课时28

如何构建一个高性能、易扩展的 Redis 集群

1. Client 端分区
2. Proxy 端分区
3. Redis Cluster 分区

Redis 集群

拉勾
LAGOU.COM

Redis 集群方案

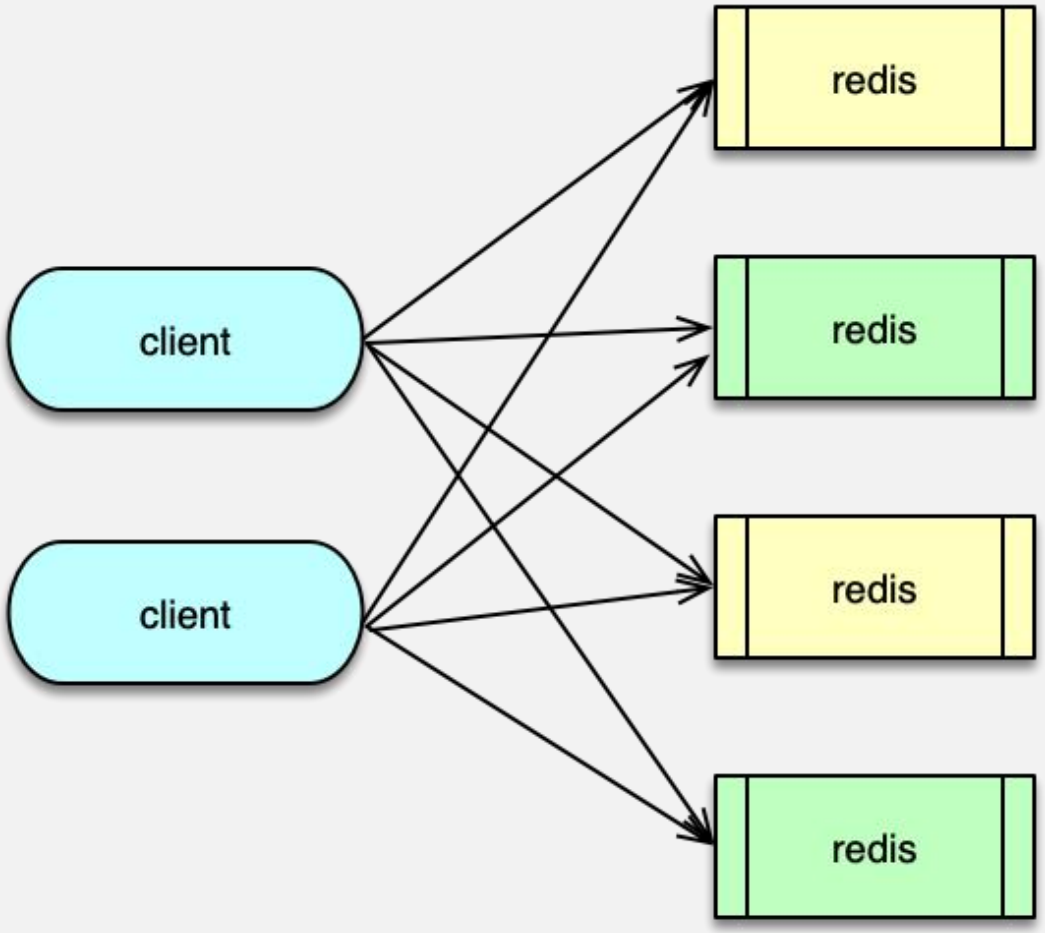
- Redis 集群：大容量，高性能，高可用
- Redis 集群的分布式方案
  - Client 端分区
  - Proxy 分区
  - Redis-cluster 分区

Redis 集群

拉勾 LAGOU.COM

Client 端分区

- Client 端决定存储或读取key的节点
- 单key请求 ➔ 按哈希选择节点
- 单个请求多个key ➔ 按哈希分解到多个请求
- 哈希算法将数据进行分布
  - 取模哈希
  - 一致性哈希
  - 区间分布哈希

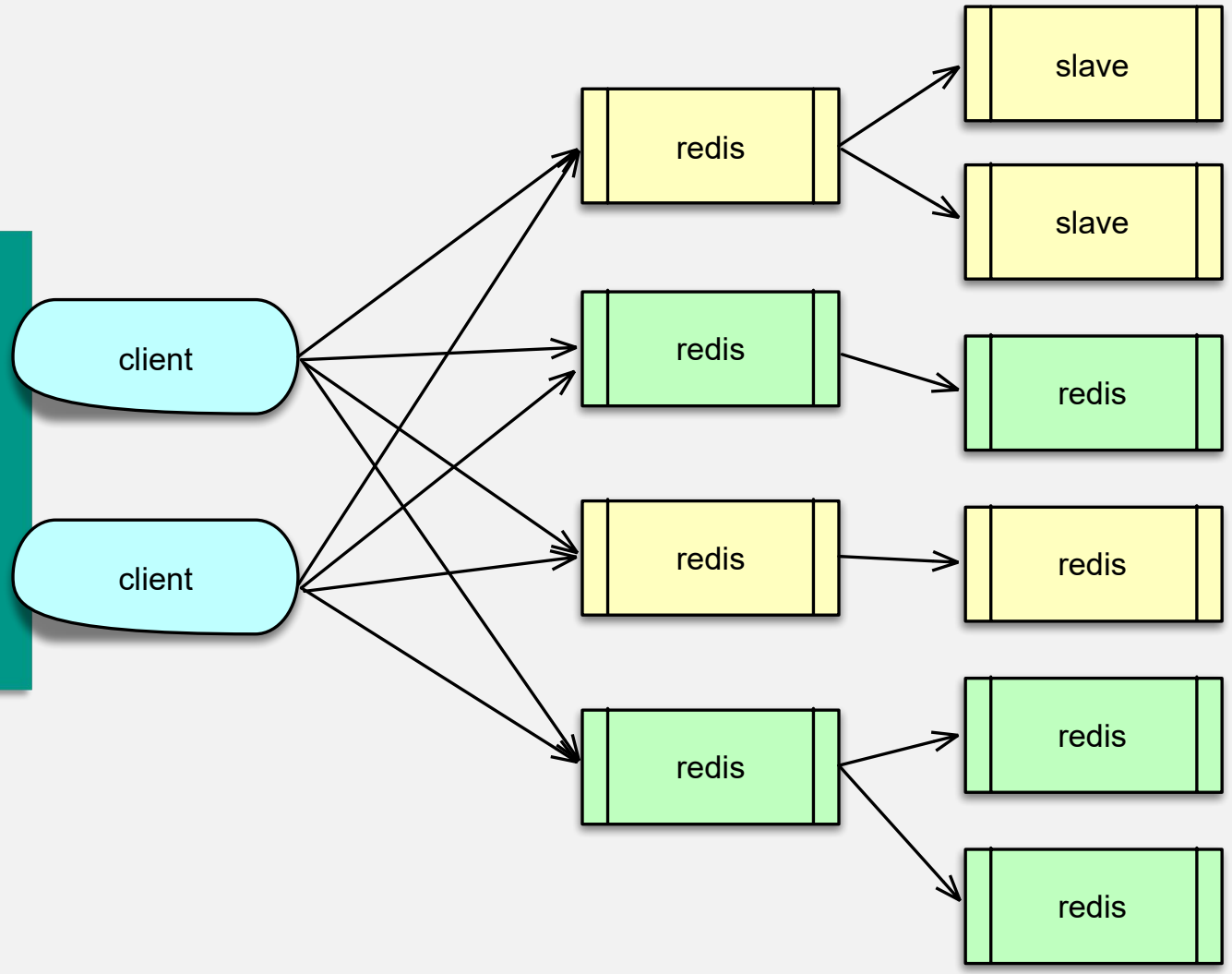


Redis 集群

拉勾 LAGOU.COM

Client 端分区

- 通过DNS访问及管理主从
- 主从变化，Client 配置无需变更
- Client 对多个从库需要负载均衡
- Client 定时探测DNS下IP变化
  - 主库故障切换
  - 从库增减

