QCon
全球软件开发大会

喜马拉雅百亿级API网关
——南天门架构演进

平台架构—彭荣新

（排名不分先后）

“我们在 DIVE 全球基础软件创新大会上等你”

DIVE

全球基础软件创新大会

议/题/提/交

大/会/官/网

深入基础软件，打造新型数字底座

2021.11.26-27 / 北京·悠唐皇冠假日酒店

目录

喜马为啥要做网关

喜马用户数增长达到6亿的级别

Web服务个数达到600+

1、通用技术升级推广难?

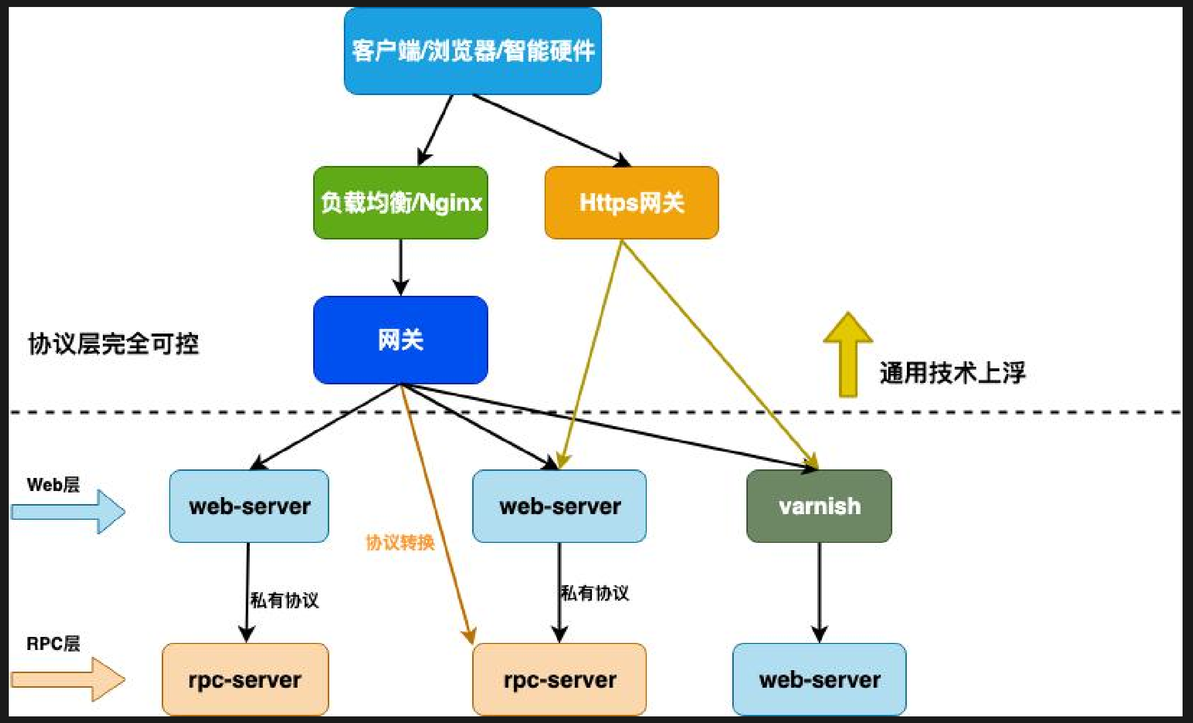
登录校验、安全认证、WAF....

2、统一 ， 灵活，个性化的流量管控

流控、降级、调度、监控、报警 .....

3、统一标准和规范

协议标准，API标准.....

