万亿级流量转发软件BFE的开源之路

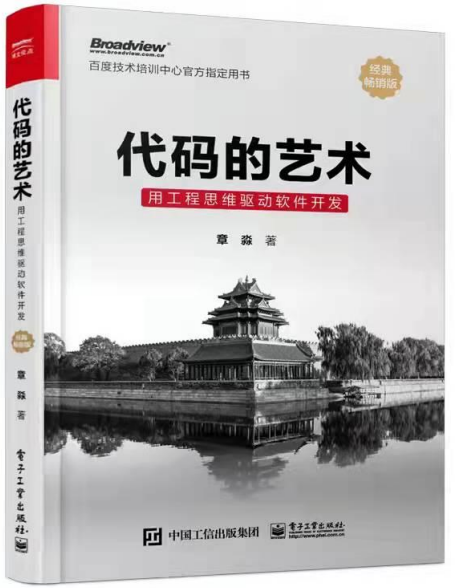
章淼 博士

瑛菲网络

创始人 & CEO

个人简介

- 1994-2004，清华大学计算机系，博士
- 2004-2006，清华大学网络中心，助理研究员
- 1997-2006，清华大学，互联网协议 / 网络体系结构研究
  - 曾参与中国第一代核心路由器研发工作
- 2006-2012，多家公司（搜狗、腾讯等），用户产品研发
- 2012.11 – 2023.4, 百度，网络基础架构，软件工程
  - BFE团队负责人
  - 2018.1- 2021.10, 百度代码规范委员会主席
  - 2021.10 -，百度代码规范委员会荣誉主席
- 2023.4 -，**瑛菲网络**，创始人&CEO



BFE历史背景

百度统一的七层流量转发平台**开始建设**
2012

BFE => Baidu Front End

基于**Go语言**重构，2015年1月在百度**全量上线**　　2014 - 2015

BFE亮相美国Velocity大会，成为Go领域**标杆项目**

顺利完成对百度**春晚红包项目**的支持
2019

核心转发引擎**对外开源**，被**央视网**、360选用

BFE成为**网络方向中国首个CNCF官方开源项目**
2020

BFE => Beyond Front End，被**招商银行**选用

每日转发请求**超1万亿**，日峰值超过1000万QPS

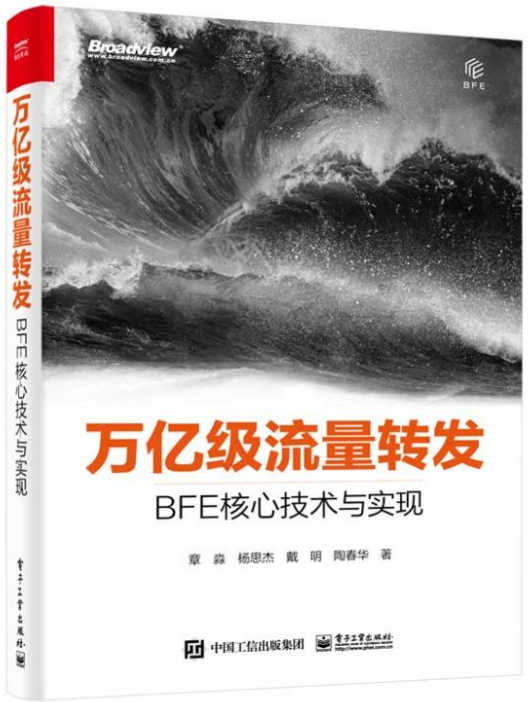
《万亿级流量转发：BFE核心技术与实现》正式出版
2021

BFE控制面和BFE ingress开源

在招商银行**大规模部署**
2022

被**广发银行**选用

BFE开源项目被**海外(北美，非洲)用户采用**



https://github.com/bfenetworks/bfe

第一部分

为什么要基于Go开发BFE?

