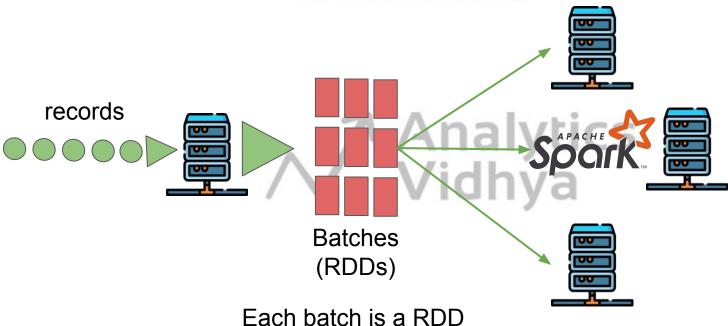
Benefits of DStream Processing





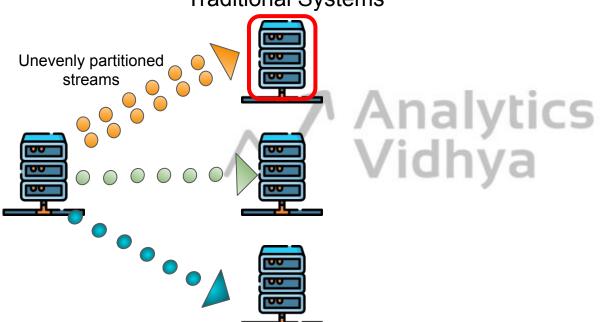
discretized stream processing



(partitioned dataset)



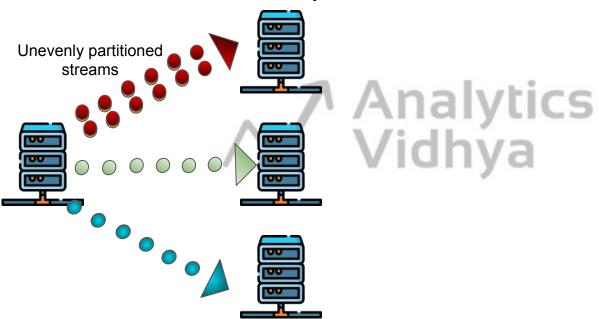
Traditional Systems



Static scheduling to nodes can cause bottlenecks

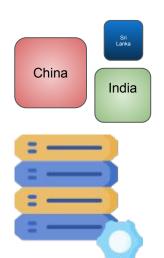


Traditional Systems



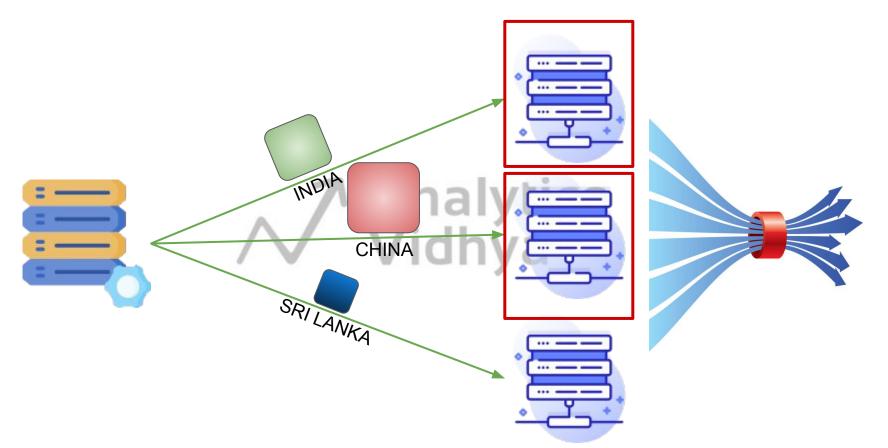
Static scheduling to nodes can cause bottlenecks



