本小节将展示如何退役、永久移除集群当中一个或多个节点，适用于需要缩小集群规模、处理硬件故障等情景。

关于暂时停止节点的详细介绍，可以查阅[停止节点](https://www.cockroachlabs.com/docs/stable/stop-a-node.html)章节。

概要

术语

* **Range** CockroachDB将所有用户数据和几乎所有的系统数据存储在巨型的K-V有序映射中。这个Key值空间都切分为“Range"（包含连续的Key值块），所以每一个key都会落在一个Range上。
* **Range副本冗余** CockroachDB对每个Range做数据冗余（默认情况下是3份副本），并分别存储在不同的节点上。
* **Range租约** 对于每个Range来说，其中一个副本持有“Range租约”。这个副本（即“租约持有者”）接受和协调该Range的所有读写请求。

实现机制

当用户主动退役节点的时候，CockroachDB将让节点完成所有正在执行的请求，拒绝任何新的请求，同时会将该节点上的Range副本和Range租约迁移到其他节点上，从而保证节点可以安全下线。

退役后的节点可根据情况选择移除节点或是重新服役节点。

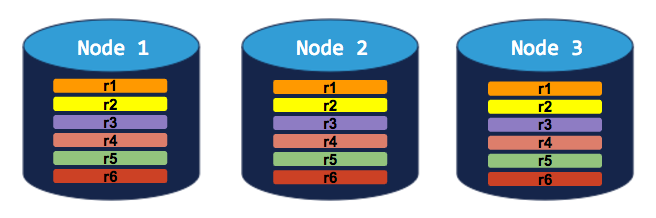
用户须知

当退役节点时，需要确保有其他节点可用于从该节点接管Range副本。如无可用的其他节点，该退役操作会无限期挂起。

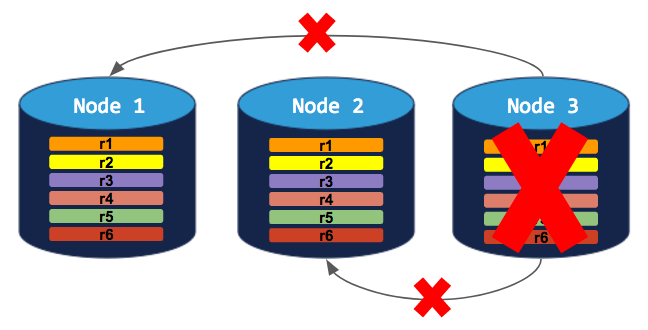
示例

3数据副本的3节点集群

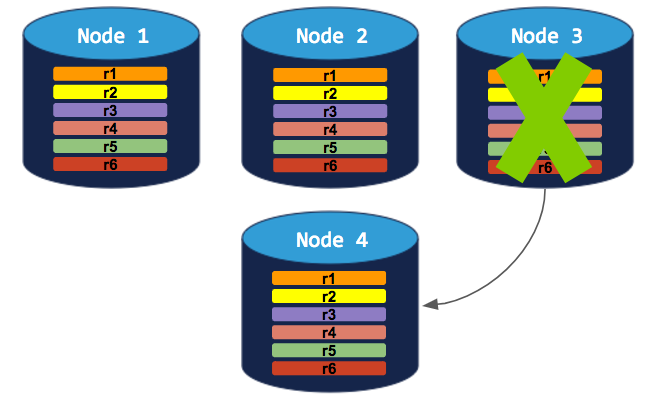
该场景下每个Range复制3份，且每个副本位于不同的节点上。



此时若用户退役其中一个节点，退役操作将无限期挂起。因为其他节点已经拥有一份完整的Range副本，无法接收退役节点上的Range副本。

