

百度万人协同规模下的代码管理架构演进

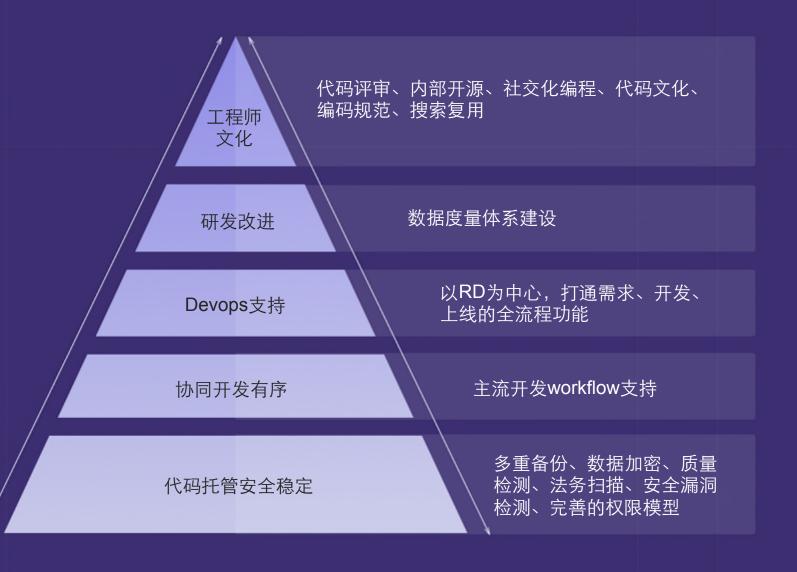
廖超超 百度工程效率部 2016.11.20



大纲

- >> 百度代码管理概况
- >> 百度代码管理遇到的挑战
- >> 百度代码管理架构演进的历程
 - 产品初创时期
 - 产品发展时期
 - 产品成熟时期
- >> 企业级SaaS服务下代码管理架构实践

百度代码管理概况



百度代码管理概况



百度研发工具链

主要功能

Request & Define Plan & Track

Code Hosting Change Request Build Test Deploy



项目管理平台

Retro



代码开发协作平台



持续交付平台

User Story Mapping
Product Backlog
Priority
Estimate
Card Board
Burn down

SCM
Feature flow
Code Analysis
Code Review

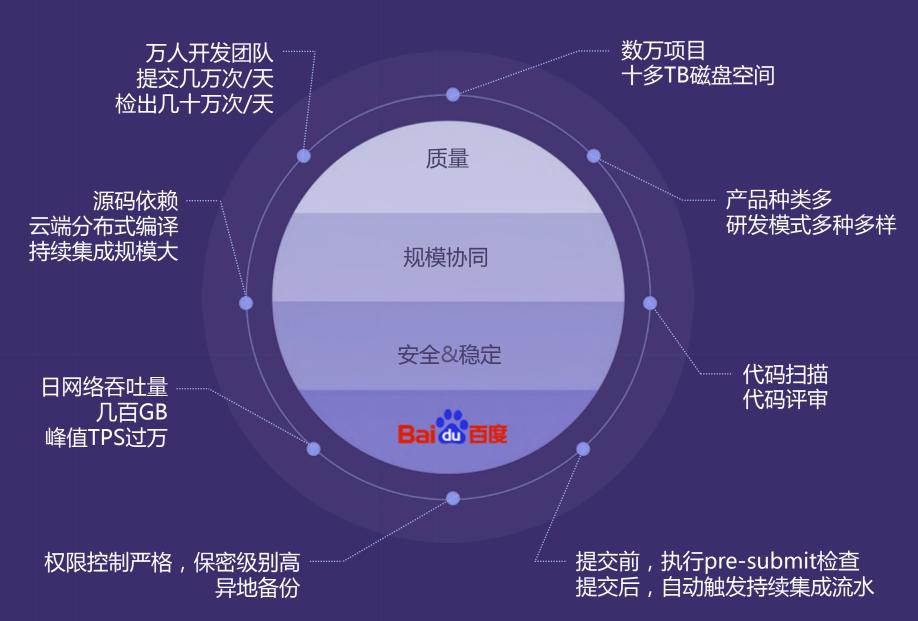
Compile
Unit Test
Auto Test
Daily Build
Deploy

