In [1] Jorge Veiga has proposed about the "Performance evaluation big data frameworks for large data analytics". In this paper author has proposed about the, MapReduce frameworks such as Hadoop are being replaced by emerging ones like Spark or Flink, which improve both the programming APIs and performance. This paper addresses this issue by performing a comparative evaluation of Hadoop, Spark and Flink using representative Big Data workloads and considering factors like performance and scalability. The analysis of the results has shown that replacing Hadoop with Spark or Flink can lead to a reduction in execution times.

In this paper Lidong Wang has proposed about the "Machine learning in big data". In this paper author has introduced the challenges and application of machine learning in big data. The advantage of this paper is that, using the big data frameworks it meets the memory constraints. The dis-advantage of this paper is that, some of the big data framework is not suitable for real time transaction and also for unstructured data.

In this paper author has used frameworks are compared on a 4-node cluster using real-world datasets. Results show that Spark outperforms Flink by up to 2x for WordCount, while Flink is better than Spark by up to 3x for K-Means, 2.5x for PageRank and 5x for a relational query. In this paper author has discussed about the processing of some operators like *groupBy* or *join* and the pipelining of data between operators are more efficient in Flink, and that the Flink optimizer provides better performance with complex algorithms.

In this paper author Marc Kaepke has proposed about the "A comparitive evaluation of big data frameworks for graph processing".In this paper, author has evaluated and compared GraphX based on Spark and Gelly based on Flink as two prominent graph processing frameworks. Author has performed some experiments with different graph algorithms and both real world data and artificially generated data. For this, author has implemented a new algorithm using both Flink and Gelly.

In this paper ziya karakaya has proposed about the "A comparision of stream processing frameworks". In this paper, author has compared the performance of Big Data Stream Processing frameworks including Apache Spark, Flink, and Storm. In this paper has discussed about the Flink outperforms both Spark and Storm under equal constraints. The advantage of this paper is that Spark can be optimized to provide the higher throughput than Flink with the cost of higher latency.