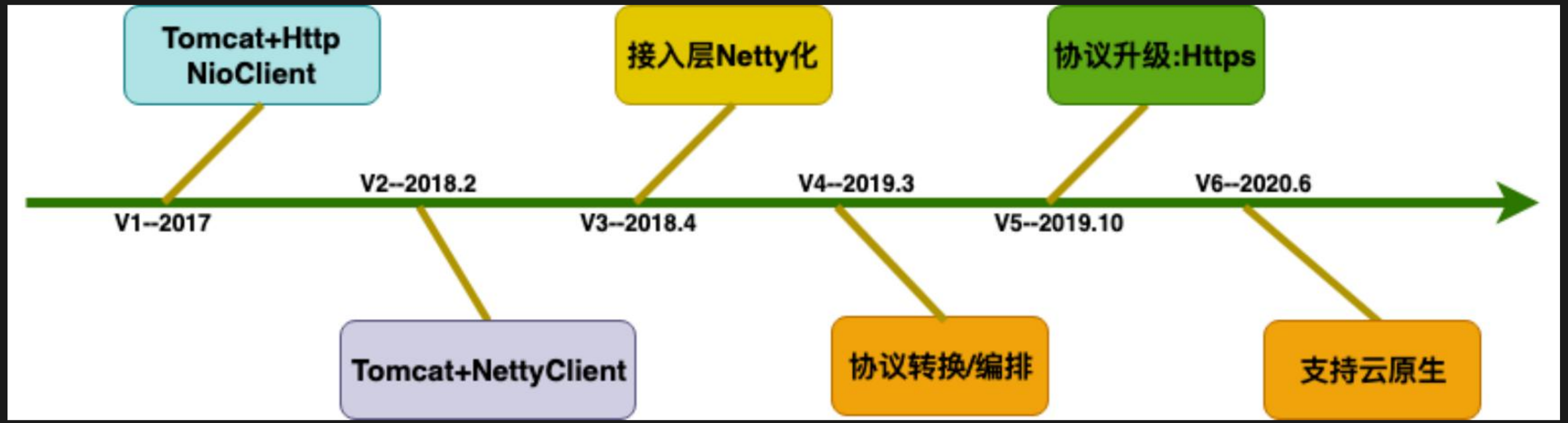


喜马拉雅南天门网关—Southgate

自研：

- 自主完全可控

版本演化史：



Southgate 能力概览

- 基础能力
- 业务统一接入能力
- 监控报警

V1:四层架构:Tomcat NIO+HttpNioClient

接入层 Tomcat

- Tomcat nio

通用逻辑层

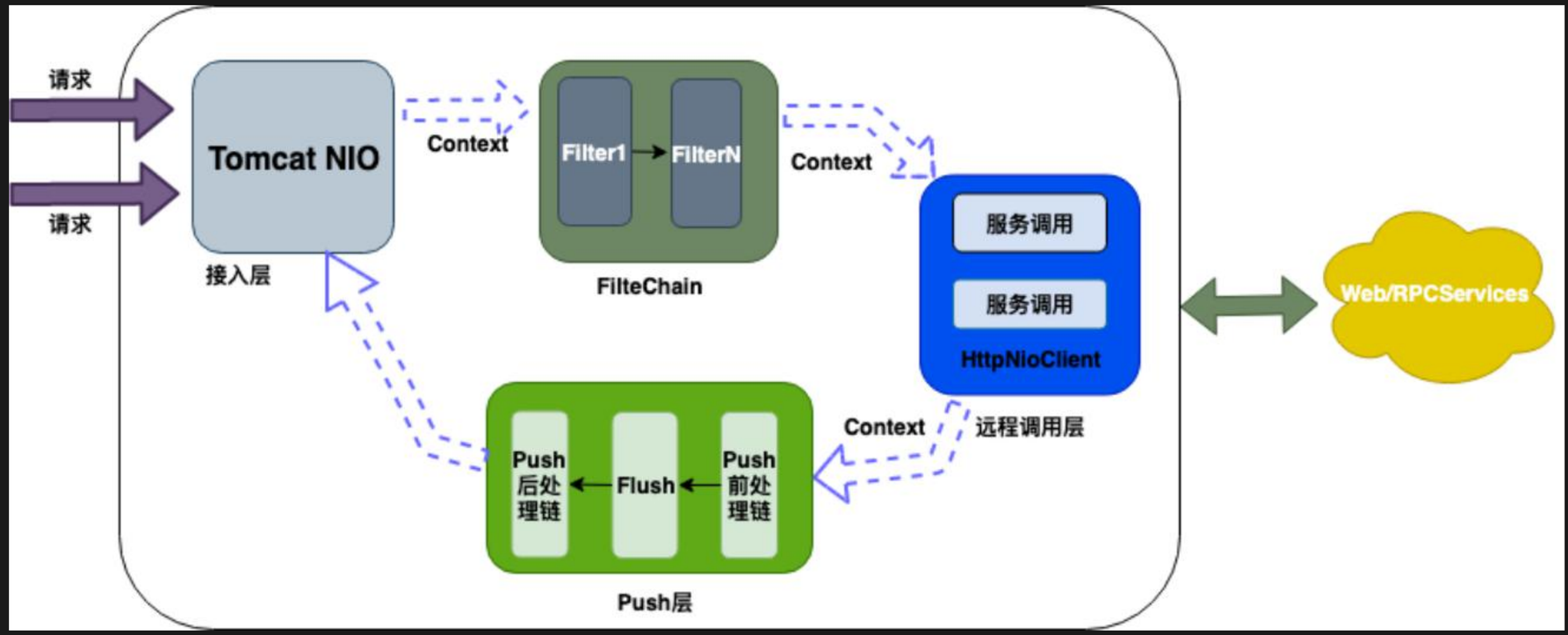
- FilterChain模式
- Servlet3.0实现异步
- Catalina work Pool