流量管理平台的需求和发展趋势

负载均衡要为 现代应用 服务

现代应用(Modern App) 的特征

- Scalability: 容量可扩展
- Portability: 支持多云和混合云
- Resilience：高可用，快速恢复
- Agility：敏捷迭代，快速更新

为了支持现代应用，需要 **新一代** 的

面向 **云原生** 的 **流量管理平台**

新一代流量管理平台的特征

① **软件化** （容量可扩展，支持多云部署）

② **四七层分离**（容量可扩展）

③ **多主集群**（容量可扩展）

④ **数据平面和管理平面分离**（容量可扩展）

⑤ **多云/多集群调度**能力（多云，高可用，敏捷）

⑥ **强大的路由管理**能力（微服务，敏捷）

⑦ **流量洞察**能力（高可用，敏捷）

⑧ **安全能力**（高可用）

⑨ **多租户** 能力（敏捷）

⑩ **平台化 / API接口** (高可用，敏捷)

七层负载均衡生态对比

生态	代表项目	说明	性能	安全性/稳定性	开发效率	开源生态	转发延迟
Nginx ／ OpenResty 生态	Nginx，OpenResty，Kong APISIX	• OpenResty是对Nginx的一种扩展，可以利用Lua语言对Nginx功能做扩展。OpenResty开源项目由中国工程师章亦春创建。 • Kong和APISIX均为API网关开源项目，详情见GitHub	高	低	低	强	低
Envoy 生态	Envoy	• Envoy是基于C++开发的七层开源软件。最早由美国Lyft公司技术团队开发并开源，后Google加入。目前Envoy已经成为服务网格(Service Mesh)中Sidecar网关的重要候选系统。	高	低	低	强	低
Go语言 生态	BFE，Traefik，Tyk	• Traefik为一家法国创业公司推出的七层负载均衡开源软件，详情见 https://github.com/traefik/traefik。 • Tyk为一家英国创业公司推出的API网关开源软件，详情见 https://github.com/TykTechnologies/tyk。	低	高	高	强	低，但有少量长尾
Rust语言 生态	Linkerd	• Linkerd为一家美国创业公司，专注于服务网格方向。其中包含一个使用Rust语言开发的七层负载均衡软件。	高	高	低	弱	低

第二部分

BFE相比Nginx有哪些优势?

应用层路由

- **技术**
  - -对每个租户提供独立的分流转发表
  - -分为**基础转发表、高级转发表、默认集群**
  - -自研的**条件表达式**描述转发条件
  - -内置40多种条件原语，可支持 与/或/非 组合
- **优势**
  - 兼具**高性能**和**强大描述能力**
  - 优先级清晰
  - 相比**正则表达式**
    - (1) 具有更好的**可维护性**
    - (2) **无性能退化（恶性回溯）**的隐患

基础转发表

匹配条件	目标集群
www.a.com/a/*	Demo-A
www.a.com/a/b	Demo-B
*.a.com/	Demo-C
www.c.com	ADVANCED_MODE

