QingStor™ 对象存储架构设计及产品演进

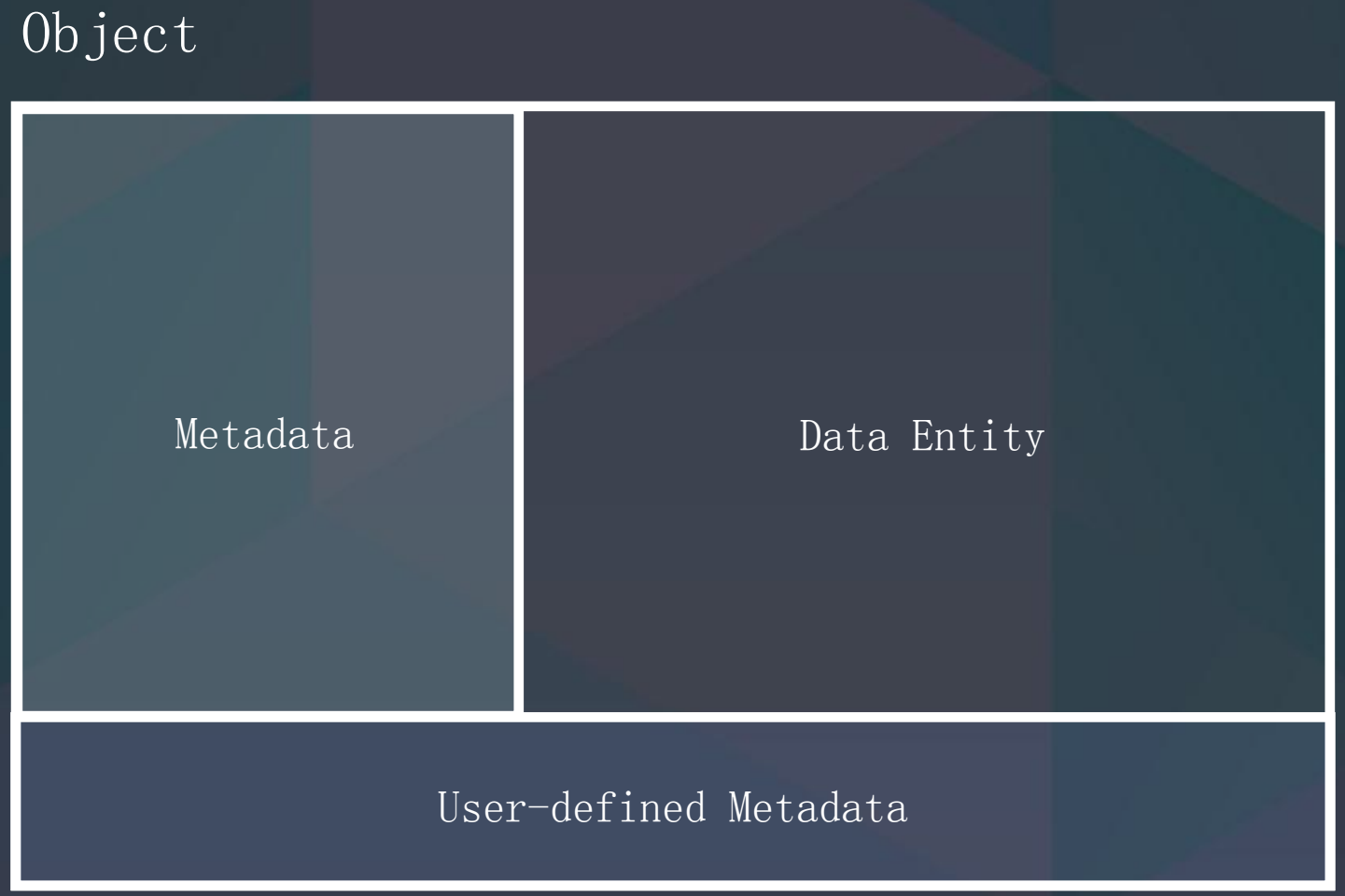
Osier Yang

目录

- QingStor 对象存储产品及架构
- QingStor 对象存储产品体系

术语介绍

- Zone - 区域
- Bucket - 存储空间
- Object - 对象



产品核心定位

通用

海量

非结构化

平台

无区域 or 多区域？

- Nearby Computing Resource
- Hybrid Cloud
- On-premise Cloud
- 用户根据业务决策地理位置

Why RESTful?