服务调用层

- 请求/响应异步
- Nio Thread Pool

Push层

- Push Thread Pool



V1: Copy And Block

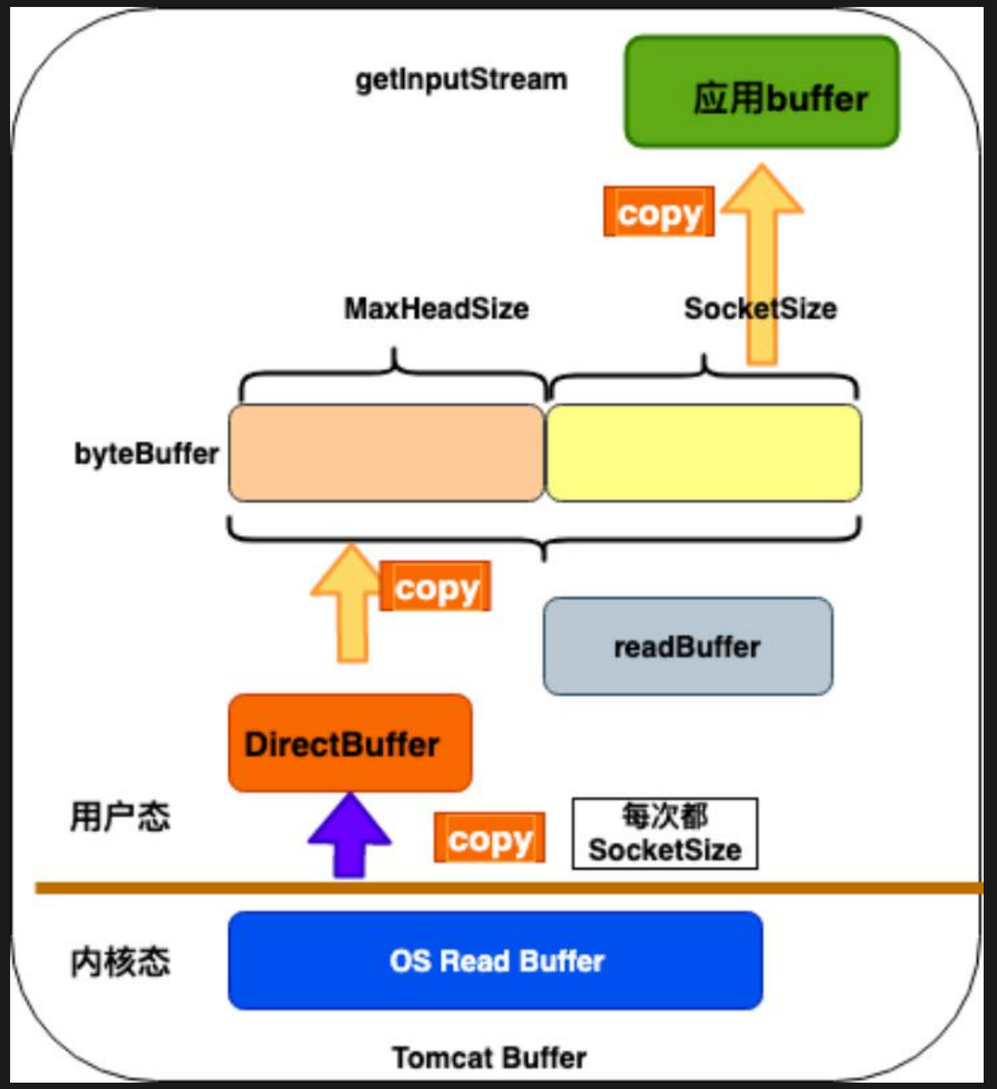
默认堆Buffer

读3次Copy

- Os —>DirectBuffer
- DirectBuffer—>ByteBuffer
- ByteBuffer—>应用Buffer

读body阻塞

- Block catalina work 线程
- keep alive timeout 默认20秒



V3:四层架构:Netty全异步

接入层：NettyServer

- EPoll Event Loop
- EPoll ET

通用逻辑层：

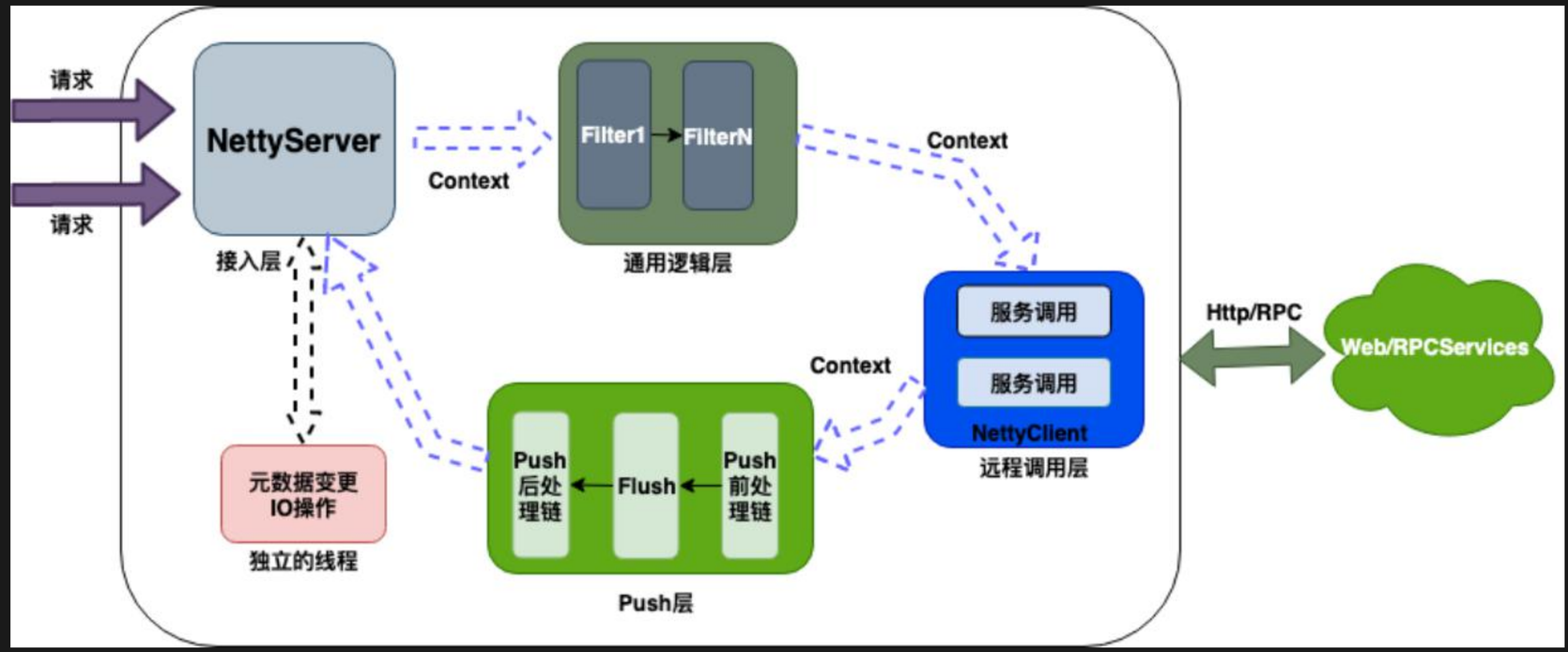
- Worker Thread Pool

调用层: NettyClient

- 连接获取和释放异步化
- 无锁操作

元数据变更

- 独立的线程池
