开放原子校源行 MEETUP 聚缘于校 开源共行

云原生时代的 PikiwiDB(Pika)

肖毅-PikawiDB(Pika) Contributor

目录

CONTENTS

开放原子校源行 MEETUP 聚缘于校 开源共行

PikiwiDB(Pika) 发展历程

PikiwiDB(Pika) 解决的问题及应用场景

Pika 的出现并不是为了替代 Redis, 而是 Redis 的场景补充。Pika 力求在完全兼容 Redis 协议、继承 Redis 便捷运维设计的前提下，通过持久化存储的方式解决 Redis 在大容量场景下的问题，如：

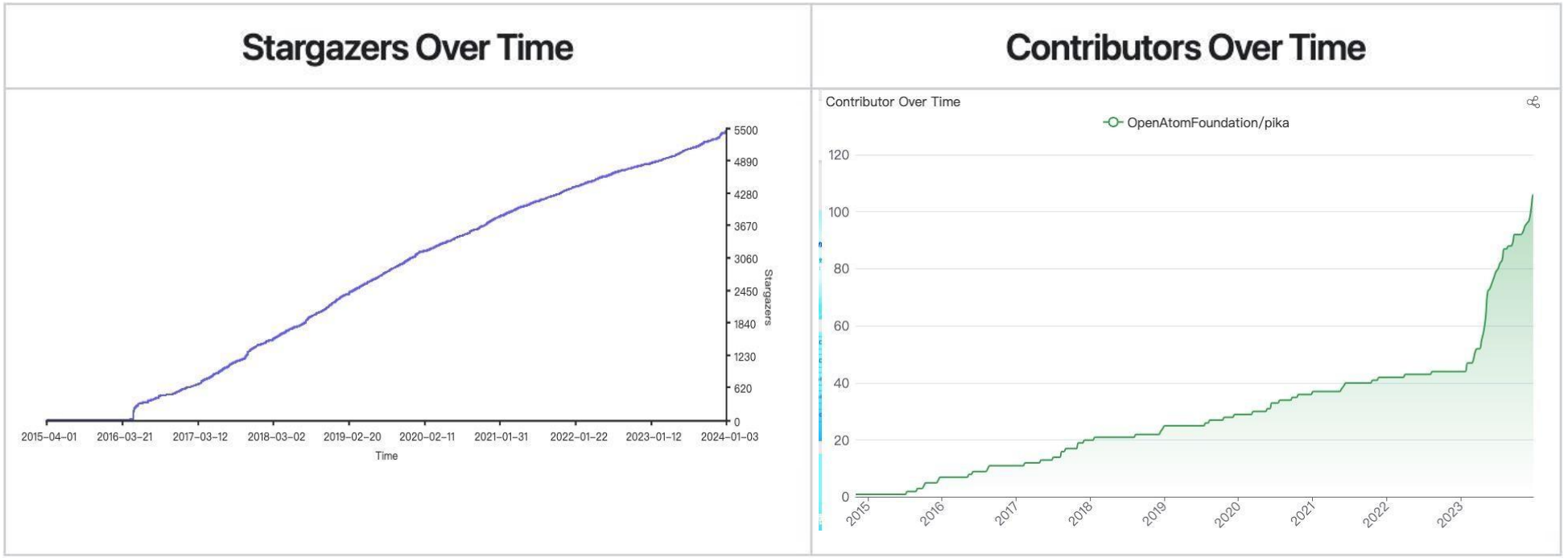
- 单线程易阻塞
- 容量有限
- 加载数据慢
- 故障切换代价高

key-string
高性能 KV
搜索推荐、机器学习

key-hash
复杂在线业务
用户信息、好友关系、对象存储元数据

key-list
简单高效的消息中间件
分布式任务系统

发展历程



2015.04 项目启动 → 2015.11 发布 1.0 → 2016.02 开源 → 2016.04 发布 2.0 → 2018.08 发布 3.0 → 2020.08 申请加入 OpenAtom → 2021.03 孵化运营