- 非结构化数据的非易变性
- HEAD/PUT/GET/DELETE/POST
- 数据的流转

Local Filesystem or Block Device?

- 单机时代设计
- 信息冗余
- 空间浪费
- No O_ATOMIC

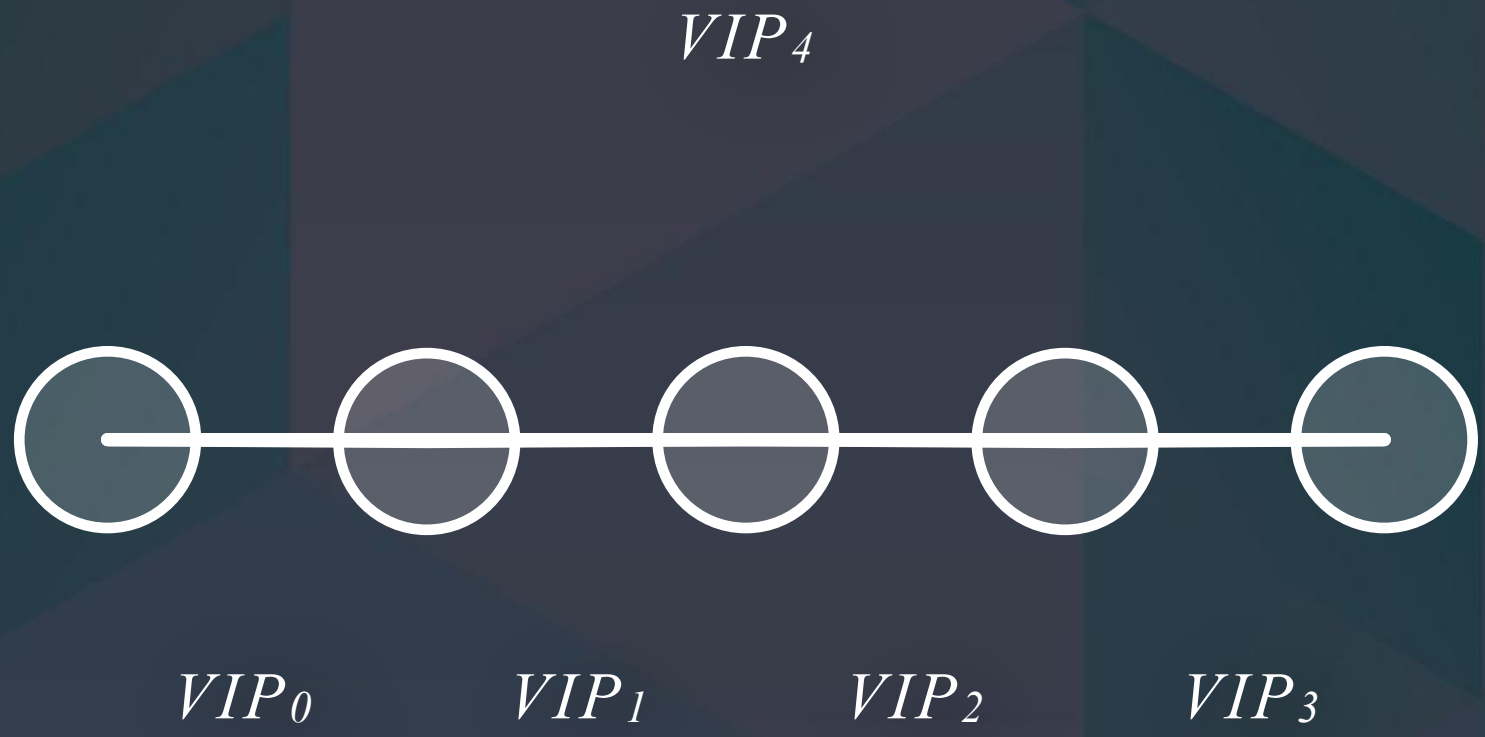
- Stable

- 开发成本

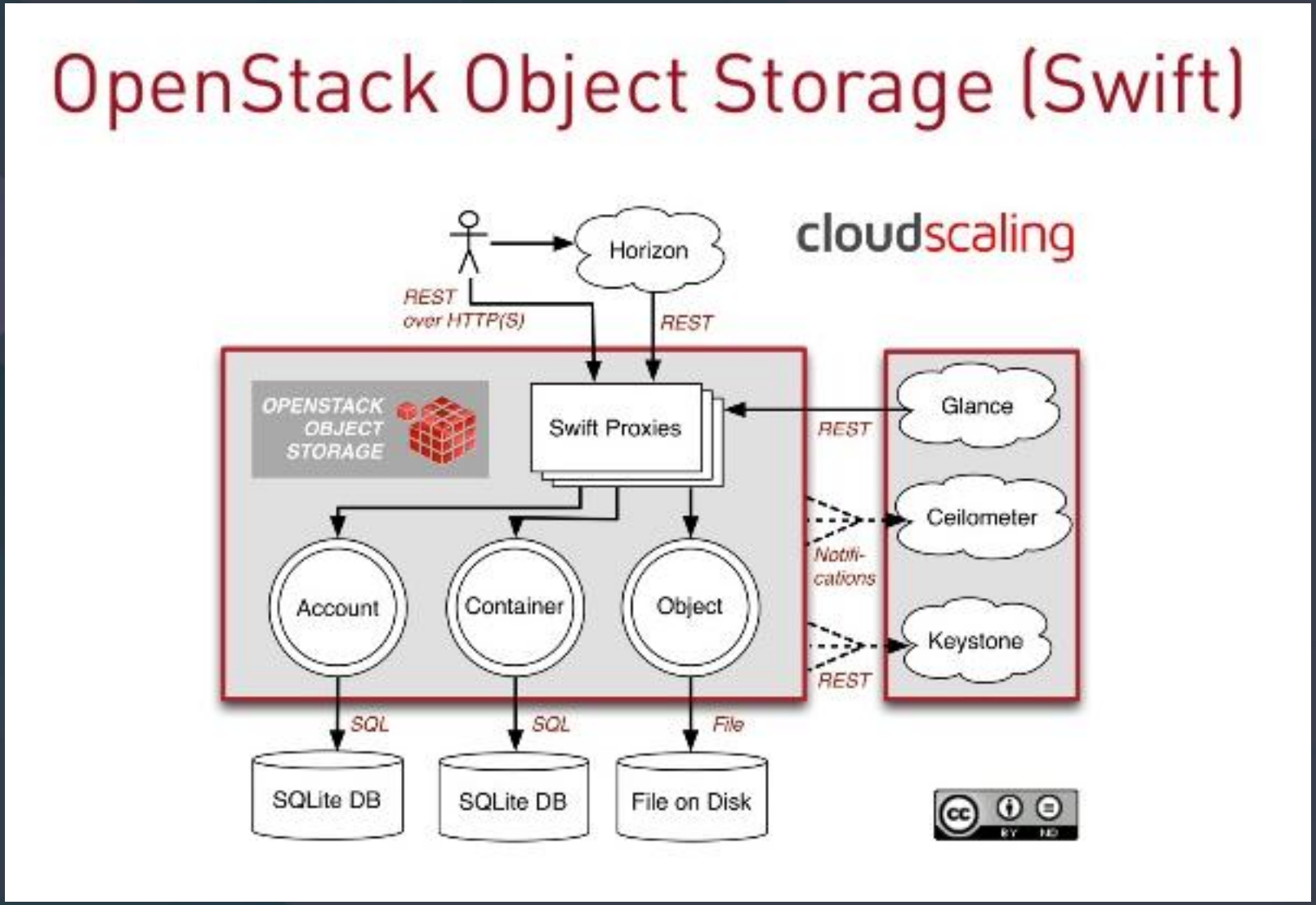
- Bluestore ?

