

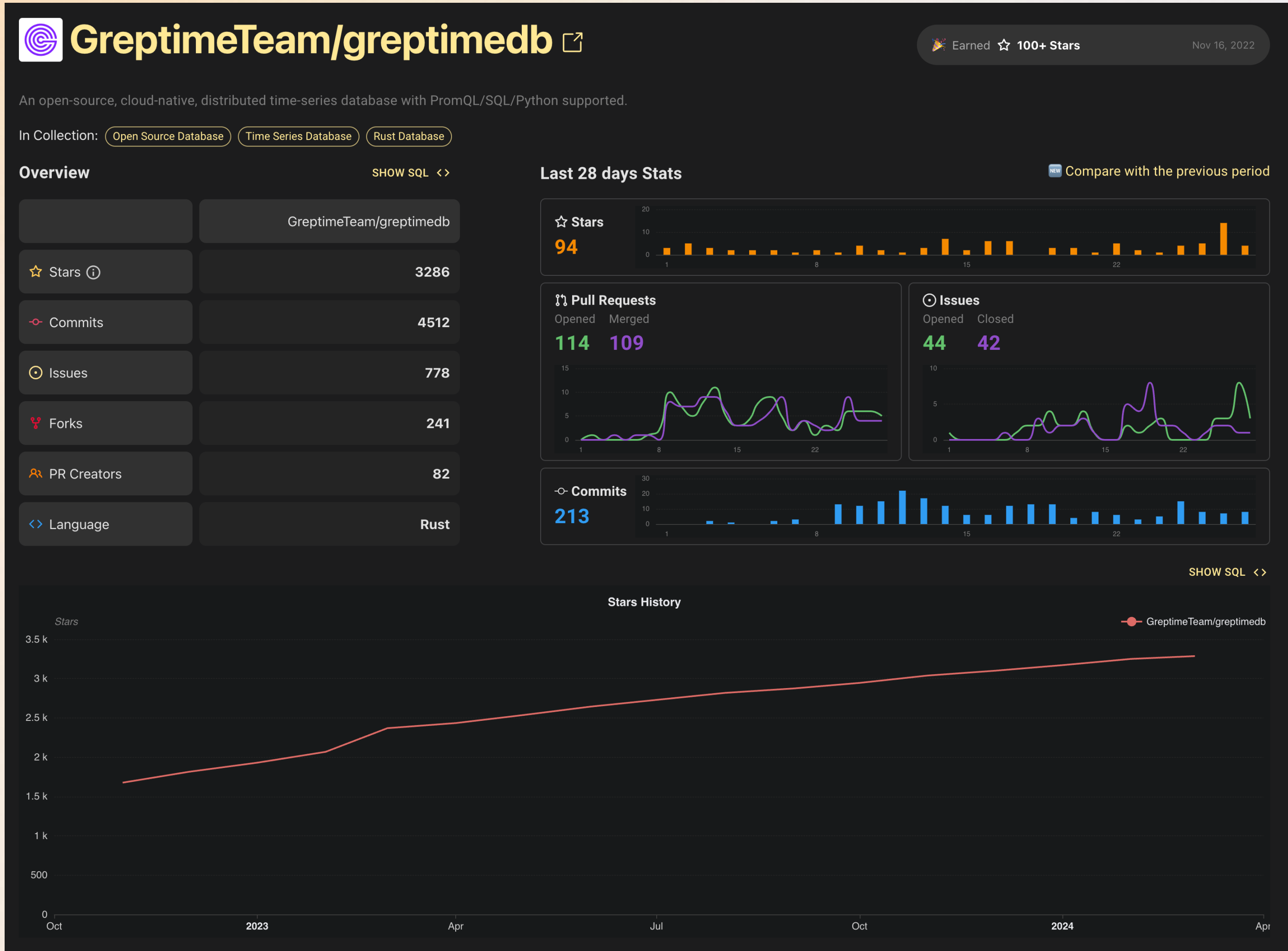
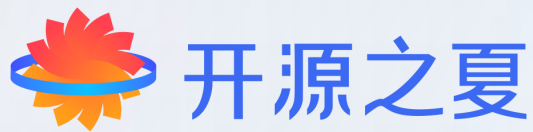