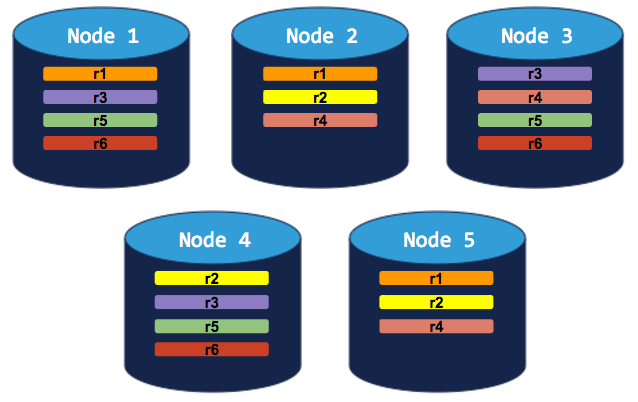


为了成功退役节点，首先需要新增第四个节点。

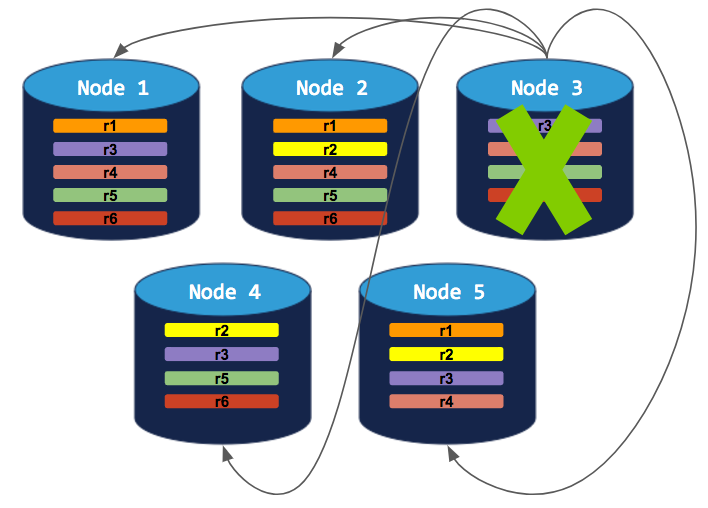


3数据副本的5节点集群

类似上面的场景，每个Range冗余3份副本，每个副本位于不同节点上。

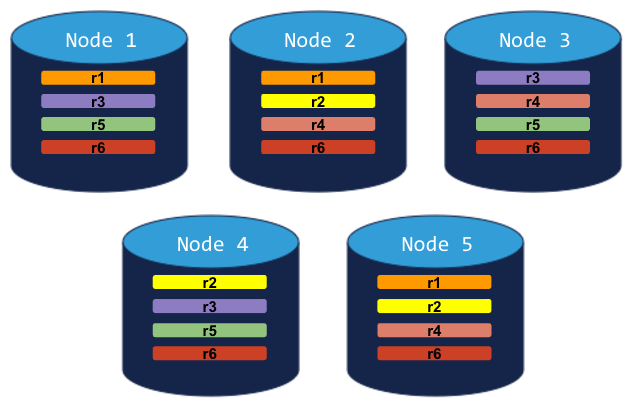


由于集群能够将待退役的节点副本迁移至其他节点而不会引起任何Range副本重复，所以用户退役节点操作可成功执行完成。

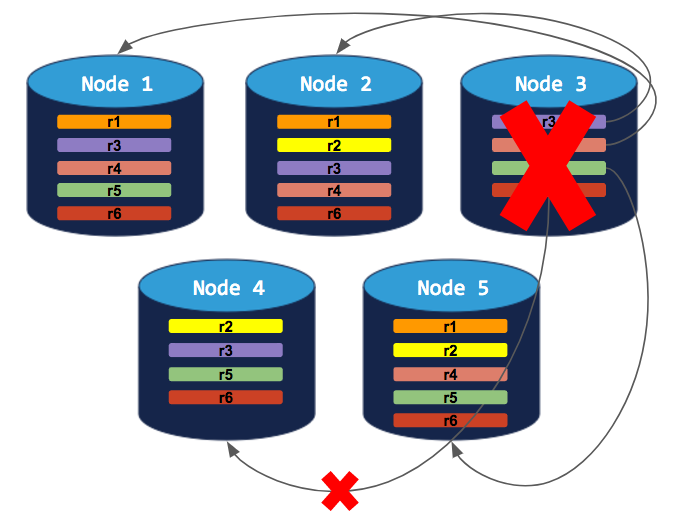


针对特定表的5份数据副本的5节点集群

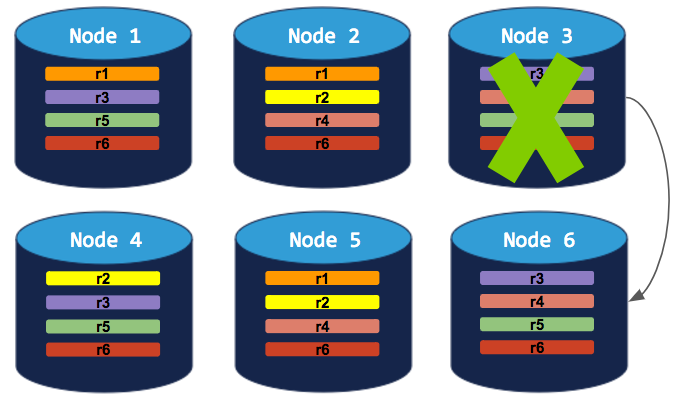
该场景下集群针对特定表数据（参照下图Range6）进行了5份数据冗余，其他数据（参照下图Range1～5）进行3份数据冗余。



当用户退役其中一个节点时，集群能够成功地将除了range 6之外的所有range重新排布。但因为range6需要5个数据副本（基础表的副本定义），且CockroachDB不允许一个节点的任意一个Range拥有超过两个的副本，所以退役操作将无限期挂起。



为了成功退役节点，首先需要新增第六个节点。



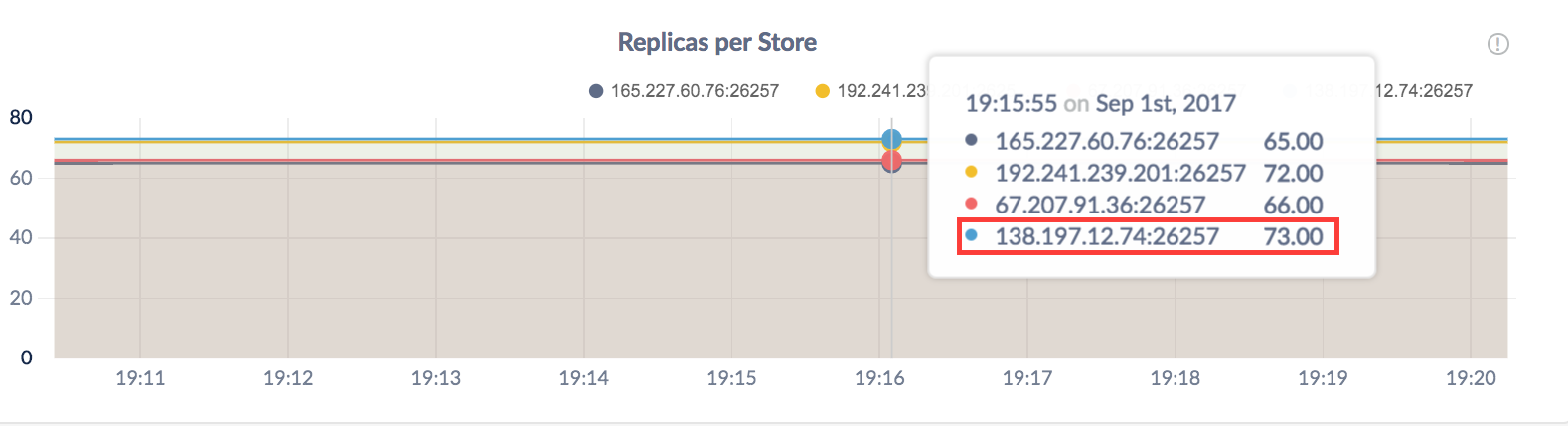
移除单个节点（存活状态）

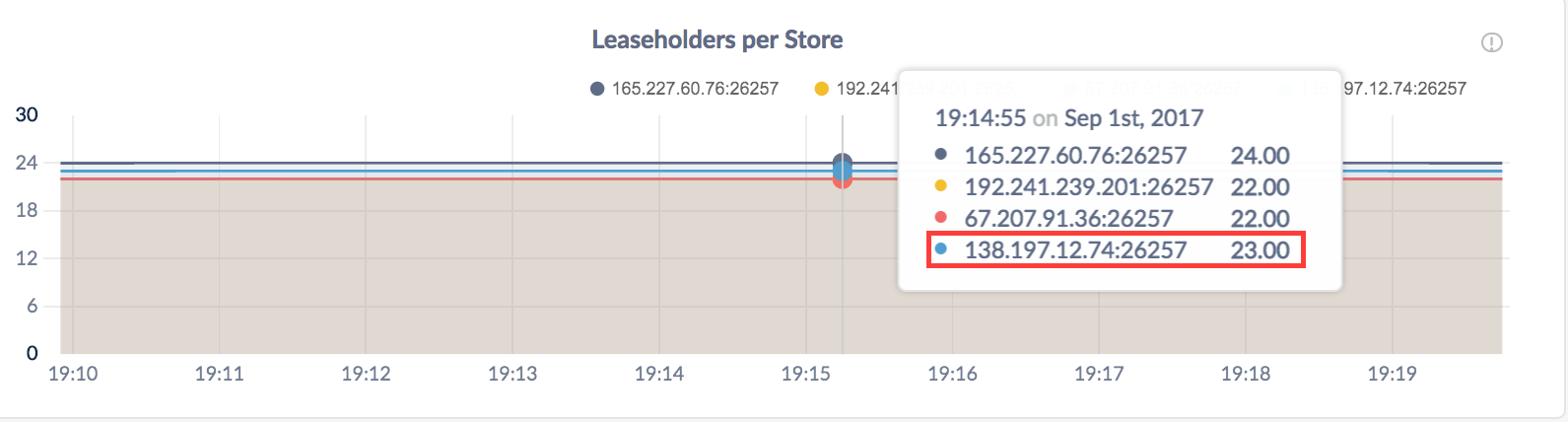
准备

务必确保移除节点的Range副本能够正常迁移到集群的其他节点，而不会导致移除操作的无限期挂起。

Step 1: 检查退役节点

登陆Admin界面，点击左方**Metrics**页，选择**Replication**仪表盘，查看**Replicas per Store**和**Leaseholder per Store**图：





