**ABSTRACT**

Our project is about analysing the Income Tax from Aadhar data set using Hadoop Big Data.Analysis of the Aadhar Data set provides valuable details. To analyze this large amount of data, we need parallel processing system and reliable data storage mechanism. So we move to Hadoop distributed file system, map reducing framework and hive database.

The Sample Aadhar Data set is taken from the Government website. Then, the dataset is to be inserted into the Hive Database table after Pre-processing using map reduce is done.

By processing the Aadhar Data set, we can Predict the total amount of Income Tax that would be collected by the government and we can eliminate the unwanted fields Registrar,District etc. The Minimized data is then stored in HDFS and processed using HIVE query Language. Finally, the data is visualized using Splunk Tool.