新人导师的开源之夏体验开箱

GreptimeDB
黄磊

2024.03.07

关于 GreptimeDB



- 基于 Rust 构建的云原生时序数据库
 - 支持时序和分析混合负载
 - 弹性扩缩容，支持无缝水平扩展
 - 兼容 MySQL、PostgreSQL、InfluxDB、Prometheus 等协议
 - 支持对象存储，降低数据存储成本
- 2022年 11 月 15 日开源
 - 连续登顶 Github Trending
 - 至今获得 3.5k star 和来自十余个国家的 61 位贡献者



GreptimeDB 公众号



GreptimeDB 小助手

目录

- 1 一个好的题目是成功的一半
- 2 “不会写 Rust 可以报名吗？”
- 3 像对待 contributor 一样对待学生
- 4 结束只是开始

01
一个好的题目是成功的一半

一个好的题目是成功的一半

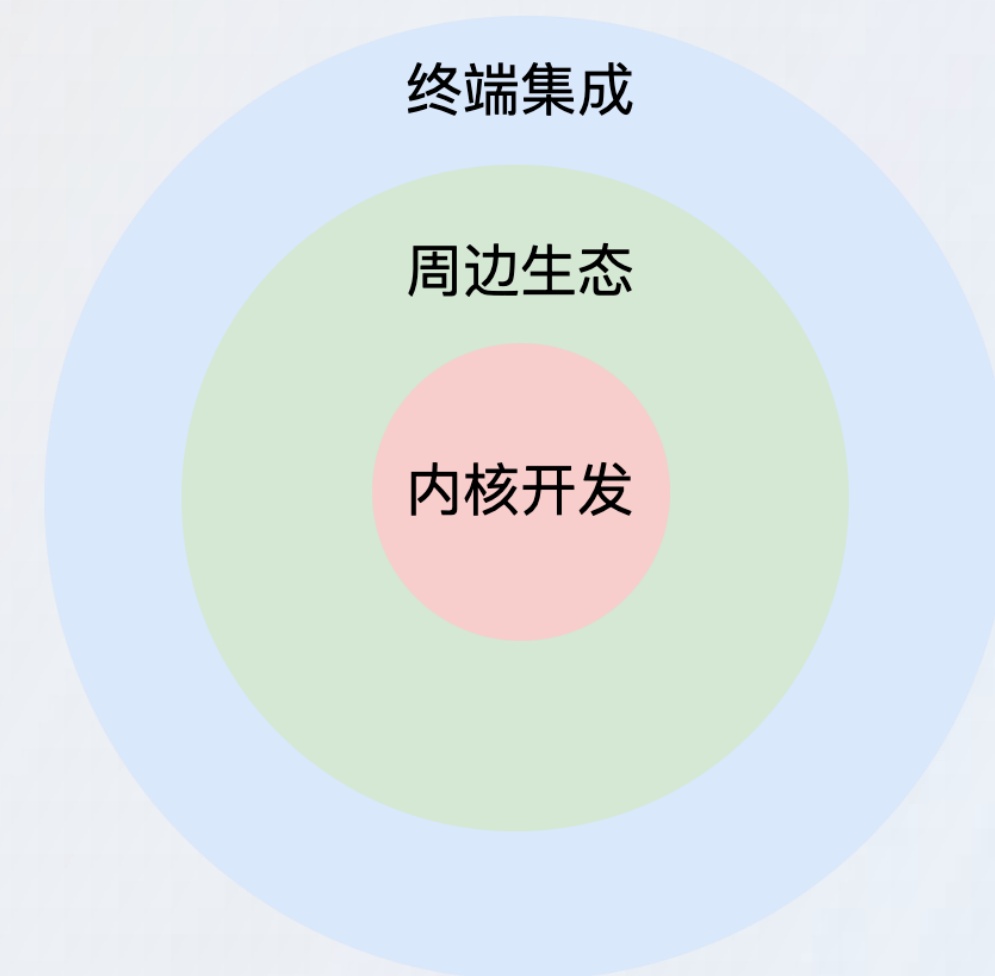
- **开放什么样的项目取决于期望吸引什么样的开发者**