2023年12月，Pika正式更名为PikiwiDB

"Pi-kiwi-DB"：

1. "Pi" 念派
2. "Pik" 恰好保留了 "Pika" 的前三个字母
3. "kiwi" 音同 "KV"，寓意几维鸟

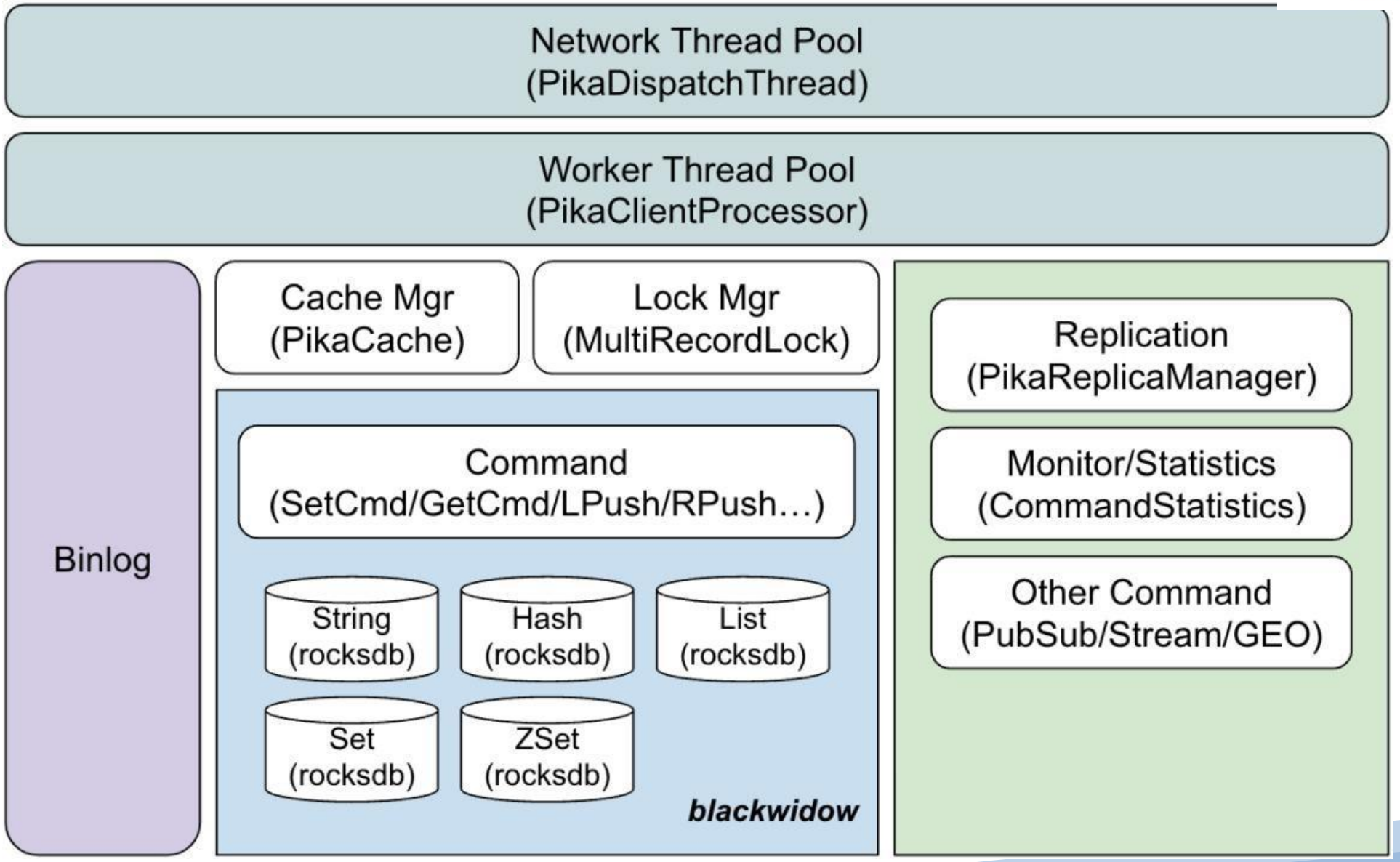
开放原子校源行 MEETUP 聚缘于校 开源共行

PikiwiDB(Pika) 产品介绍

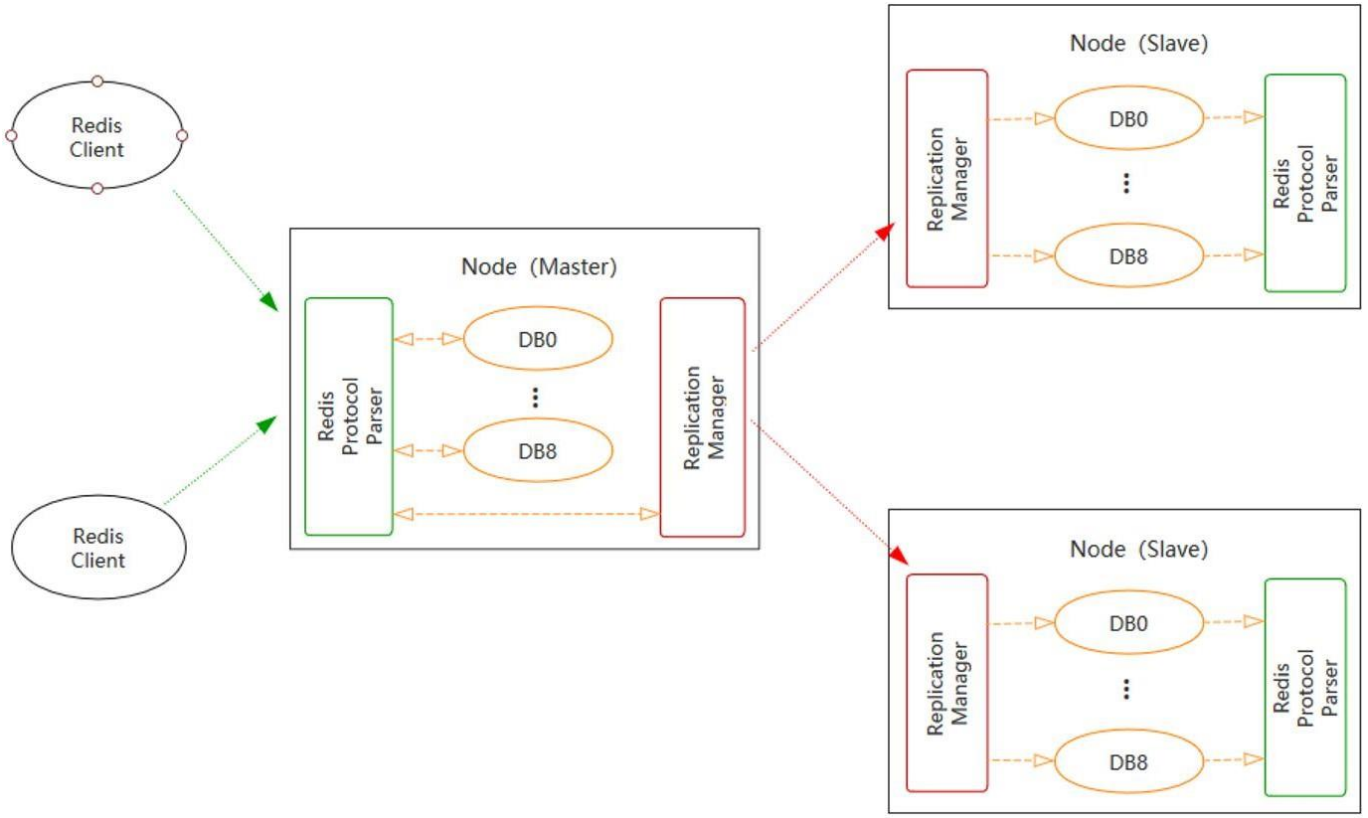
产品特点

- ✔ 在完全兼容 Redis 协议的前提下追求 高性能、大容量、低成本、大规模
- ✔ 支持 Redis 的常用数据结构 bitmap、string、hash、list、set、zset、geo、hyperloglog、pubsub、stream
- ✔ 持久化存储到 RocksDB
- ✔ 单机主从、Pika Cluster、K8s 等部署方式
- ✔ 相比于 Redis 的内存存储方式，能极大减少服务器资源的占用，增强数据的可靠性
- ✔ 支持 centOS/Ubuntu/macOS 三大 OS， Arm 和 x86 两种架构