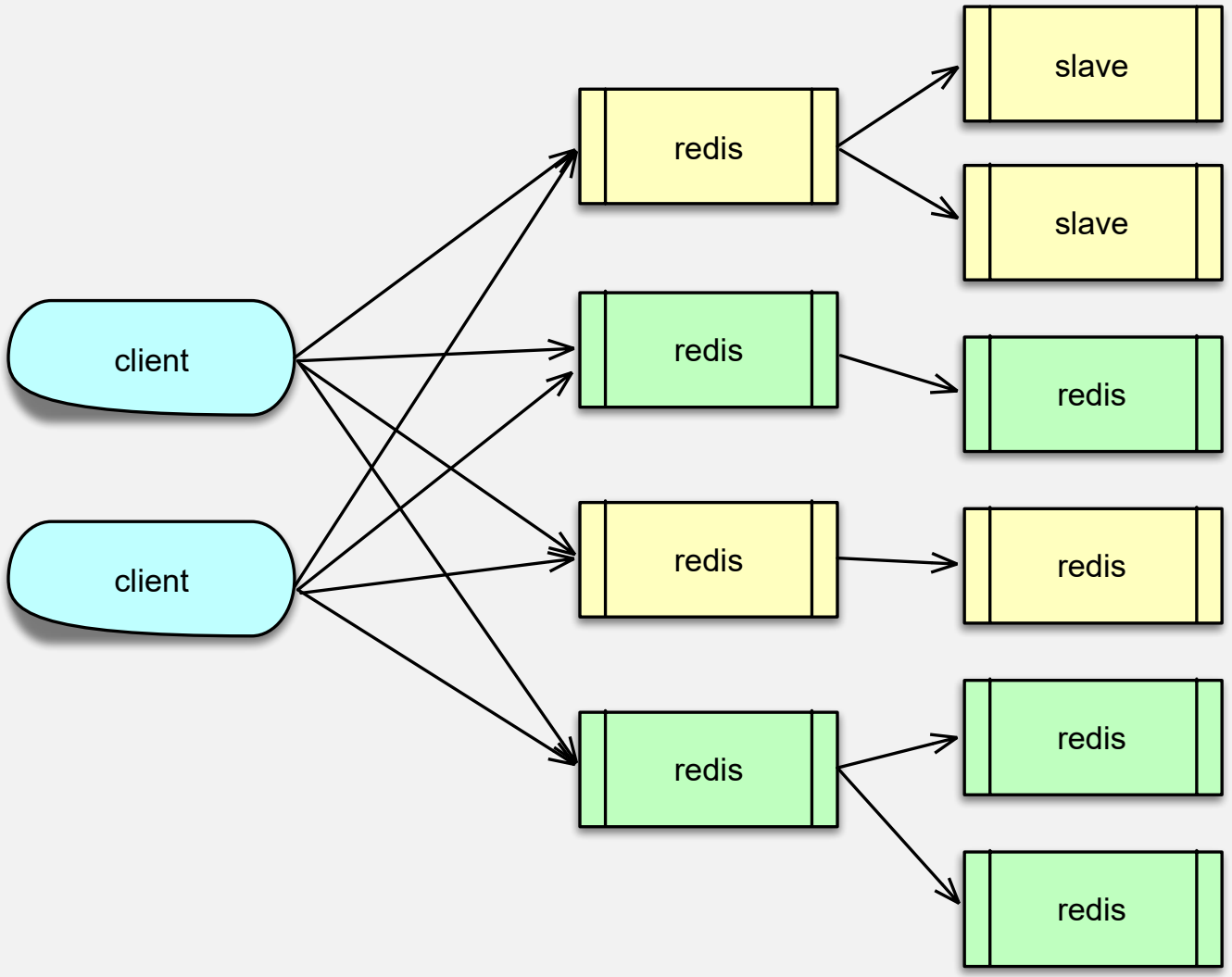


Redis 集群

拉勾 LAGOU.COM

Client 端分区

- 优势
  - 分区逻辑简单，配置简单
  - 性能高效
- 不足
  - 扩展不便：双倍扩展，预分配足够分片
  - 分片变更，业务端需要修改分发逻辑+重启

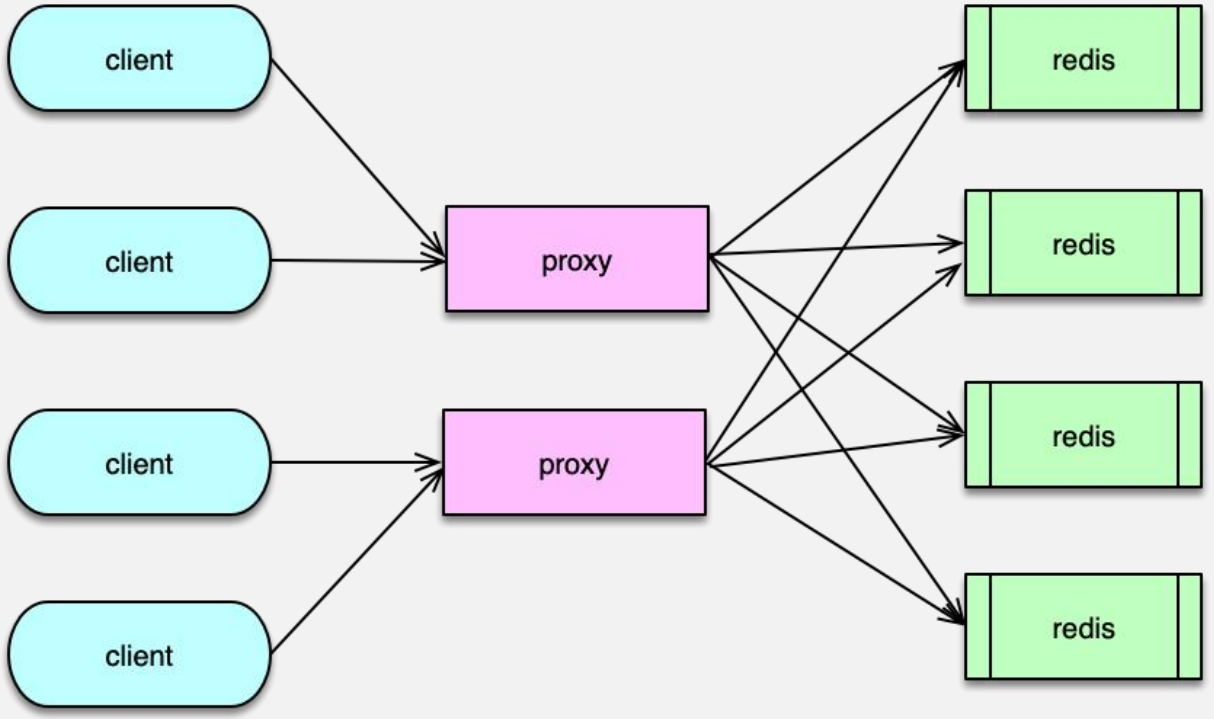


Redis 集群

拉勾 LAGOU.COM

Proxy 端分区

- Client 直接发送请求给Proxy，Proxy解析并路由key
- 单个请求多key
  - Proxy 分拆为多个请求
  - 请求不同Redis分片
  - 等待所有响应，聚合后返回
- Proxy 负责 路由逻辑、切换逻辑