Step 2: 退役和移除节点

SSH登陆想要移除的节点所在的机器，执行带--decommission及其他必需的flag的cockroach quit命令：

*#非安全模式*

cockroach quit --decommission --insecure --host=<address of node to remove>

*#安全模式*

cockroach quit --decommission --certs-dir=certs --host=<address of node to remove>

退役操作的状态信息将输出到stderr当中：

+----+---------+-------------------+--------------------+-------------+

| id | is\_live | gossiped\_replicas | is\_decommissioning | is\_draining |

+----+---------+-------------------+--------------------+-------------+

| 4 | true | 73 | false | false |

+----+---------+-------------------+--------------------+-------------+

(1 row)

+----+---------+-------------------+--------------------+-------------+

| id | is\_live | gossiped\_replicas | is\_decommissioning | is\_draining |

+----+---------+-------------------+--------------------+-------------+

| 4 | true | 73 | true | true |

+----+---------+-------------------+--------------------+-------------+

(1 row)

一旦节点完成退役并停止，将输出以下信息：

+----+---------+-------------------+--------------------+-------------+

| id | is\_live | gossiped\_replicas | is\_decommissioning | is\_draining |

+----+---------+-------------------+--------------------+-------------+

| 4 | true | 13 | true | true |

+----+---------+-------------------+--------------------+-------------+

(1 row)

+----+---------+-------------------+--------------------+-------------+

| id | is\_live | gossiped\_replicas | is\_decommissioning | is\_draining |

+----+---------+-------------------+--------------------+-------------+

| 4 | true | 0 | true | true |

+----+---------+-------------------+--------------------+-------------+

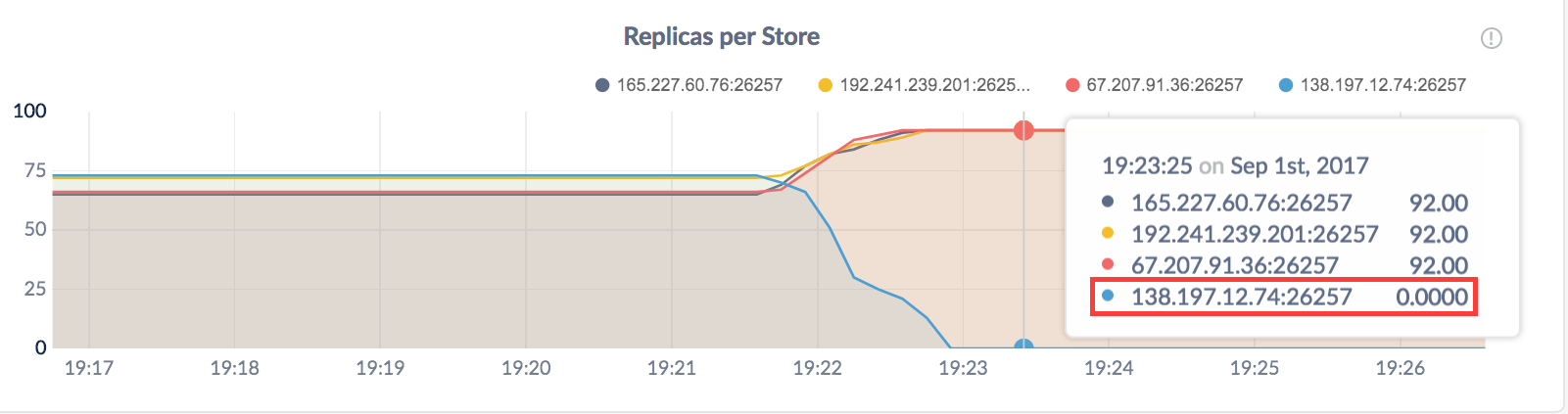
(1 row)

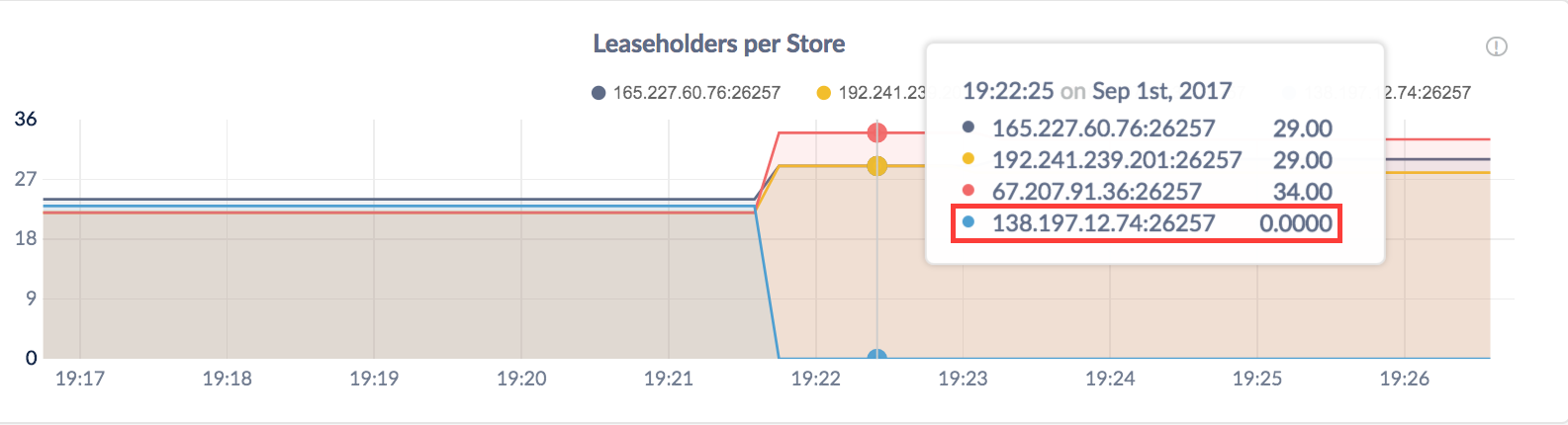
All target nodes report that they hold no more data. Please verify cluster health before removing the nodes.