- 有IO操作



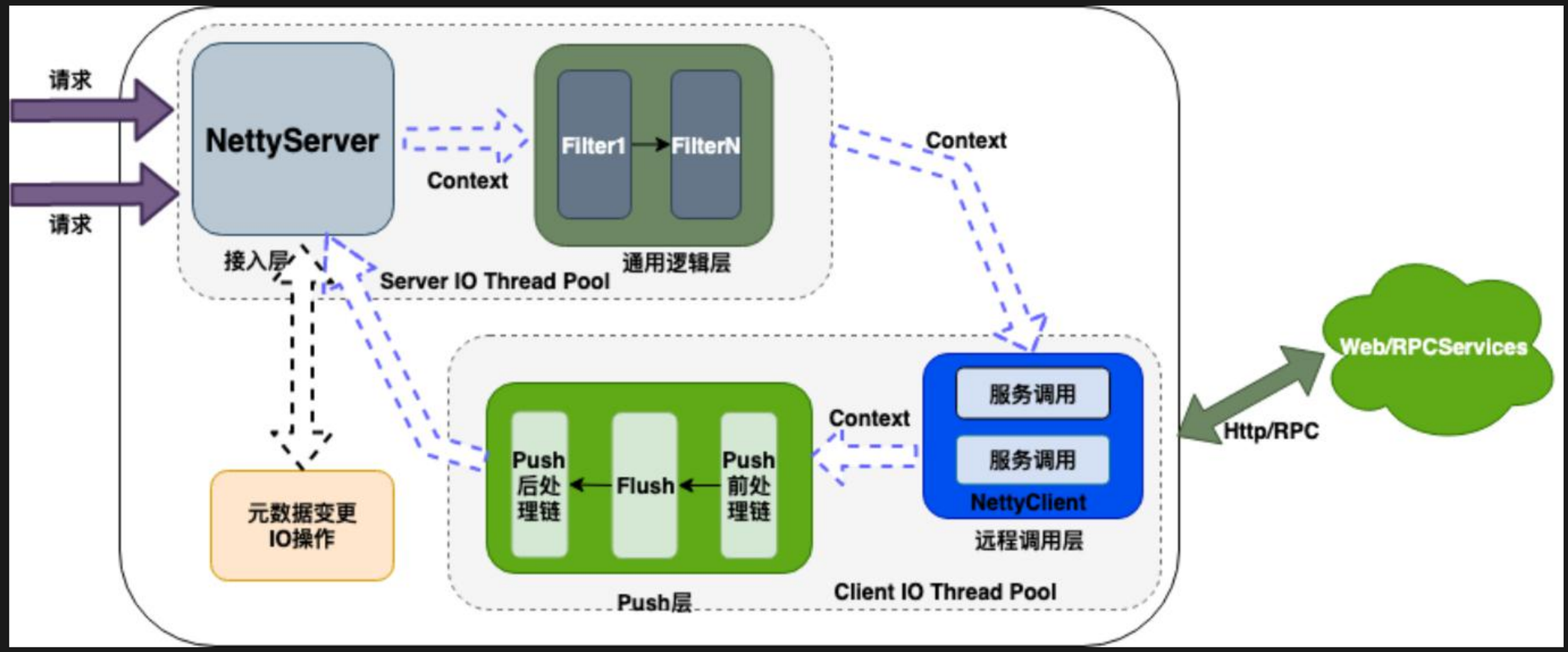
V3:Context Switch

线程模型合并

- Server IO 线程
  - 编解码+逻辑处理
- Client IO 线程
  - 连接获取/释放
  - 编解码+Push

共用条件

- 不能有IO操作
- 刷日志要异步/批量



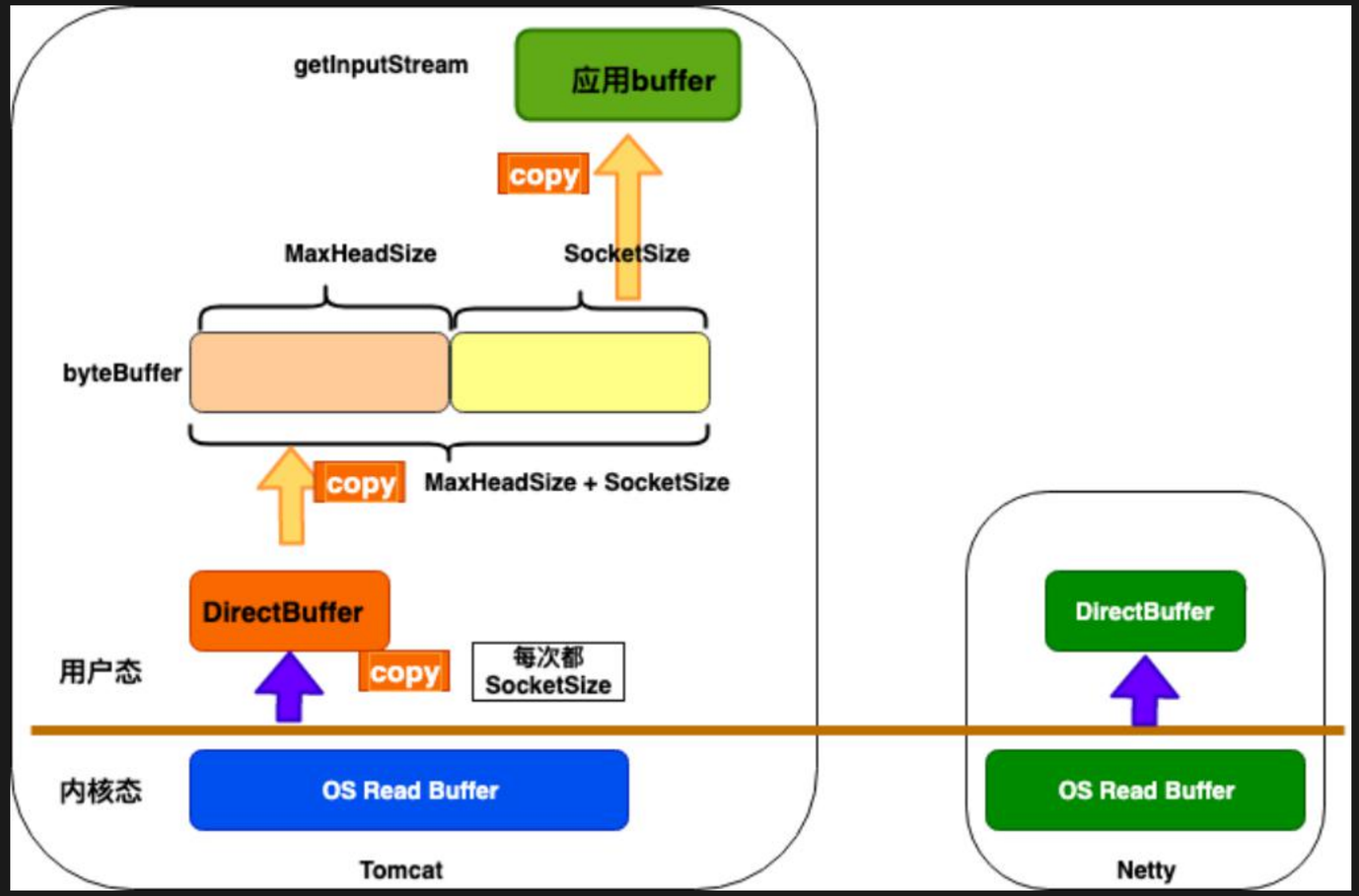
V3: 数据Copy

数据包

- os—>DirectBuffer
- 1次读1k，自动适配

Body

- 不解析body，没有Copy



链接管理

链接模式

- HTTP1.1 一个请求独占一个连接
- RPC 一个连接

长连接

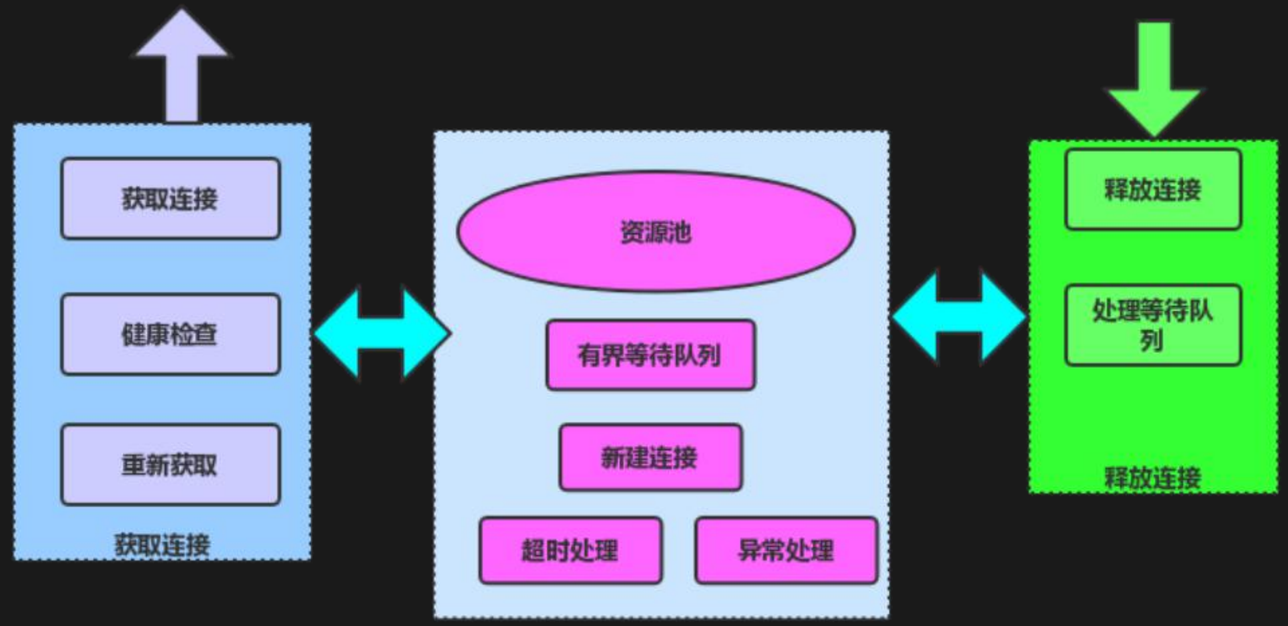
- 重用次数被限制
  - Tomcat 默认100次
  - 控制Connection:close

连接预热

- 高并发的应用

关闭链接

- Connection:close
