第十节：网络分区---进阶篇：

每当集群的配置发生更改时，集群内的所有节点就需要达成共识。常规事务数据通过这样的共识来达到集群内数据的最终一致性，但集群内节点的角色变更和一些使集群保持一致的内部消息传递也是同样如此。共识要求大多数集群节点同意既定的决定。当大多数集群节点失联时，该集群将无法继续平稳运行，因为它无法确保达到法定节点数来实现有效的决策仲裁。

当出现多个节点非自然因素故障时，集群内的仲裁机制可能会失效，从而导致大多数server突然从集群中被剔除。例如，在一个5节点的集群中，如果其中3个节点立即变为静默状态，则大多数server将受到影响，因此无法实现仲裁。实际上，其余两个server无法确定其他3个节点是否崩溃，或者网络分区是否单独隔离了这2个节点，因此无法自动重新配置集群。

另一方面，如果有server自动退出集群，则它们会提示自动重新配置集群。实际上，这意味着要退出集群的server会告诉其他节点它要离开集群。集群内其他online状态的节点会自动重新配置集群，维护集群内所有节点的一致性，并重新计算仲裁决策需要的大多数online节点数。例如，在上述5节点的集群中3个节点离开集群的情况下，如果离开的3个节点逐个通知集群要离开它们要离线，则成员资格可以将自己从5节点调整为2节点，确保达到法定的仲裁节点数。

ps:仲裁丢失本身就对集群具有不良的副作用。根据预期的故障数量计算集群的仲裁节点的数量非常重要（无论它们是连续的，一次发生的还是零星的）。

以下各节说明了如果集群如果出现网络分区的情况而无法实现自动仲裁时，（集群中的server无法自动实现仲裁）正确的操作如下：

ps：在失去多数节点的仲裁决策而被踢出集群之后，重新配置完成后，已从集群中排除的primary节点可能包含新集群中未包含的其他事务。如果发生这种情况，尝试从集群中添加回被剔除的节点将导致错误消息：