ok

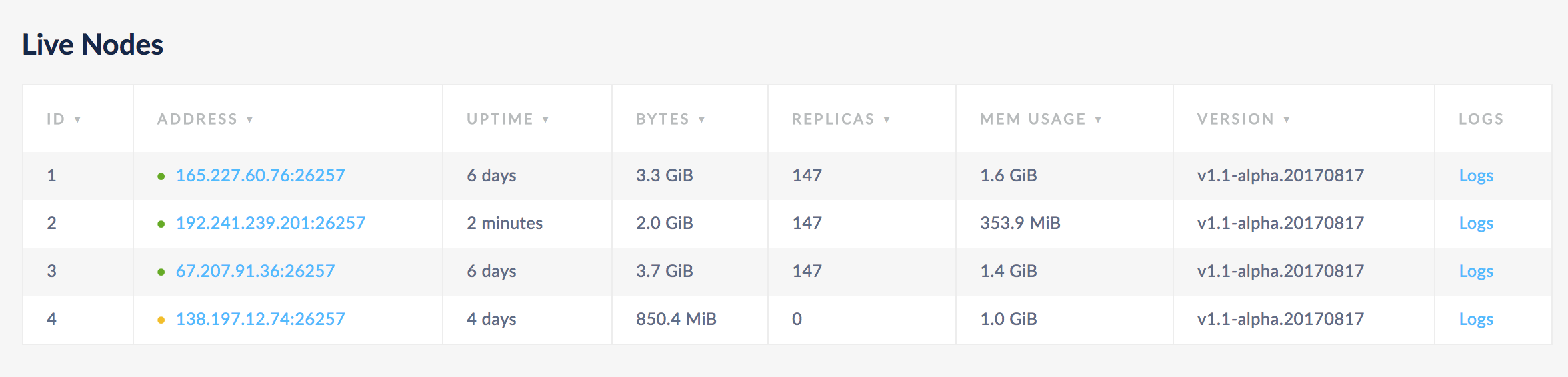
退役后检查剩余节点和集群状态

登陆Admin界面进入**Replication**仪表盘，查看**Replicas per Store**和**Leaseholder per Store**图，可以查看到被退役的节点副本数量降为0：

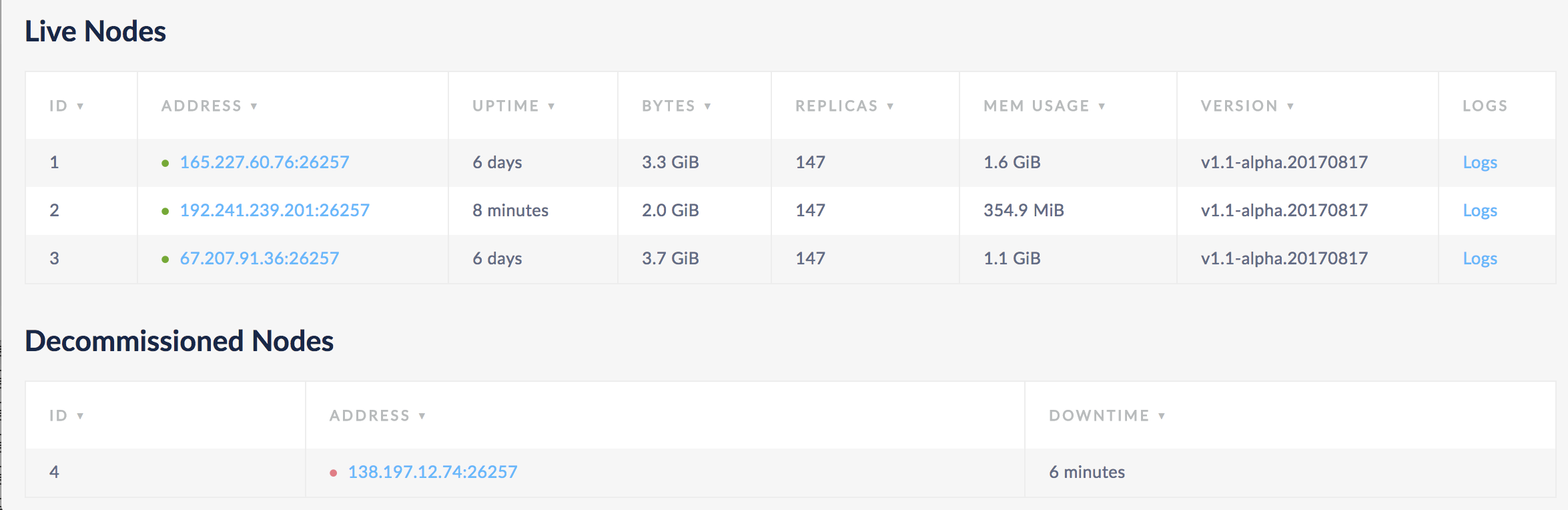




查看**Overview**页的**Node List**，可以查看到除了移除的节点以外的所有节点处于健康的状况（标注为绿色）：



大约5分钟后，被移除的节点将被移动到**Decommissioned nodes**当中：



（New in v2.0） 除非用户正在查询的时间范围包含节点正处于活跃的时间，不然节点将不再出现在时序图中。但是，该节点可以在**Decommissioned Nodes**列表查看到。