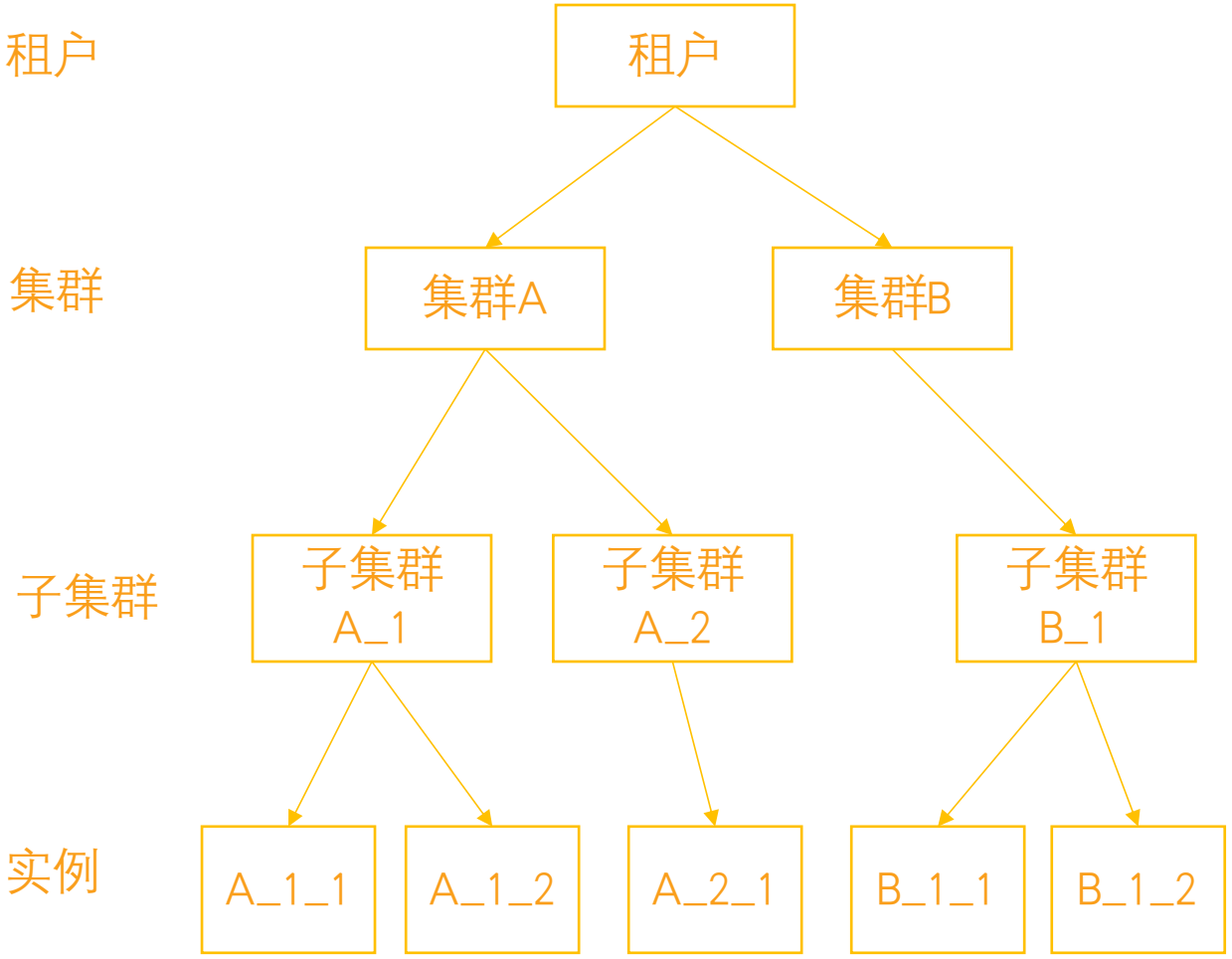
高级转发表

匹配条件	目标集群
req_host_in("www.c.com") && req_cookie_value_prefix_in("deviceid", "x", false)	Demo-D1
req_host_in("www.c.com")	Demo-D

默认集群 Demo-E

多租户的支持

- **Nginx在多租户支持方面的问题**
  - Nginx引擎未提供多租户支持
  - 配置热加载 => 长连接中断，配置加载开销大
  - 正则表达式的性能隐患 => 单租户配置风险
- **BFE的相关机制**
  - 内置多租户模型
  - 配置热加载不影响长连接
  - 多模块可单独动态加载配置
  - 条件表达式机制 避免 正则表达式的性能隐患



内网自动流量调度(GTS)