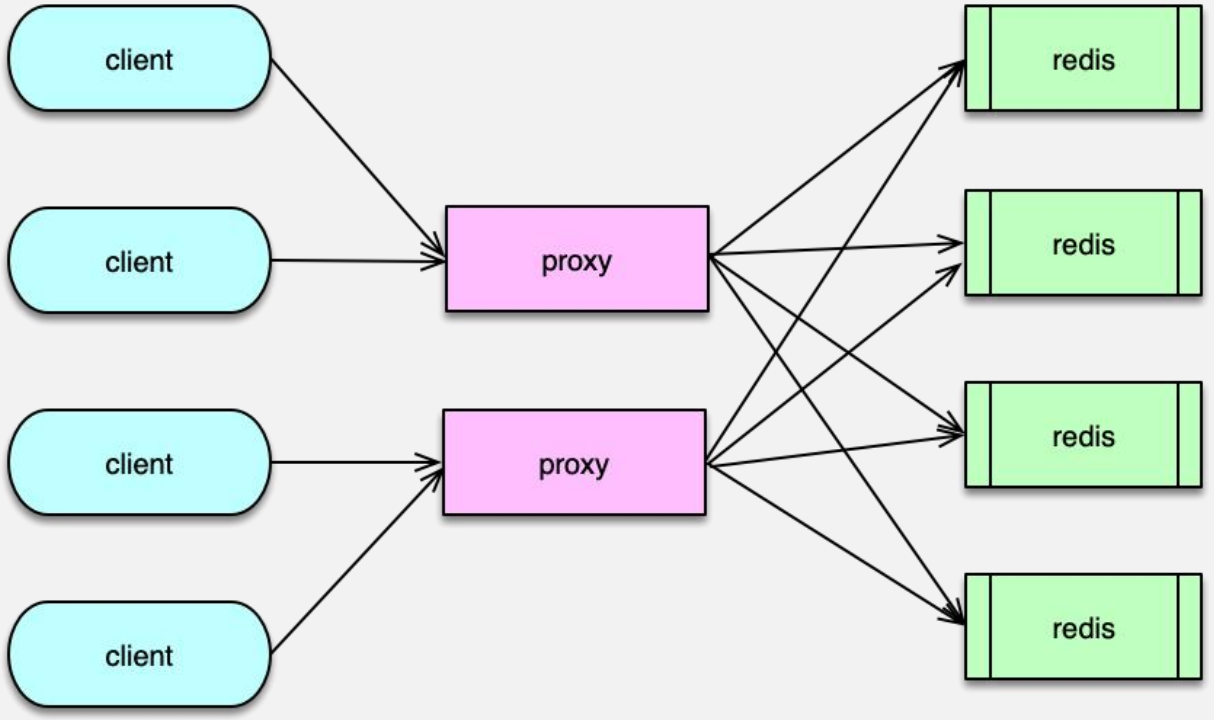


Redis 集群

拉勾 LAGOU.COM

Proxy 端分区

- 常见Proxy端分区方案
  - Twemproxy
  - Codis
- Twemproxy
  - 简单，稳定可靠
  - 对多key的multi访问性能不高
  - 不支持平滑扩缩
  - 无管理后端，运维不便利

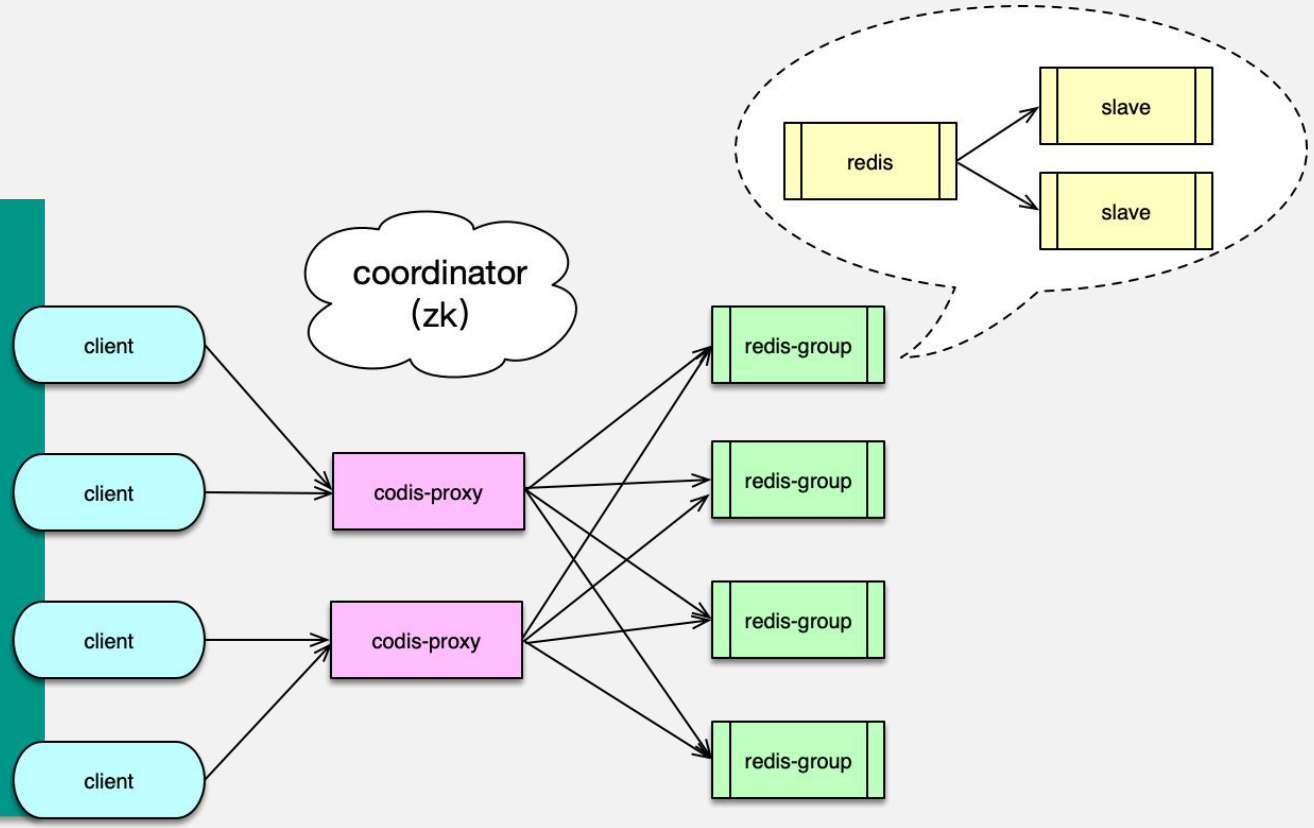


Redis 集群

拉勾 LAGOU.COM

Proxy 端分区

- Codis
  - Codis-server 基于redis扩展存储，支持slot及数据迁移
  - Codis-proxy 代理client访问，解析并路由
  - Zookeeper 维护codis集群节点，存储元数据
  - Codis-dashboard 管理Redis存储节点及proxy节点
  - 管理后台，方便监控及运维