- 对 contributor 的分类
 - 内核开发：涉及存储、查询引擎的核心功能
 - 周边生态：周边工具开发和共建（如 SDK、connector 等）
 - 终端集成：GreptimeDB 对各类系统的适配
- 我们参加开源之夏的初心
 - 吸引更多的新人进入数据库内核开发的领域

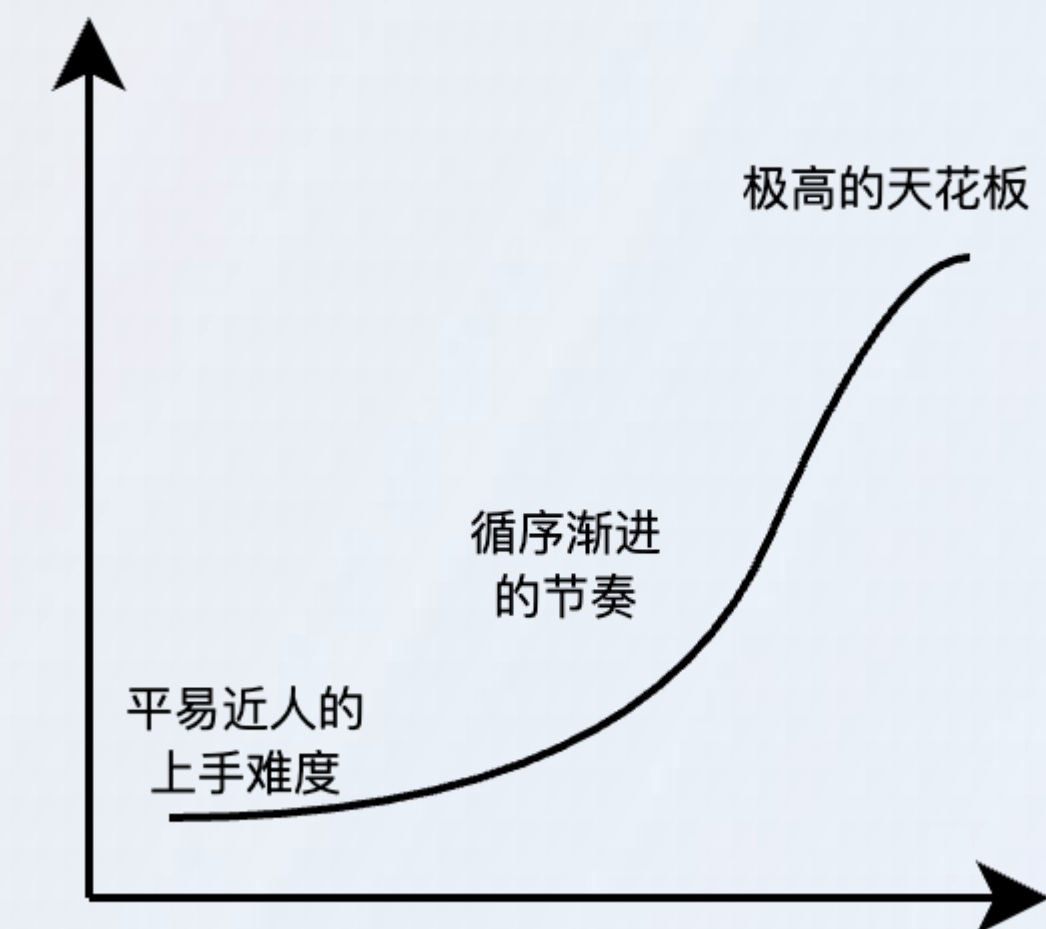
- **什么是一个好项目？**

- 平易近人的上手难度
- 循序渐进的节奏
- 极高的天花板

——（重要但不紧急）——



开源社区的 contributor 分类



理想的项目难度曲线

一个好的题目是成功的一半

GreptimeDB 的题目

- **实现对 Duration 和 Interval 数据类型的支持 (#1886)**

- Duration 数据类型是度量特定类型的时间，可以用秒、毫秒、微秒或纳秒表示。Interval 数据类型通过计算两个时间点之间的天数表示时间差异。项目需要为 GreptimeDB 实现这两个数据类型的存储已经相应的计算功能。