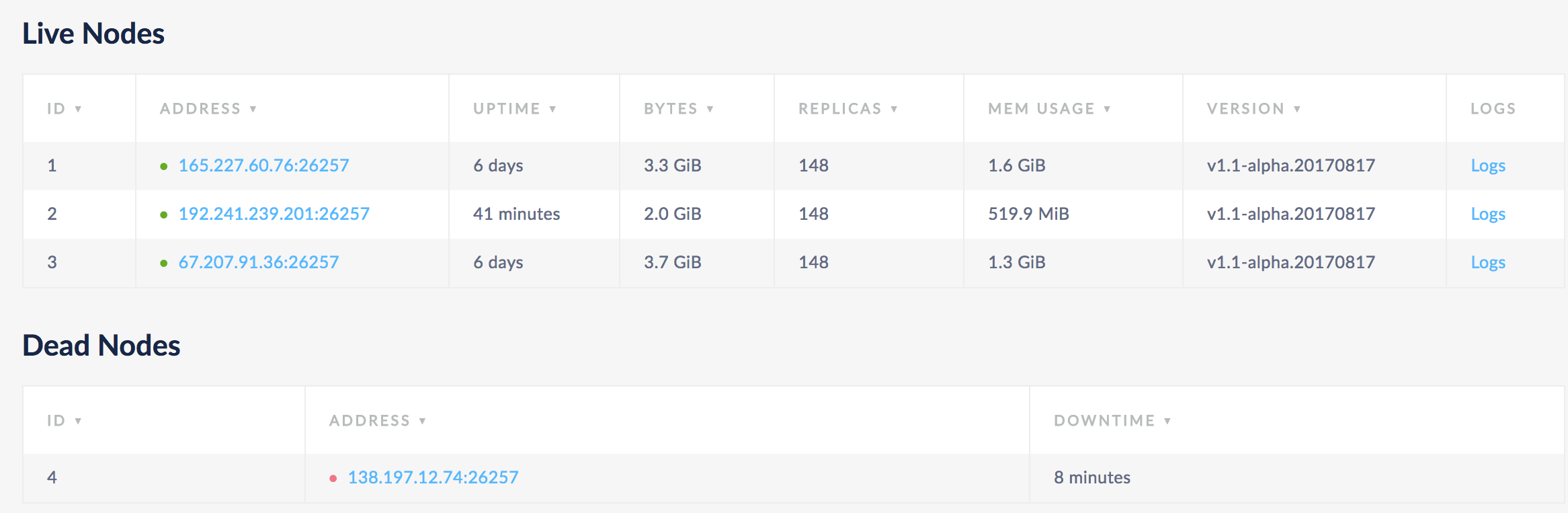
如果被移除的节点要再次加入集群，在节点[重新服役](http://doc.cockroachchina.baidu.com/#maintain/remove-nodes/#recommission-nodes)之前，集群无法使用该节点进行Range副本均衡分布操作，该节点也无法接收来自客户端的连接。

移除单个节点（死亡状态）

一旦节点失联超过5分钟，集群自动将该节点上的Range副本和Range租约迁移到其他存活的节点上。而当该节点重新连入集群，集群将重平衡Range副本并再次启用该节点。

为了避免集群重平衡数据到一个“死而复生”的节点，请执行以下操作：

Step 1: 获取死亡节点的ID

登陆Admin界面选择**Node List**视图，在**Dead Nodes**列表上获取死亡节点的ID： 

Step 2: 标记死亡节点为退役

SSH登陆集群的任意可用节点，根据获取到的死亡节点ID，执行[cockroach node decommission](https://www.cockroachlabs.com/docs/stable/view-node-details.html)命令：

*#非安全模式*

cockroach node decommission 4 --insecure --host=<address of live node>

*#安全模式*

cockroach node decommission 4 --certs-dir=certs --host=<address of live node>

+----+---------+-------------------+--------------------+-------------+

| id | is\_live | gossiped\_replicas | is\_decommissioning | is\_draining |

+----+---------+-------------------+--------------------+-------------+

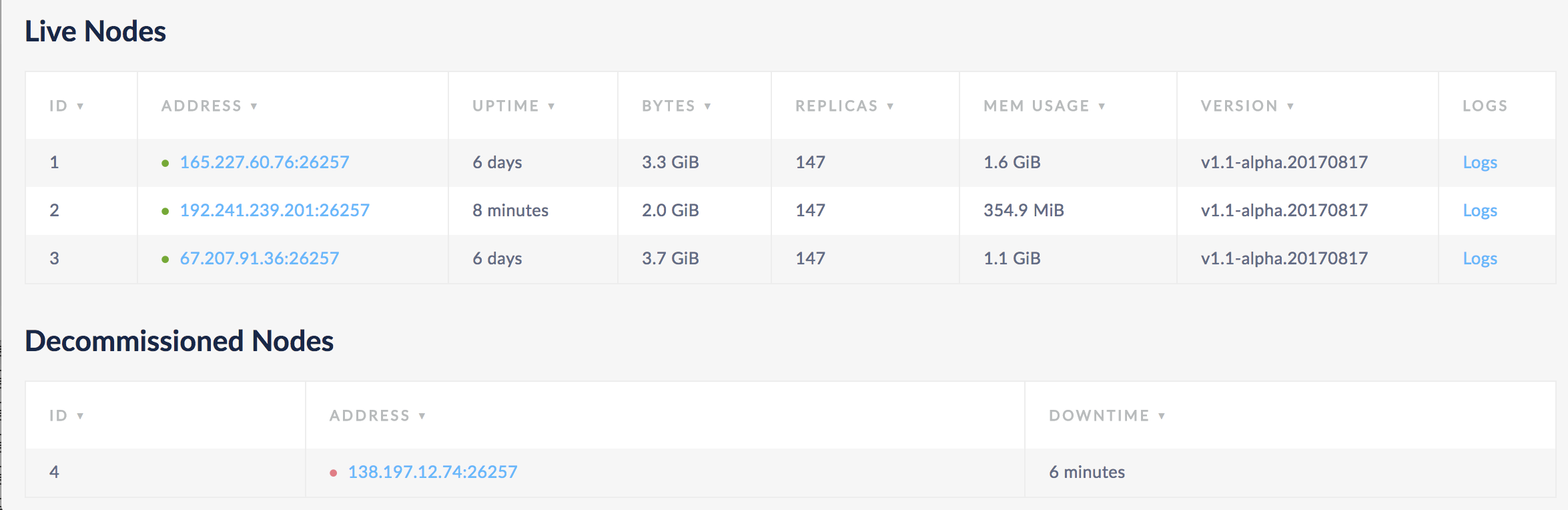
| 4 | false | 12 | true | true |

+----+---------+-------------------+--------------------+-------------+

(1 row)

Decommissioning finished. Please verify cluster health before removing the nodes.

（New in v2.0）在**Node Lists**页，在上述命令执行完成5分钟后，用户将查看到被退役的节点从**Dead Nodes**列表移动到**Decommissioned Nodes**列表。除非用户正在查询的时间范围包含节点正处于活跃的时间，不然节点将不再出现在时序图中。但是，该节点可以在**Decommissioned Nodes**列表查看到。



此外，即使节点重启了，它也不会接受任何客户端的连接，同时集群也不会重平衡数据到该节点。要使集群再次启用该节点，必须[重新服役](http://doc.cockroachchina.baidu.com/#maintain/remove-nodes/#recommission-nodes)它。

