

Tools	Evaluation	Learning	Selected
Web Crawling and data collection			
Twitter API (streaming) V1.1	Done	Done	Yes
Tweepy	Done	Done	Yes
Apache Flume			
Apache Kafka	Done	Required	Not
Apache Sqoop	Done	Done	Yes
Batch Processing - MapReduce Toolkits			
Apche Hadoop	Done	Done	Yes
Cloudera CDH4	Done	Done	Optional
Apache Pig	Done	Done	Yes
Apache Hive	Done	Done	Yes
Databases			
MongoDB	Done	Done	Yes
mySQL	Done	Done	Yes
Online Query Engine			
Cloudera Impala	Done	Done	Yes
Apache Spark	Done	Required	For Future Use
Cloudera Storm	Done	Required	Not
Presto	Limited	Required	Not
Data Mining			
Apache Mahout	Done	In Progress	Yes
Data Visualization			
Tableau	Done	Done	Yes
Google Maps	Done	Done	Yes
D3.JS	Done	Done	Optional
Advanced Analytics			
R	Done	Done	Yes
Apache Mahout	Done	Done	Yes
Cloudera Oryx	Done	Done	Yes
MLBase	Not Started	Required	Not