- 空闲写超时，关闭链接
- 读超时关闭链接
- Fin,Reset



容灾降级

快速失败

- 超时机制
- 智能熔断
  - 自动熔断
  - 自动恢复
- 有界队列
  - 链接池等待队列
  - Netty io event queue

自动降级

- 自动重试
  - 获取连接失败，最多重试3次
  - 发送失败，根据异常重试
- 自动下线
  - 定时健康检查
  - 连接拒绝

流量防护

接入层

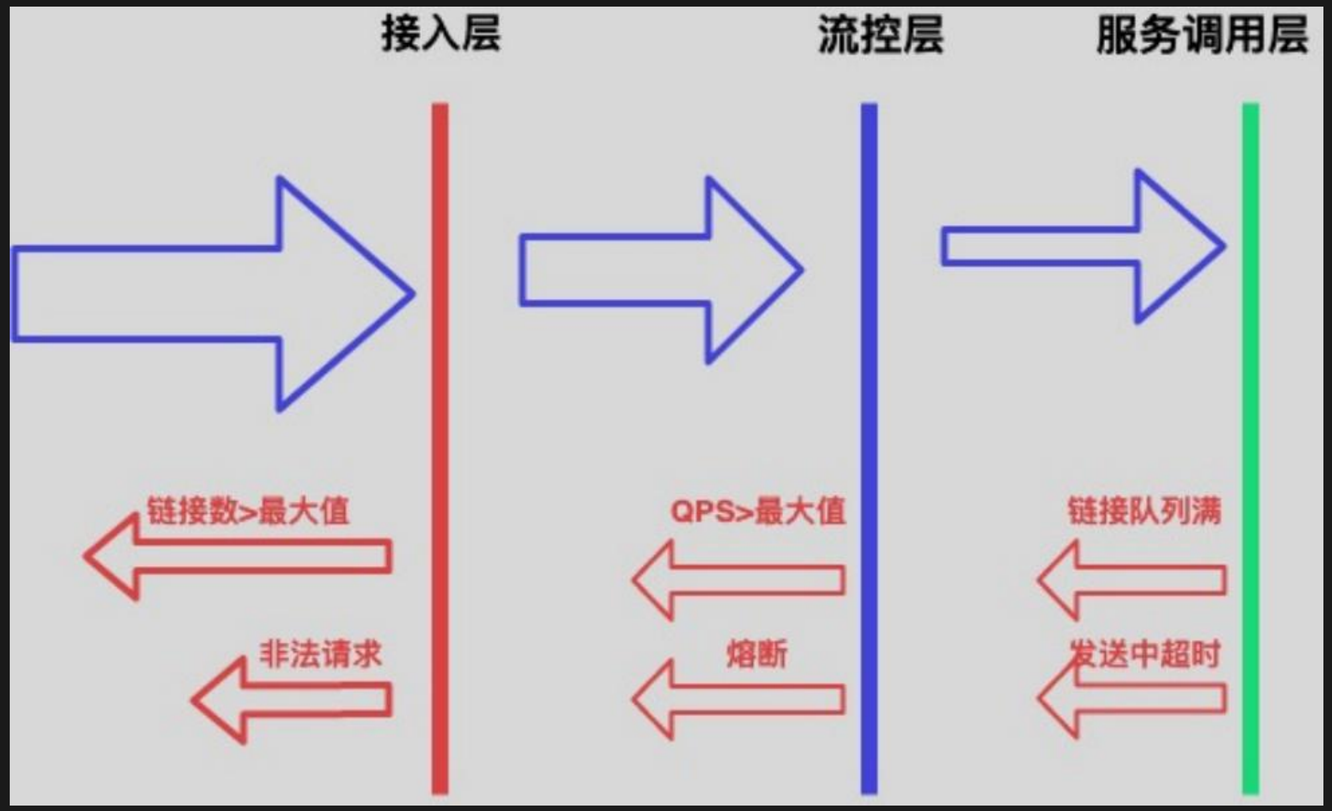
- 最大链接数限制 —全局
- 非法请求快速拦截
- 大带宽请求自动拦截

业务逻辑层

- 流控自动削峰填谷—令牌桶算法
- 统一鉴权
- 个性化的黑白名单

服务调用层

- 超时机制
- 连接数限制



协议安全Https

服务端

- Openssl　性能高
- ALPN
  - HTTP1.0
  - HTTP2.0 可配
- Session Ticket
  - 客户端存储
  - 集群统一部署
  - 定时更换