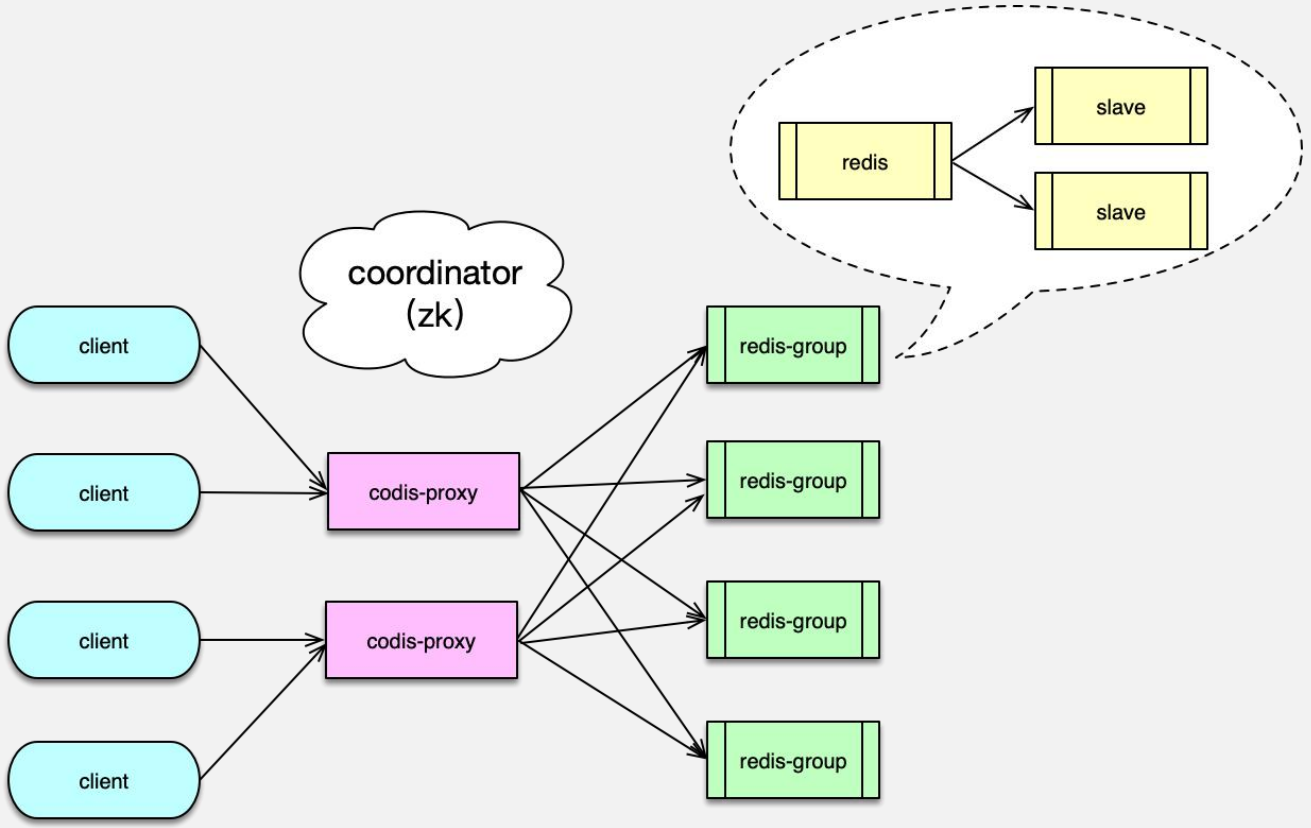


Redis 集群

拉勾 LAGOU.COM

Proxy 端分区

- 优势
  - Client访问逻辑和Redis分布逻辑解耦，业务访问简单
  - 资源变化、扩缩容，只用修改有限的Proxy，业务无需升级重启
- 不足
  - 访问多一跳，性能损耗
  - 多一代理层，系统架构复杂

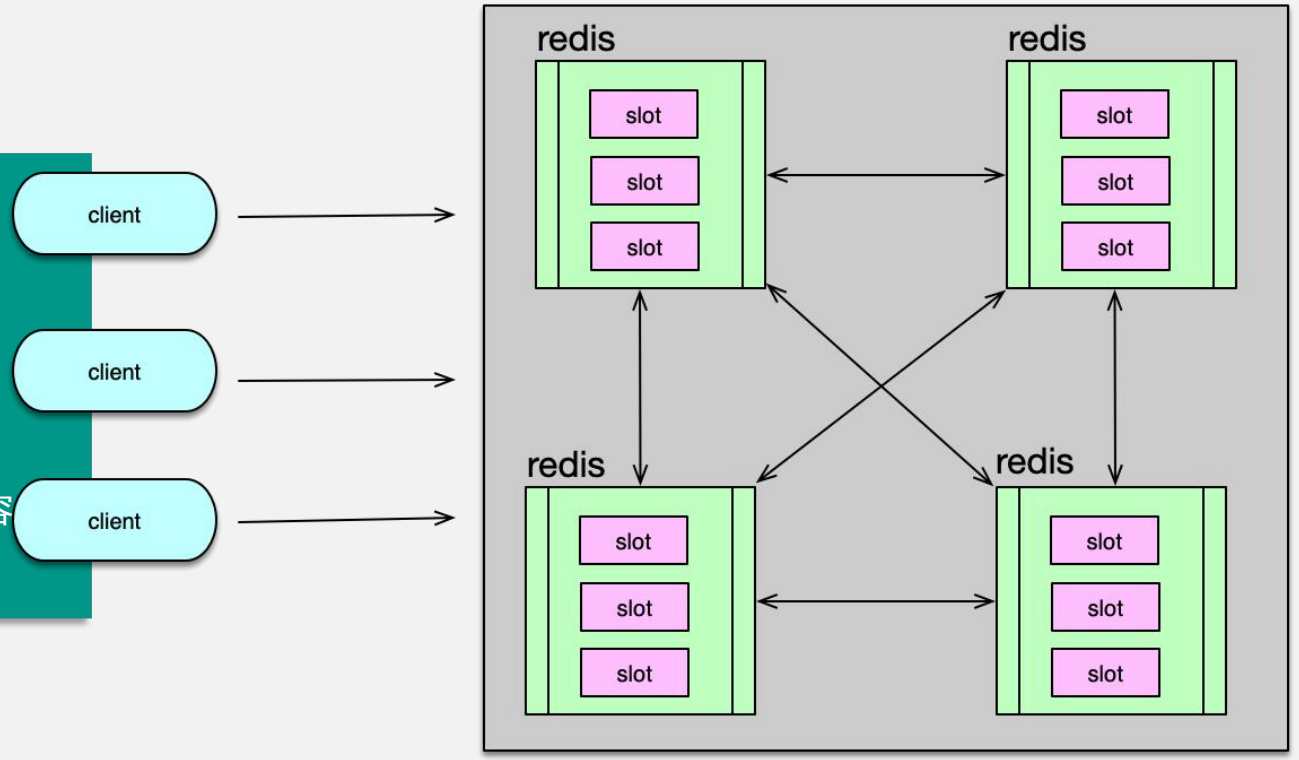


Redis 集群

拉勾 LAGOU.COM

Redis-Cluster 原理

- Redis-Cluster 按slot进行数据读写和管理
- 一个redis 集群包含 16384 个slot
- Slot 按需分配到不同redis节点
- Redis节点的slot可内部迁移，以均衡访问
- Redis节点的slot可迁移到新节点，以进行扩容