- **技术**

  BFE支持按照给定的权重在多个子集群间分配流量

  GTS支持根据流量、容量、机房间距离等因素持续自动计算分流权重

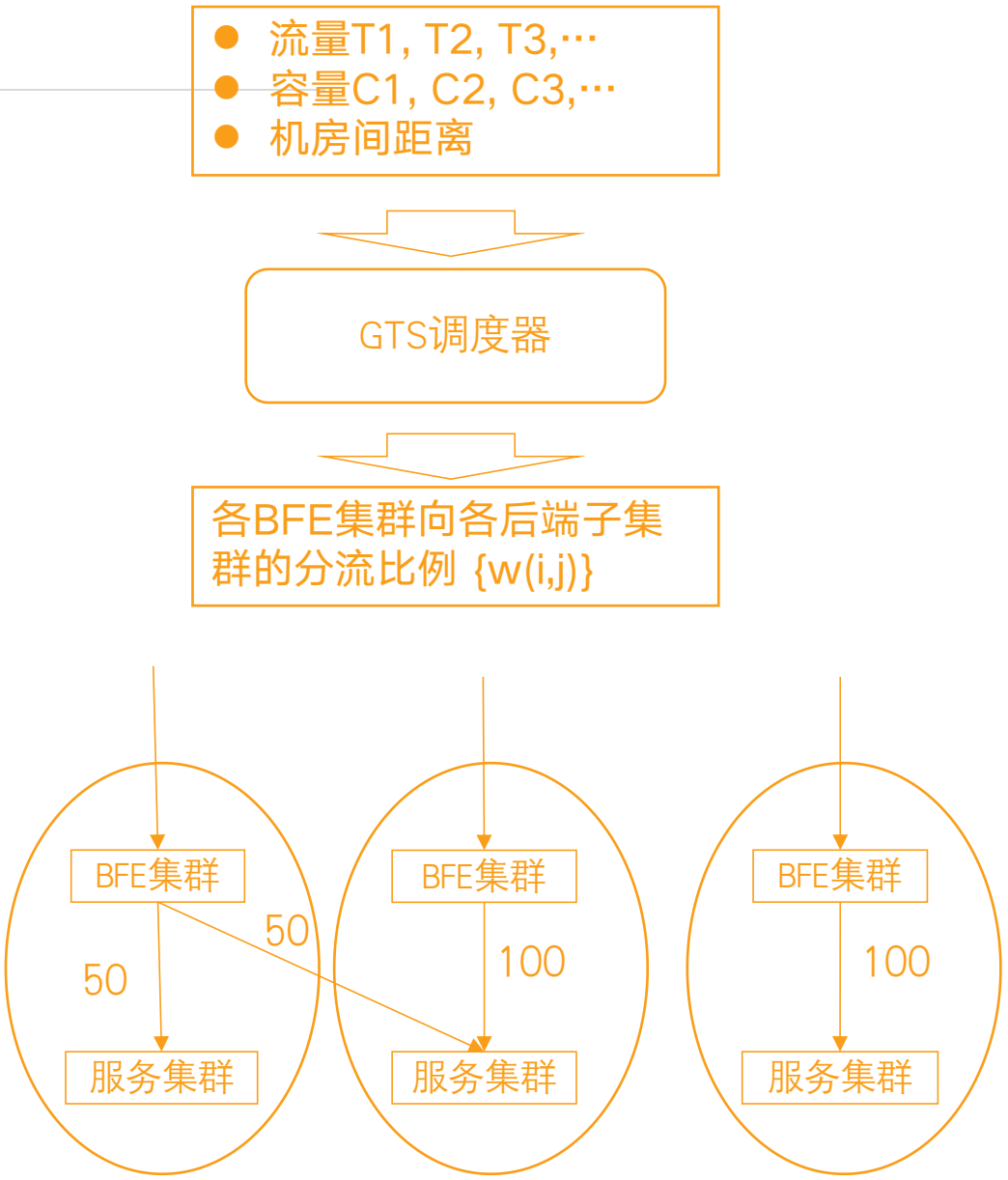
- **优势**

  在流量、容量等发生变化后，可在**20秒内**完成调整

  和传统的DNS的方案相比

  (1) 实现子集群间流量的**精确分配**

  (2) 调整时间**大幅缩短**（10分钟 => 20秒）



新一代的安全架构

- **原有问题**

  本地计数，限流不准确

  RSA加速卡本地调用，限制调度能力

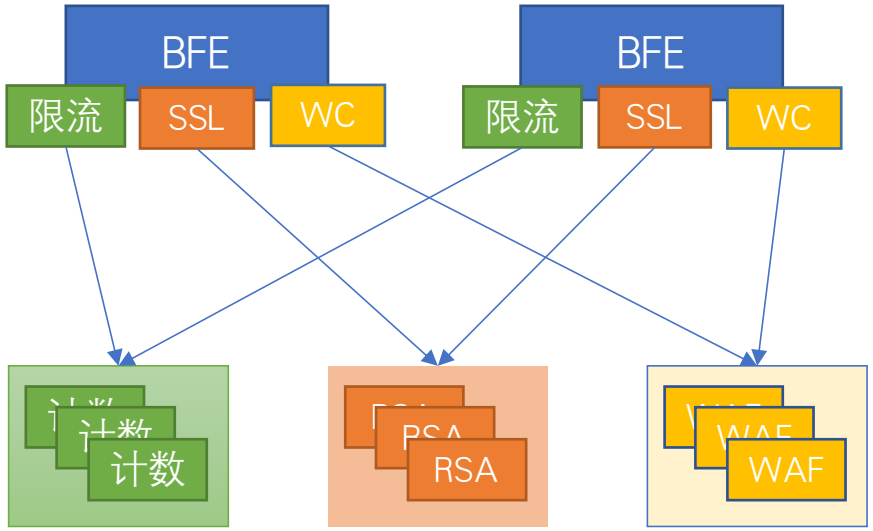
  WAF内嵌，**变更**易导致**性能下降**、甚至**服务中断**

- **优化思路**

  全局计数器

  RSA加速卡形成**资源池**

  WAF处理成为**外部调用** => 保证转发延迟和可靠性



第三部分

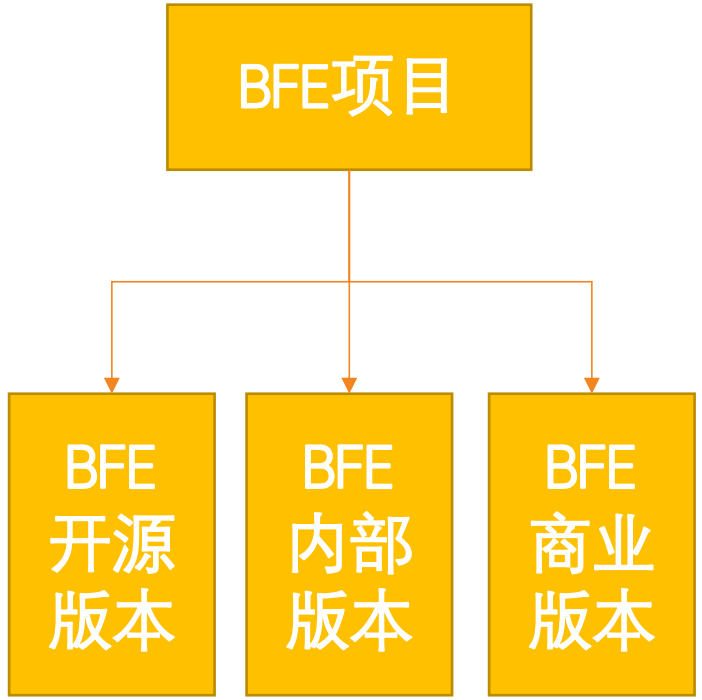
BFE开源和商业化的关系

为什么BFE要开源

- 2018年底，商业化遭遇挫折，决定开源
  - 让客户看看代码质量
- 2020年，加入开源基金会
  - 其实还搞不清楚几个开源组织的差异
- 2021年，需要让用户更容易使用
  - 将控制面开源
- 2021年，需要建立使用者社区和开发者社区
  - 建立BFE开源公众号
  - 出版《万亿级流量转发: BFE核心技术与实现》
- 2021年，需要融入K8s生态
  - 研发和开源BFE Ingress

BFE开源和商业的关系

- 商业化是BFE开源的主要目的之一
  - 便于潜在客户了解产品
  - 便于和相关开源/闭源系统建立生态
  - 在云原生领域，开源就是标准
- BFE开源项目的模式：Open Core
  - 企业级功能不会主动开源
  - 短期内，完全开源的产品很难让中国客户接受付费
  - 长期看，服务才是客户付费的主要理由
  - 和二进制可执行程序相比，开源产品的质保也是问题
- 免费并不意味着最好
  - 商业客户的诉求：以合理的成本获得高质量的产品和服务
  - 免费 => 项目难以持续 / 没有服务 => 客户更大的损失