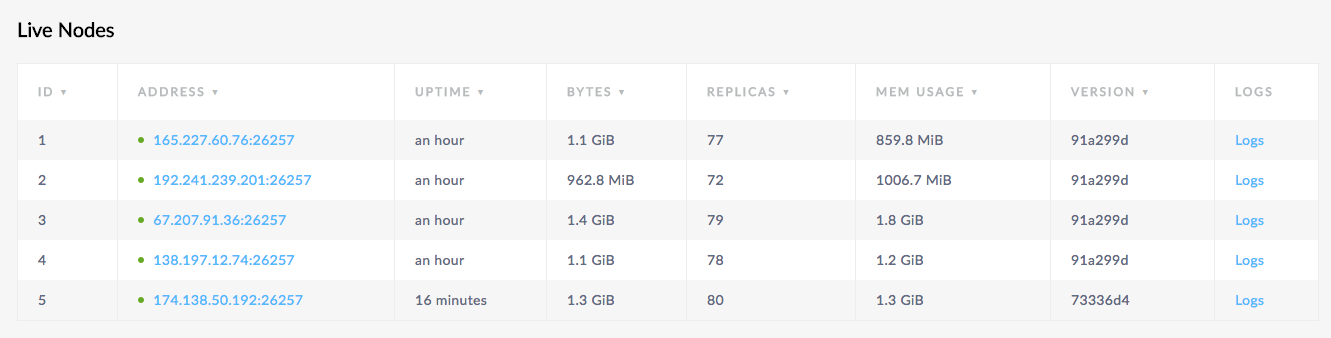
移除多个节点

准备

务必确保移除节点的Range副本能够正常迁移到集群的其他节点，而不会导致移除操作的无限期挂起。

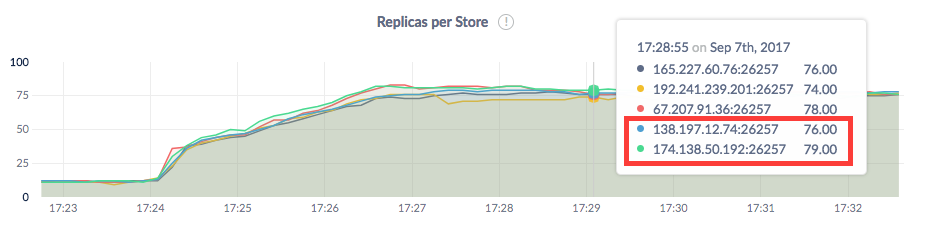
Step 1: 获取将要退役的节点的ID

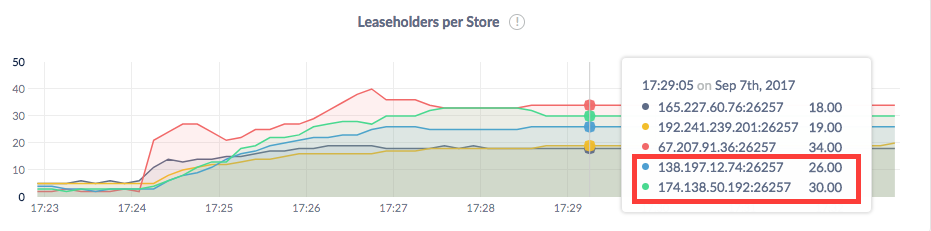
登陆Admin界面，点击**Node List**视图，或是在左方**Metrics**页的**Summary**区域点击**View nodes list**。查看用户想要退役的节点的ID：



Step 2: 检查节点状态

点击**Replication**仪表盘，查看**Replicas per Store**和**Leaseholders per Store**图：





Step 3: 退役节点

SSH登陆集群任意可用节点，基于获取到的节点ID执行cockroach node decommission命令来正式退役：

cockroach node decommission 4 5 --certs-dir=certs --host=<address of live node>

退役操作的状态信息将输出到stderr当中：

+----+---------+-------------------+--------------------+-------------+

| id | is\_live | gossiped\_replicas | is\_decommissioning | is\_draining |

+----+---------+-------------------+--------------------+-------------+

| 4 | true | 8 | true | false |

| 5 | true | 9 | true | false |

+----+---------+-------------------+--------------------+-------------+

(2 rows)

+----+---------+-------------------+--------------------+-------------+

| id | is\_live | gossiped\_replicas | is\_decommissioning | is\_draining |

+----+---------+-------------------+--------------------+-------------+

| 4 | true | 8 | true | true |

| 5 | true | 9 | true | true |

+----+---------+-------------------+--------------------+-------------+

(2 rows)

一旦节点完成退役，将输出以下信息：

+----+---------+-------------------+--------------------+-------------+

| id | is\_live | gossiped\_replicas | is\_decommissioning | is\_draining |

+----+---------+-------------------+--------------------+-------------+

| 4 | true | 0 | true | true |

| 5 | true | 0 | true | true |

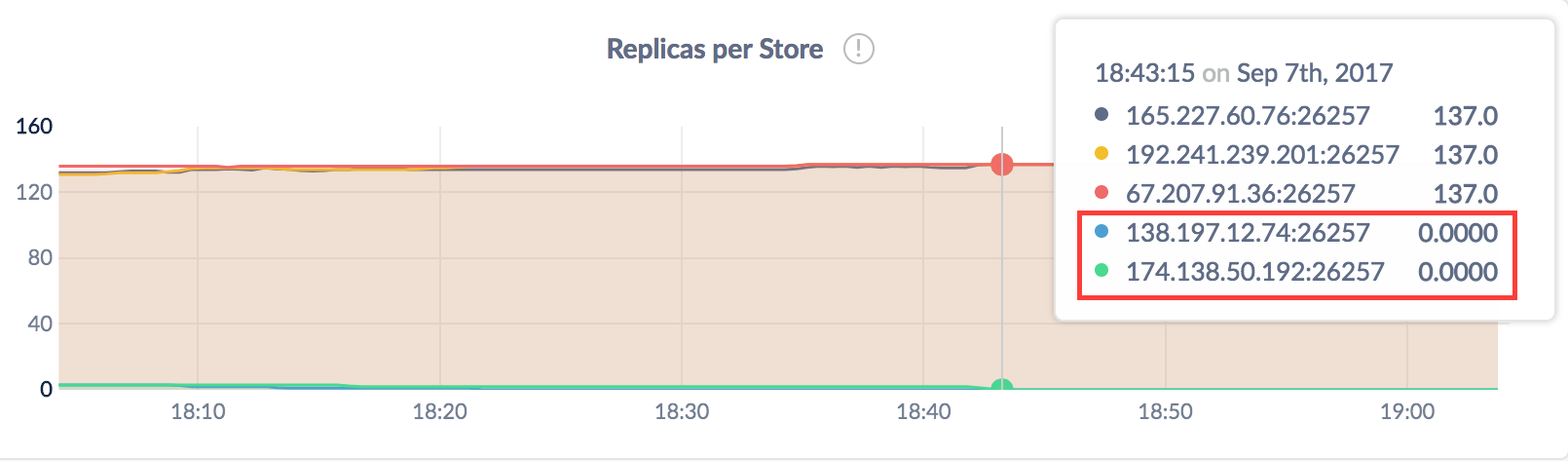
+----+---------+-------------------+--------------------+-------------+