Load balancing

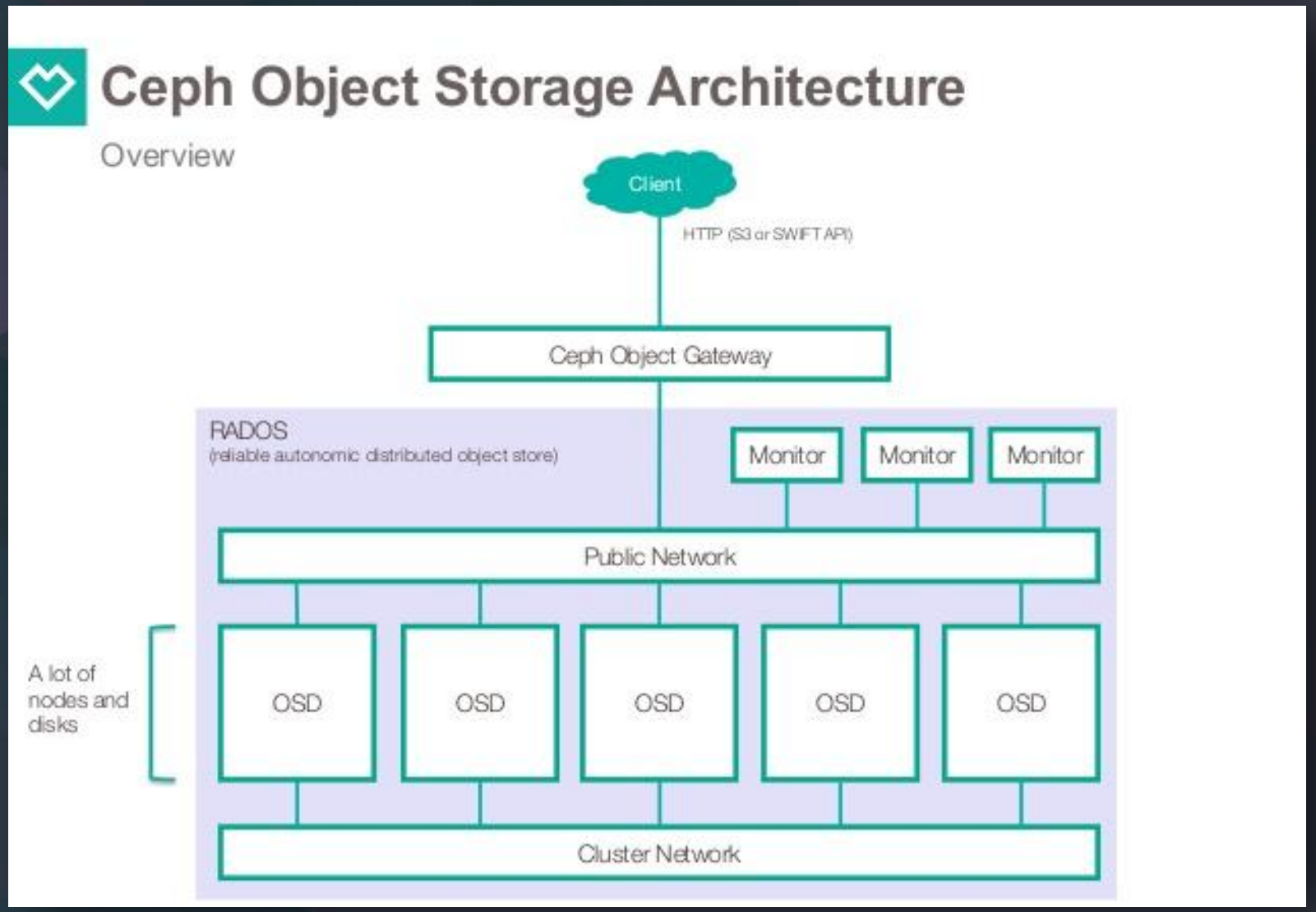
- 控制流和数据不可分离
- HAProxy
- LVS
- Maglev
- DNS Round Robin + 环形链路 VIP