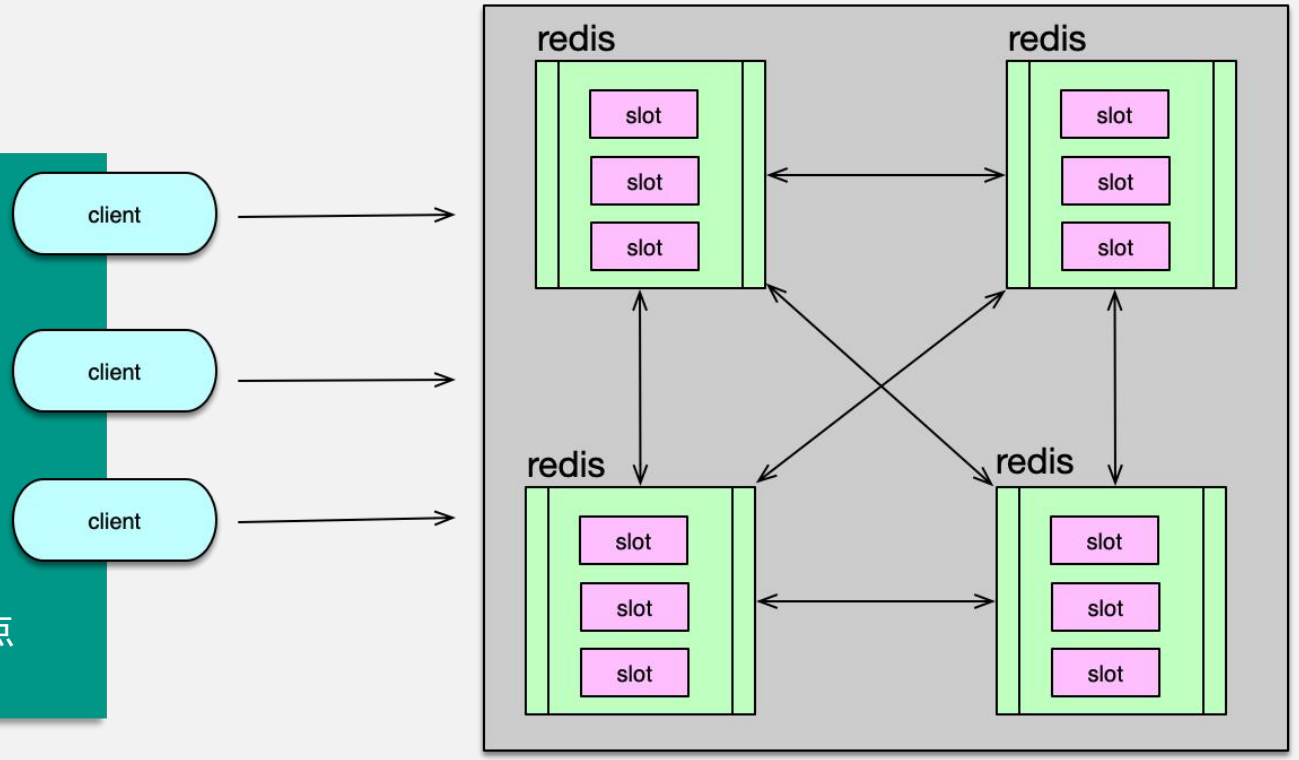


Redis 集群

拉勾 LAGOU.COM

Redis-Cluster 原理

- Redis 节点通过cluster的 addslots、delslots增减slot
- Redis 节点 通过cluster flushslots 清空已有slot信息
- 去中心化，Redis 每个节点存储全部slot拓扑分布
- 不同Redis节点通过gossip协议进行互联
- 节点通过发送cluster meet将新节点加入到集群，存在一条节点链即可，无需meet 所有节点

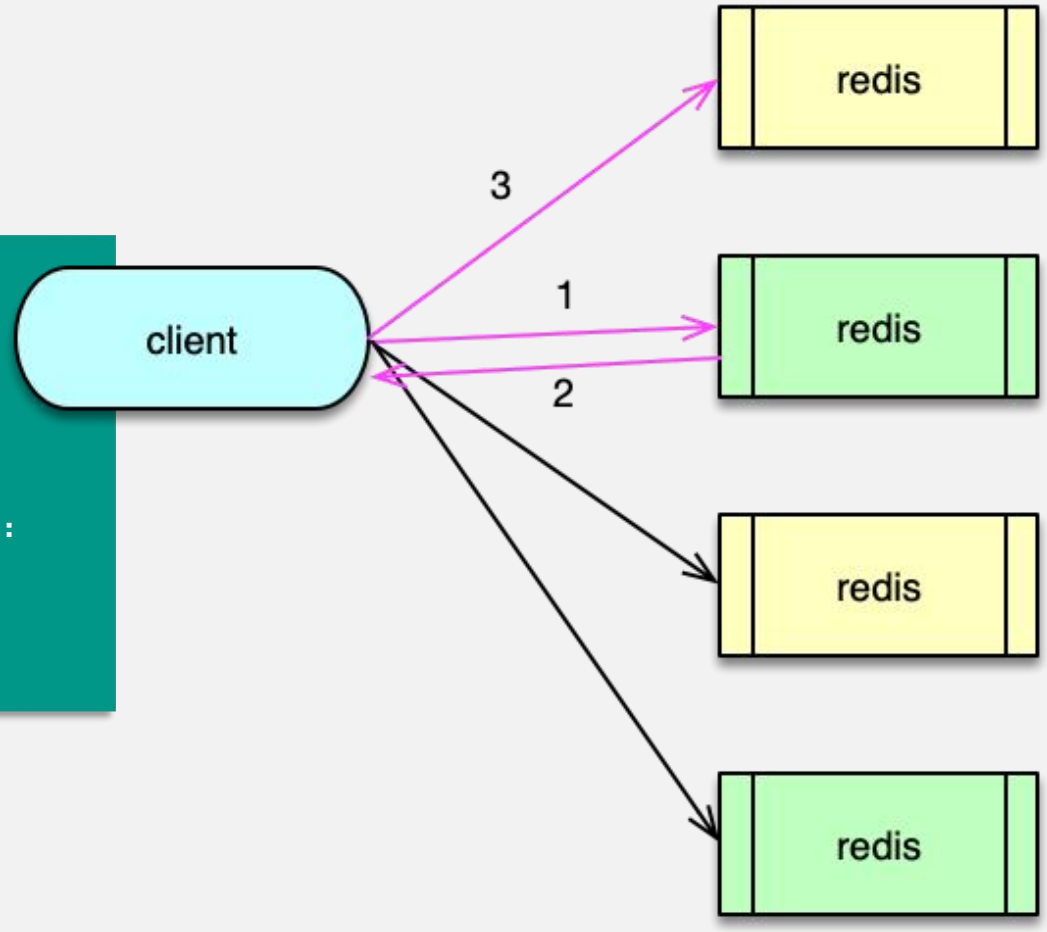


Redis 集群

拉勾 LAGOU.COM

Redis-Cluster key访问

- Key访问需要mart client配合
- Cluster 模式下key访问流程
  - Client 选择一个Redis节点发送请求
  - Redis解析命令后，对 Key进行 slot hash 定位：crc16(key) & 0x3FFF
  - Redis 发现 key对应slot在本地，直接处理后返回

