of consecutive data items of a respective sequence.

4. The system according to any of claims 1 to 3, wherein each of the plurality of sequences includes a pre-defined minimum number of the consecutive data items.

5. The system according to any of claims 1 to 4, wherein the processor is adapted to designate at least one large sequence of the plurality of sequences as the variable size segment in case a number of the consecutive data items of the at least one large sequence exceeds a pre-defined maximum value before detecting compliance of the rolling hash value with the at least one segmentation criterion.

6. The system according to any of claims 1 to 5, wherein the processor is adapted to sequentially segment the input data stream in case the synchronization fails.

7. A method for segmenting an input data stream, comprising:

splitting an input data stream to a plurality of data sub-streams such that each of the plurality of data sub-streams has an overlapping portion with a consecutive data sub-stream of the plurality of data sub-streams;

creating a plurality of segmented data sub-streams by concurrently segmenting each of the plurality of data sub-streams, whereby each of the plurality of data sub-streams are segmented independently using a corresponding processing pipeline of a plurality of processing pipelines of the processor; and

joining the plurality of segmented data sub-streams to create a segmented data stream having a same order as the input data stream by synchronizing a sequencing of each of the plurality of segmented data sub-streams according to at least one overlapping segment in the overlapping portion of each two consecutive data sub-streams of the plurality of data sub-streams.

8. The method according to claim 7, wherein each of the plurality of data sub-streams is segmented by calculating a rolling hash value for a rolling sequence applied over each of a plurality of sequences of consecutive data items of each of the plurality of data sub-streams and designating each of the plurality of sequences as a variable size segment when the rolling hash value complies with at least one segmentation criterion, wherein the rolling sequence comprises a subset of consecutive data items of a respective sequence.

9. The method according to any of claims 7 or 8, wherein each of the plurality of sequences includes a pre-defined minimum number of the consecutive data items.

10. The method according to any of claims 7 to 9, wherein at least one large sequence of the plurality of sequences is designated as the variable size segment in case a number of the consecutive data items of the at least one large sequence exceeds a pre-defined maximum value before detecting compliance of the rolling hash value with the at least one segmentation criterion.

11. The method according to any of claims 7 to 10, wherein the input data stream is sequentially segmented in case the synchronization fails.