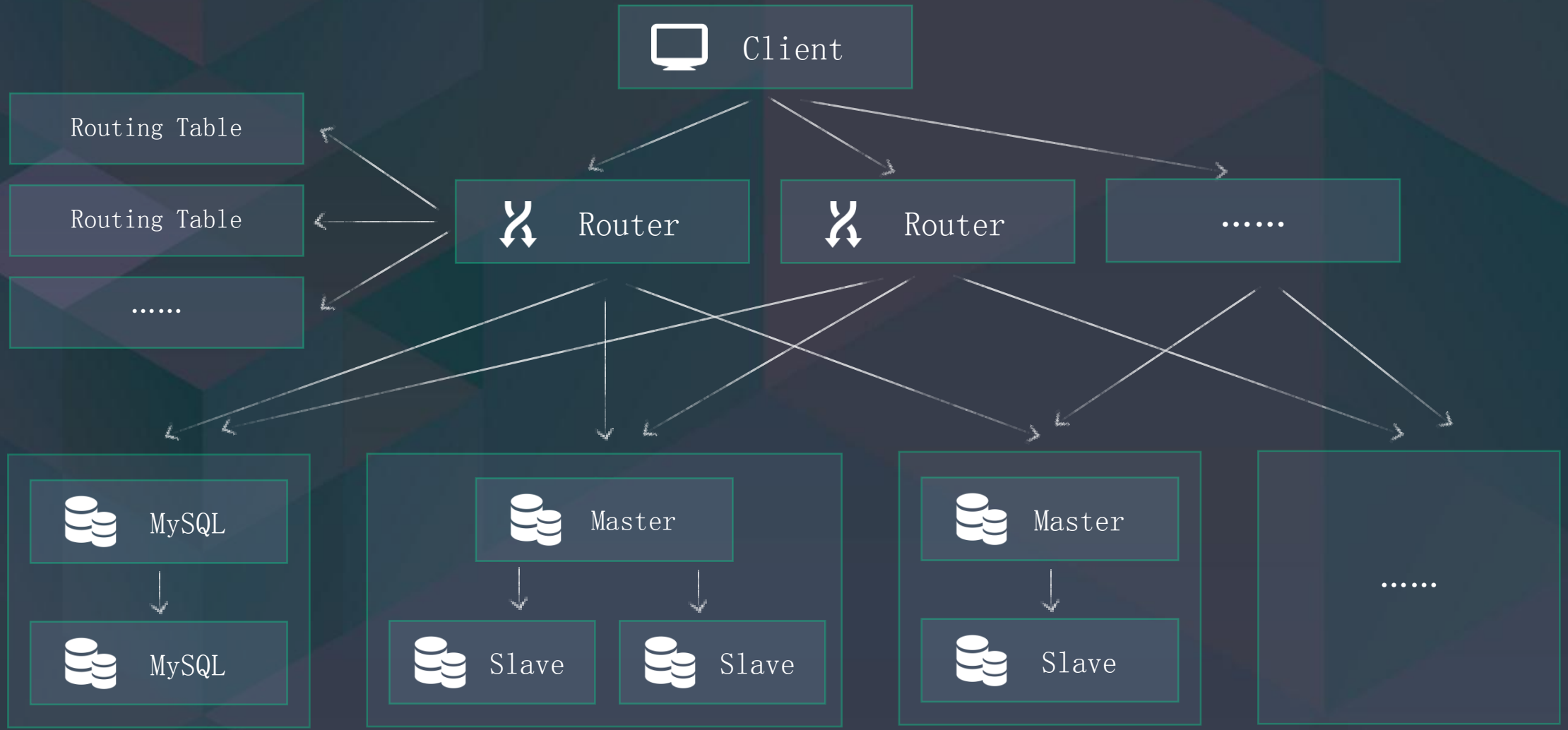
独立索引层？


OpenStack Object Storage (Swift)
cloudscaling
Horizon
REST
over HTTP(S)
REST
OPENSTACK
OBJECT
STORAGE
Swift Proxies
REST
Glance
Ceilometer
Notifi-
cations
Keystone
REST
Account
Container
Object
SQL
SQL
File
SQLite DB
SQLite DB
File on Disk

独立索引层？


Ceph Object Storage Architecture
Overview
Client
HTTP (S3 or SWIFT API)
Ceph Object Gateway
RADOS
(reliable autonomic distributed object store)
Monitor
Monitor
Monitor
Public Network
A lot of nodes and disks
OSD
OSD
OSD
OSD
OSD
Cluster Network

独立索引层？


Client
Routing Table
Routing Table
.....
Router
Router
.....
MySQL
MySQL
Master
Slave
Slave
Master
Slave
.....