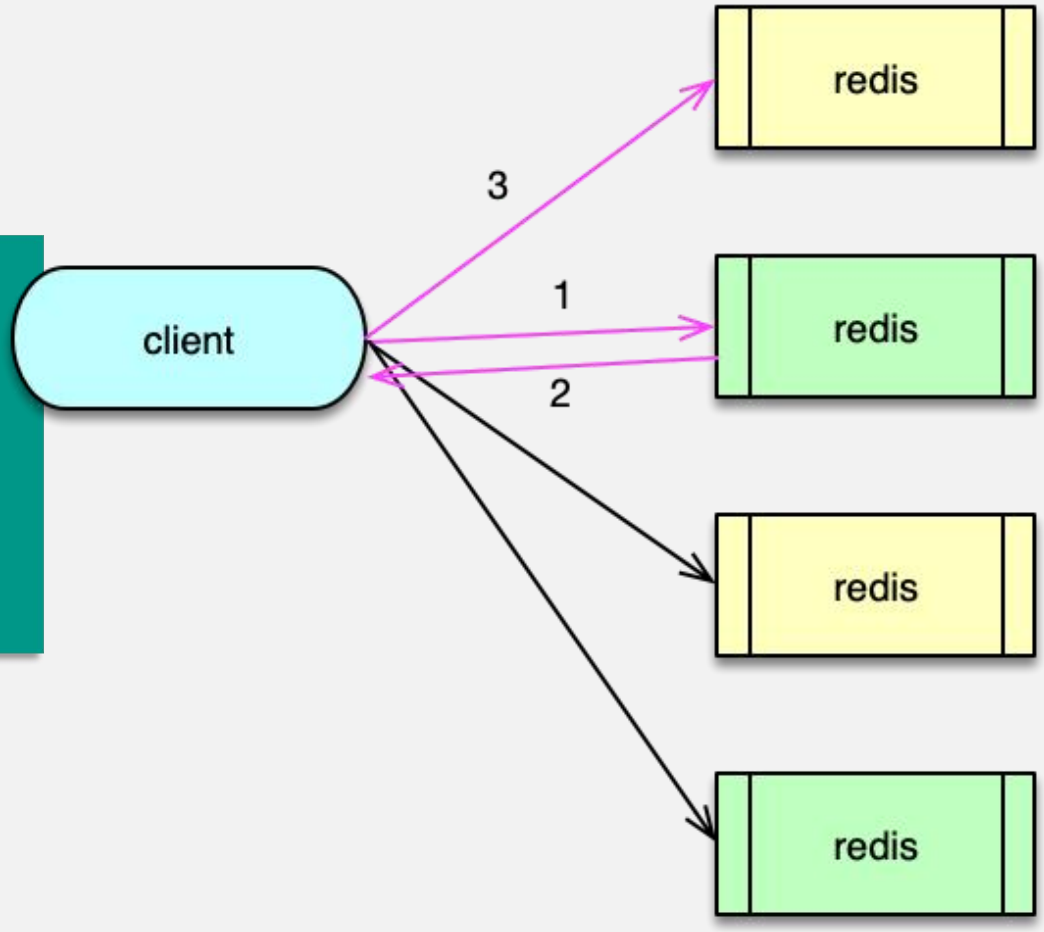


Redis 集群

拉勾 LAGOU.COM

Redis-Cluster key访问

- Cluster 模式下key访问流程
  - Redis 发现key对应slot不在本地，返回错误响应，附带上key的slot、正确的host/port
  - Client 根据返回的host/port信息，重定向请求
  - Client 可以缓存slot与节点对应关系，加速访问

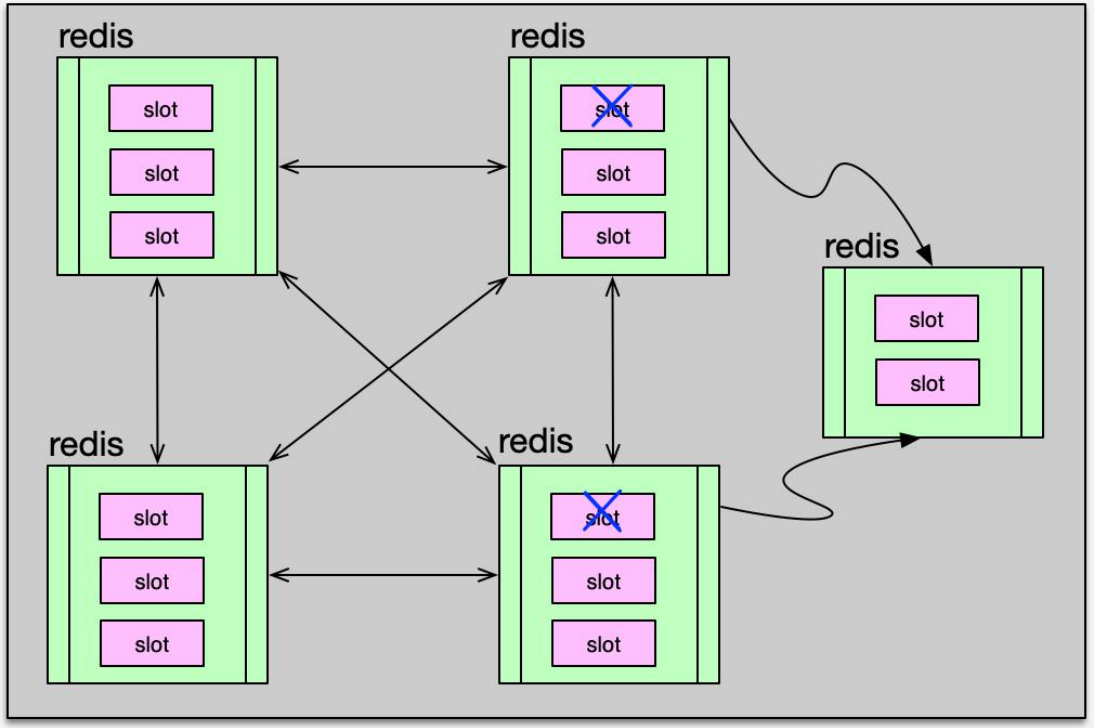


Redis 集群

拉勾 LAGOU.COM

Redis-Cluster 扩缩

- 集群的扩缩容通过迁移slot实现
- Redis-Cluster 扩容流程
  - 准备待迁入slot的Redis节点，开始集群模式
  - 将新节点加入到集群（cluster meet）
  - 将新节点的待迁入slot设为importing状态（cluster setslot）
  - 将待迁出的节点的slot设为migrating状态（cluster setslot）

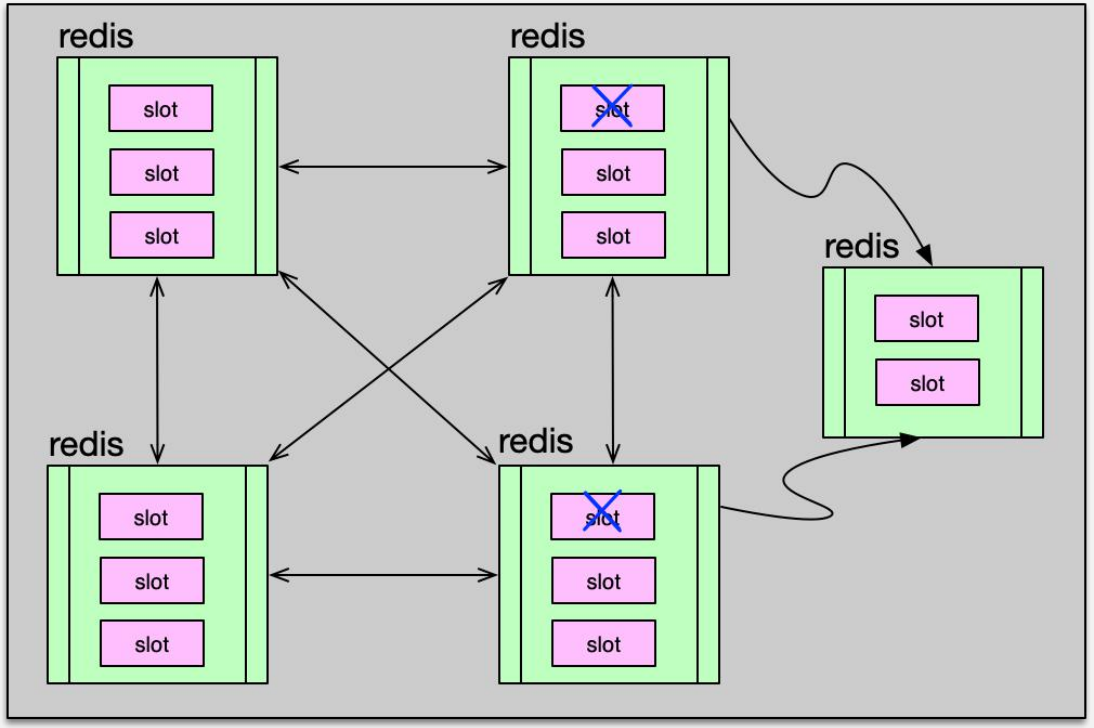


Redis 集群

拉勾 LAGOU.COM

Redis-Cluster 扩缩

- Redis-Cluster 扩容流程
  - 从待迁出节点获取slot里的N个key（cluster getkeysinslot）
  - 将N个key依次或批量迁移到新节点（migrate）
  - 循环前面2个步骤，迁移待迁移slot的所有数据
  - 将迁移的slot指派给新节点（cluster setslot）
  - 新节点增加从库（cluster replicate）

