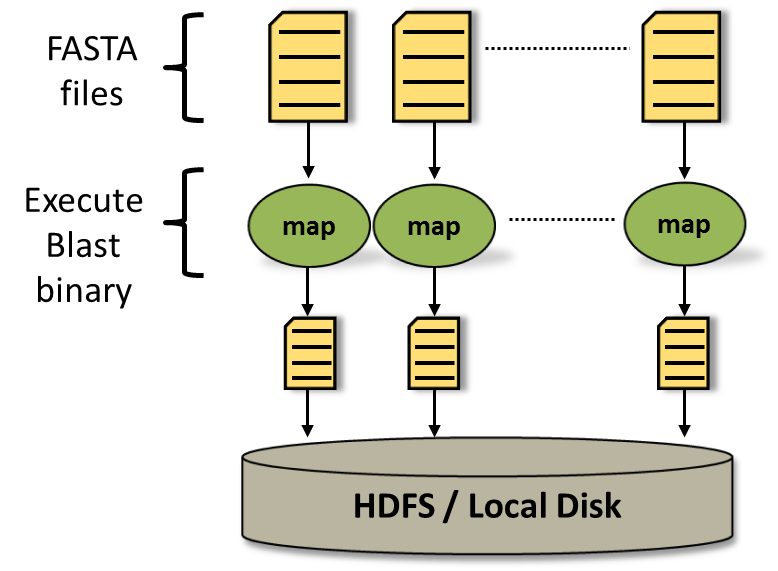
**PROJECT 2**

Basic Logical Alignment Search Tool (Blast) is a search tool used in bioinformatics for comparing primary biological sequence information. Hadoop Blast helps blast to utilize the capability of Hadoop.

In Hadoop Map reduce program the input data is loaded to HDFS before it is given as input to the mapper to generate output in the form of <key,value> pairs. BLAST input data is set of FASTA files that contain the search data and is uploaded to HDFS. The sequence from the FASTA file is search in the database which is an archive file. DataFileInputFormat.java generates <key, value> pairs which contain the file name and file path. BLAST fits properly in Hadoop MapReduce framework. Hadoop Blast is an advanced Hadoop program which helps blast to utilize the computing capability of Hadoop.



**Fig. 1: Hadoop Blast Data Flow**

Inputs to the program:

FASTA files (which are actual search queries), BLAST Binary which is executable file (runs as an external process), Optimized database

Output of the program:

Generated output is stored in HDFS file system.

There is no need of Reduce function since the Map function just compares the FASTA files with the BLAST database and returns the similar protein sequences.