- **实现滑动采样窗口函数的语法和计算 (#1662)**

- 查询并聚合一个给定长度的时间范围的数据是时序数据常见的一种查询模式，例如 PromQL 中的 Range Selector，而 GreptimeDB 计划在 SQL 界面中实现类似的能力。与 WINDOW 语法不同，滑动采样窗口函数会按固定的步长及范围进行计算，例如：查询每日气温的周平均值；该查询的步长是一天（每日气温），每一次计算都使用过去一周的数据求平均。该项目需要从拓展 SQL Parser 开始，定义并实现滑动采样窗口函数的语法，并拓展逻辑计划器来实现相关语句的计划生成，最后实现对常见聚合算子的支持。

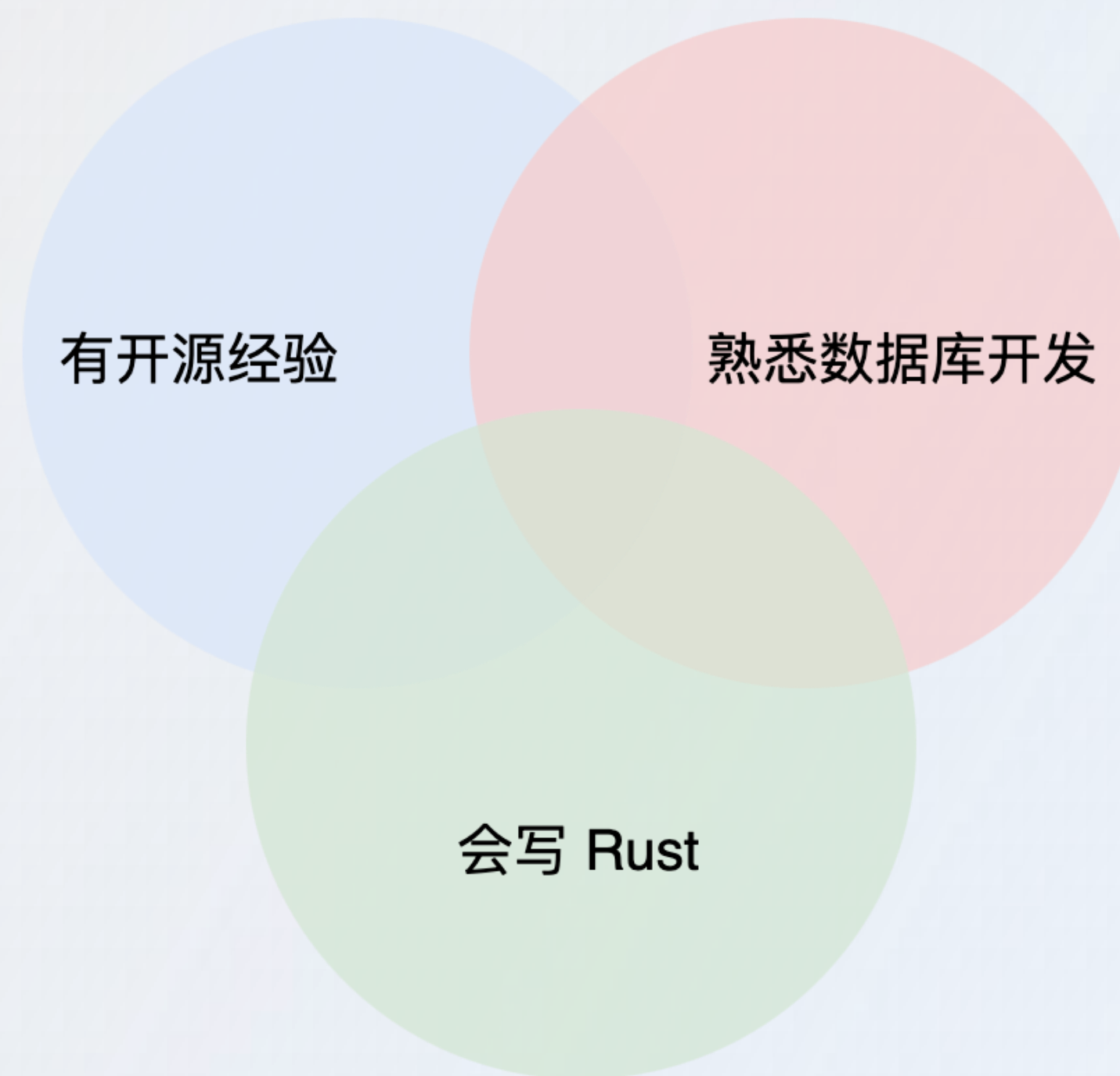