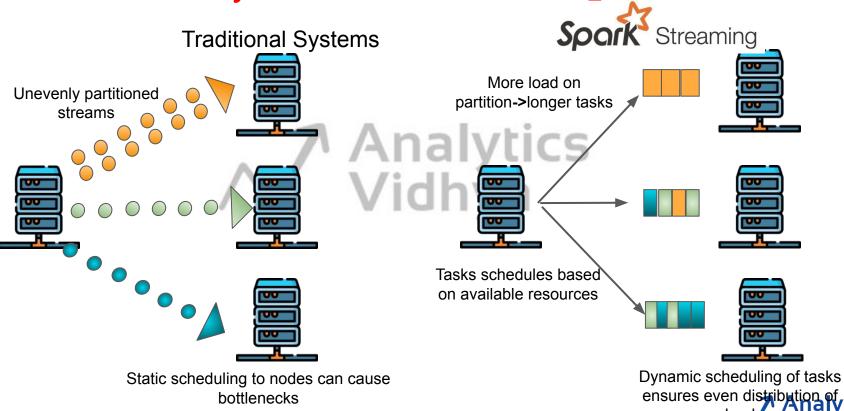




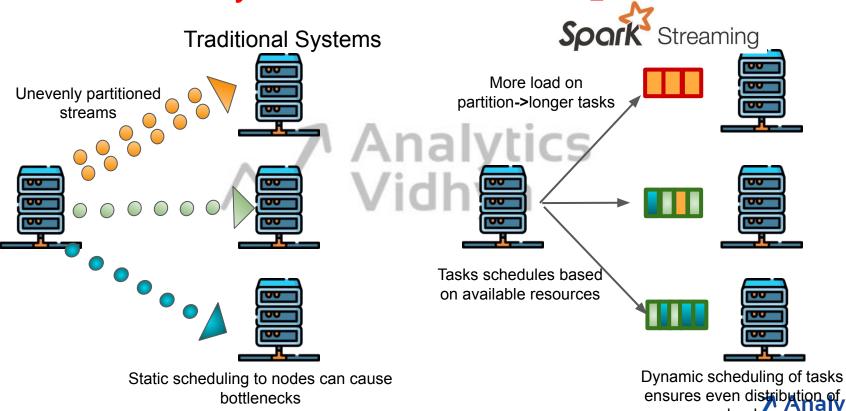






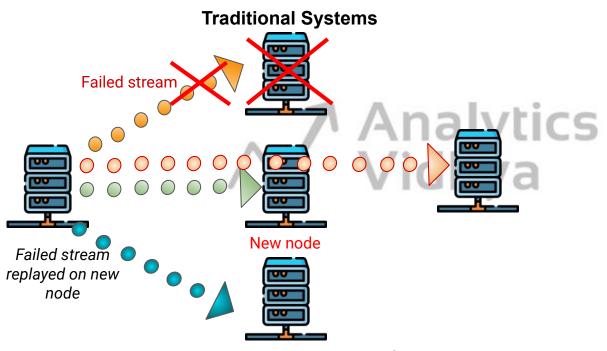


bottlenecks



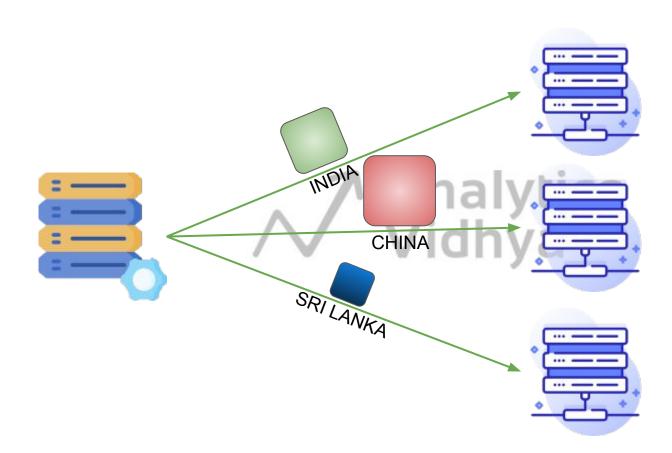
bottlenecks

Fast failure and straggler recovery

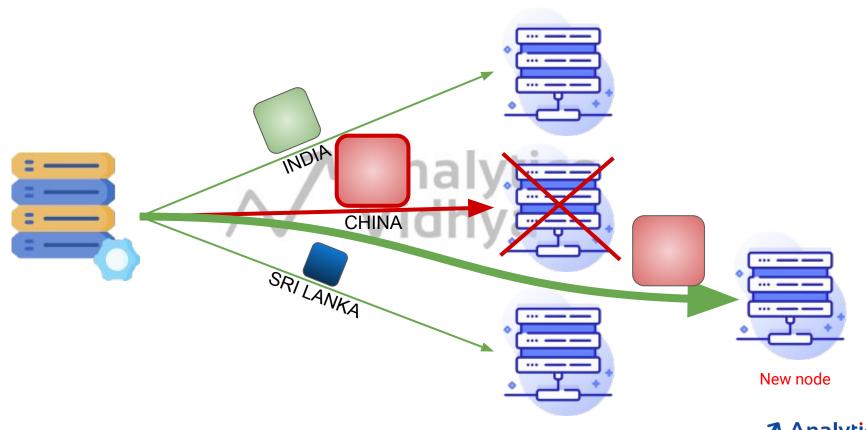


Slower recovery by using single node for recomputation



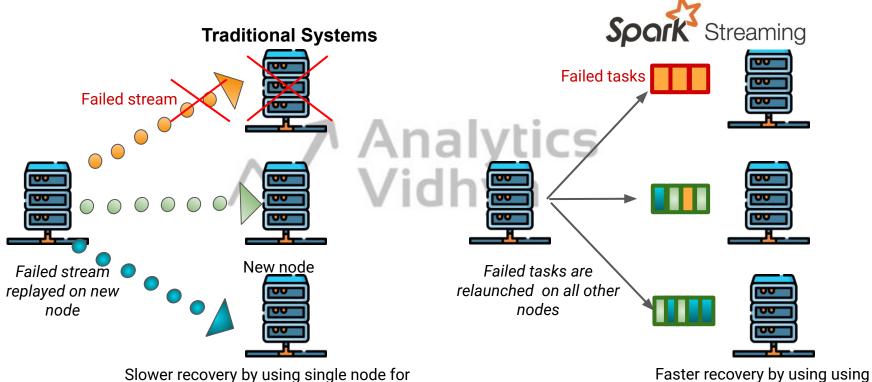








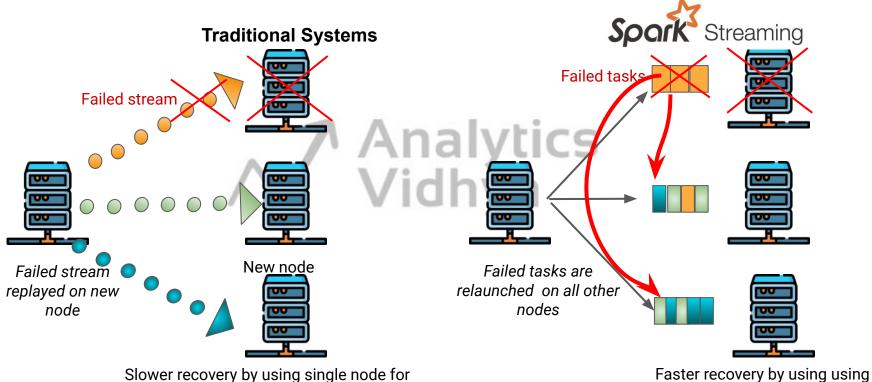
Fast failure and straggler recovery



recomputation

multiple nodes for recomputations

