Flume

One of the basic components of any system of big data pipeline is the export layer and data ingestion. Commercial tools or common enterprise utilized for this reason include tools such as Data Stage, Informatica, and more. In the world of Hadoop, the most famous choice is the Flume. Based on documented information, Flume is known as a distributed, trusted, and accessible service for efficiently gathering, aggregating, and moving a lot of log data. Like a ton of different tools in the Hadoop environment, Flume is intended in its design to be fault tolerant. It has an easy architecture on the basis of streaming data flows, despite the fact that not all data requires to be streaming data.

https://flume.apache.org/