产品架构

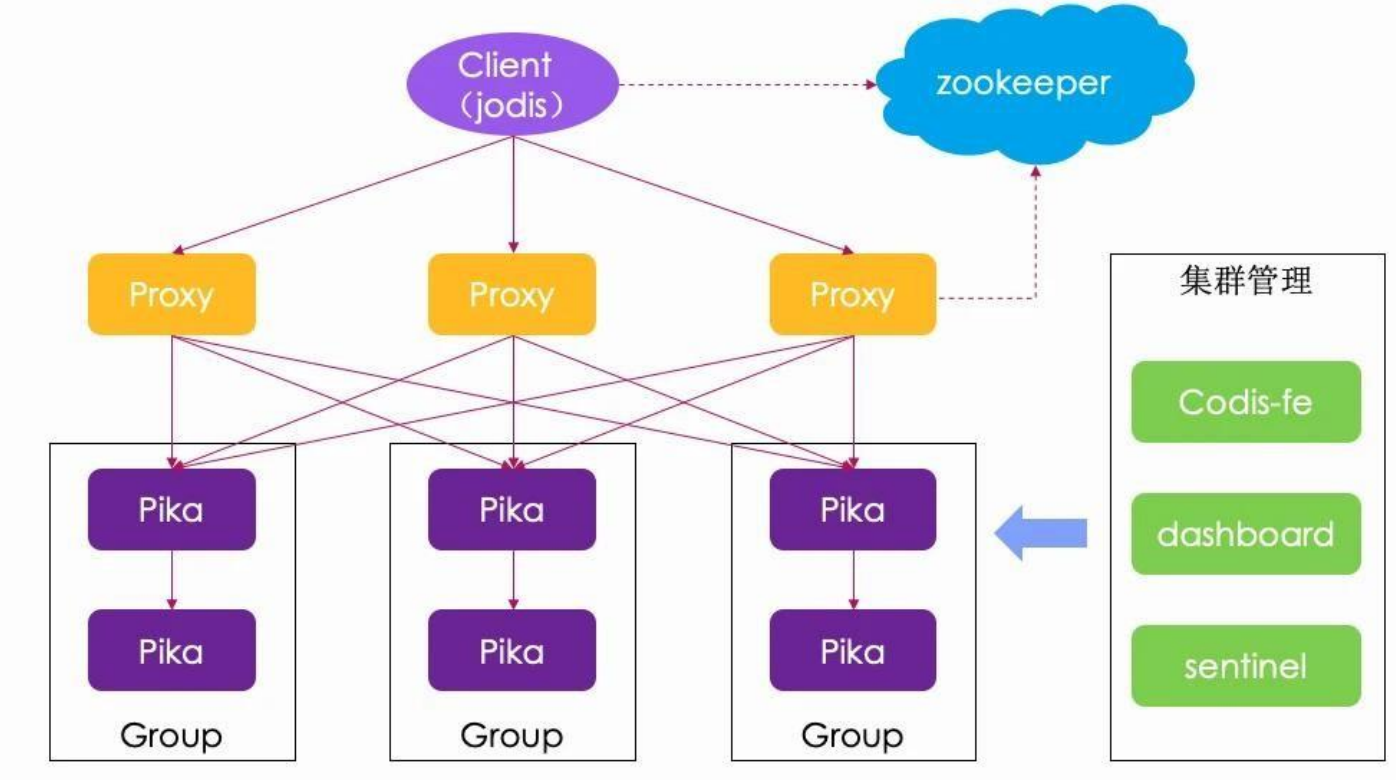


主从模式



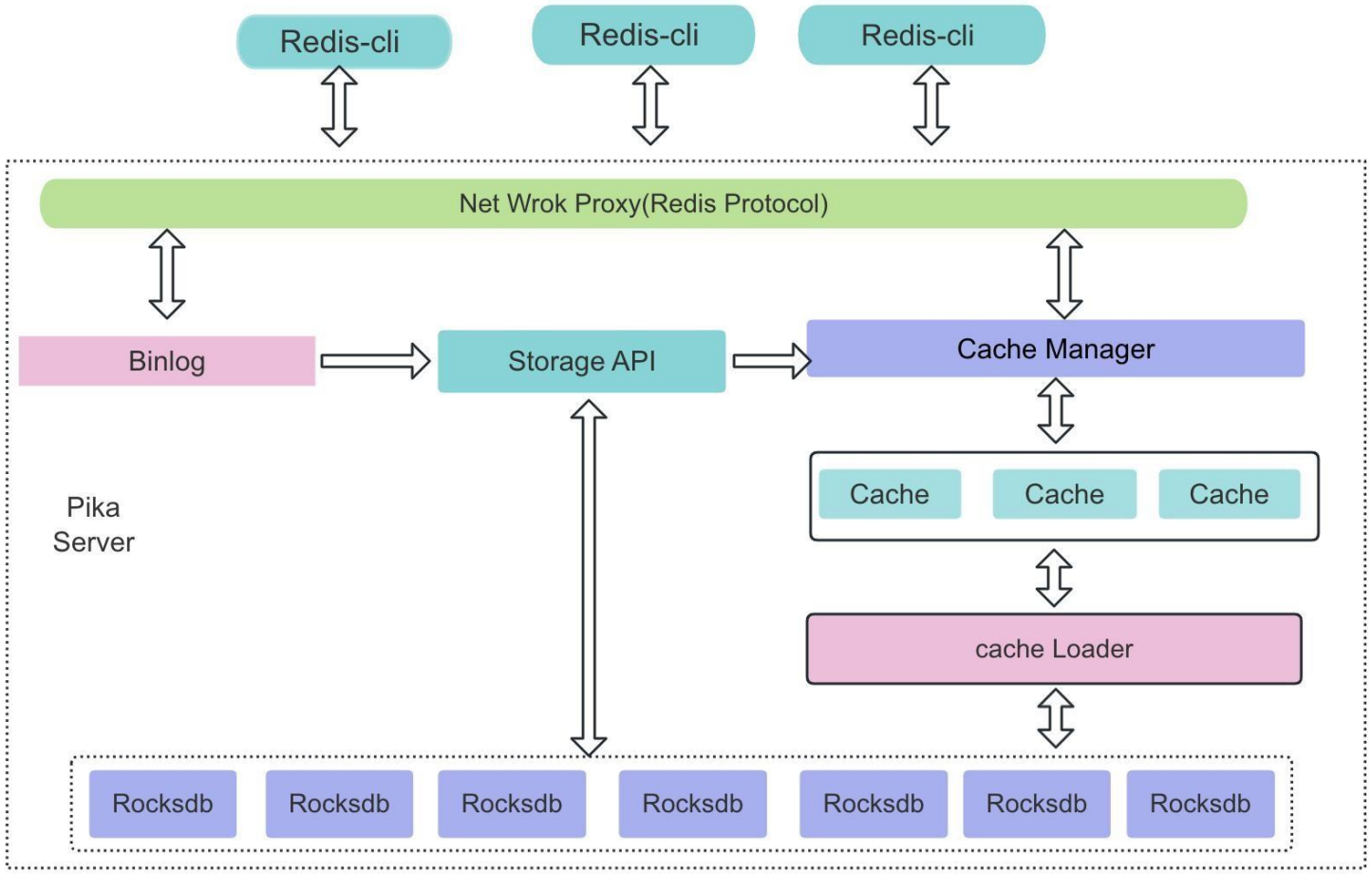
- 架构与 Redis 类似
- 与 Redis 协议和数据结构兼容性好
- 每种数据结构使用一个 RocksDB 实例
- 主从采用 Binlog 异步复制方式