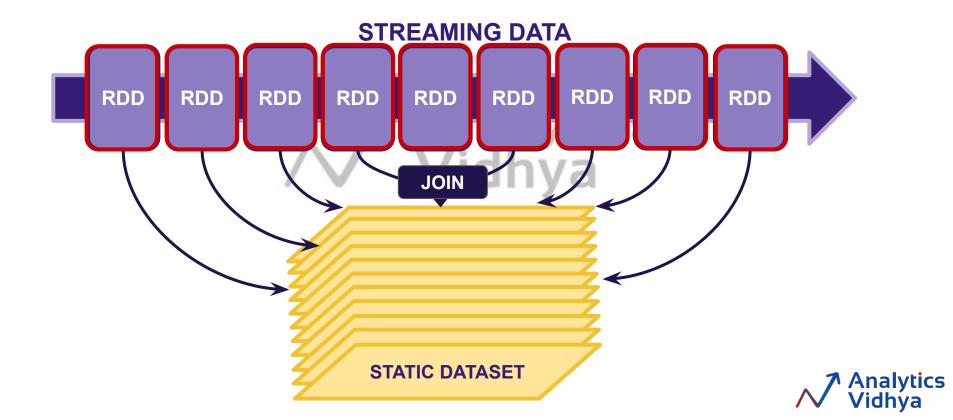
Fast failure and straggler recovery



recomputation

multiple nodes for recomputations

Unification of Various Workloads



Advanced Analytics

STREAMING + SQL and DataFrames **RDD RDD RDD RDD RDD RDD RDD RDD RDD** wordCountsDStream.foreachRDD { lambda rdd: // Convert RDD to DataFrame and register it as a SQL table wordCountsDataFrame = rdd.toDF(|"word", "count"|) wordCountsDataFrame.registerTempTable("word counts") DF DF DF DF DF DF DF DF DF Analytics Vidhya

Thank You/tics Vidhya

