Rutgers Course Recommender

Simon Gao, Joshua Devasagayaraj {simongao, joshuad}@rutgers.edu
In Collaboration With Rutgers OIT-ESS, Registrar, and Scheduling Office
Capstone Advisor: Prof. Parashar ESS Advisors: Fei Luo, Martin Zhao

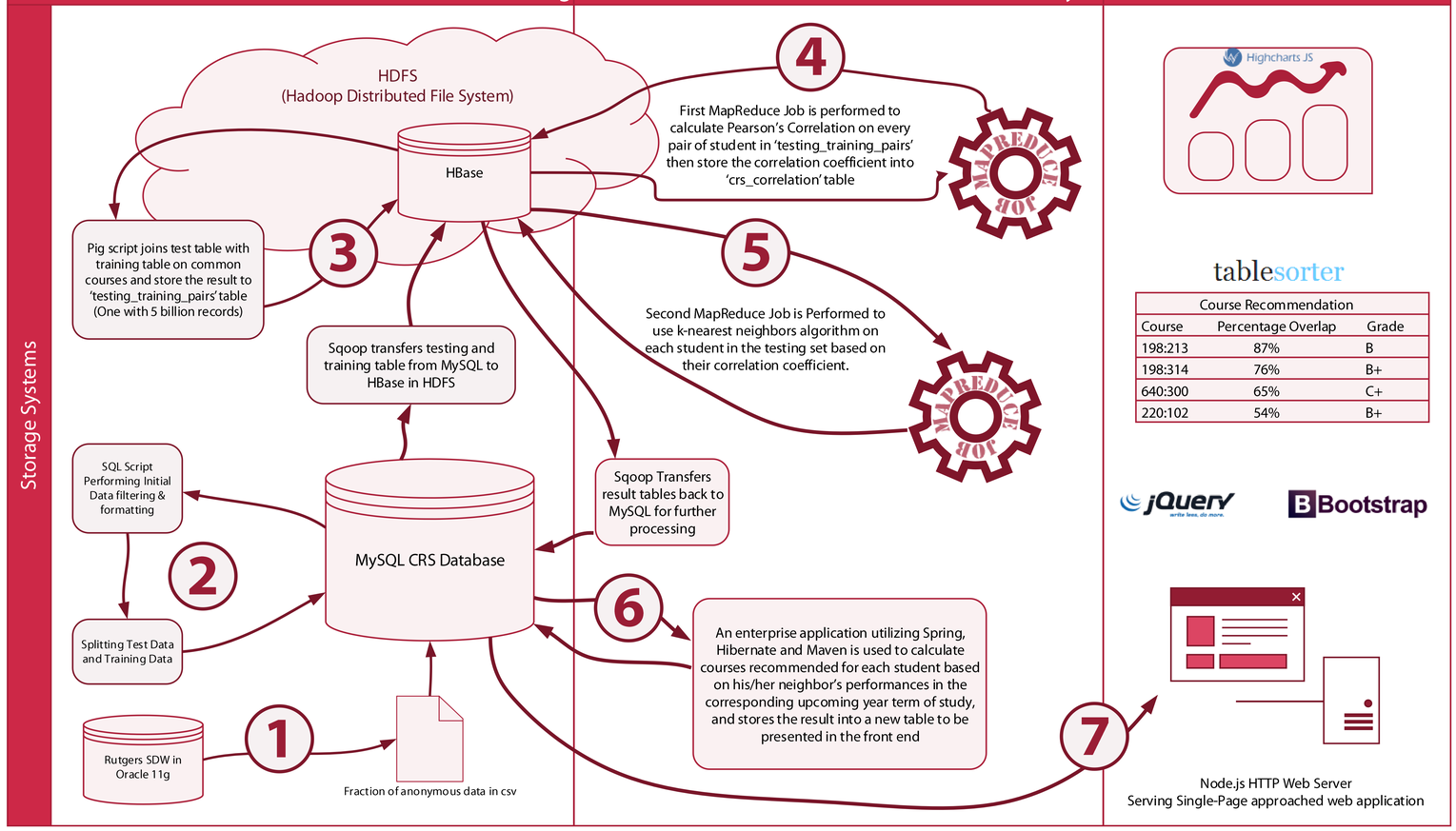
Motivation

- ❑ Overwhelming number of courses.
- ❑ Hard time choosing courses that fit interest, major, skills.
- ❑ Students need to quickly change decision if course fill up.
- ❑ Some courses fill up too fast.
- ❑ Hard to predict course demand which leads to overcrowded or scarcely populated classrooms.
- ❑ Course recommendations provide students with guidance and would predict courses that fit their interest, major, and skills.
- ❑ Recommendations can predict course demand and scheduling office can use this to provision classrooms and buses to alleviate congestions.

Objective

The goal of this project is to create a **recommender system** to recommend courses to currently enrolled students for the upcoming semester. The system will **recommend courses** to students based on courses that students most similar to them in prior years have taken in their next corresponding semester. The system will also use these recommendations to predict what courses all students will likely take and **project course demands** for the scheduling office to use to then provision buses and classrooms.

Course Recommender System Flow Diagram



Research Challenges

- ❑ Finding correlations in the data is hard since it is hard to decide which parameters to use and their weight in calculating the Pearson correlation coefficient.
- ❑ Limited resources and time. Only a 4-node Hadoop[1] cluster with makes optimizing parameters inefficient.
- ❑ The complexity of the problem might require us to discover new machine learning techniques instead of just the k-nearest using the Pearson correlation coefficient.

Results

As a Software System:

- ❑ Used HBase[8] and MapReduce[1] to store and calculate recommendations, enabling us to perform large joins, complex calculations and aggregations efficiently.
- ❑ Used Pig[2] and Sqoop[3] to transfer data between components of the system since data was stored, processed, and presented in different formats.
- ❑ Used Node.js[6] and MySQL[7] to efficiently query final recommendations to display to the user since multiple users would require asynchronous IO.
- ❑ **Future Work:** Automate and integrate with Rutgers.

As a Recommender System:

- ❑ Recommendations were better for students with stricter curriculums (e.g.. ECE) than those with relaxed options for courses. (e.g. Psychology)
- ❑ Students with more courses taken obtained better neighbors and thus better recommendations.
- ❑ Current Rating used for Pearson correlations were weighted average of term and grade received in common courses.
- ❑ **Future Work**: Use different weights to see If results improve. Use a different technique such as Euclidean Distance and test as an alternative to nearest neighbors or as a second layer of filtering.

Acknowledgement

We would like to thank Fei Luo and Martin Zhao of the office of information technology for providing us the guidance, the resources and the data to complete this project. We also thank professor Parashar for being our capstone advisor.

References

[1] **Hadoop** - http://hadoop.apache.org/
[2] **Pig** - https://pig.apache.org/
[3] **Sqoop** - https://sqoop.apache.org/
[4] **Spring** - http://spring.io/
[5] **Hibernate** - http://hibernate.org/
[6] **Node.js** - http://nodejs.org/
[7] **MySQL** - https://www.mysql.com/
[8] **HBase** - https://hbase.apache.org/
[8] **Maven**- https://maven.apache.org/