群集架构



- 采用 Codis 架构，支持多 group
- 单 group 内是一个主从集群
- 以 group 为单位进行弹性伸缩

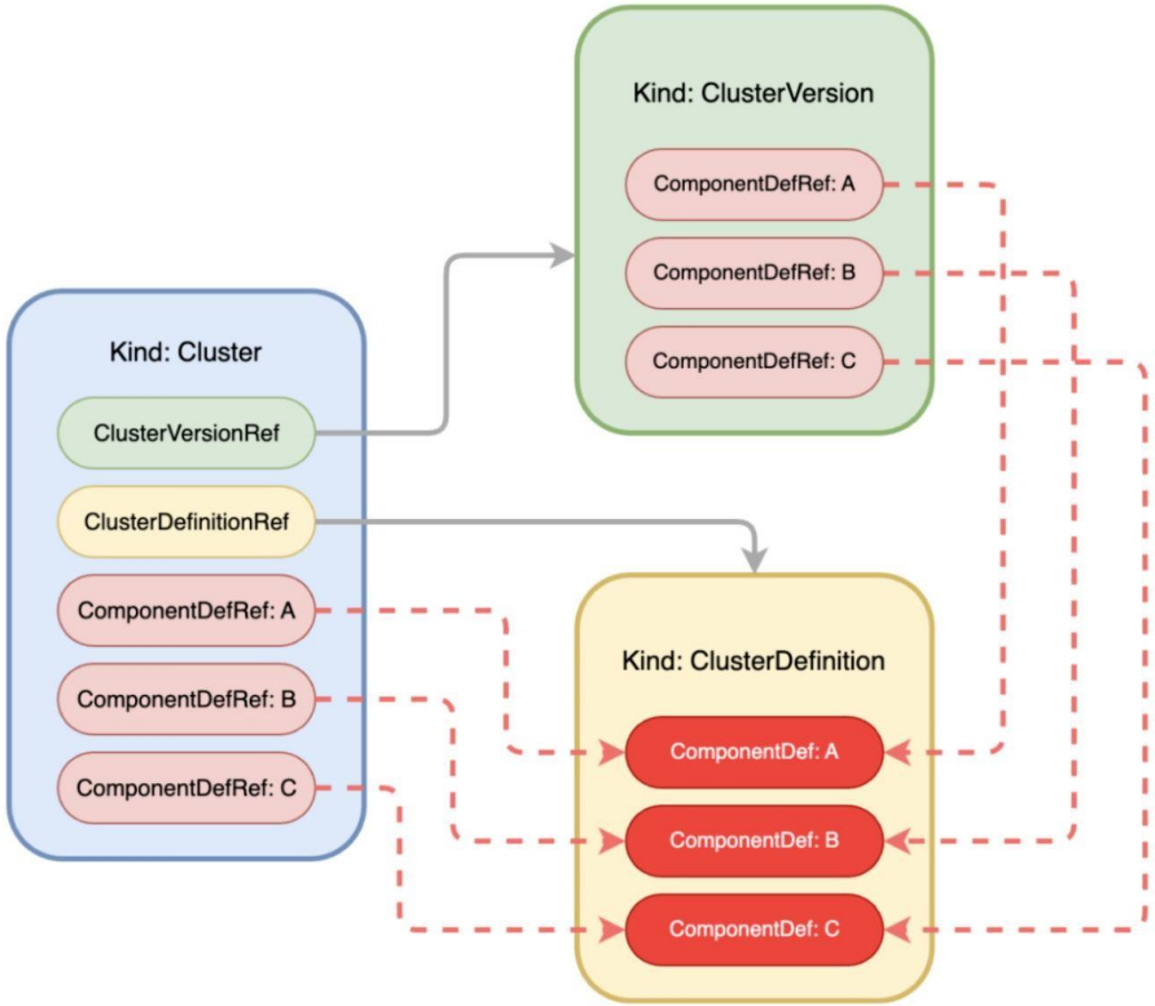
缓存层（3.5.1版本特性）



采用 Codis 架构，支持多 group
单 group 内是一个主从集群
以 group 为单位进行弹性伸缩

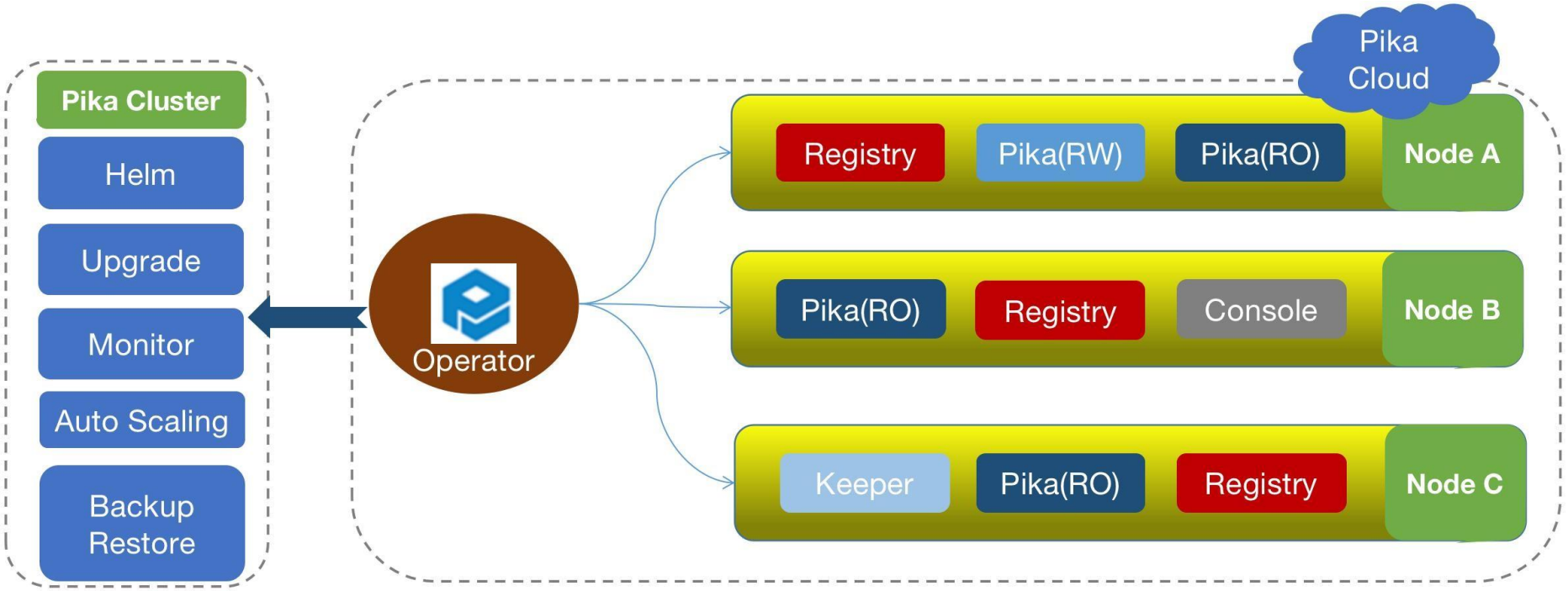
PikiwiDB(Pika) Cloud

(PikiwiDB)pika-operater 详解



Pika-opeartor 可以方便地将 Pika Cluster 部署在 Kubernetes 上，用户给出一个描述集群的拓扑、版本和资源信息 Yaml 即可

PikiwiDB(pika)-operater 详解



- Pika 集群部署
- curl -fsSL https://kubeblocks.io/installer/install_cli.sh | bash // 安装 kbcli
- kbcli kubeblocks install // 安装 kubeblocks
- helm install pika ./pika // 安装 Pika
- helm install pika-cluster ./pika-cluster // 安装 Pika-cluster