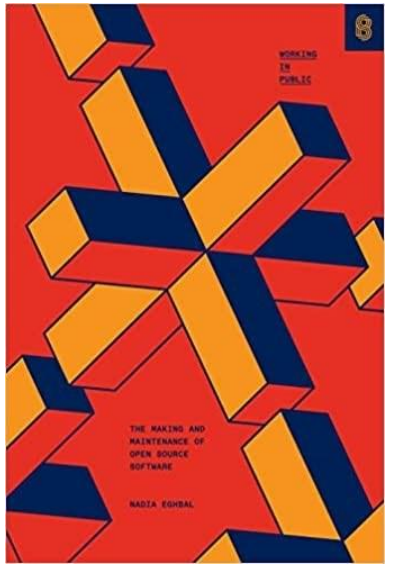


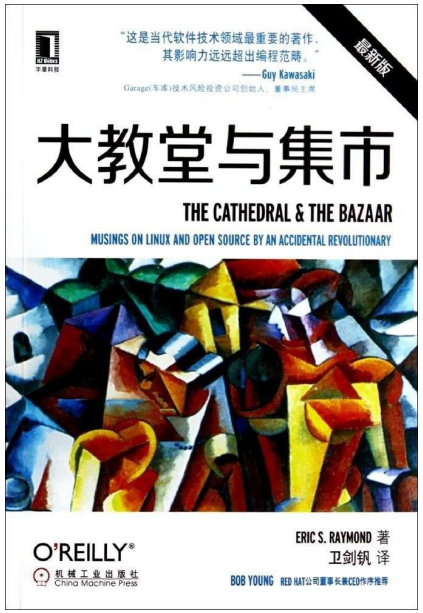
第四部分

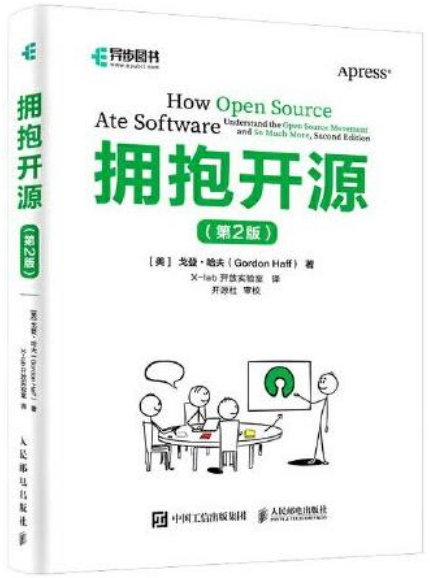
开源对于社会发展的意义

软件的开源是一种社会活动

- 软件开发需要自然科学，也需要人文社会科学
- 软件开发本质上是 人的活动
- 开源 是对 软件社会关系 的一种重构
  - 项目内的合作生态：使用者，开发者
  - 项目间的合作生态
  - 商业生态
- 需要关注开源协议的长期影响
  - 例如：GPL vs Apache-2.0
- 参考书籍
  - Working in the Public
  - 大教堂与集市
  - 拥抱开源

