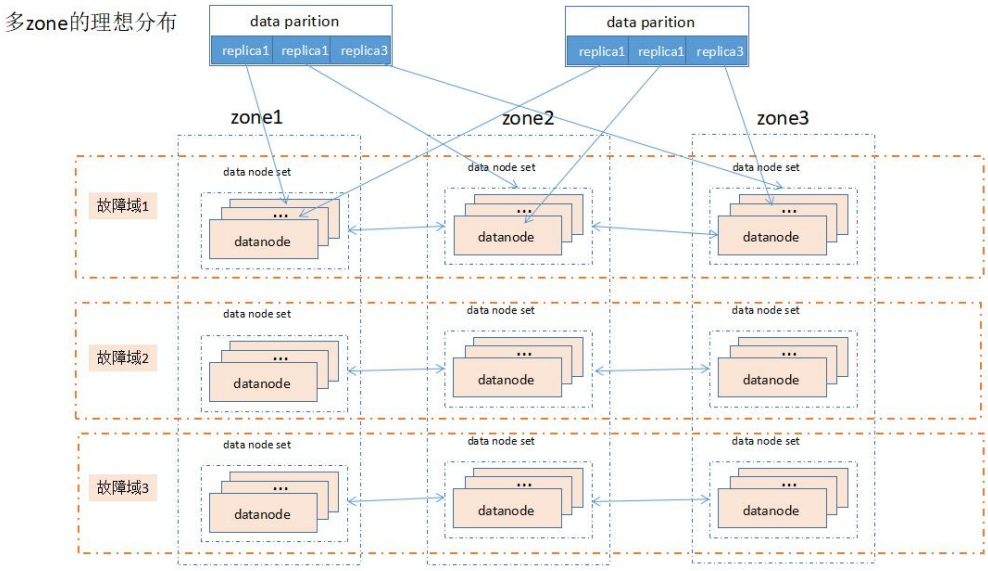
需求

1. 支持 zone 配置和展示
2. zone 对应多个 nodeset 的关系
3. nodeset 配置和展示
4. nodeset 资源不足后管控资源调配的支持
5. node 所在多个 zone 内 nodeset 结对
6. volome 级别配置 zone 上数据分布策略，默认三个 zone 分布。
7. 平滑升级



Nodeset 结对

增加 host 时，创建 nodeset，继而自动创建 nodeSetGroup
Nodesetgrp 创建后，才提供给 partition 创建

1. Nodeset 创建或者加载时，提交给管理器待结对的 nodeset 信息，触发结对
2. 持久化到 rocksdb，重启时加载结对信息

Zone

配置和展示

1. 获取 zone 下 nodeset 及其使用率，告警显示
2. 整体 zone 的资源使用率
3. Zone 和 volume 的关系，多对多的关系

封存

1. 原有 zone 使用到空间不足，需要新建立集群，让 meta partition 和 data partition 分配到新建集群，需要监控非新建 3zone 的所有 zone 的使用情况，nodeset 的剩余情况，达到阈值，再启用 3zone 故障域
2. 新建 partiton 在 3zone 故障域中分配
3. 扩展资源，自动结对逻辑判断为 nodeset 所在 zone 为封存 zone，走普通资源添加逻辑

资源监控

1. Cluster 内启用一个协程，定时 check nodeset，是否增加资源
2. Cluster 内启用一个协程，定时 check old zone，老业务是否启动故障域
3. 管控面定时拉去 zone 的需要增加的 nodeset 显示
4. Nodeset 内资源不足，监控等待管控添加资源，节点支持指定 nodeset（动态也可以）

升级及配置项

Cluster 级别配置

默认是要支持原有的配置，启用故障域需要一个特殊的配置，**增加 cluster 级别的配置，是否支持故障域**

FaultDomain bool // 默认 false

否则无法区分，新增 zone 是故障域 zone 还是归属于原有 cross_zone

Volume 级别配置

保留：

crossZone bool //跨 zone

新增：

default_priority 为 true 在生效，优先选择原有的 zone，而不是从故障域里面分配

故障域 zone 识别

1. cluster 配置故障域选项，如果此时 master 重启，zoneset 重新加载，Putnodeset 到 group 里面，如何判断 nodeset 是属于故障域呢？？
2. 如果 cluster 不配置，直接添加 zone，cluster 如何区分 zone 属于故障域？
3. 全部掉电，新增 zone 的机器和普通的机器无法区分，除了部分已经持久化

解决方案：

1. 配置当前 master 为 crosszone，master 重启，之后再添加新的 zone；
2. 重启为了将当前的 zone 持久化为非故障域 zone（持久化层没有该信息，默认当前 zone 应该全部持久化为旧的 zone（default zone））；
3. 重启后加载，后面添加新的 zone，则默认为新的故障域 zone；并持久化；

小结

1. 现有的 cluster，无论是自建的，还是社区的，无论是单个 zone，还是跨 zone，如果需要故障域启用，需要 cluster 支持，master 重启，配置更新，同时管控更新现有 volume 的策略。否则继续沿用原有策略。
2. 如果 cluster 支持，volome 不选择使用，则继续原有 volome 策略，需要在原有 zone 中按原有策略分配。原有资源耗尽再使用新的 zone 资源，
3. 如果 cluster 不支持，volome 无法自己启用的故障域策略
4. 在新的配置下，对应以下策略分配资源（灰色部分表示配置不起作用）

	Cluster：故障域	Vol：cross Zone	Vol：default priority	volume 策略
1	N			不支持故障域，维持原有策略
2	Y	N		先写旧，再写故障域
3	Y	Y	N	直接写故障域
4	Y	Y	Y	先写旧，再写故障域