

Project <ul style="list-style-type: none">● Team Number : Team 20● Project Title: <i>Distributed Streaming Platform</i>● Project description: We wish to develop a system that ingests live streams, process and structures the data, which would be used for further data analysis. The structured data will be stored on AWS for future use. The structured data would be used to find co-occurrence of words, sentiment and other data analysis. The analysis would be displayed on interactive dashboard.● Team members Ankith Divakar Raja - araja2 Madhumitha Raveendar - mraveen Neeharika Sompalli - nsompal	Deliverables <ul style="list-style-type: none">● D1 : Workload generation and Real time data stream Generator<ul style="list-style-type: none">○ Get the streaming data from Twitter API and process data for a particular rate using real time data stream generator● D2 : Pipeline on VCL<ul style="list-style-type: none">○ Install and configure NiFi, Kafka, Spark to work together and structure the data in required format● D3 : Kinesis<ul style="list-style-type: none">○ Build the similar pipeline that is built on VCL, in AWS using Kinesis● D4 : Data Visualization<ul style="list-style-type: none">○ The structured data is used for sentiment analysis and co-occurrence of words and the results are displayed in Kibana interactive dashboard.
Dependencies <ul style="list-style-type: none">● Twitter Streaming API, Python stream generator● Configure Apache NiFi, Apache Spark, Apache Kafka to build the pipeline● Configure and use AWS technology stack to build similar pipeline	Issues <ul style="list-style-type: none">● Building a workload generator which can mimic real world twitter streaming workloads.● Evaluate the system on AWS with minimal experimental cost.