

开源软件课程报告

韩明聪 10165101243

本文将总结本学期我学习开源课程过程中的收货，主要以参与开源项目 TiDB 为主，除了参与 TiDB 功能的开发，我还围绕着“开源”这个主题，在 TiDB 中做了一些其他的相关工作。

一、参与 TiDB 开发工作

截止到 2019 年底，我在 TiDB 中共合并 40 个 Pull Request（下文简称 PR），这其中只包括在 master 分支合并的代码，不包含在 release 分支的 cherry-pick。在 master 分支的改动约为 6000+，2000-。主要参与的开发组件是 TiDB 查询优化器，绝大部分是在做 TiDB 新的优化器 Cascades Optimizer 的开发，少部分包括旧优化器的 bug fix 以及功能提升。

除了开发工作，我还在 TiDB 中 Review 了超过 70 个 PR，其中 approve（或评论 LGTM）的 PR 超过一半。

从课程中学到的很多东西都对我在 TiDB 中的开发工作起了很大的作用，也帮助我深入理解了开源项目的一些原理和机制。最典型的是我在课程中学习了开源协议，然后我去查看了 TiDB 的开源协议使用的是 Apache 2.0 协议。之后又听到建盛老师关于“配置文件及代码”的一些概念，我发现在 TiDB 中，包括 Makefile、Dockerfile 和一些编译脚本中都没有包含 Apache 2.0 协议的声明。不仅仅是 TiDB，包括 TiKV 等其他 PingCAP 公司主导的项目，都没有在脚本文件中添加 License 声明。因此我首先在 TiDB 中通过一个 Pull Request 填补了缺少的 License，然后我又在 TiDB Community 里建议大家检查各个 repo 的 License，并且正在推进使用 CI 来检查各个文件（尤其是新添加文件）的

License 改动。

二、 协助同学参与 TiDB 项目

开学前几周，我在课堂中对 TiDB 项目进行了宣讲，尤其是对前一阵子开展的表达式向量化工作进行了推广，吸引了三十多位同学参与。其中有超过二十位同学在 TiDB 中有 PR 被合并，总计合并的 PR 超过六十个。我也帮助了其中十个左右的同学，解决他们开发过程中遇到的问题，同时帮助推进他们的 PR 合并。

大多数同学都是第一次参与开源项目，其中也有很多同学是第一次与他人合作“写代码”。从我帮助同学的过程来看，很多同学对 Git 并不熟悉，导致出现 Conflicts 等问题时，不知道该怎么处理，从而关掉 PR 又重新开一个新的。这一点其实是 Git 这个工具在本科教学过程中没有很好的做普及，在我看来 Git 与 Makefile 等工具同等重要，是开发过程中基础工具。另外很多同学看不懂 Review 中的常用语，包括但不限于“LGTM”、“nit”、“ditto”和“PTAL”等，其实我们的这门课程也可以在实践角度上带大家熟悉这些基本的东西，比如在课堂上带领大家一起看几个典型 PR Review 内容，帮助大家快速上手。

在帮助同学的过程中，我也发现很多同学平时 coding 的时候很不注意自己的 code style，大部分的 review comment 都与 code style 相关，希望这次参与开源项目的经历能够提醒大家 code style 的重要性，给大家养成一个好的习惯。

三、 参与开源社区运营

除了开发工作外，我还参与了开源社区运营的一些工作。前文介绍过我的主要开发工作的 TiDB 的一个新优化器 Cascades Optimizer，从我接手这项

Project 开始的三个月里，一直都是只有我一个人做 Coding，其他的同事最多就是给我进行一些 Code Review 和一些大型改动的讨论。这导致这个项目的开发进度其实并不是很快，直到 2019 年 11 月底，优化器的整体框架基本完成，因此我们写了两篇文章（其中一篇是我写的）发表在 PingCAP 公司的博客里，分别来介绍这个优化器的原理以及邀请社区来参与开发工作，于是在 2019 年 12 月，来自全世界各地的开发者（国内的学生、程序员，还有美国、波兰等国家的开发者）都进入到这个项目中，开始为这个工作提 PR，为优化器添加优化规则。

在运营这项工作的过程中，我为了吸引社区同学参与，也精心设计了很多的工作：

1. 为 new-contributor 推荐容易上手的 issue，详细讨论的过程可以参加 <https://github.com/pingcap/tidb/issues/13709>。
2. 简化并记录下整个开发的 Workflow，降低开发者的参与门槛。
3. 定期在 Slack 里询问大家的开发进度，是否需要帮助等。
4. 每过一段时间，都在 Slack 里同步一下近期开发的进度，并且鼓励、感谢各位社区的开发者。

四、 课程建议

整体来说在这门课上学到的内容非常多，在以前我只认识到开源是把源代码开放出来，让整个社会收益。现在我知道了开源不仅仅是源代码开发，也是一种协作的方式，是软件开发的方式。同时在课上还学到了开源软件与自由软件的区别，和一些开源软件的发展历史等。有一些不足的是，整个课程缺乏带有指导性的实践内容，例如课程需要同学们参与开源，是不是就可以带同学们开一个测试用的 repo，然后熟悉一下 PR 怎么开，怎么 Review 之类的。