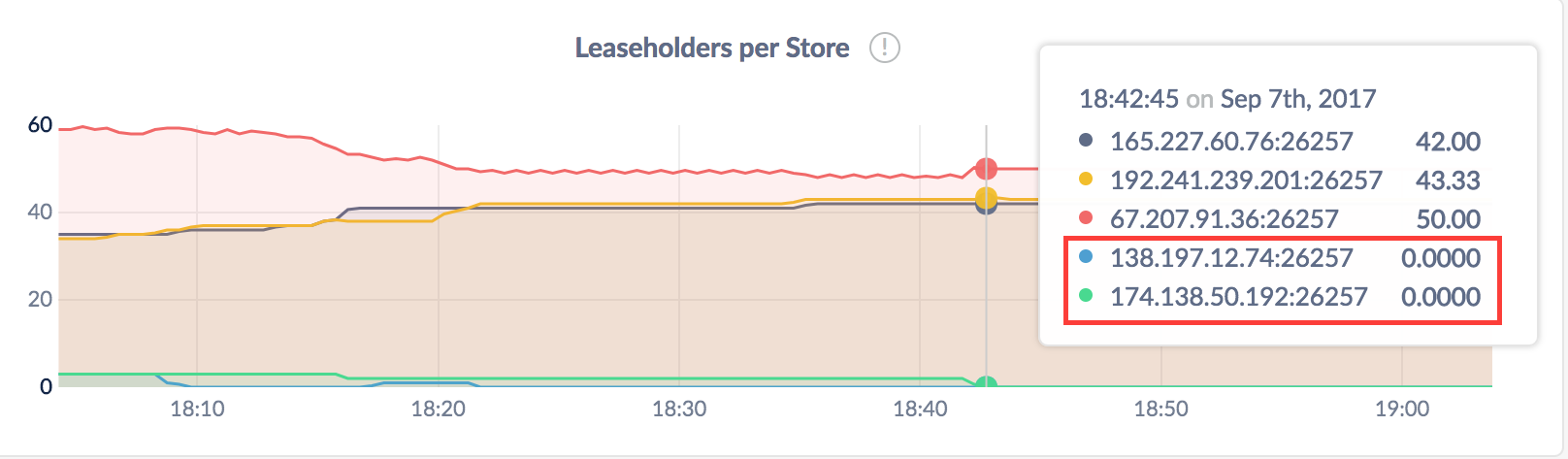
(2 rows)

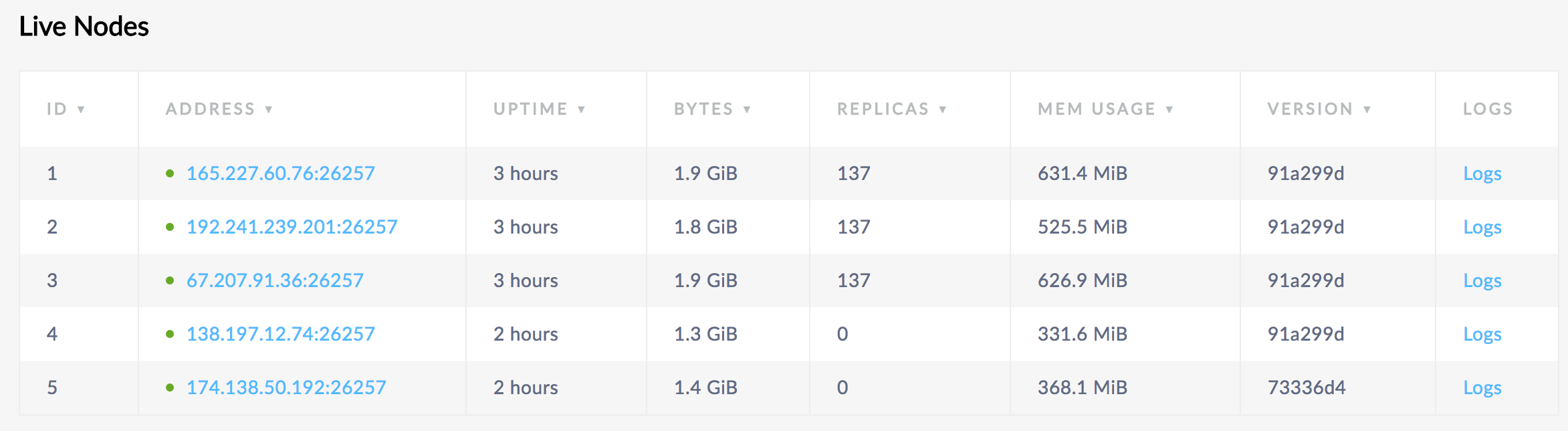
Decommissioning finished. Please verify cluster health before removing the nodes.

Step 4: 查看退役后剩余节点和集群状态

登陆Admin界面选择**Replication**仪表盘，查看**Replicas per Store**和**Leaseholder per Store**图，可以查看到被退役的节点副本数量降为0：





点击**View node lists**页的**Summary**区域，可以查看剩余节点处于健康的状况（标注为绿色）： 

（New in v2.0）在大概5分钟过后，被退役的节点将被移动到**Decommissioned nodes**列表当中。目前用户无法在退役节点的时刻后，在时间序列图中查看到被退役的节点，需要通过**Decommissioned Nodes**列表查看该节点。该节点也不会主动从**Decommissioned Nodes**列表中清除。

Step 5: 移除退役节点

退役后的节点将不再接受客户端连接请求，集群也不会将数据副本迁移到该节点上。 对于每个退役节点，用户需要SSH登陆到对应的机器，执行以下命令来从集群中移除节点：

*#非安全模式*

cockroach node recommision 4 --insecure --host=<address of live node>

*#安全模式*

cockroach quit --certs-dir=certs --host=<address of decommissioned node>

+----+---------+-------------------+--------------------+-------------+

| id | is\_live | gossiped\_replicas | is\_decommissioning | is\_draining |

+----+---------+-------------------+--------------------+-------------+

| 4 | false | 12 | false | true |

+----+---------+-------------------+--------------------+-------------+

(1 row)

The affected nodes must be restarted for the change to take effect.