小文件合并？

- 可配置的，及可浮动的合并文件大小
- 标记删除
- 异步回收

大文件拆分？

- Striping，可配置的 Striping Size
- 客户端分段

多集群调度

A Readonly

B Readwrite

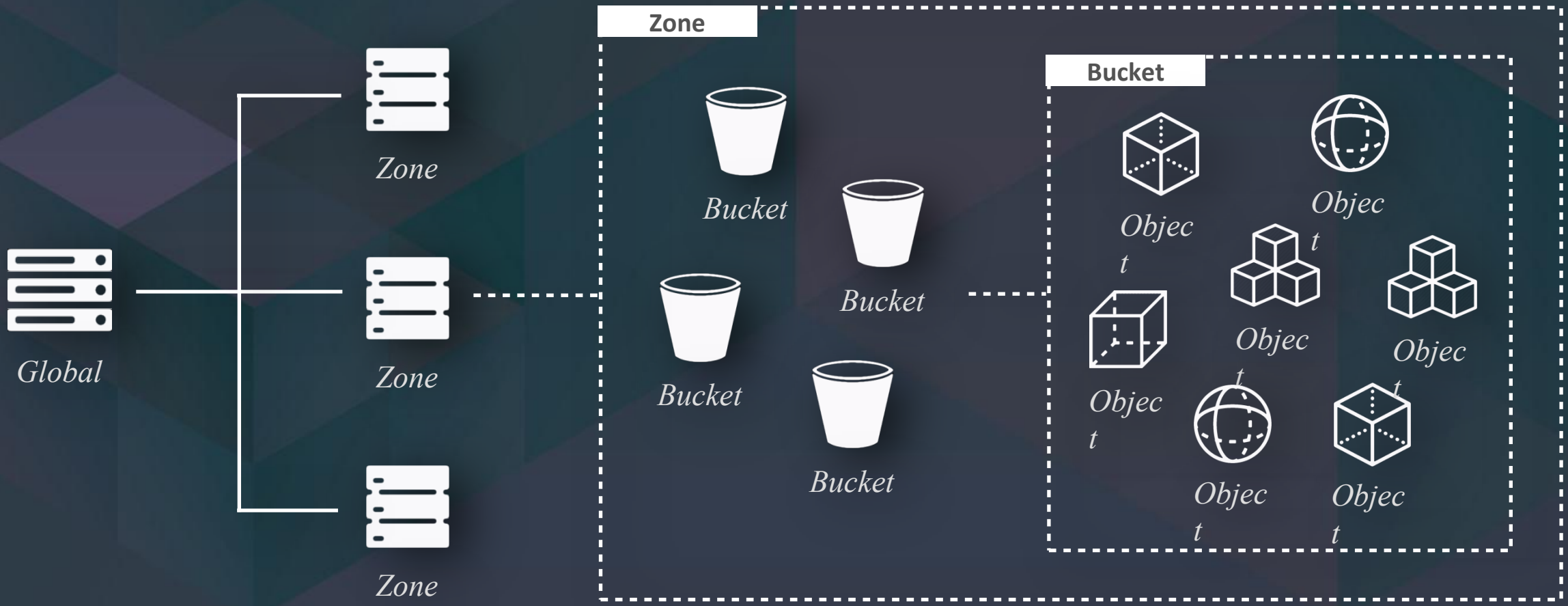
C Readwrite

A Standard

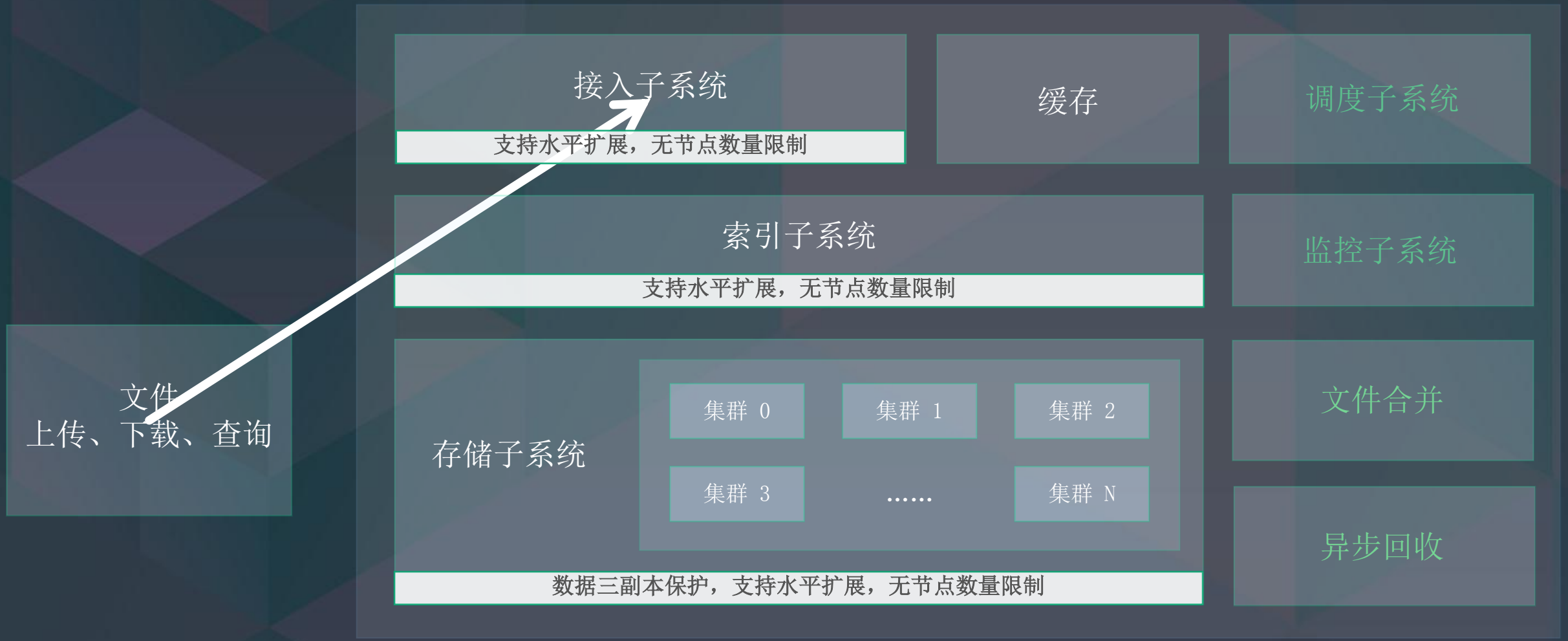
B Infrequent

C NearLine

全局架构

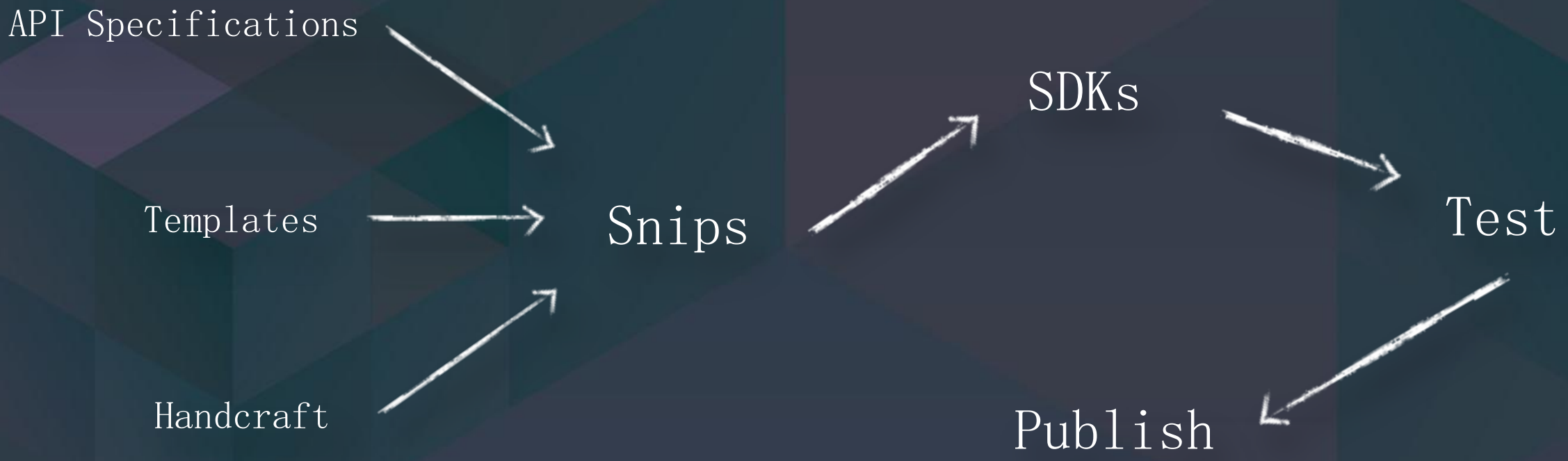

Global
Zone
Zone
Zone
Zone
Bucket
Bucket
Bucket
Bucket
Bucket
Object
Object
Object
Object
Object
Object
Object
Object

区域架构


接入子系统
支持水平扩展，无节点数量限制
缓存
调度子系统
索引子系统
支持水平扩展，无节点数量限制
监控子系统
文件
上传、下载、查询
存储子系统
集群 0
集群 1
集群 2
集群 3
.....
集群 N
文件合并
异步回收
数据三副本保护，支持水平扩展，无节点数量限制

