

03、TiDB⽣产环境⾼可⽤分享

--2022-12-03 春雷

1、前⾔

当前公司有TiDB集群100套+，稳定运⾏了3年+,此处分享下⽇常如何保证集群的⾼可⽤

2、基本信息

2.1、机房信息

⽬前公司在⽤的两个机房，A，B机房，因为容量问题，在建设第三个机房C，可能会腾退机房 B，即还是保留2个机房的模式

|  |  |
| --- | --- |
| 机房 | 描述 |
| A机房 | 主机房 同城 |
| B机房 | 从机房 同城 |
| C机房 | 在建设的机房，同城 |

2.2、TiDB信息

|  |  |
| --- | --- |
| 分类 | 具体 |
| 集群数 | 100套+ |
| 版本 | 5.0，5.1，5.2，5.4 |
| 业务 | ⽤于⼤表、⼤流量、⽇志、分析等业务的使⽤ |

3、⾼可⽤



3.1、TiDB架构

当前所有节点都部署在⼀个机房，A机房没有使⽤同城双中⼼的⾼可⽤架构

减少⽇常的运维复杂度

规范：TiDB没有部署影响收⼊的⼀级业务，即使机房不可⽤，也不会影响收⼊

集群的备份(天级别备份)在多个机房都存储了，如果A机房不可⽤，利⽤备份恢复也可以紧急恢复

VIP为使⽤的腾讯的TGW： Tencent GateWay

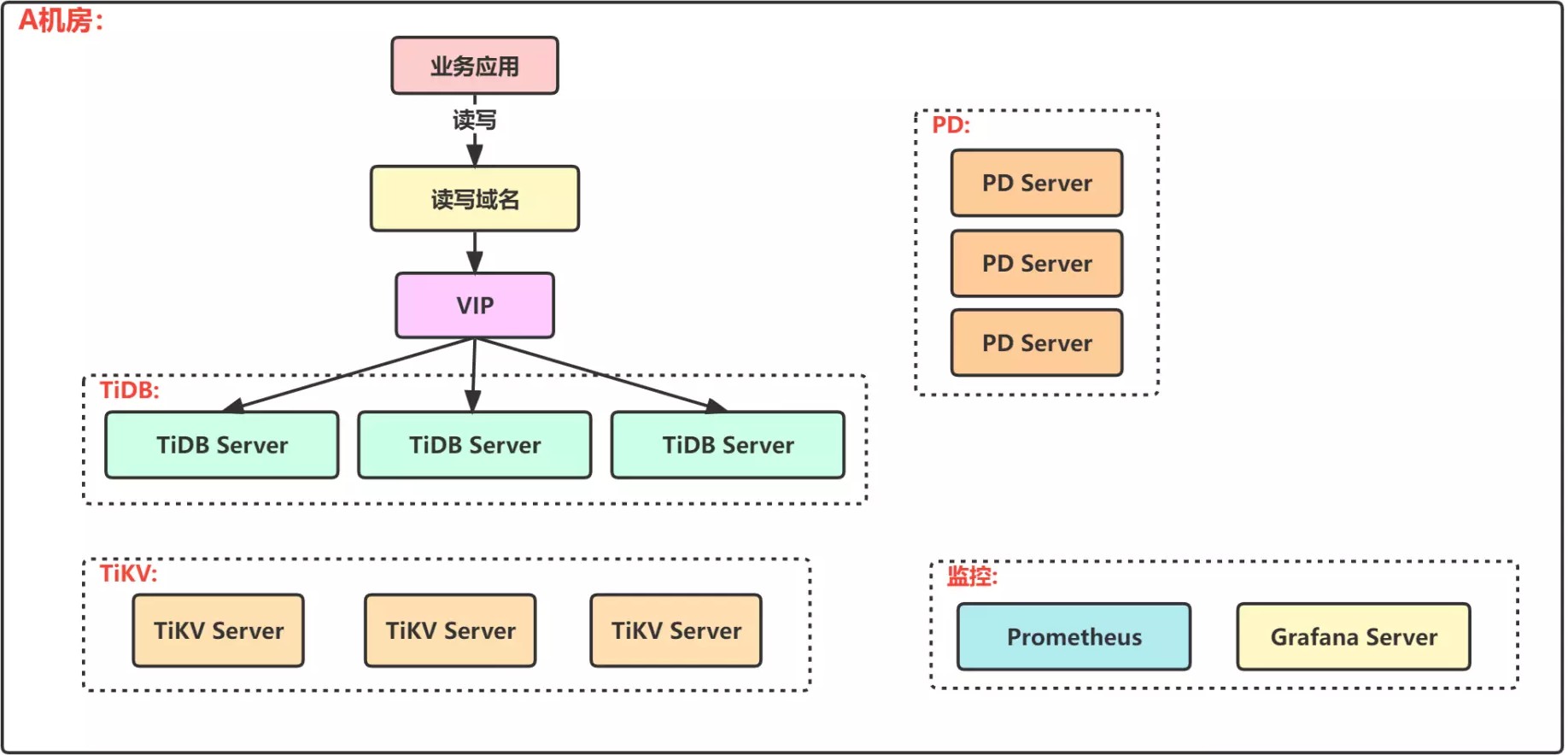
可以实现轮询访问后端多个节点，负载均衡

可以实现端⼝不通的节点不访问默认集群架构：

3个TiDB Server，每个TiDB Server不同机器 3个PD Server，每个PD Server不同机器 3个TiKV Server，每个TiKV Server不同机器