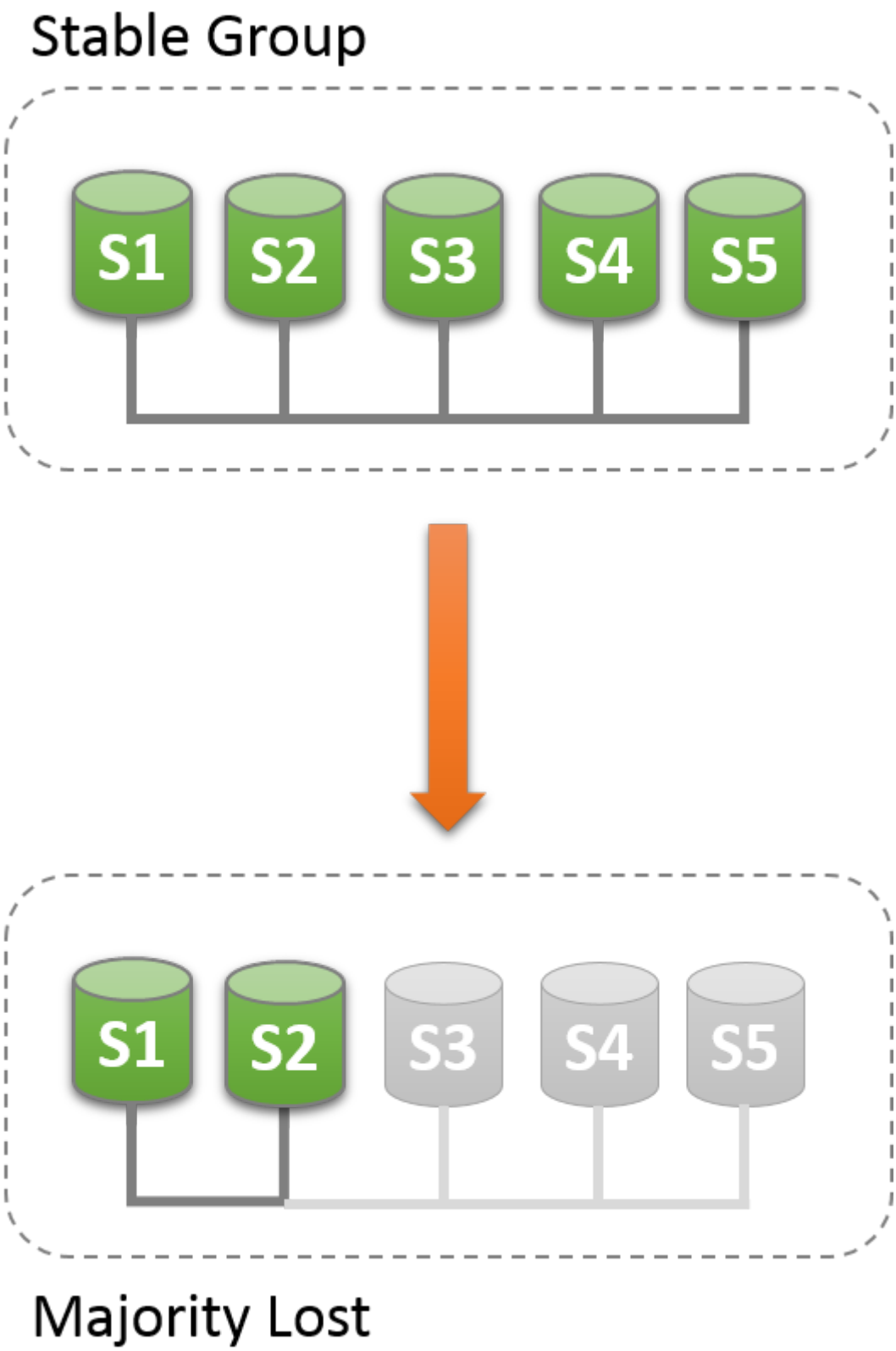
此节点具有比集群中存在的事务更多的已执行事务。

1.分区检测：

performance_schema.replication_group_members表从该server的角度显示了当前视图中集群内每个server的状态。在大多数情况下，系统不会遇到分区，因此该表显示的信息在集群中的所有节点之间都是一致的。换句话说，该表中每个server的状态在当前视图中被所有节点同意。但是，如果存在网络分区，并且仲裁策略失效时，那么该表将为无法连接的那些server显示状态UNREACHABLE。该信息由组复制中内置的本地故障检测器来提供。

下图表示集群中从5个活跃成员变成2个活跃成员，少于仲裁所需的超过半数成员时（少于3个成员），对于2个成员的分区来说，它的仲裁策略处于失效状态。

示意图如下：



为了了解这种类型的网络分区，以下描述了一个场景，其中最初有5个节点正常工作，然后只有2个节点处于online状态，然后对集群进行了更改。

假设有一个如上图所示的5节点集群：

```
- Server s1  `199b2df7-4aaf-11e6-bb16-28b2bd168d07`
- Server s2  `199bb88e-4aaf-11e6-babe-28b2bd168d07`
- Server s3  `1999b9fb-4aaf-11e6-bb54-28b2bd168d07`
- Server s4  `19ab72fc-4aaf-11e6-bb51-28b2bd168d07`
- Server s5  19b33846-4aaf-11e6-ba81-28b2bd168d07`
```

最初，集群正常运行，节点之间彼此通信。可以通过登录s1并查看其Replication_group_members表来验证这一点。例如：

```
mysql> SELECT MEMBER_ID,MEMBER_STATE, MEMBER_ROLE FROM
performance_schema.replication_group_members; +--------------------------------
-----+--------------+-------------+ | MEMBER_ID                             |
MEMBER_STATE |-MEMBER_ROLE | +--------------------------------------+------------
--+-------------+ | 1999b9fb-4aaf-11e6-bb54-28b2bd168d07 | ONLINE       |
SECONDARY    | | 199b2df7-4aaf-11e6-bb16-28b2bd168d07 | ONLINE       | PRIMARY
  | | 199bb88e-4aaf-11e6-babe-28b2bd168d07 | ONLINE       | SECONDARY    | |
19ab72fc-4aaf-11e6-bb51-28b2bd168d07 | ONLINE       | SECONDARY    | | 19b33846-
4aaf-11e6-ba81-28b2bd168d07 | ONLINE       | SECONDARY    | +--------------------
-----------------+--------------+-------------+
```