好的架构是设计出来的还是演进出来的？

Snips & SDK


API Specifications
Templates
Handcraft
Snips
SDKs
Test
Publish

Snips & SDK

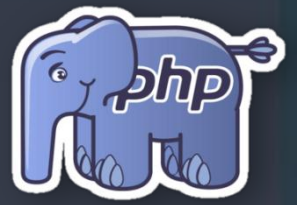
Go

Ruby

Swift

Java

JavaScript



PHP

Python

平台

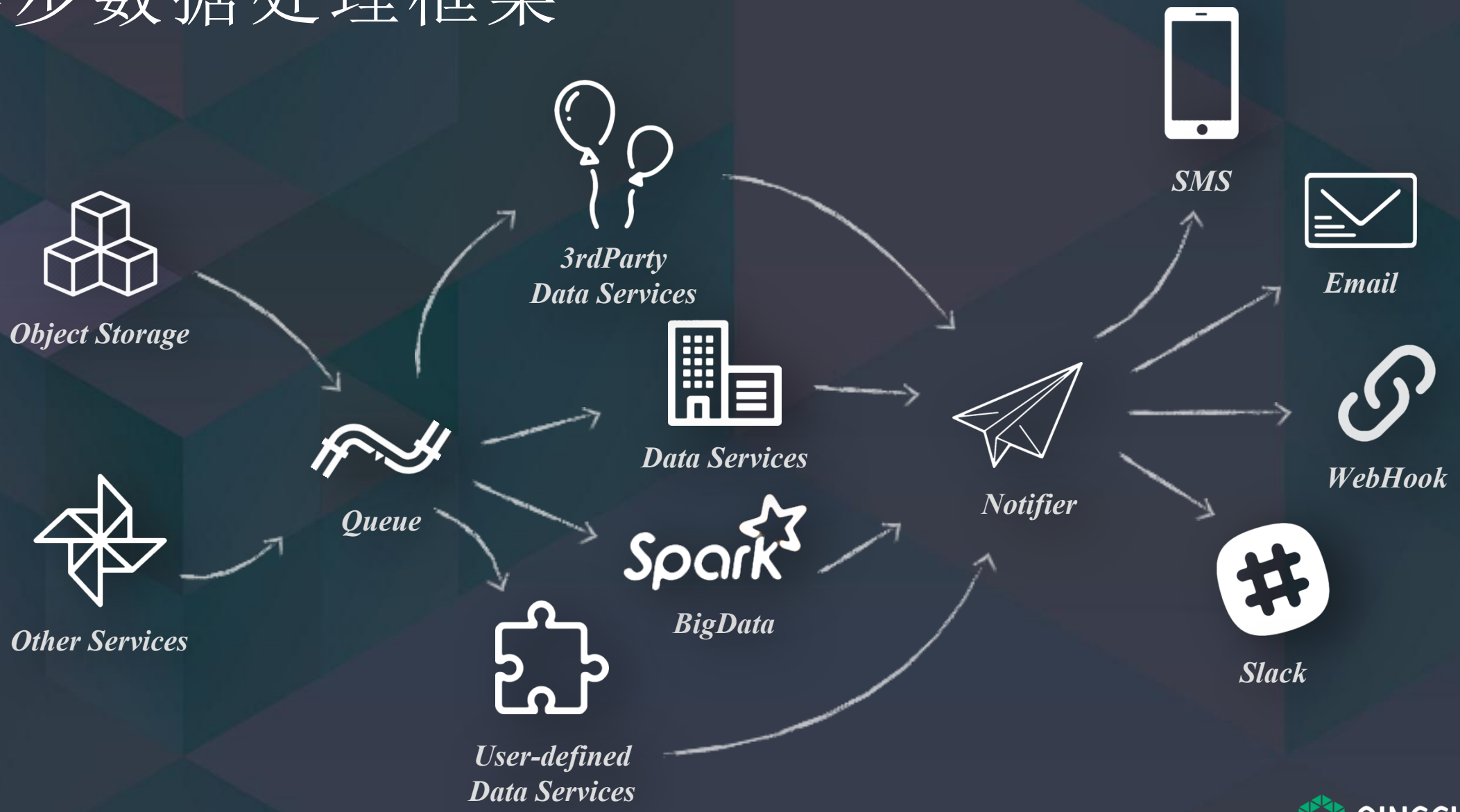
「狭义」

「广义」

产品体系策略

- Public cloud, Hybrid Cloud, On-premise cloud
- 场景化
- 走向用户
- 行业化

异步数据处理框架



产品体系


CDN
点播 / 直播
Event Bus
存储网关
大数据
数据迁移
（软 / 硬）
对象存储
Open
Source
数据处理
Portal
AppCenter
SDK / Utilities
qsftpd

Thank you.