客户端加速

- OCSP Stapling
  - 优点
    - 服务端返回证书查询结果
    - 减少客户端下载证书的时间
  - 缺点
    - 增加带宽

流量调度

调度模式

- 独占模式
- 共享模式
- 流量Copy
- 流量转发
- 跨机房

调度规则

- 接口
- Tag
  - Header+Cookie
- Params

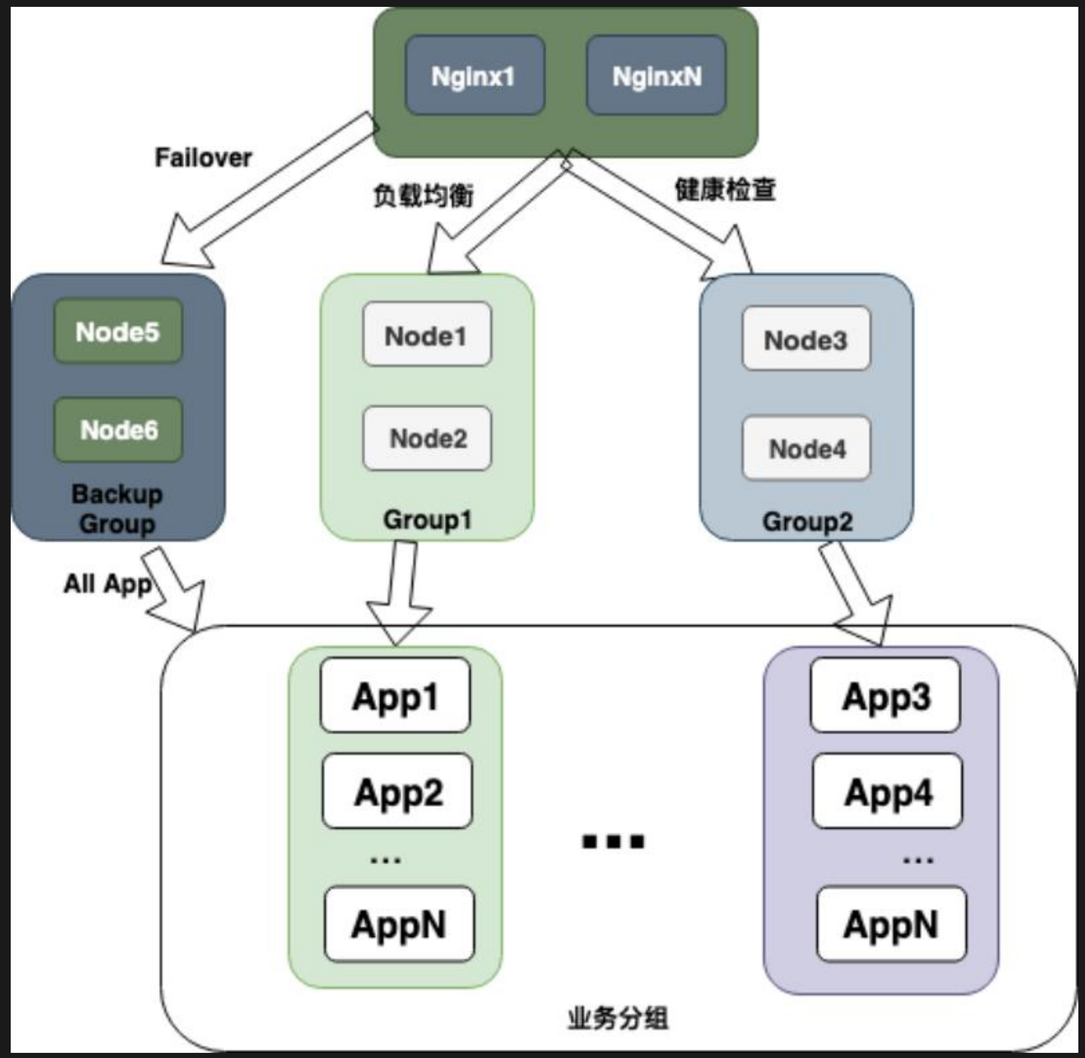
高可用保障

分组部署

- 集群部署
- 业务线物理隔离
- 每个分组只加载该分组的应用
- 限制openfile数量

健康检查

- Nginx侧健康检查
- 网关可以无损下线
- 独立的backup
  - 网关故障，backup分组生效



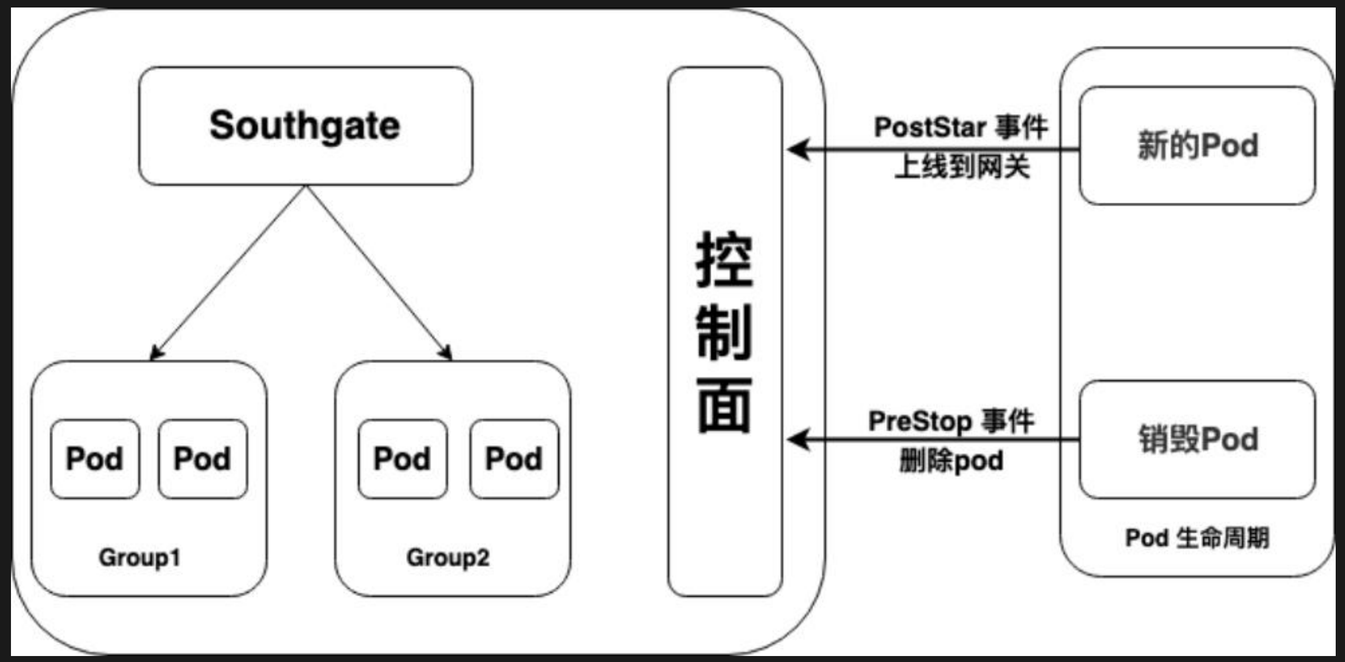
云原生发布

先上线

- Pod hook PostStar通知网关
- 网关做healthcheck
- 流量慢启动

后下线

- Pod hook 通知网关
- 网关下线路由
- 删除对应的pod



网关生态—运行时无强依赖

元数据存储