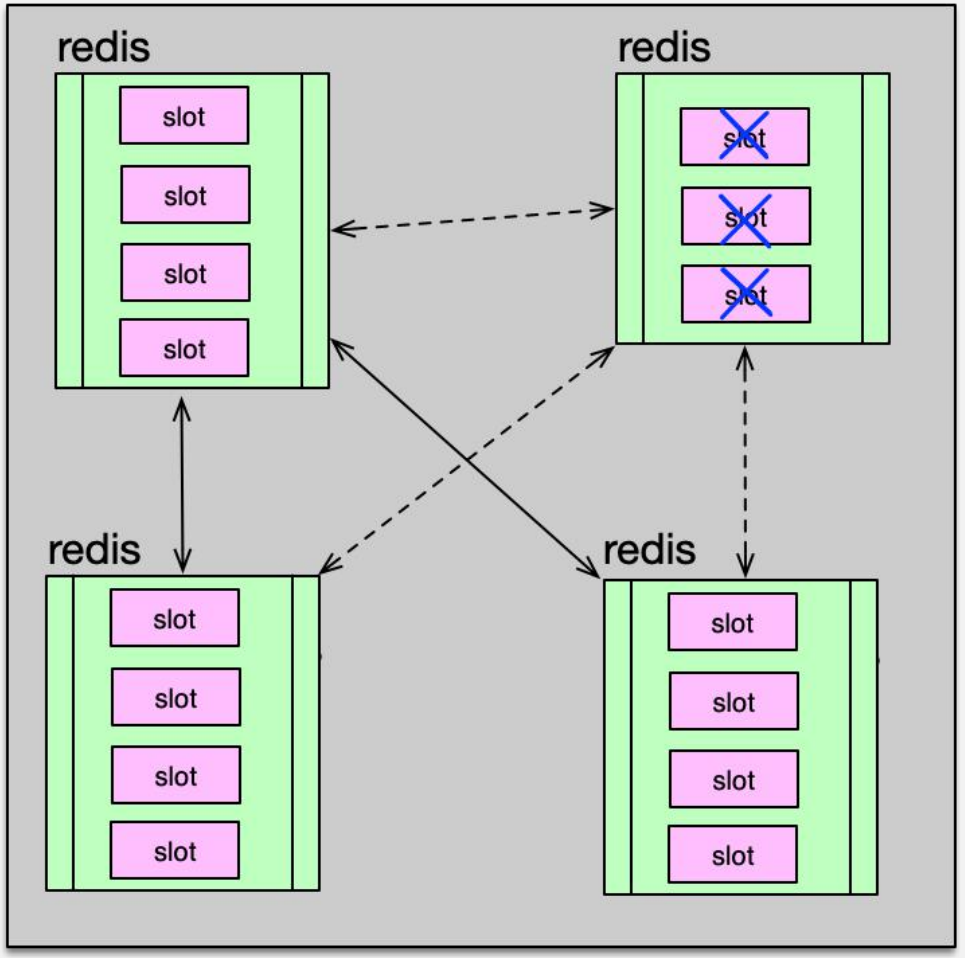


Redis 集群

拉勾 LAGOU.COM

Redis-Cluster 扩缩

- 缩容流程
  - 与扩容流程类似，只是节点变少
  - 用cluster forget 将去掉的节点从集群移除
- Redis 附带 redis-trib.rb 工具，设置迁移计划，由 redis-trib 执行扩缩中的命令操作

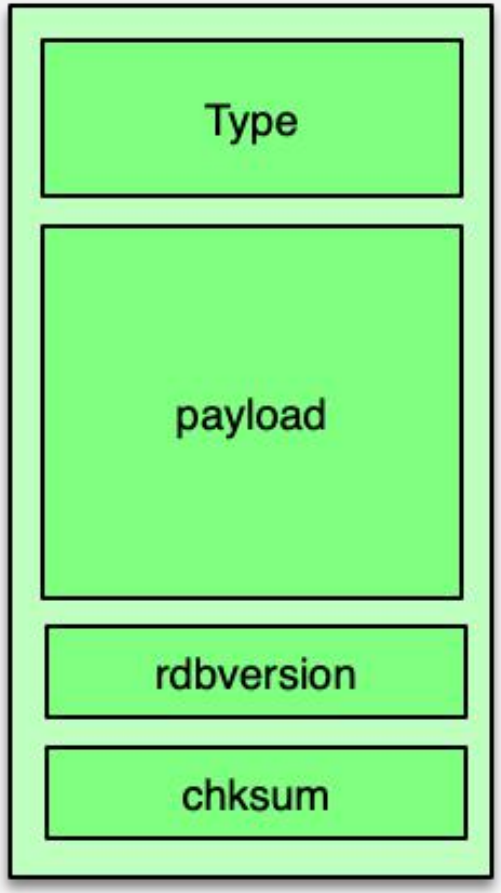


Redis 集群

拉勾 LAGOU.COM

Redis-Cluster key迁移

- Slot 迁移过程不影响正常访问，但key迁移是阻塞进行
- slot内的key迁移通过migrate进行
- Migrate key 处理流程
  - 源节点构建与迁移目标节点的socket连接
  - 发送select $dbid 指令设置key所在的db

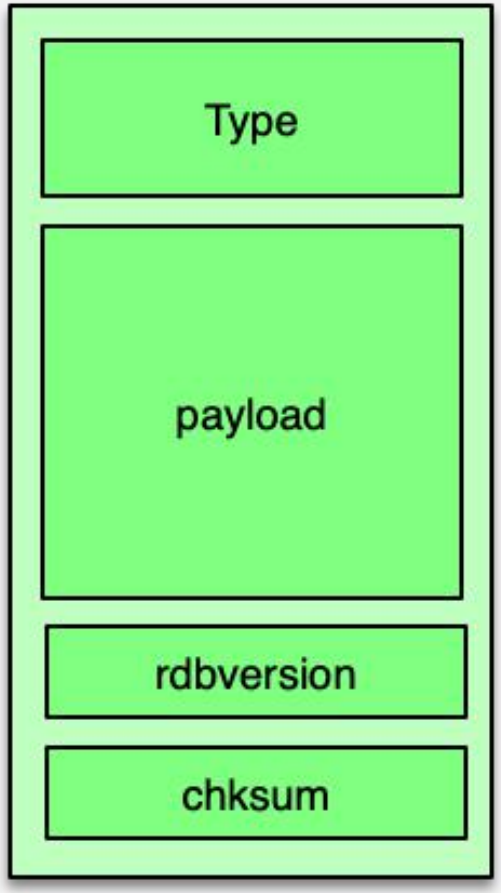


Redis 集群

拉勾 LAGOU.COM

Redis-Cluster key迁移

- Migrate key 处理流程
  - 将待迁移的value 进行dump成类rdb的二进制格式
    - 头部：value对象type
    - Body：value对象的实际数据
    - 尾部：rdb 版本 及 CRC64校验码
  - 源节点通过restore-asking 指令将过期时间、key、value的二进制数据发给目标节点

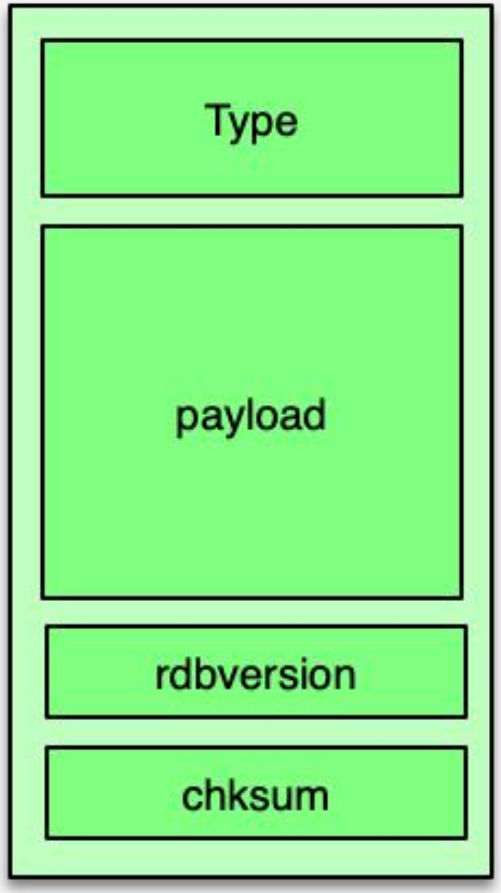


Redis 集群

拉勾 LAGOU.COM

Redis-Cluster key迁移

- Migrate key 处理流程
  - 目标节点将数据解析校验后存入redisDb，返回响应
  - 源节点收到响应，删除key，迁移完成
  - 迁移指令一次可以迁移1个或多个key
  - 迁移指令执行期间，源节点短时阻塞，直到迁移完成