- **实现 TRUNCATE TABLE 功能**

- 实现 GreptimeDB 的 TRUNCATE TABLE 功能；TRUNCATE TABLE 从表中删除所有行。它类似于没有 WHERE 子句的 DELETE 语句，但比逐行删除更快。通过该功能的实现，提升 GreptimeDB 的性能。

“我不会 Rust 可以报名吗？”

如何选择你的天命人

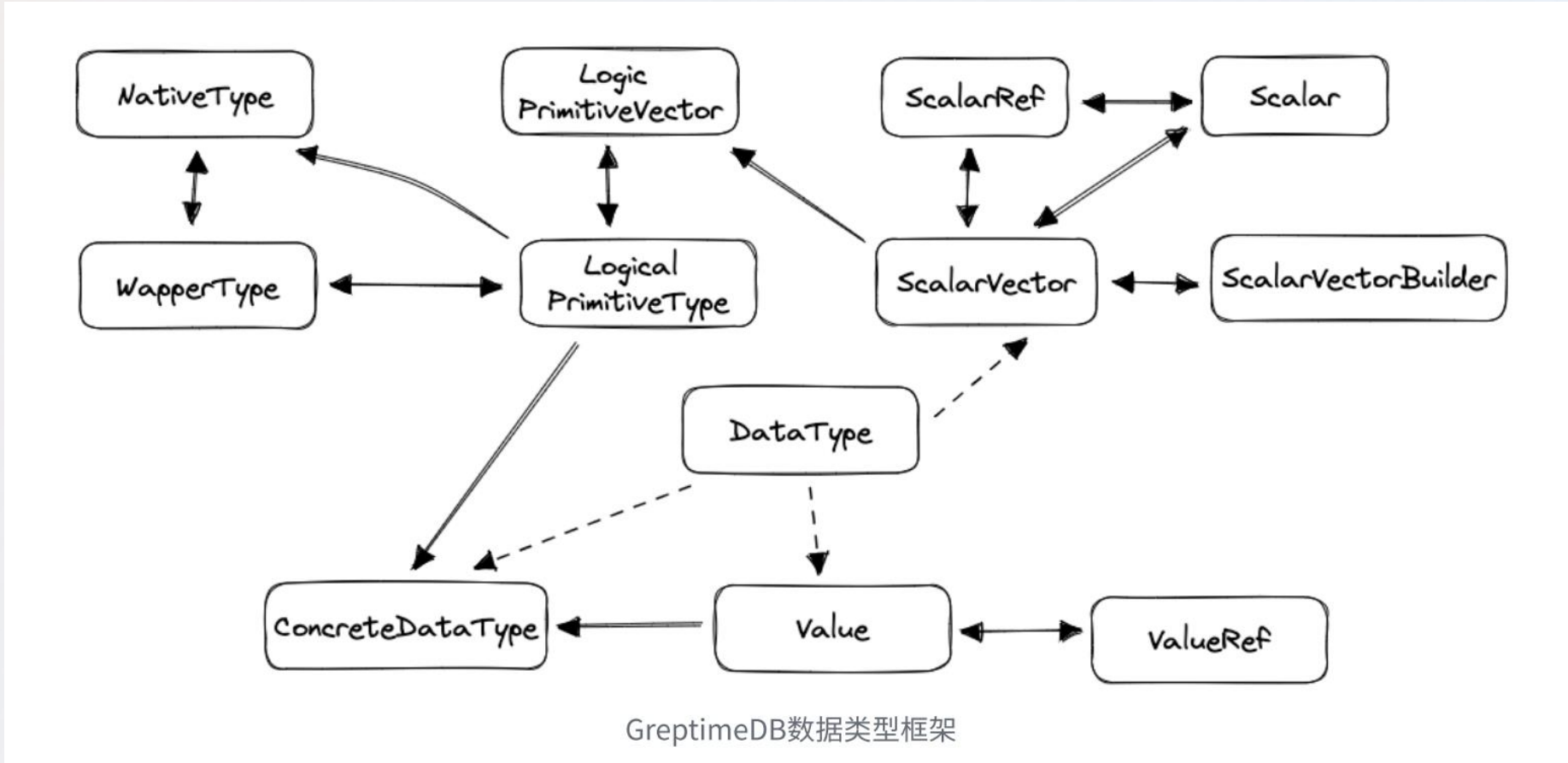
- 既有开源经验，又懂数据库开发，还会写 Rust 当然最好
 - 但是这样的他，需要参加开源之夏吗？
 - 开源之夏是双向选择
- 比起锦上添花，更重要的是雪中送炭
- 对“集邮式”参与开源项目的看法