监控节点：为⼀台物理机，混合部署多个集群的监控节点，减少节点的浪费读写域名：

读写域名为直接解析到VIP，VIP后端挂在多个TiDB Server



【优点】：

运维相对简单

减少了跨机房访问的延迟单个节点的不可⽤：

PD/TiKV 可以不影响服务 TiDB有VIP访问，不可⽤则不访问，不影响服务

【缺点】：

⽆机房级别的容灾

4、⽇常故障

4.1、节点故障



4.1.1、TiDB节点故障

TiDB节点的故障，有VIP托底，不影响服务

如⽆法开启：调整VIP的配置，将故障节点进⾏强制下线，再扩容⼀个新的即可。如可以开启：再启动即可



4.1.2、PD节点故障

PD节点的故障，会导致PD节点选举，集群有1-3秒的不可⽤，切换后集群恢复默认3节点

如⽆法开启：将故障节点进⾏强制下线，再扩容⼀个新的即可如可以开启：再启动即可



4.1.3、TiKV节点故障

TiKV节点故障，会有数据leader的选举，导致部分数据不可访问，影响时间视数据量⽽定。正常⼏⼗秒即可

如⽆法开启：将故障节点进⾏强制下线，再扩容⼀个新的即可如可以开启：再启动即可

4.1.4、监控节点

不影响服务，进⾏强制下线，再使⽤新的机器扩容即可



4.1.5、TiDB中控节点宕机

当前TiDB的中控机为单节点，需要定期备份下集群的元信息

如⽆法开启：需要搭建新的机器环境，并把元信息备份放到新的机器进⾏恢复



5、备份与恢复



【⽀持备份⽅式】：

dumpling BR

元信息备份：表结构

账号授权、密码



【误操作数据】：

未开启binlog未开启ticdc

只能天级别的恢复，不⽀持定点恢复

6、TiKV部署

公司的tikv部署经历了⼤约3个阶段：



初期：

SSD磁盘

每个SSD磁盘⼀个分区每个分区⼀个tikv实例

中期：

SSD磁盘

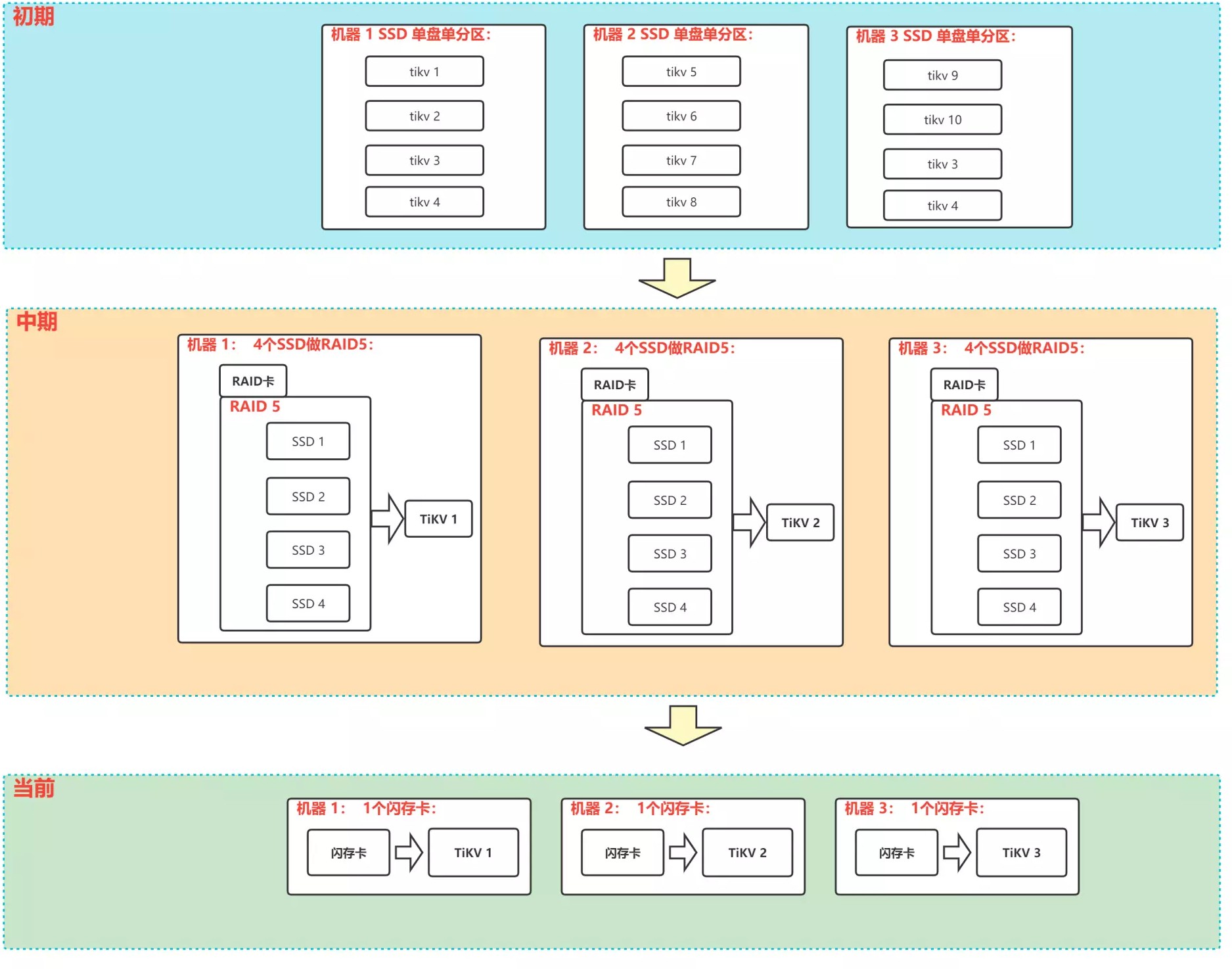
机器需要购买RAID卡

4个SSD磁盘做成⼀个RAID 5，挂载到⼀个分区

⼀个分区⼀个tikv实例

当前：

闪存卡 1个闪存卡1个分区