然而，片刻之后发生了灾难性故障，节点s3，s4和s5意外宕机。此后几秒钟，再次查看s1上的replication_group_members表显示该server仍处于online状态，而其他几个节点则处于离线状态。实际上，如下所示，它们被标记为“不可达”。而且，系统无法重新配置自身以更改成员资格，因为大多数节点已经宕机。

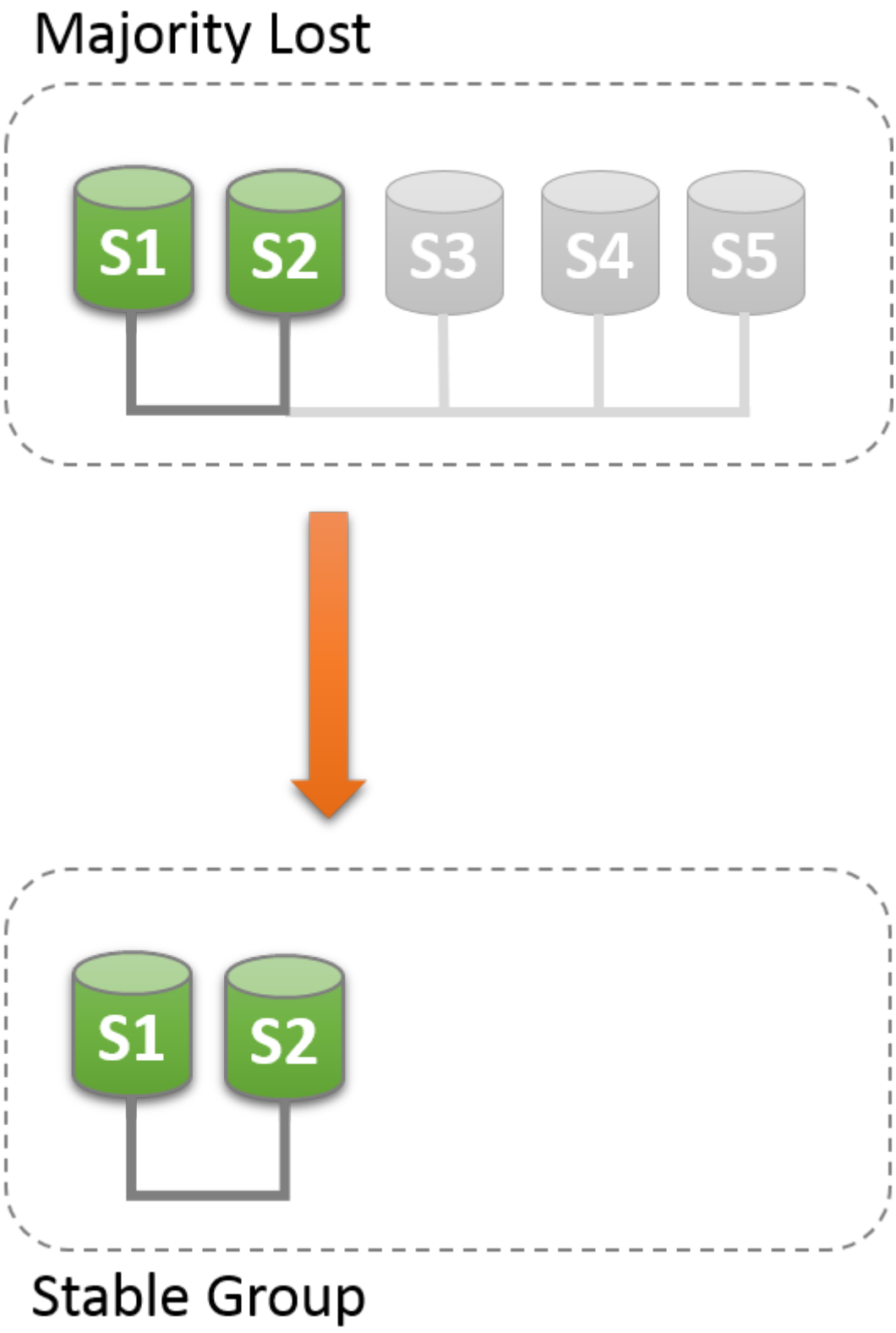
```
mysql> SELECT MEMBER_ID,MEMBER_STATE FROM
performance_schema.replication_group_members; +--------------------------------
-----+--------------+ | MEMBER_ID                             | MEMBER_STATE | +--
-------------------------------------+--------------+ | 1999b9fb-4aaf-11e6-bb54-
28b2bd168d07 | UNREACHABLE  | | 199b2df7-4aaf-11e6-bb16-28b2bd168d07 | ONLINE
    | | 199bb88e-4aaf-11e6-babe-28b2bd168d07 | ONLINE       | | 19ab72fc-4aaf-
11e6-bb51-28b2bd168d07 | UNREACHABLE  | | 19b33846-4aaf-11e6-ba81-28b2bd168d07 |
UNREACHABLE  | +--------------------------------------+--------------+
```