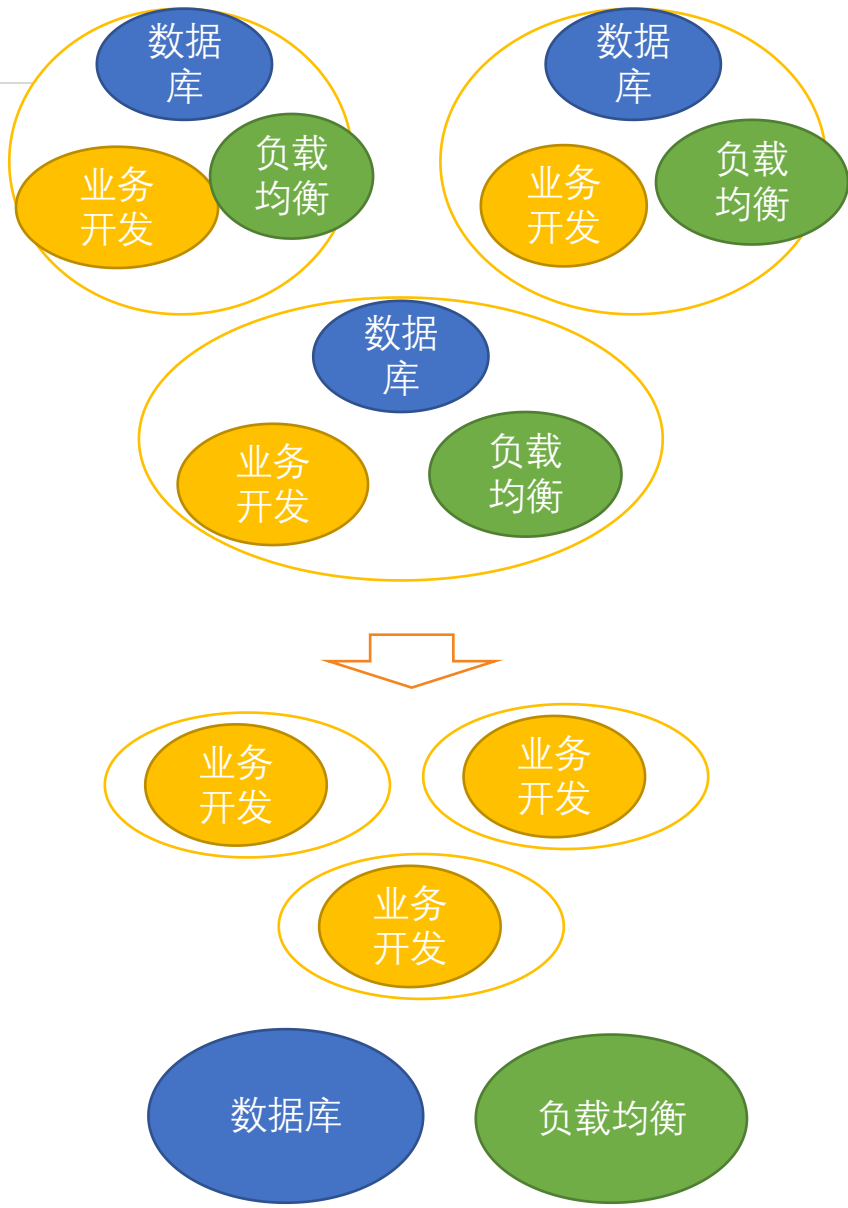






产品化和专业化是中国软件发展的必由之路

- 中国软件工程师现状
  - 数量庞大（400-800万），但专业化程度不高
  - 大量从事应用开发和定制开发
  - 各专业领域，大量公司都有人在基于开源/自研系统
- **专业化** 是软件研发生产力和效率提升的必由之路
  - 2000-2020：组织内部范围的专业化
  - 2020-2040：统一大市场范围的专业化
  - 所有组织都会将非主业 外包/使用专业软件
- **产品化** 是中国软件发展的必由之路
  - 定制化开发，对 软件企业 和 最终客户 都是一种伤害
  - 软件研发的升级：程序 => 软件 => 软件产品
  - 产品化的决定因素：优秀的产品经理



开源对于中国软件业的意义

- 促进技术的共享
- 促进组织间的技术合作
  - 减少重复造轮子
- 为年轻人提供实践的机会
  - 例：BFE开源之星实习项目

BFE开源之星Intern项目第一期圆满结束！

原创 BFE开源项目 BFE开源项目 2022-08-29 10:29 发表于上海

BFE开源之星Intern项目第二期圆满结束！

原创 BFE开源项目 BFE开源项目 2023-04-21 10:26 发表于上海

实习生总结

by loheagn@github(李楠)

*在项目开始之前，我本来以为既然在issue中已经有了基本的proposal，就可以直接进入编码阶段。但实际上在项目前期，花了很多时间进行需求细化和Annotation的重新设计。在这一过程中，我对"要做什么"有了更清楚的认识，也**拓展了我对HTTP协议的了解**，比如，重定向的Response中Header的Location字段等。*

实习生总结

by mengtao97@github(孟涛)

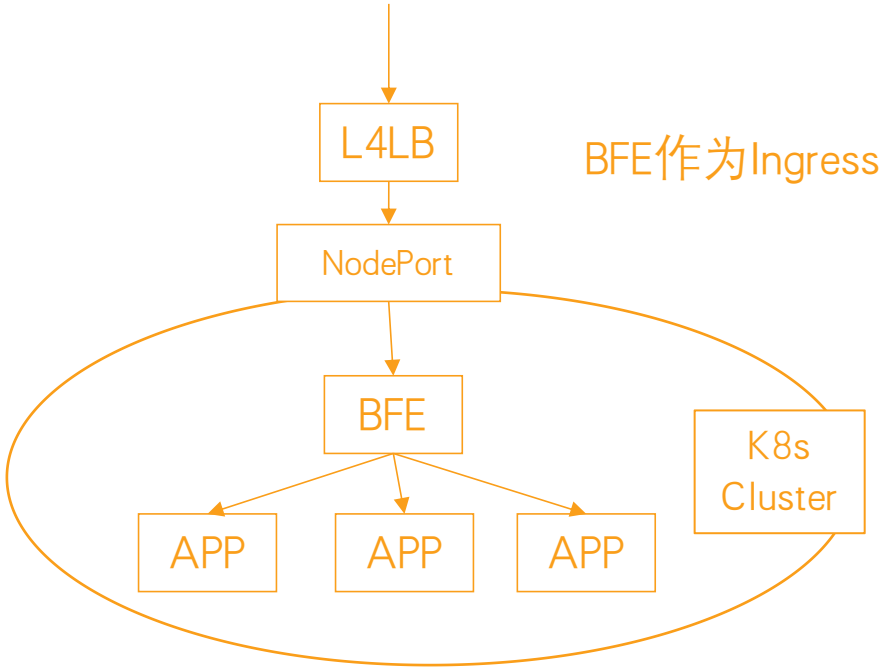
*作为一个开源新人，很荣幸能有机会参与到百度BFE Ingress 开源项目的URL重定向功能开发中。在这段过程中，**BFE团队重视软件工程的态度、严谨的代码编写规范与自动化单元测试与集成测试的实践，让我学到了很多受用的知识，收获颇丰。***

