

# GSoC2015

## Google Summer of Code 2015 for Pig

Pig is exciting! Pig provide an intuitive way to program hadoop. Inside Yahoo!, more than 80% of hadoop jobs are Pig jobs. It is heavily used in Twitter, LinkedIn and lots of other organizations (<https://cwiki.apache.org/confluence/display/PIG/PoweredBy>).

This is the forth straight year Pig participate Google Summer of Code. Last year we accepted 5 students and 4 of them complete their projects successfully.

Here we picked up a list of highly desired projects for students. Once accepted, we will assign a dedicated mentor to guide you through different stages of the program. We need your help and you will get a great experience of participating an open source project.

### Project List

**TPC-DI queries for Pig** (<https://issues.apache.org/jira/browse/PIG-4435>)

**TPC-DS queries for Pig** (<https://issues.apache.org/jira/browse/PIG-4436>)

### Other Project Ideas

You can also propose new project not listed. Please discuss with us before apply.

### Getting start

First, you need to learn PigLatin language. The best source for learning PigLatin is:

- [Pig Latin Reference](#)
- [Pig Latin paper at SIGMOD 2008](#)
- [Pig Tutorial](#)
- [Pig optimizor paper](#)

Be sure to sign up [pig mailing list](#).

Then checkout Pig source code using svn:  
svn co <http://svn.apache.org/repos/asf/pig/trunk>

Set up environment for Eclipse:<https://cwiki.apache.org/confluence/display/PIG/How+to+set+up+Eclipse+environment>

Learn more about Pig internal at [Pig paper at VLDB 2009](#).

Browse through Pig code. Some good start points are:

- [QueryLexer.g](#), [QueryParser.g](#), [LogicalPlanGenerator.g](#): Pig parser, LogicalPlan construction
- [LogToPhyTranslationVisitor](#): From logical plan to physical plan
- [MRCompiler](#): From physical plan to map-reduce plan
- [JobControlCompiler](#): From map-reduce plan to hadoop job
- [MapReduceLauncher](#): Hadoop launcher
- [PigMapBase](#): map class for Pig
- [PigMapReduce](#): reduce class for Pig

Try to start with something simpler: <https://issues.apache.org/jira/browse/PIG-3371?jql=project%20%3D%20PIG%20AND%20resolution%20%3D%20Unresolved%20AND%20labels%20%3D%20newbie>

### How to Apply

- Follow [GSoc](#) instruction to apply. Please apply to Apache Software Foundation.
- A competent application should include:
  1. A little about your background
  2. Your experience level in hadoop/MapReduce/NoSQL or related areas, Java programming skill
  3. Project understanding
  4. Proposed project schedule
- Keep [timeline in mind](#).
- It is highly recommend to discuss your interest before you apply. The best way to discuss is to comment on individual Jira or send mail to dev list.