如何选择你的天命人



来自东南大学的邹伟同学

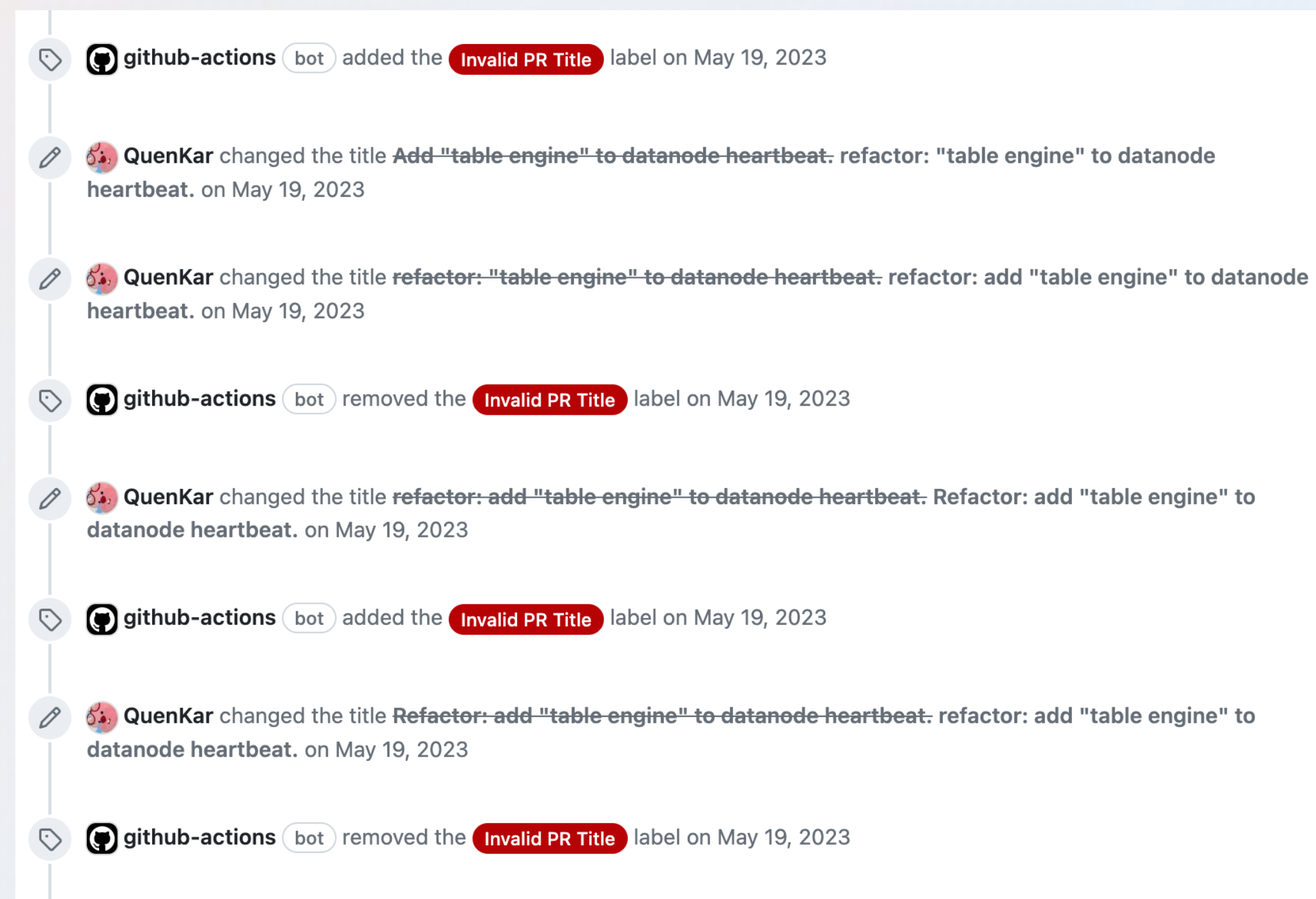


项目申请书

像对待 contributor 一样对待学
生

像对待 contributor 一样对待学生

- 从 [CONTRIBUTING.md](#) 开始
- 学生和开源项目的 gap
 - Code of conduct
 - 隐秘的角落：测试、日志
- 开源之夏也是一个检视开源社区的机会
 - 新人 onboarding 流程是否丝滑？
 - 对社区贡献者的 CR 是否及时？



和 bot 反复纠缠的第一个 PR

04

结束只是开始

结束只是开始

- 从开源之夏到实习生，从类型系统到存储引擎
 - 邹伟同学总共为 GreptimeDB 提交了 [48 个 PR](#)，其中 47 个已合并，覆盖类型系统、查询引擎、存储和缓存等模块
- 关于开源之夏的一点建议
 - 和开源项目的开发流程略有割裂
- 期待与各位在今年夏天再见！

错过开源之夏不要紧，我们还在招聘实习生！



• Rust 实习工程师（杭州）

- 使用 Rust 从事分布式数据存储及分析系统研发，直接参与设计和研发高性能的分布式 KV、弹性查询和计算引擎等，工作内容包括架构设计、功能迭代、bug 修复和性能优化等工作，持续交付一流的功能模块或产品
- 计算机基础扎实，熟悉常用数据结构和算法，掌握 C/C++/Rust/Go 语言中的至少一门，愿意学习和使用 Rust 语言
- 对数据库系统（如MySQL/SQLite/TiDB 等）、数据仓库系统（如 ClickHouse、Hive）或者其他大数据计算框架了解者优先
- 对编写优秀的代码有卓越的追求和品味对技术领域有持久的热情，具备持续学习的能力
- 熟悉 Linux 系统，有丰富的系统编程经验者优先
- 一周能到岗 4 天，连续实习 3 个月以上者优先

• Golang 实习工程师（北京）

- 熟悉 Go、Rust 其中一门编程语言
- 熟悉 Prometheus、Grafana 生态
- 熟悉 Linux 系统和 Docker 使用
- 计算机/软件工程或相关领域 25 届应届毕业生优先
- 一周能到岗 4 天，连续实习 3 个月以上者优先

谢谢观看！