Step 3: 重启节点

SSH登陆退役节点所在的机器，执行cockroach start命令启动节点：

*#非安全模式*

cockroach start --insecure --host=<address of node to restart> --join=<address of node 1>:26257 --background

*#安全模式*

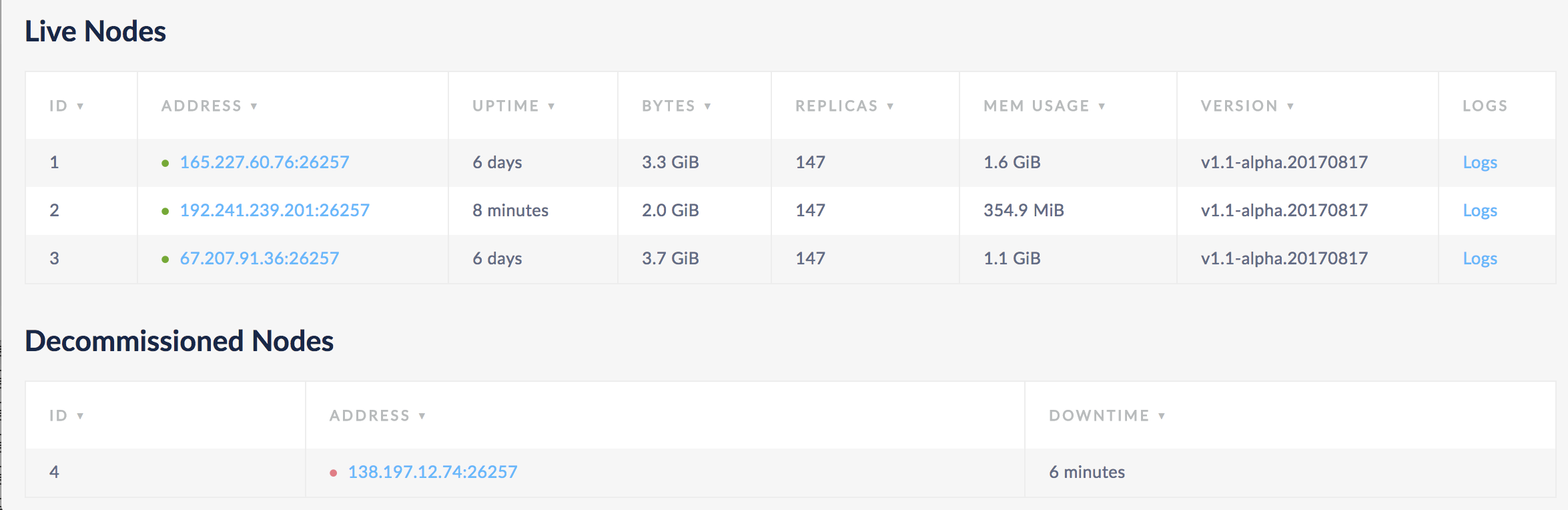
cockroach start --certs-dir=certs --host=<address of node to restart> --join=<address of node 1>:26257 --background

登陆Admin界面查看在**Nodes List**页，用户将看到服役节点被移动到**Live Nodes**列表当中，不久后集群会将Range副本重新均衡分布到该服役节点上。

节点重新服役

如果用户不小心退役了节点，或者希望将退役节点重新添加到集群当中，可以进行以下操作：

Step 1: 获取退役节点ID

登陆Admin界面，点击**Node Lists**视图，查看**Decommissioned Nodes**列表获取需要的ID： 

Step 2: 服役节点

SSH登陆集群中可用节点，基于需要重服役的节点ID，执行cockroach node recommission命令：

*#非安全模式*

cockroach node recommision 4 --insecure --host=<address of live node>

*#安全模式*

cockroach node recommission 4 --certs-dir=certs --host=<address of live node>

+----+-----------------------+---------+---------------------+---------------------+---------+-------------------+--------------------+-------------+

| id | address | build | updated\_at | started\_at | is\_live | gossiped\_replicas | is\_decommissioning | is\_draining |

+----+-----------------------+---------+---------------------+---------------------+---------+-------------------+--------------------+-------------+

| 1 | 165.227.60.76:26257 | 91a299d | 2017-09-07 18:16:03 | 2017-09-07 16:30:13 | true | 134 | false | false |

| 2 | 192.241.239.201:26257 | 91a299d | 2017-09-07 18:16:05 | 2017-09-07 16:30:45 | true | 134 | false | false |

| 3 | 67.207.91.36:26257 | 91a299d | 2017-09-07 18:16:06 | 2017-09-07 16:31:06 | true | 136 | false | false |

| 4 | 138.197.12.74:26257 | 91a299d | 2017-09-07 18:16:03 | 2017-09-07 16:44:23 | true | 1 | true | true |

| 5 | 174.138.50.192:26257 | 91a299d | 2017-09-07 18:16:07 | 2017-09-07 17:12:57 | true | 3 | true | true |

+----+-----------------------+---------+---------------------+---------------------+---------+-------------------+--------------------+-------------+

(5 rows)```

+----+---------+-------------------+--------------------+-------------+

| id | is\_live | gossiped\_replicas | is\_decommissioning | is\_draining |

+----+---------+-------------------+--------------------+-------------+

| 4 | false | 12 | false | true |

+----+---------+-------------------+--------------------+-------------+

(1 row)

The affected nodes must be restarted for the change to take effect.

Step 3: 重启节点

SSH登陆退役节点所在的机器，执行cockroach start命令启动节点：

*#非安全模式*

cockroach start --insecure --host=<address of node to restart> --join=<address of node 1>:26257 --background

*#安全模式*

cockroach start --certs-dir=certs --host=<address of node to restart> --join=<address of node 1>:26257 --background

在**Nodes List**页上，用户将看到重服役节点被移动到**Live Nodes**列表中，不久后集群会将Range副本重新均衡分布到该服役节点上。

检查退役节点状态

执行带--decommissionFlag的cockroach node status命令查看退役节点状态：

*#非安全模式*

cockroach **node** status --decommission --insecure --**host**=<address of any live **node**>

*#安全模式*

cockroach **node** status --decommission --certs-dir=certs --**host**=<address of any live **node**>

+----+-----------------------+---------+---------------------+---------------------+---------+-------------------+--------------------+-------------+

| id | address | build | updated\_at | started\_at | is\_live | gossiped\_replicas | is\_decommissioning | is\_draining |

+----+-----------------------+---------+---------------------+---------------------+---------+-------------------+--------------------+-------------+

| 1 | 165.227.60.76:26257 | 91a299d | 2017-09-07 18:16:03 | 2017-09-07 16:30:13 | true | 134 | false | false |

| 2 | 192.241.239.201:26257 | 91a299d | 2017-09-07 18:16:05 | 2017-09-07 16:30:45 | true | 134 | false | false |

| 3 | 67.207.91.36:26257 | 91a299d | 2017-09-07 18:16:06 | 2017-09-07 16:31:06 | true | 136 | false | false |

| 4 | 138.197.12.74:26257 | 91a299d | 2017-09-07 18:16:03 | 2017-09-07 16:44:23 | true | 1 | true | true |

| 5 | 174.138.50.192:26257 | 91a299d | 2017-09-07 18:16:07 | 2017-09-07 17:12:57 | true | 3 | true | true |

+----+-----------------------+---------+---------------------+---------------------+---------+-------------------+--------------------+-------------+