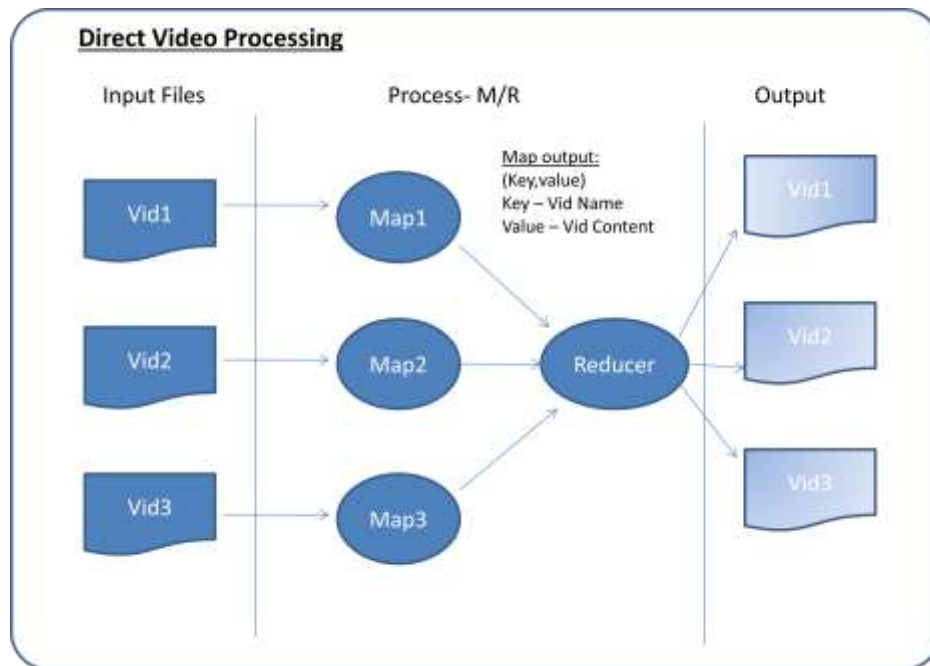
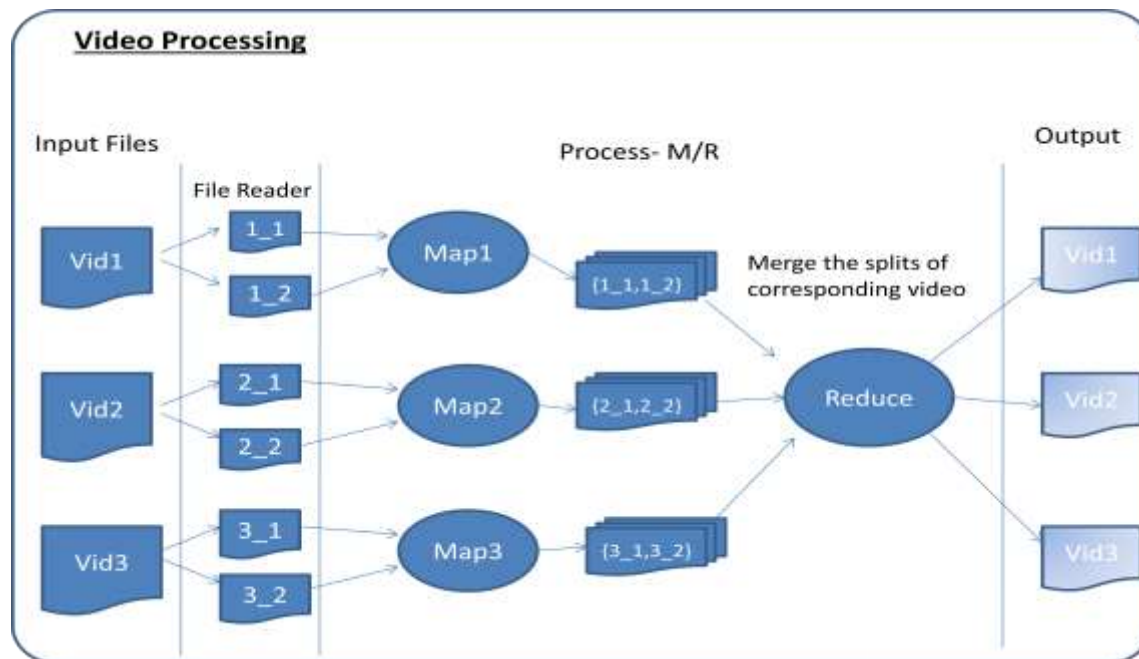


The scalable video processing solution has been implemented in two ways, and either can be used, depending on the use-case under consideration. The two directories “video processing” and “direct video processing” contain the 2 implementations for scalable video analytics.

The difference between video processing and direct video processing is the manner in which the input video files are read and split prior to processing. In “direct video processing”, a file reader reads the video file and sends the entire input video file to a mapper as a (key,value) pair where file name is the key and the file content is the value. This implementation lends itself well to situations where each video file is of a reasonable size that each can be processed on a single node. As a result, many large numbers of video collections can be processed simultaneously across multiple nodes in a Hadoop cluster.



On the other hand, in the directory “video processing”, each input video file is split into a number of partial video clips, each of which is at or below a pre-specified size or time segment. The number of splits of each video is specified by a user-supplied argument and can be dynamically changed based on system configuration and speed/accuracy requirements. The split video segments are then sent to a mapper task as (key,value) pairs. The key is a filename\_<partnum> and value is the split content. In the reduce phase, the processed split video segments are then merged back to a single video based **on the key (split segments with the same file name are merged back into a single video file)**. The resultant output after Mapreduce therefore contains the same number of output videos as input videos. This Mapreduce implementation for unstructured video data analytics is useful when a few large videos (or even a single massive video) needs to be processed swiftly, and therefore parallel processing via splitting of a large file into numerous small ones is desirable for a scalable solution or to obtain near real-time results.



**Note:** At the current time, it is recommended that the direct video processing project be used since it does not involve splitting of video files. The splitting of videos in video processing currently is achieved with FFMPEG. During testing, a minor bug was detected in the tools to split the videos using FFMPEG.