

Lab instructions for “Computing with R and Hadoop”

Lab outline

GOAL: Make a data-driven business decision for an insurance company to start a new program.

1. Fit a logistic regression model to a small amount of data obtained from a pilot study of customer feedback data.
2. In a virtual cloud, use the **RHadoop** package to run a MapReduce job that summarizes which other customers the model predicts will provide positive feedback.
3. (Optional) Modify the MapReduce job to refine the analysis.

Monitor Hadoop:

- Open “The Hadoop UI” (HUE): <http://192.168.1.105:8888/>
- Username and password: **cloudera**
- View status of MapReduce jobs: <http://192.168.1.105:8888/jobbrowser/>
- View contents of **hdfs**: <http://192.168.1.105:8888/filebrowser/#/>