(PikiwiDB)pika-operater 详解

Scale out 后，将 Pika 自动添加至 Codis 集群，并 Reblance 。

Scale in 开始后，先进行 slot 搬迁，再减少实例。

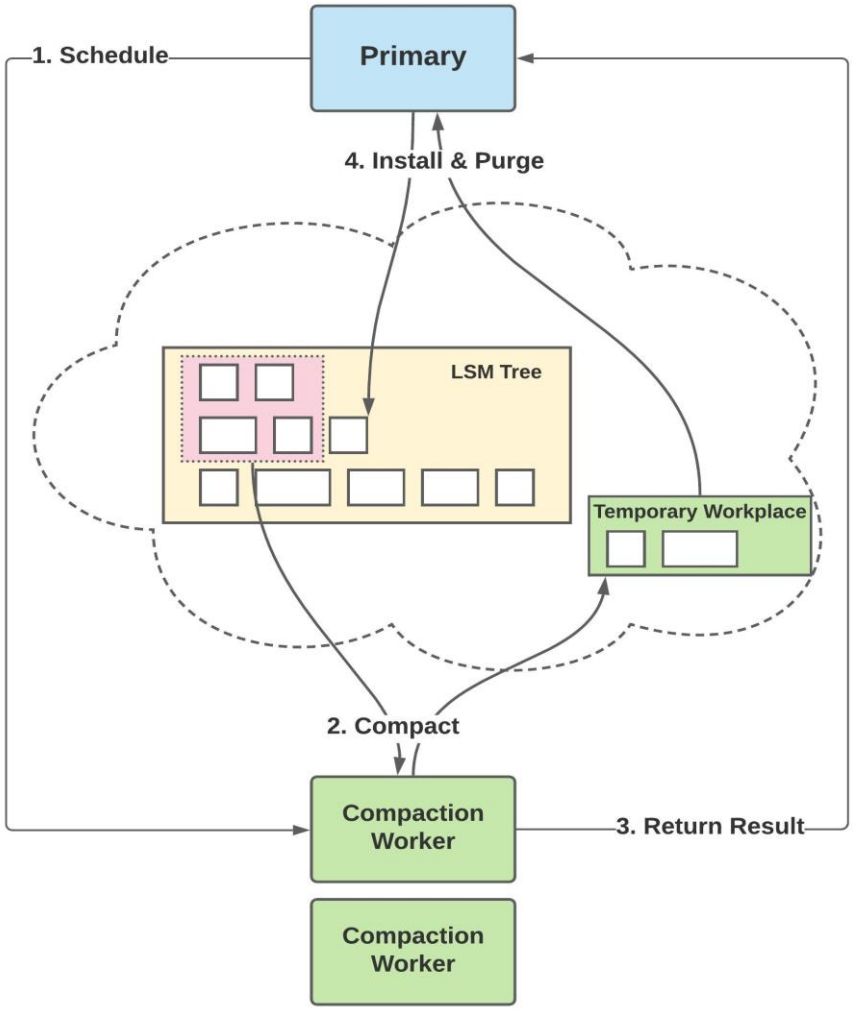
Scale up 实现集群磁盘的 scale up，实现集群内存的 scale up。

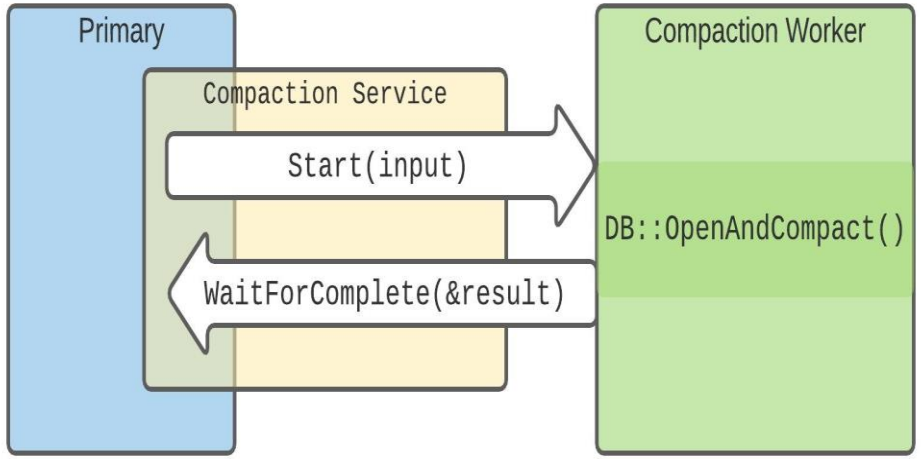
