We claim:

1. A computer implemented method of processing data blocks in a data analytics system, comprising:

storing a plurality of sequential data blocks in a block cache, each data block containing one or more data records, each data record containing one or more values and a record marker marking the end of the data record;

allocating a data block in the block cache to a worker thread;

parsing, by the worker thread, the data block to identify the one or more data records contained within the data block;

determining whether the data block was correctly parsed, wherein determining whether the data block was correctly parsed comprises:

storing an estimated first record marker location of the data block found by the worker thread parsing the data block,

determining, by another worker thread, a true first record marker location in the data block,

determining whether the stored estimated first record marker location matches the true first record marker location, and

determining that the data block was correctly parsed if the stored estimated first record marker location matches the true first record marker location; and

providing the one or more data records contained within the data block to a data analytics workflow executing on the data analytics system responsive to determining that the data block was correctly parsed.

2. The method of claim 1, wherein storing the plurality of sequential data blocks in the block cache comprises:

issuing a plurality of input/output (I/O) requests to a data source, the plurality of I/O requests requesting a plurality of sequentially-ordered data blocks;

receiving the plurality of sequentially-ordered data blocks from the data source in response to the plurality of I/O requests, wherein the plurality of sequentially-ordered data blocks are received from the data source in an order different than the sequential order of the data blocks; and

storing the received plurality of sequentially-ordered data blocks in the block cache.

3. The method of claim 2, wherein the plurality of sequentially-ordered data blocks are stored at locations in the block cache determined responsive to the sequential order of the data blocks.

4. The method of claim 1, wherein there are a plurality of worker threads and allocating a data block in the block cache to a worker thread comprises:

allocating a plurality of data blocks to a plurality of worker threads, the plurality of worker threads adapted to concurrently parse the allocated plurality of data blocks.

5. The method of claim 1, wherein parsing, by the worker thread, the data block to identify the one or more data records contained within the data block comprises:

scanning the data block to identify a first record marker location within the data block;

parsing data records in the data block located after the first record marker location; and

parsing, by the worker thread, a data record in a next sequential data block occurring before a first record marker location within the next sequential data block.

6. The method of claim 1, further comprising:

determining that the data block was incorrectly parsed in response to that the estimated first record marker location does not match the true first record marker location; and

re-parsing the data block using the true first record marker location in the data block responsive to determining that the data block was incorrectly parsed.

7. A computer-implemented data analytics system for processing data blocks, comprising:

a computer processor for executing computer program instructions; and

a non-transitory computer-readable memory storing computer program instructions executable by the computer processor to perform operations comprising:

storing a plurality of sequential data blocks in a block cache, each data block containing one or more data records, each data record containing one or more values and a record marker marking the end of the data record;

allocating a data block in the block cache to a worker thread;

parsing, by the worker thread, the data block to identify the one or more data records contained within the data block;

determining whether the data block was correctly parsed, wherein determining whether the data block was correctly parsed comprises:

storing an estimated first record marker location of the data block found by the worker thread parsing the data block,

determining, by another worker thread, a true first record marker location in the data block,

determining whether the stored estimated first record marker location matches the true first record marker location, and

determining that the data block was correctly parsed if the stored estimated first record marker location matches the true first record marker location; and

providing the one or more data records contained within the data block to a data analytics workflow executing on the data analytics system responsive to determining that the data block was correctly parsed.

8. The system of claim 7, wherein storing the plurality of sequential data blocks in the block cache comprises:

issuing a plurality of input/output (I/O) requests to a data source, the plurality of I/O requests requesting a plurality of sequentially-ordered data blocks;

receiving the plurality of sequentially-ordered data blocks from the data source in response to the plurality of I/O requests, wherein the plurality of sequentially-ordered data blocks are received from the data source in an order different than the sequential order of the data blocks; and

storing the received plurality of sequentially-ordered data blocks in the block cache.

9. The system of claim 8, wherein the plurality of sequentially-ordered data blocks are stored at locations in the block cache determined responsive to the sequential order of the data blocks.

10. The system of claim 7, wherein there are a plurality of worker threads and allocating a data block in the block cache to a worker thread comprises:

allocating a plurality of data blocks to a plurality of worker threads, the plurality of worker threads adapted to concurrently parse the allocated plurality of data blocks.

11. The system of claim 7, wherein parsing, by the worker thread, the data block to identify the one or more data records contained within the data block comprises:

scanning the data block to identify a first record marker location within the data block;

parsing data records in the data block located after the first record marker location; and

parsing, by the worker thread, a data record in a next sequential data block occurring before a first record marker location within the next sequential data block.

12. The system of claim 7, wherein the operations further comprise:

determining that the data block was incorrectly parsed in response to that the estimated first record marker location does not match the true first record marker location; and

re-parsing the data block using the true first record marker location in the data block responsive to determining that the data block was incorrectly parsed.

13. A non-transitory computer-readable memory storing executable computer program instructions for processing data blocks in a data analytics system, the instructions executable to perform operations comprising:

storing a plurality of sequential data blocks in a block cache, each data block containing one or more data records, each data record containing one

or more values and a record marker marking the end of the data record;

allocating a data block in the block cache to a worker thread;

parsing, by the worker thread, the data block to identify the one or more data records contained within the data block;

determining whether the data block was correctly parsed, wherein determining whether the data block was correctly parsed comprises:

storing an estimated first record marker location of the data block found by the worker thread parsing the data block,

determining, by another worker thread, a true first record marker location in the data block,

determining whether the stored estimated first record marker location matches the true first record marker location, and

determining that the data block was correctly parsed if the stored estimated first record marker location matches the true first record marker location; and

providing the one or more data records contained within the data block to a data analytics workflow executing on the data analytics system responsive to determining that the data block was correctly parsed.

14. The non-transitory computer-readable memory of claim 13, wherein storing the plurality of sequential data blocks in the block cache comprises:

issuing a plurality of input/output (I/O) requests to a data source, the plurality of I/O requests requesting a plurality of sequentially-ordered data blocks;

receiving the plurality of sequentially-ordered data blocks from the data source in response to the plurality of I/O requests, wherein the plurality of sequentially-ordered data blocks are received from the data source in an order different than the sequential order of the data blocks; and

storing the received plurality of sequentially-ordered data blocks in the block cache.

15. The non-transitory computer-readable memory of claim 13, wherein there are a plurality of worker threads and allocating a data block in the block cache to a worker thread comprises:

allocating a plurality of data blocks to a plurality of worker threads, the plurality of worker threads adapted to concurrently parse the allocated plurality of data blocks.

16. The non-transitory computer-readable memory of claim 13, wherein parsing, by the worker thread, the data block to identify the one or more data records contained within the data block comprises:

scanning the data block to identify a first record marker location within the data block;

parsing data records in the data block located after the first record marker location; and

parsing, by the worker thread, a data record in a next sequential data block occurring before a first record marker location within the next sequential data block.

17. The non-transitory computer-readable memory of claim 13, wherein the operations further comprise:

determining that the data block was incorrectly parsed in response to that the estimated first record marker location does not match the true first record marker location; and

re-parsing the data block using the true first record marker location in the data block responsive to determining that the data block was incorrectly parsed.