工程实践

百度代码开发协作平台架构



百度代码管理遇到的挑战



百度代码管理架构演进的历程

1000 活跃用户 4000 Repositories 1TB 磁盘空间 3000 活跃用户 1.5万 Repositories

2.5TB 磁盘空间

8万 Repositories 12TB 磁盘空间

1万 活跃用户

企业级SaaS

产品成熟时期

- 产品发展时期
 - 数据分片
 - 智能路由
 - 负载均衡

- · 专属云
 - 弹性伸缩

容器化

• 自动化运维

- 产品初创时期
- 单机冷备
- 自动failover

- 读写分离
- 实时同步
- 三重备份
- 高可用

产品初创时期

产品初创时期,在代码库服务设计上,重点考虑了如下几个方面:

服务稳定性

• 代码服务,稳定是基石。

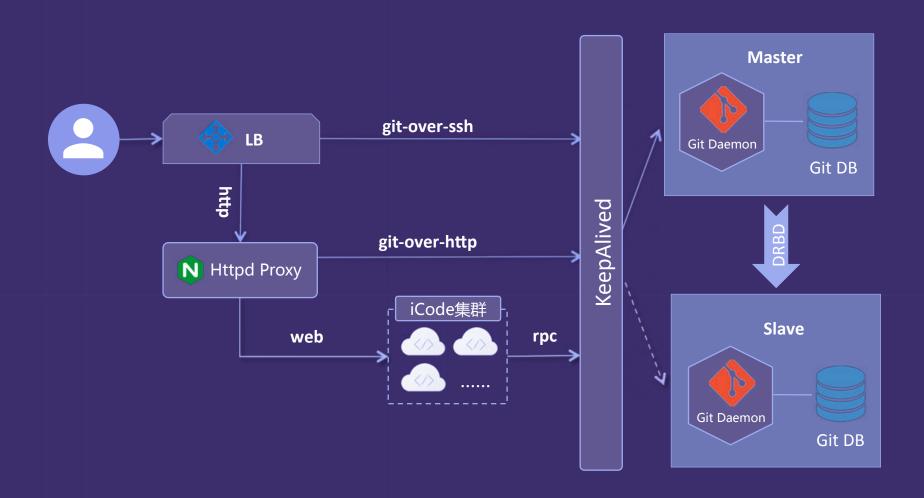
代码安全性

- 代码不丢, 异常情况下可恢复。
- 细粒度的代码访问控制。

代码质量

- 提交前进行自动化检查。
- 简便易用的代码评审。
- 提交后进行持续集成。

产品初创时期 — 架构



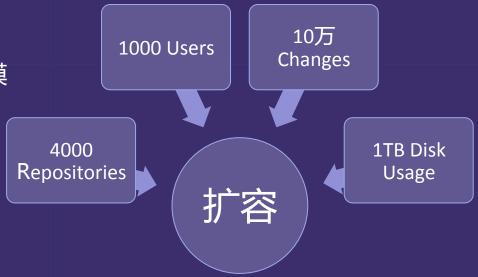
产品发展时期

在不断尝试优化单机性能的努力后,单机容量已经无法满足并发和吞吐量的需求。

发展初期:

(评估:半年内公司代码服务的规模

将是当前的15倍)



产品发展时期 — 可选方案

分布式存储 🗶

优点:

- 存储容量可以横向伸缩。
- 架构简单,计算资源与存储资源相互独立。
- 数据多重备份。

缺点:

- I/O性能大打折扣。
- · 需要优化缓存,甚至扩展Git底层存储方式。

数据分片 🗸

优点:

- 使用原生Git。
- 应用层控制灵活性较强,便于扩展。
- 使用本地存储 , I/O性能有保障。

缺点:

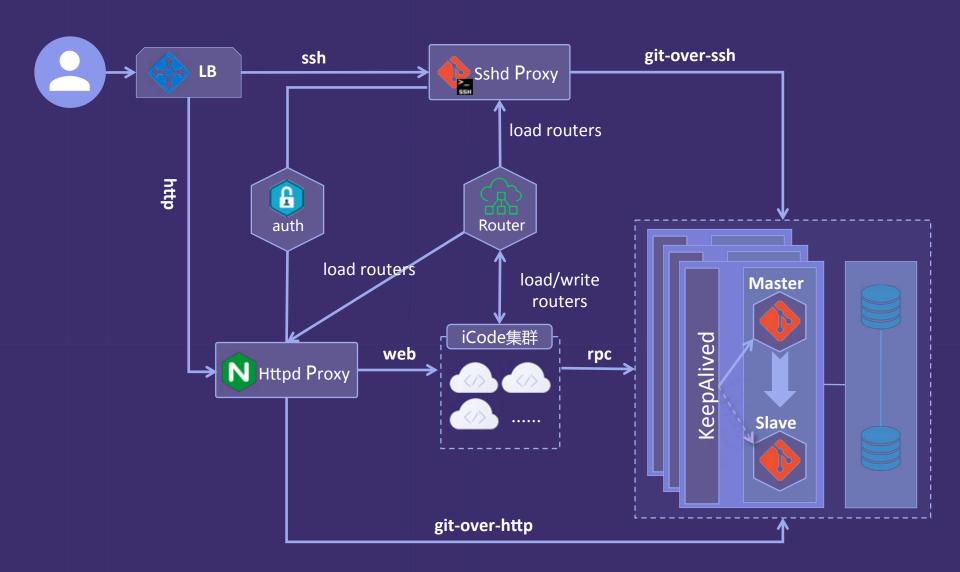
- 需要在应用层做路由,架构改变大。
- 跨分片的操作需要较高成本实现。

产品发展时期一可选方案

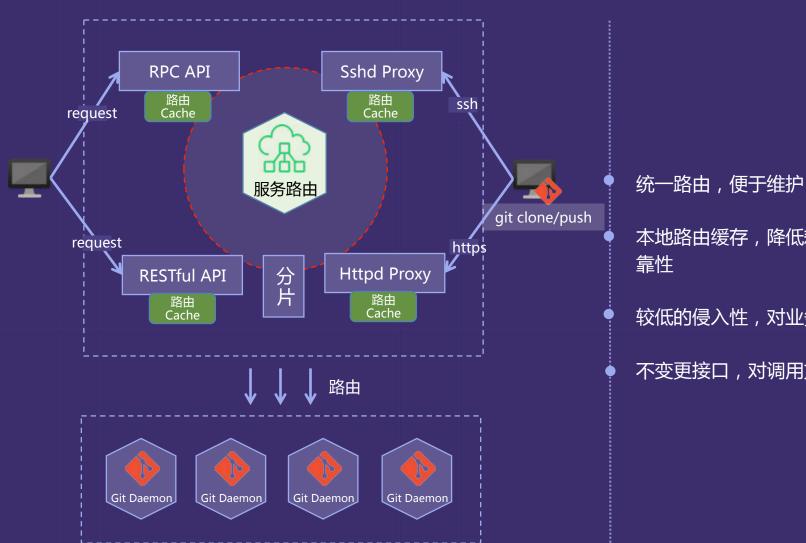
Distributed Storage Benchmarks:

	Local Disk	CCDB FS	Gluster FS
git clone 130M	13.405s	16.920s	14.873s
git fetch 28M	2.213s	4.846s	3.197s
git push 22M	10 s	116.833 s	99.002s

产品发展时期 — 架构



产品发展时期 — 路由设计



本地路由缓存,降低耦合,提升可

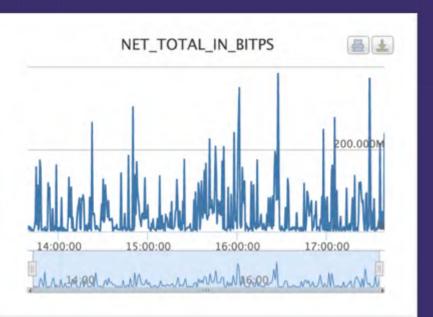
较低的侵入性,对业务代码透明

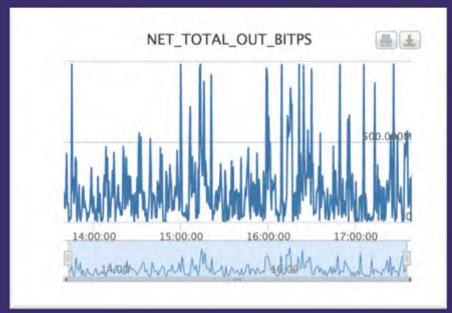
不变更接口,对调用方透明

产品成熟期

增长规模

- Git活跃用户数增长200%
- 代码库规模增长超过3倍
- 读超过30万次/天,写超2万次/天
- 高峰时段 , TPS近1000





单机网络流量情况(干兆网卡)