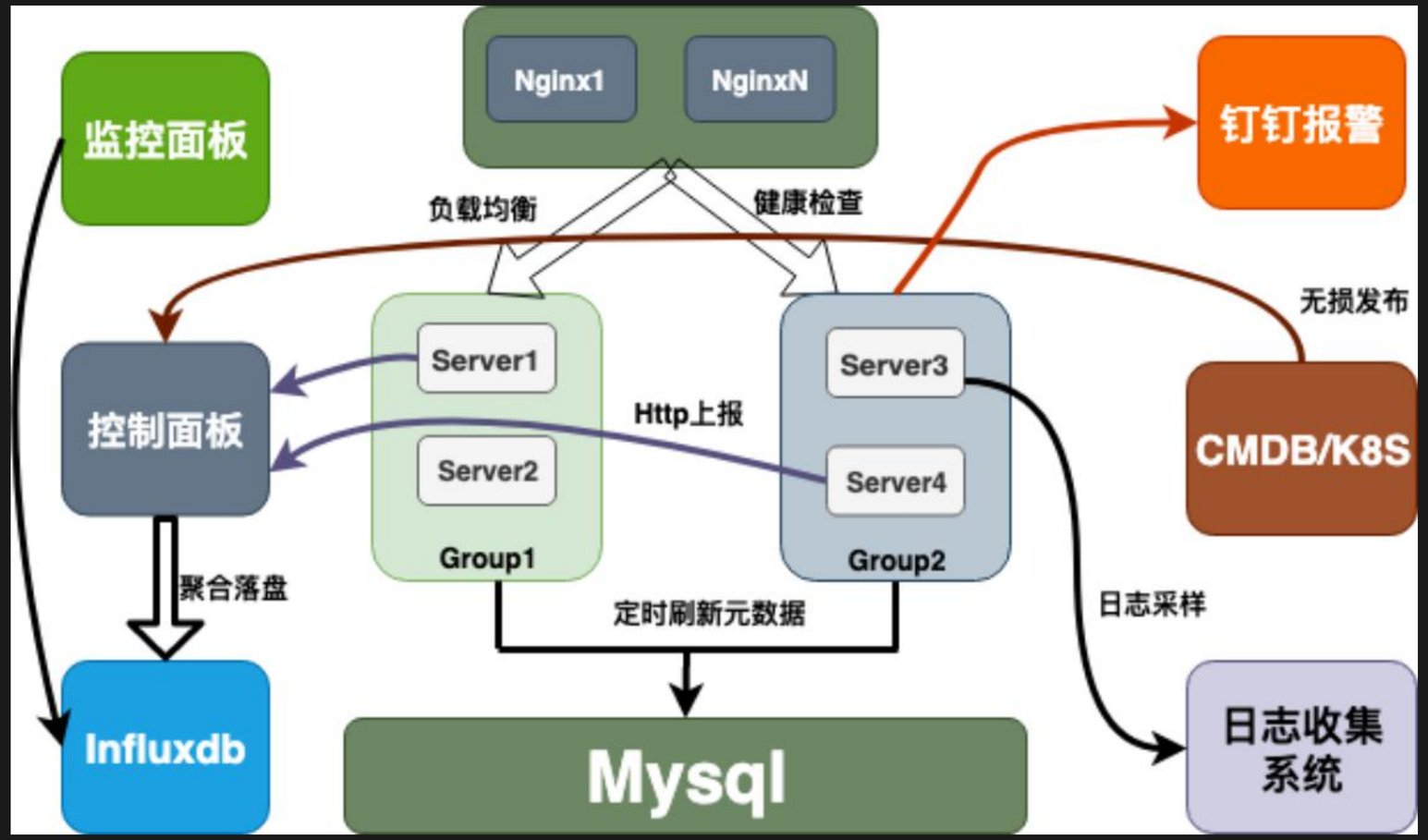
- MySql

监控Metrics

- 定时HTTP上报
- Influxdb 存储
- 秒级监控

钉钉报警

- 网关Metrics收集报警
- 控制面板定时报警



性能优化实践

线程池

- 尽量共享
- 减少线程池切换的开销

对象池

- Request Context And Event
- Thread Pool Task
- String Buffer Global

Netty内存池

- numDirectArenas 越大，分配内存锁竞争越小

Tomcat

- 连接队列大小 acceptCount 默认100
- **keepAliveRequest 默认100**

踩过的一些坑—Netty

ByteBuf 释放

- 写操作自动释放，异步处理记得retain
- 读操作，不走后面的piple，需要主动释放
- 关闭主动检查
- 一定要监控内存池的引用和释放次数

PoolThreadCache

- **默认启用,最好关闭**
- **申请和释放一一对应**

Connection:close

- **一定要关闭链接**

