

PROJECT PROPOSAL

SONIA KARGUTKAR

Problem to solve: Career Prediction

Background: Analyzing student grades in various subjects and predicting a prospective career field.

Preferred Database: MYSQL

Preferred tool for Data Analysis: Hadoop

Hadoop is designed to handle the three V's of Big Data: volume, variety, velocity.

- Hadoop was designed to **scale out**, and it is very **cost effective** to grow the system.
- Hadoop allows you to store data in any format, be that **structured or unstructured** data. This means that you will not need to alter your data to fit any single schema before putting it into Hadoop.
- Hadoop will allow you to **process** massive amounts of data very **quickly**.

Resources: Youtube videos