GoogleSummerOfCode Ideas

NUTCH-1936 GSoC 2015 - Move Nutch to Hadoop 2.X

Description

The Nutch PMC discussed ideas for a good 2015 GSoC project. It appears that porting the (trunk) codebase to Hadoop 2.X seems to an attractive option and one which would present an excellent learning experience for a summer student. Essentially this project will involve porting ALL aspects of every class which utilizes the Hadoop 1.X codebase to Hadoop 2.X. A good starting point would be to understand what has already been discussed in the existing JIRA Issues.

Interested Mentors

Lewis John McGibbney

Student Proposals

- Name:
- University:
- Short Description of Interests in GSoC:
- Proposal: Please attach as a PDF/Word Document and simply link to it from here.

Reports

TODO

Documentation

TODO

Jira Issues

- NUTCH-1936 - GSoC 2015 - Move Nutch to Hadoop 2.X (Parent Issue)
- NUTCH-1219 - Upgrade all jobs to new MapReduce API