第五部分

BFE开源项目的后续计划

增强对Kubernetes的支持

场景1：在K8s集群之内

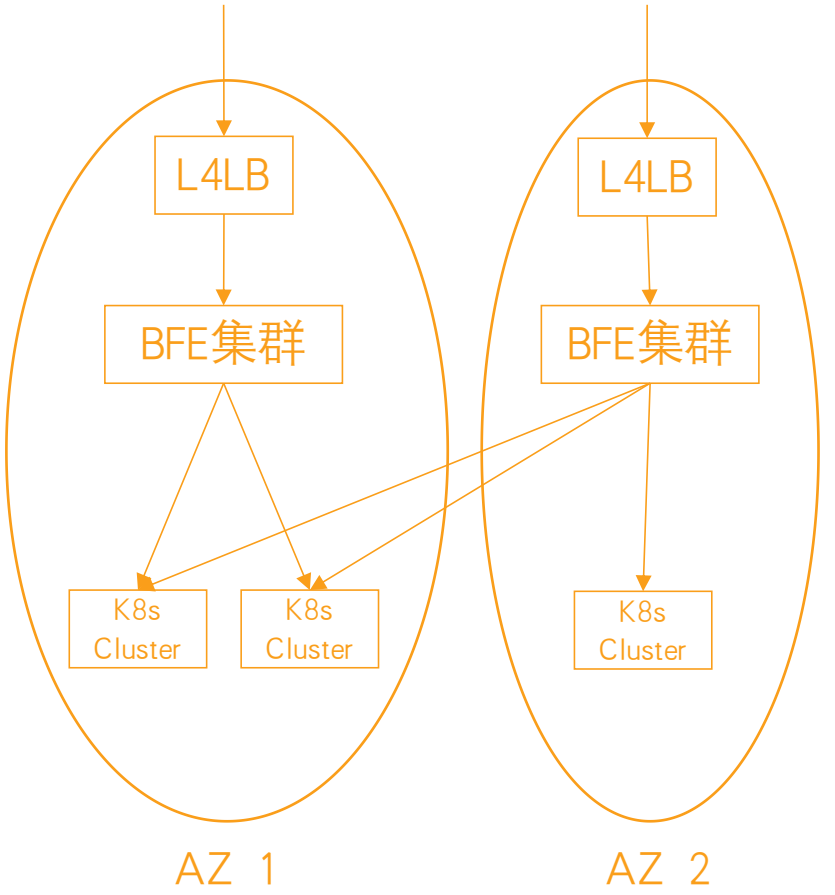


BFE作为Ingress

趋势：

- Ingress => Gateway API
- 增强安全、流量洞察能力

场景2：在K8s集群之外



BFE作为
全局负载
均衡器

关于瑛菲网络

- 2023年4月成立
- 聚焦研发为现代应用服务的下一代面向云原生的流量管理平台
- 基于 BFE 开源项目 构建大规模负载均衡核心引擎，覆盖数据中心全域流量管理、流量调度、安全等功能，包括完善的运维管理和数据报表能力

瑛菲网络

以人为本 心向自然

总结

- 负载均衡要为 **现代应用** 服务
- 传统的负载均衡技术已**无法满足**现代应用的需求
- 为了支持现代应用，需要 **新一代** 的技术，即面向 **云原生** 的 **流量管理平台**
- 相比Nginx，BFE在安全性、稳定性、研发效率、平台化支持等方面有明显优势
- 软件的开源是一种**社会活动**
- **产品化和专业化**是中国软件发展的必由之路

https://github.com/bfenetworks/bfe