



Anatomy of a Read Operation

We'll cover the following ^

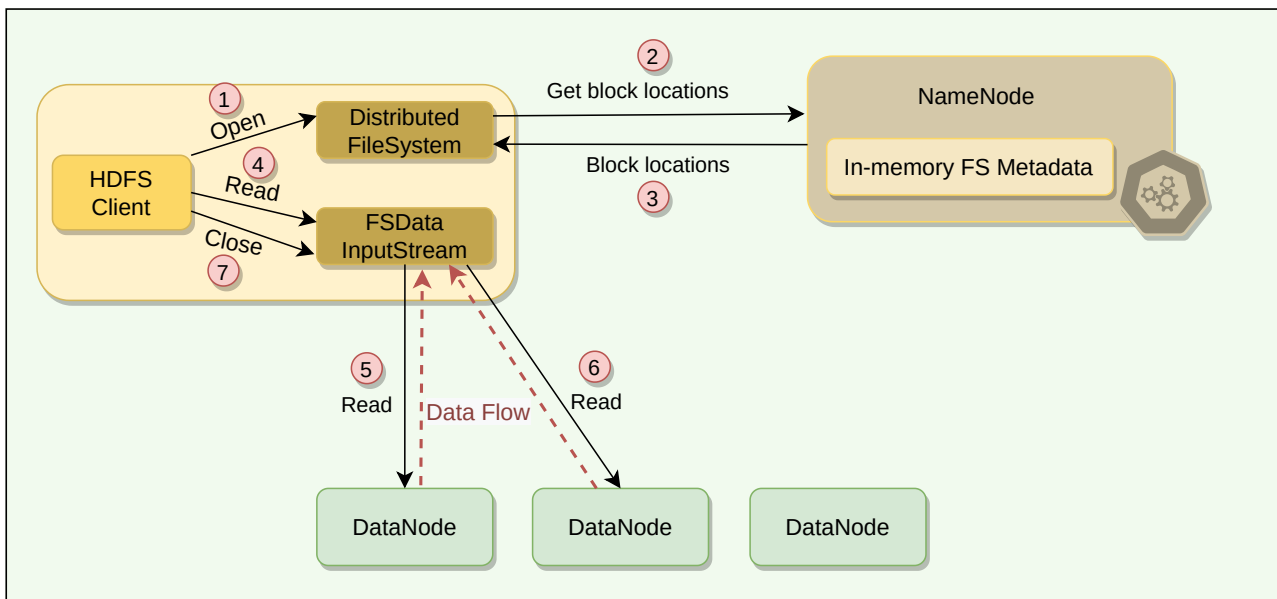
- HDFS read process
- Short circuit read

HDFS read process#

HDFS read process can be outlined as follows:

1. When a file is opened for reading, HDFS client initiates a read request, by calling the `open()` method of the `Distributed FileSystem` object. The client specifies the file name, start offset, and the read range length.
2. The `Distributed FileSystem` object calculates what blocks need to be read based on the given offset and range length, and requests the locations of the blocks from the `NameNode`.
3. `NameNode` has metadata for all blocks' locations. It provides the client a list of blocks and the locations of each block replica. As the blocks are replicated, `NameNode` finds the closest replica to the client when providing a particular block's location. The closest locality of each block is determined as follows:
 - If a required block is within the same node as the client, it is preferred.
 - Then, the block in the same rack as the client is preferred.
 - Finally, an off-rack block is read.

4. After getting the block locations, the client calls the `read()` method of `FSDaata InputStream`, which takes care of all the interactions with the `DataNodes`. In step 4 in the below diagram, once the client invokes the `read()` method, the input stream object establishes a connection with the closest `DataNode` with the first block of the file.
5. The data is read in the form of streams. As the data is streamed, it is passed to the requesting application. Hence, the block does not have to be transferred in its entirety before the client application starts processing it.
6. Once the `FSDaata InputStream` receives all data of a block, it closes the connection and moves on to connect the `DataNode` for the next block. It repeats this process until it finishes reading all the required blocks of the file.
7. Once the client finishes reading all the required blocks, it calls the `close()` method of the input stream object.



The anatomy of a read operation

Short circuit read#

As we saw above, the client reads the data directly from DataNode. The client uses TCP sockets for this. If the data and the client are on the same machine, HDFS can directly read the file bypassing the DataNode. This scheme is called short circuit read and is quite efficient as it reduces overhead and other processing resources.

[← Back](#)[Deep Dive](#)[Next →](#)[Anatomy of a Write Operation](#)[Mark as Completed](#)[Report an Issue](#)