⼀个分区⼀个tikv实例

【图示】：



【初期部署⽅式的优缺点】：

【优点】：

提⾼了磁盘的空间利⽤率

⽆需RAID卡(但实际机器上还是有RAID卡)

【缺点】：

机器宕机会存在部分数据不可⽤的问题

分区盘损坏，有些⽆法热插拔修复，需要重启机器，影响⽐较⼤



【中期部署⽅式的优缺点】：

【优点】：

使⽤RAID后，⽀持⼀个磁盘的损坏不影响服务解决了单个物理机宕机部分数据不可⽤的问题

【缺点】：

需要使⽤RAID卡，有⼀定成本问题

存在⼀些集群的性能⽆法满⾜，只能添加多个tikv机器，有⼀定成本问题



【当前部署⽅式优缺点】：

【优点】：

物理机⽆需购买闪存卡，节约成本

闪存卡性能优于SSD做RAID5后的 ⼏倍，满⾜了业务性能需求解决了单个物理机宕机部分数据不可⽤的问题

【缺点】：

闪存卡的价格⽐单块盘的价格⾼，但是不购买RAID卡与多块SSD，整体应该相差不⼤

存储容量可能会⽐SSD RAID后的下降⼀些，不过闪存卡也⽀持3.5T,5T,7T,8T等⼤容量的

注：具体的采购价格没有详细数据，此处不详细说明了

7、服务

TiDB集群的所有节点的所有⻆⾊，都是使⽤systemd来保证服务的存活的，如果服务异常宕掉，会被systemd⾃动拉起，保证了⼀定的⾼可⽤。

例如配置⽂件：

1

cd /home/tidb/.tiup/storage/cluster/clusters/xxxx/config-cache tidb-10.1.1.1-6666.service

pd-10.2.2.2-7777.service tikv-10.3.3.3-8888.service grafana-10.4.4.4-9999.service

prometheus-10.5.5.5-5555.service node\_exporter-10.6.6.6-4444.service alertmanager-10.7.7.7-3333.service blackbox\_exporter-10.8.8.8-2222.service

2

3

4

5

6

7

8

9