**CS-498 Cloud Computing Applications Final Project Proposal**

**Team Members:** Nischol Antao, Mike Kresca, Rob Rupprath

**Background**: When developing a new Product (Automotive/Construction/Mining Equipment), Engineers collect a large volume of time series data to analyze the Performance of their Product. These data sets include Channels broadcast by Electronic Control Modules (ECMs) on Communication buses, as well as Channels that can recorded directly from ECM memory by specialized devices. Some of thes devices include, low frequency black box recorders, high frequency engineering dataloggers, and Off the Shelf Equipment for Data Bus Acquisition. All these sets of data can usually be made available to engineers to analyze a specific issue, but there usually isn’t a centralized repository for the data. Additionally there is usually no means for an engineer to easily search/query for a specific machine condition, as each of the different datalogging sources names channels differently, and it is not always clear which datalogger file contains the specific condition an engineer is looking for.

**Proposal:** Our proposal is to build a cloud based system that can perform the following actions

1. Process Engineering Data from multiple data sources, such as high speed engineering data loggers, off the shelf commercial data loggers, and black box recorders. The system would perform a map action on channels that have different names, but physically correspond to the same channel, so as to create a single, user specified channel name. Data for all the channels would then be stored in easily searchable data repository for a Machine. In the interest of not violating any proprietary data formats, we will build the system to read in csv files, each of which might have a slightly different format. We will use data obtained directly from actual machines, but may obfuscate the names of the Machines, and/or randomly offset some of the data, so as not to reveal any confidential information.
2. Provide a Turn-Key Analytics Interface that would allow Engineers to Batch process data for specific Machine, specific time period, or a specific set to machines, to identify a specific condition. The system could be built to provide a scalable time slice around all the instances of that condition, to allow for analysis pre & post condition.
3. Provide a scalable cloud interface that allows for data from several machines to be fed in for analysis, and process data sets of varying sizes (Few Hundred Kb to Several Hundred/Thousand MB)

The Intellectual Merit and Broader Impact requirements for this project would be served by building a system that would allow Engineers to quickly and efficiently search through Machine Performance Data. Easily being able to explore Machine data would allow them to better analyze failures, and subsequently design future systems to avoid machine performance issues.

The system we build could conceptually be expanded to store and analyze data for any set of devices/sources (IOT Sensors, Stock Exchange Data etc), however for the purposes of this project we will be focusing on a Machine Performance Data, that we will obtain directly from Machines.