各个组件可独立配置存储，实现集群数据 backup，实现集群从备份拉起，实现集群恢复备份

将 pika-exporter 集成入集群，实现集群监控在 grafana 中插入 Pika 监控 dashboard

实现从 API 创建 Pika 集群 demo，建立 E2E Test

Rocksdb Remote Compaction





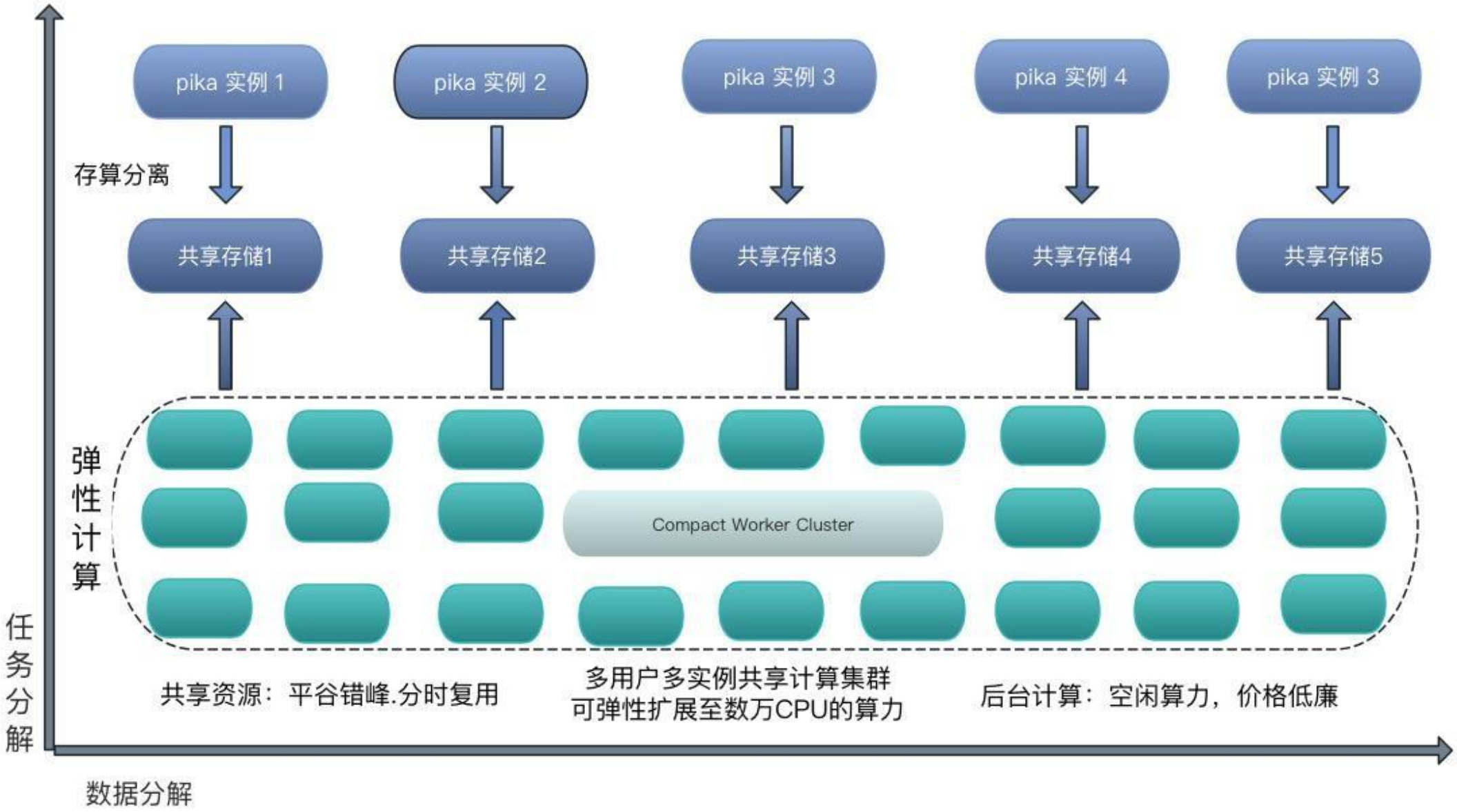
1Pika(Compute) 执行 R/W 任务时非常轻量，P99 延时平稳，几乎无毛刺