吞吐量不足,网络带宽成为瓶颈。

产品成熟期 — 可选方案

带宽扩容 🗸

• 千兆网卡换成万兆网卡。

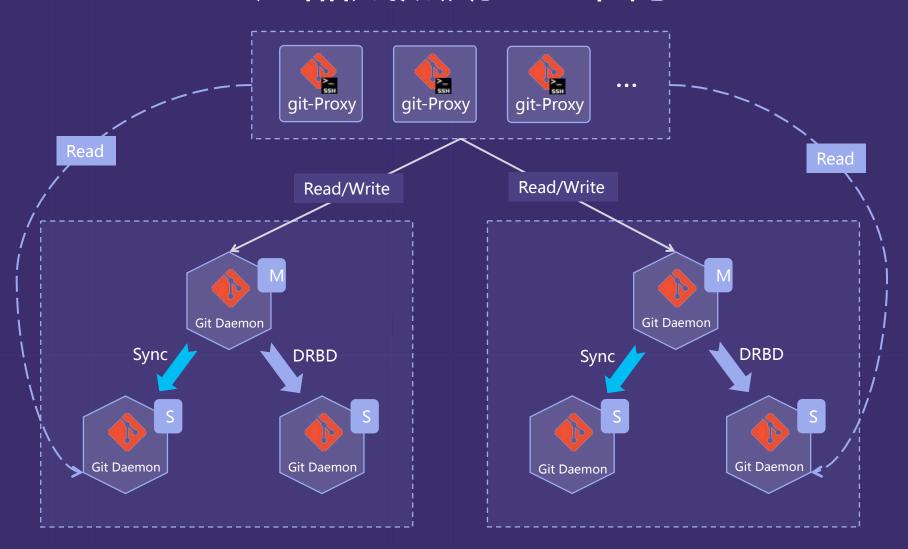
机器扩容 🗸

- 拆分更小的实例(带宽压力分摊)
- 有效利用备份机器(冷备机器闲置)
- 增加每组实例的只读节点数(读请求远大于写)