EventLoop TaskQueue

- **默认是Integer Max,需要手动设置下**

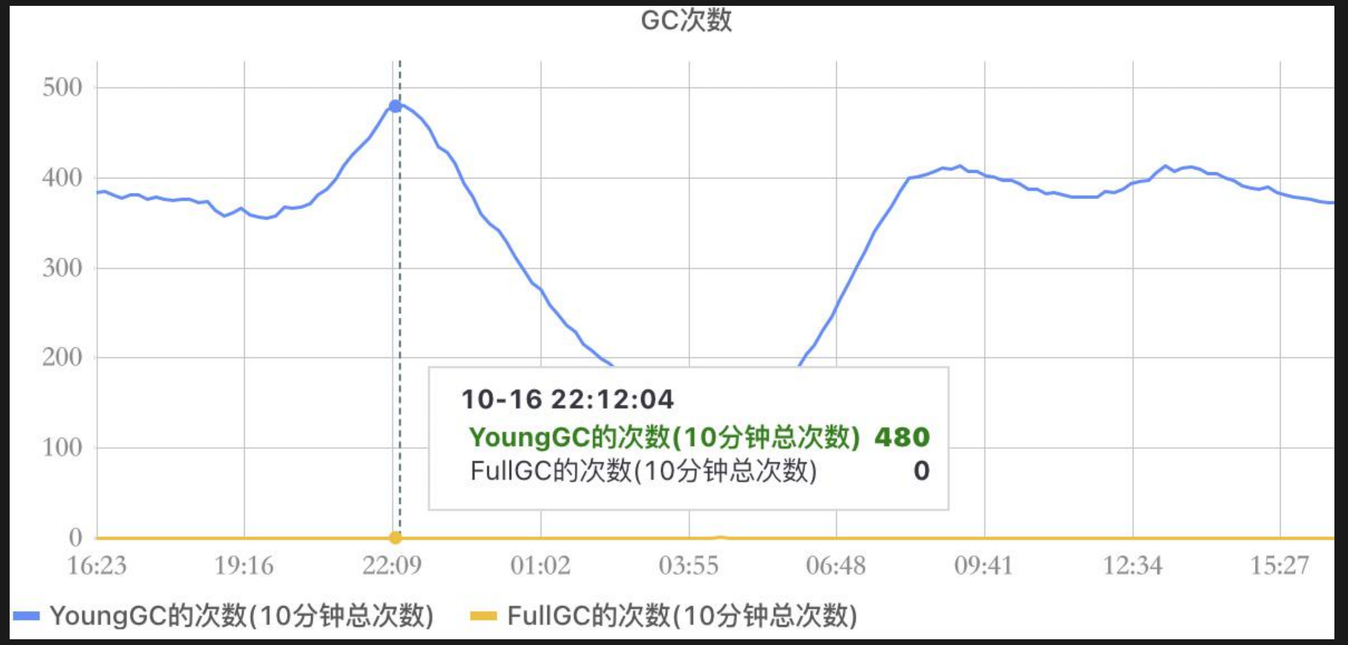
运行指标

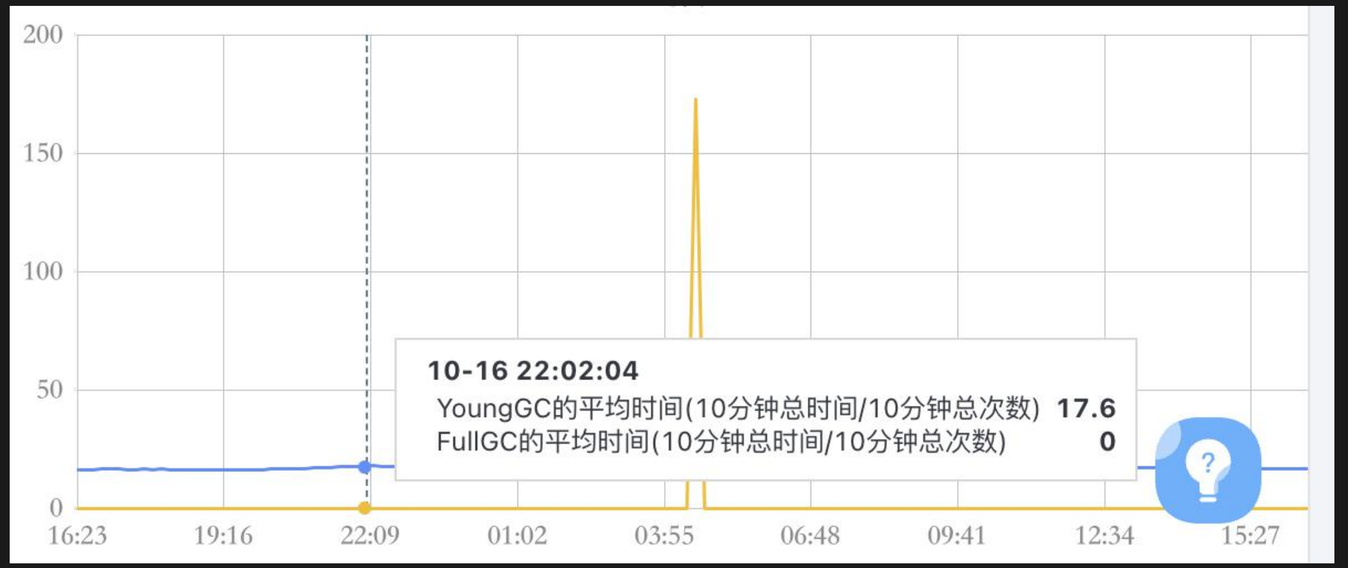
流量

- 日访问300亿+次
- 单机高峰2.5w+

GC

- Minnor gc 1秒一次
- Gc 耗时 13—18ms 总体20ms以下
- 每天低峰主动full gc





未来规划

- 容器化服务发现
- 流量网关和API网关统一，节省成本
- 支持自定义插件

QCon+ 案例研习社

扫码学习大厂案例

学习前沿案例，向行业领先迈进

40 个
热门专题
—
行业专家把关内容筹备，
助你快速掌握最新技术发展趋势

200 个
实战案例
—
了解大厂前沿实战案例，
为 200 个真问题找到最优解

40 场
直播答疑
—
40 位技术大咖，每周分享最新
技术认知，互动答疑

365 天
持续学习
—
视频结合配套 PPT
畅学 365 天

THANKS

《喜马拉雅百亿级 API 网关南天门架构演进》

扫描二维码，提交议题反馈，