2LSM Tree 相对于 Worker 只读，对于 Pika(Compute) 是 Append Only，unmutable

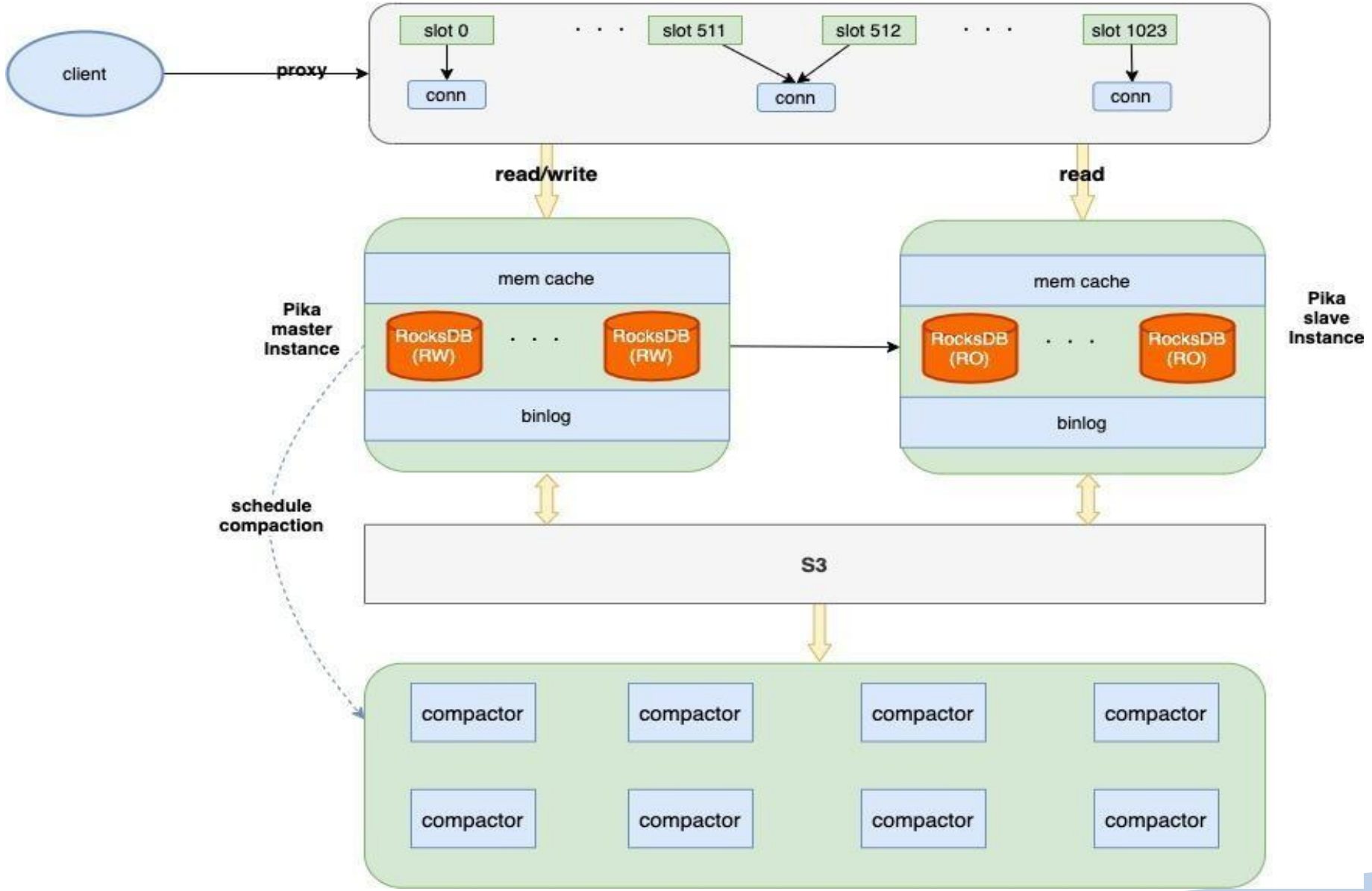
3 Pika(Compute) 瞬时启动，Worker 节点可弹性伸缩

4Compaction 任务分离后，Storage 层可使用廉价硬件实现 Share-storage，还可以通过 EC 方式实现极大容量

(PikiwiDB)Pika 分布式 compact



(PikiwiDB)Pika Serverless 架构思考



Pika Serverless 详细规划

强内核

- 向Redis接口靠拢
- 单机性能提升
- 多租户
- 更多平台支持
- 支持 CP

高质量

- 单测
- 集成测试
- 混沌压测
- ChatGPT Code Review

工具集

- 基于 Prometheus 的 exporter 接口
- Redis 与 Pika 迁移

大社区

- 开源大赛
- 和 OpenAtom 协作，提升项目知名度

云原生

- 存算分离：S3
- 弹性： Operator
- 故障自愈
- 多租户
- 资源隔离

PikiwiDB(Pika) 实战案例

Pika 用户

用户展示

360公司内部部署使用规模 10000+ 实例，单实例数据量 1.8TB；

微博公司内部部署实例 10000+；

喜马拉雅(X Cache)实例数量 6000+，数据量 120TB+；

个推公司内部部署300+实例，总数据量30TB+；

迅雷公司用于用户存储个性化推荐数据, 目前使用100台机器

小米公司目前已经上线

其他更多用户：

360 Codis 集群实战经验

- 部署方式：灵活搭配。
- 搜索部门节点配置：
- 40C 192G，1.8T NVMe 磁盘，千兆网卡

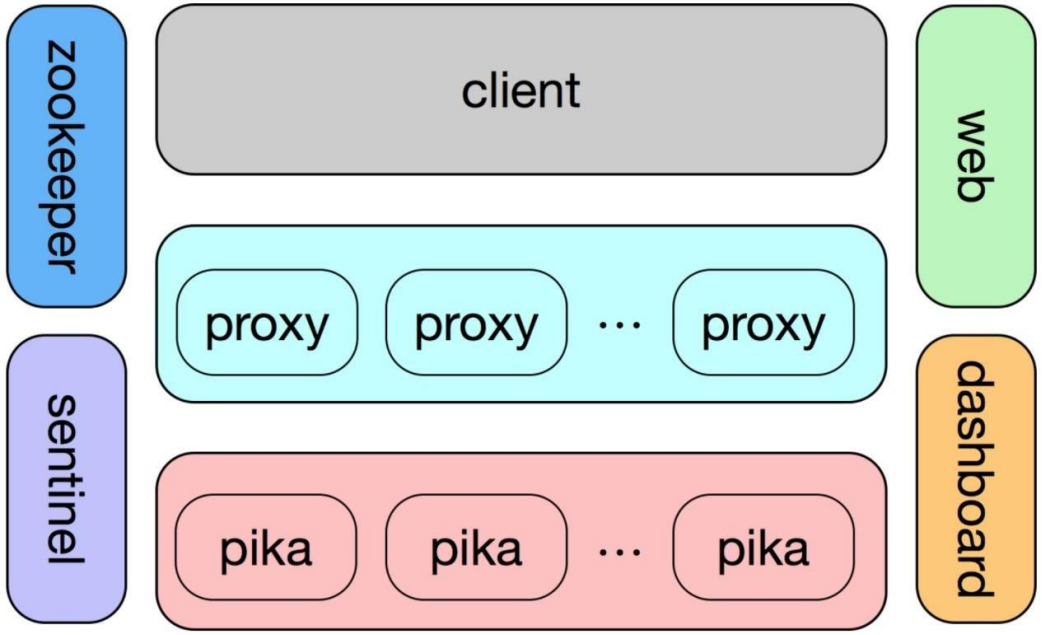
组件	节点个数	实例规格