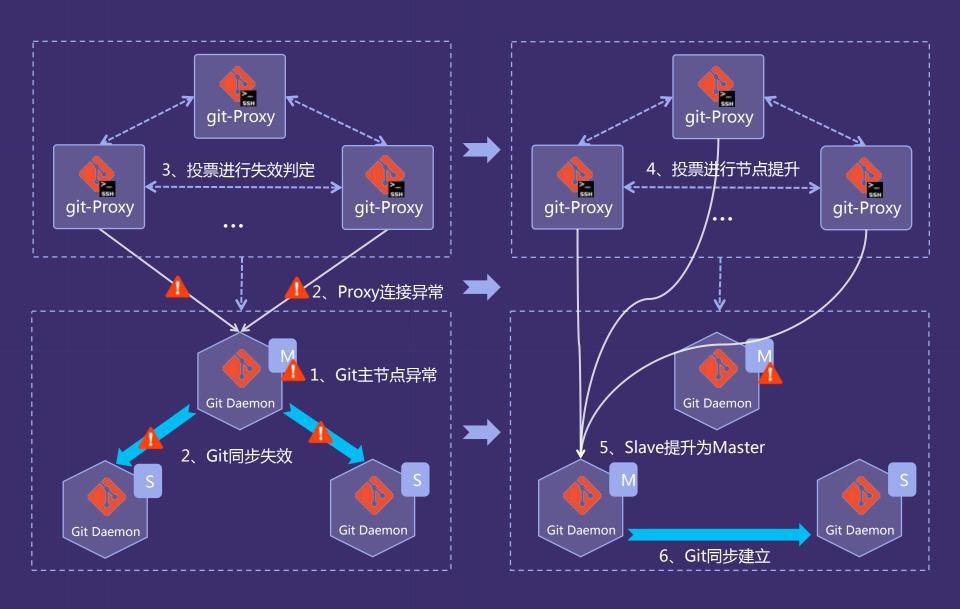
数据压缩 🗶

• Git已经对原始文件压缩,再压缩的效果不明显,而且耗CPU资源

产品成熟期 — 架构

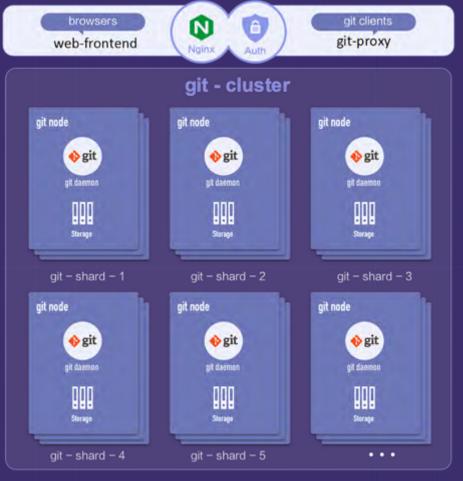


产品成熟期 — 架构优化

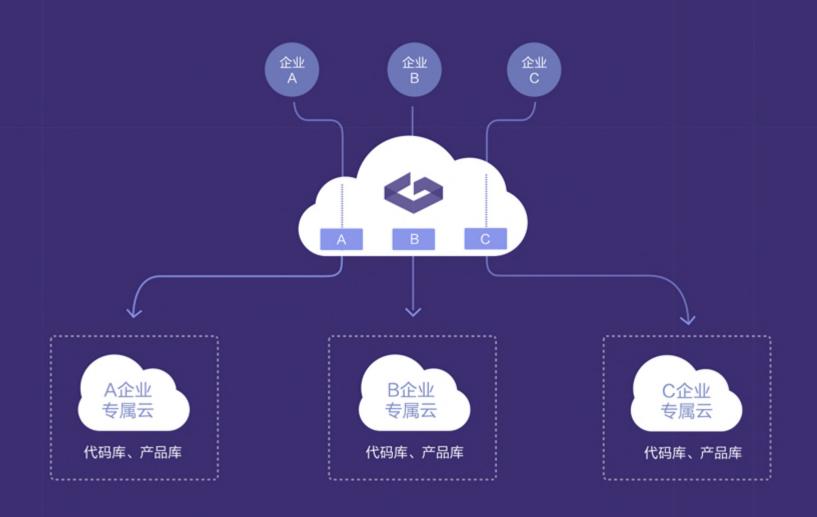


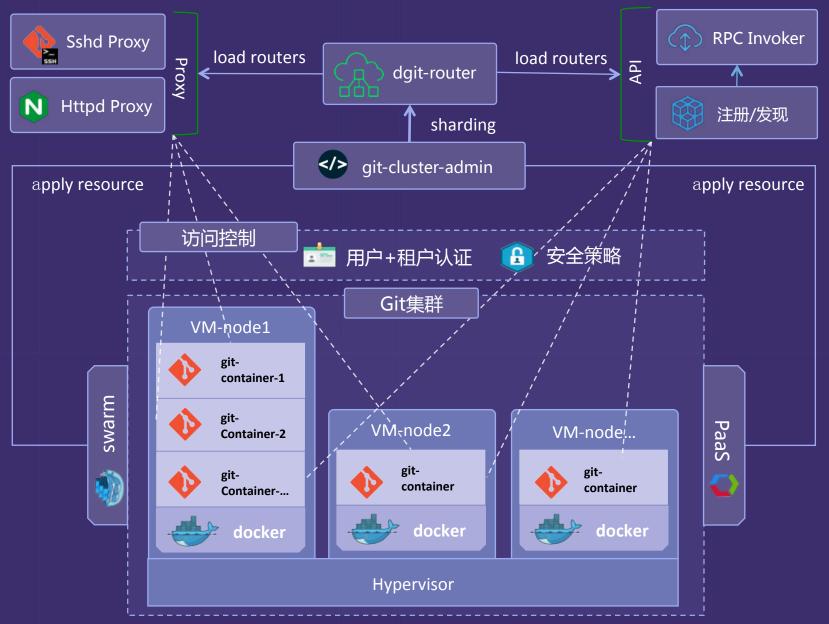
百度效率云是为企业客户打造的新一代软件研发SaaS解决方案,代码管理是软件研发的核心环节之一。企业级SaaS服务模式给代码管理提出了新的挑战。

安全性	• 确保企业代码资产不泄漏、不丢失 	
容量&性能	• 更多的用户意味着更多的Repositories和更高的并发	
弹性伸缩	• 企业不同发展阶段,对代码库容量和性能的需求不同	
自动化运维	支持企业快速接入方便大规模集群管理,提升系统稳定性	



- 更安全 Proxy集成安全认证
- 服务容量 基于repository分片,支持动态扩容
- 高可用
 Git Server三重备份,数据永不丢失
 Git Server服务Failover, Proxy负载均衡
 支持客户本地备份
- 易运维 集群监控、资源管理 提供可视化的dashboard
- 支持专属云,物理隔离







咨询邮箱: xiaolvyun@baidu.com

QQ群: 317759963

门户地址: https://xiaolvyun.baidu.com