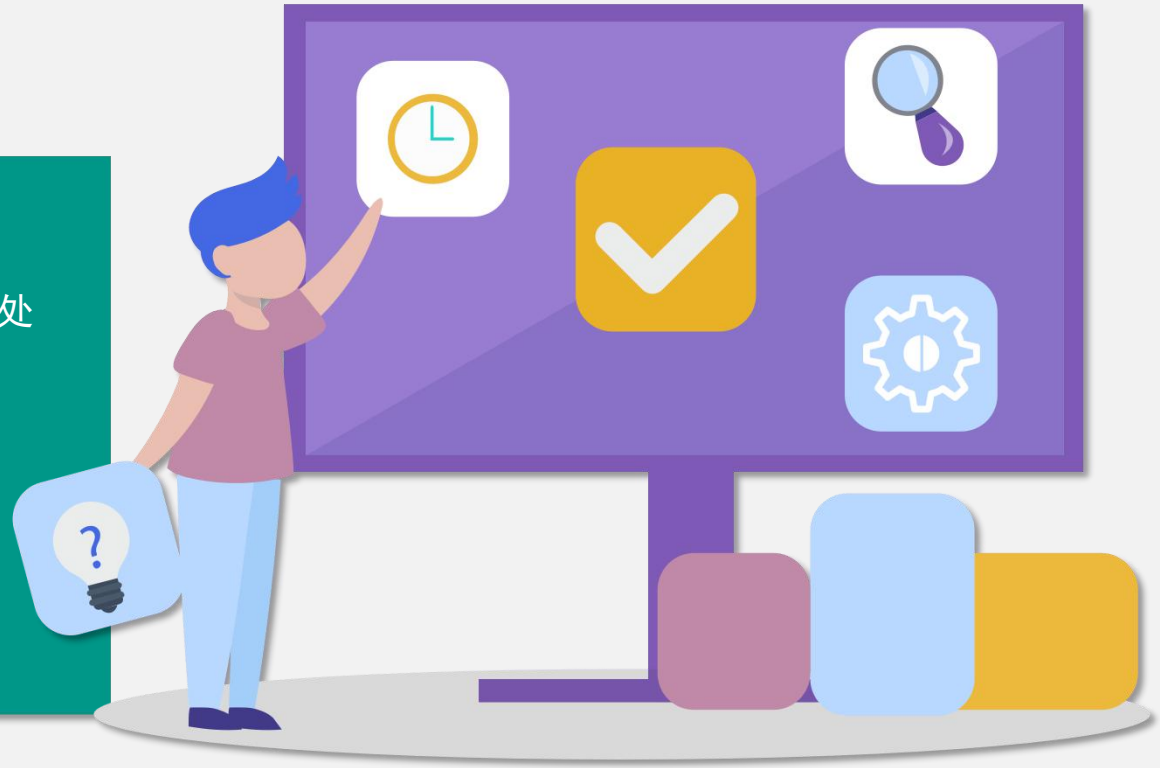


Redis 集群

拉勾 LAGOU.COM

Redis-Cluster 迁移中读写处理

- Slot迁移过程中，该slot中的key仍然可读写
- Key迁移为阻塞模式，key迁移过程中，对应节点不处理任何请求
- Slot 迁移过程中，待读写的key只有3种存在方式
  - 尚未被迁移
  - 已被迁移
  - 不在集群中，属于新key

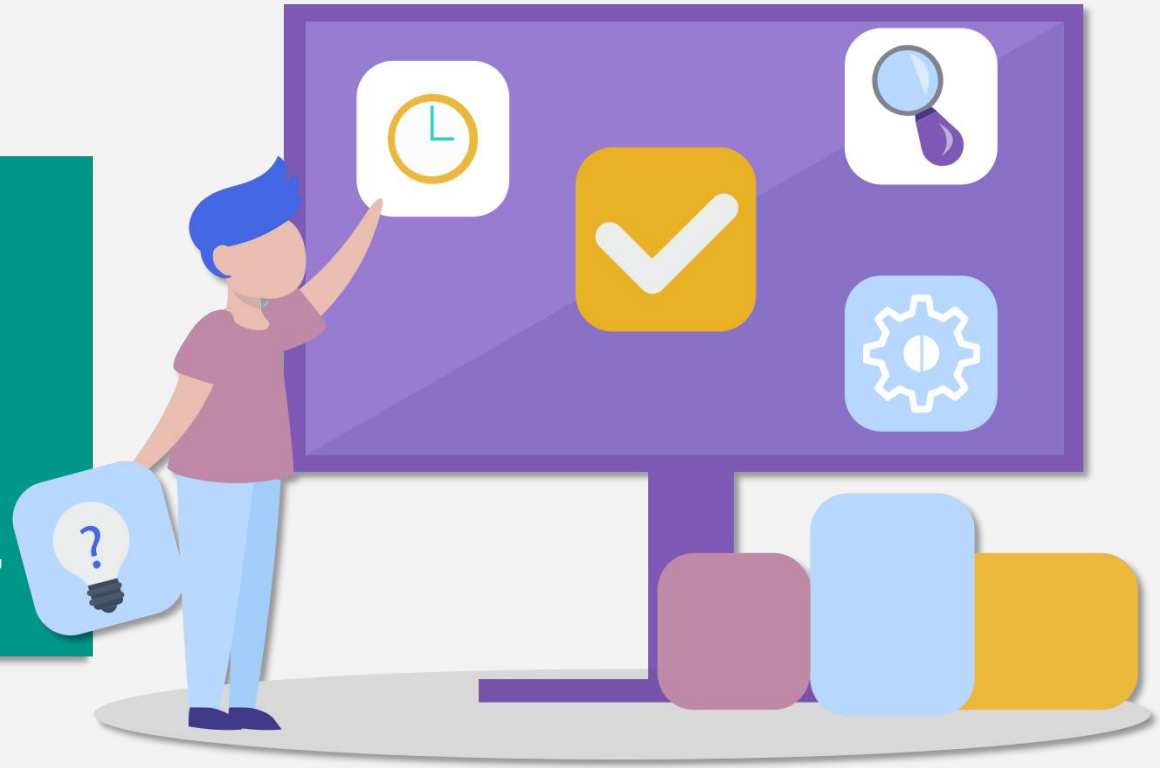


Redis 集群

拉勾 LAGOU.COM

Redis-Cluster 迁移中读写处理

- Slot 迁移过程中，对该slot的key访问处理方式
  - Key 尚未被迁移，直接在本地进行读写
  - Key 不存在，且属于本节点正迁移的slot，Redis 返回ask，附带目标节点的ip/port
  - Client 根据Ask响应，重定向请求到新节点
  - Key 不存在，key 所在的slot不属于本节点，Redis返回 moved，附带目标节点的ip/port



Redis 集群

拉勾 LAGOU.COM

Redis-cluster模式分析

- 优势
  - 社区官方实现，并有相关辅助工具
  - 支持扩缩，集群状态随时可查
- 不足
  - 存储数据与集群逻辑耦合，逻辑复杂
  - slot映射还需额外内存占用
  - key迁移阻塞式，迁移大value 导致短时不可用
  - 迁移效率低
  - 集群复制只可从master复制，不支持嵌套

Next：课时29《从容应对亿级 QPS 访问，Redis 还缺少什么》