Pika Server	12 主 12 从	每个实例：20 核，32G 内存，200G 磁盘
Codis FE	1 个节点	1 个节点 2 核 4G 虚拟机
Codis Dashboard	1 个节点(虚拟机)	1 个节点 2 核 4G 虚拟机
Codis Etcd	3 个节点	3 个节点 2 核 4G 虚拟机
Codis Proxy	4 个节点	4 个节点 2 核 4G 虚拟机

360 Codis 集群实战经验

关键数据：

Read QPS > 20万，P99.9延时 < 5ms，P99 < 2ms
Write QPS > 30万，P99.9延时 < 4ms，P99 < 2ms

喜马拉雅



使用规则：

1.把 Pika 当做缓存使用时，扩缩容时不迁移数据,扩缩容速度快，节点宕机恢复时不加载旧数据

2.把 Pika 当做数据库使用时，实例扩缩容时需要 迁移数据，速度比较慢

3.存储海量数据：Pika实例数量6000+，日承载 业务请求量超过千亿次

新浪微博

- 通用 KV 存储解决方案，作为Redis 数据库的有效补充
- 当前实例规模近 10000+
- 每日访问量近万亿，峰值QPS高达 3000万+
- 磁盘数据规模达百 TB+
- 支撑微博平台、搜索、机器学习等核心业务
- 降本增效的一种方案

每日互动

机器	Codis	Pika
1主1从 192GB 内存 480GB*6 磁盘	192GB	480GB*6
可存储的数据量对比	1	15
相同数据量需要的机器数量对比	15	1

- HBase 迁移过来，延迟要求 p99<50ms，Pika 可以达到要求
- 最早是在 2021年的时候上的 Pika，2022 年开始大规模推广使用
- 目前 Pika 数据 30TB+，存储了上百亿的 key

欢迎加入我们社区

开放原子校源行 MEETUP 聚缘于校 开源共行

PIKA

Pika DB 开发群

请关注 Pika 官方微信公众号

开放原子校源行 MEETUP 聚缘于校 开源共行

Thanks