* Full installation of Hadoop using Linux Command line on both local and cloud based systems
* Cleaned data to prepare for loading into database
* Implement Map reduce jobs to aggregate information across nodes in a HDFS cluster
* Imported data into the Hadoop ecosystem using Sqoop
* Created Sqoop Jobs to automate data importation
* Query Data within HDFS using Hive
* Integrated HBase and Hive to provide both scalable storage as well data manipulation
* Perform near real time data ingestion using Flume to find current trends in video games
* Perform cluster analysis using Model based clustering to assign users into various groups to analyze separate trends within each group
* Classified new users into each various groups bases on metrics to predict possible interests of new users
* Predict the probability of future actions of new individuals using logistic regression to assist in business decisions
* Model the effect of various features on the rating of a new game using regression
* Evaluate models using ANOVA