该表显示，由于集群内的大多数节点无法访问，因此s1现在位于没有外部干预就无法进行恢复的集群中。在这种特殊情况下，需要重置组成员身份列表以允许系统继续运行，这在本节中进行了说明。或者，也可以选择停止s1和s2上的组复制（或完全停止s1和s2），找出s3，s4和s5发生了什么，然后重新启动组复制。

2.消除分区：

组复制可以通过强制进行特定配置来重置组成员资格列表。例如，在上面的示例中，其中s1和s2是唯一的online节点，可以选择强制仅由s1和s2组成的成员身份配置。这需要检查有关s1和s2的一些信息，然后使用group_replication_force_members变量。

示意图如下：



假设又回到了s1和s2是集群中剩下的唯一online节点的情况。节点s3，s4和s5意外离开了该集群。为了使节点s1和s2继续运行，需要强制指定仅包含s1和s2的成员资格配置。

注意：此过程使用group_replication_force_members，应视为万不得已的补救措施。必须格外小心地使用，并且仅用于使仲裁无效。如果使用不当，它可能会创建人为裂脑情景或完全阻塞整个系统。

假设系统已被阻塞，并且当前配置如下（如s1上的本地故障检测器所感知）：

```
mysql> SELECT MEMBER_ID,MEMBER_STATE FROM
performance_schema.replication_group_members; +--------------------------------
-----+--------------+ | MEMBER_ID                             | MEMBER_STATE | +--
-----------------------------------+--------------+ | 1999b9fb-4aaf-11e6-bb54-
28b2bd168d07 | UNREACHABLE  | | 199b2df7-4aaf-11e6-bb16-28b2bd168d07 | ONLINE
    | | 199bb88e-4aaf-11e6-babe-28b2bd168d07 | ONLINE       | | 19ab72fc-4aaf-
11e6-bb51-28b2bd168d07 | UNREACHABLE  | | 19b33846-4aaf-11e6-ba81-28b2bd168d07 |
UNREACHABLE  | +-------------------------------------+--------------+
```

要做的第一件事是检查s1和s2的组通信标识符是什么，登录到s1和s2并按以下方式获取该信息：

```
mysql> SELECT @@group_replication_local_address;
```

一旦获取到s1（127.0.0.1:10000）和s2（127.0.0.1:10001）的组通信地址，就可以在两个节点之一上使用它来注入新的成员资格配置，从而覆盖具有丢失仲裁。要在s1上执行此操作：

```
mysql> SET GLOBAL
group_replication_force_members="127.0.0.1:10000,127.0.0.1:10001";
```

通过强制执行其他配置，可以解除对集群的阻塞。在此更改之后，请同时检查s1和s2上的plication_group_members以验证集群节点身份。

首先在s1上：

```
mysql> SELECT MEMBER_ID,MEMBER_STATE FROM
performance_schema.replication_group_members; +--------------------------------
-----+--------------+ | MEMBER_ID                             | MEMBER_STATE | +--
-----------------------------------+--------------+ | b5ffe505-4ab6-11e6-b04b-
28b2bd168d07 | ONLINE       | | b60907e7-4ab6-11e6-afb7-28b2bd168d07 | ONLINE
    | +-------------------------------------+--------------+
```

然后在s2上：

```
mysql> SELECT * FROM performance_schema.replication_group_members; +------------
-------------------------+--------------+ | MEMBER_ID
| MEMBER_STATE | +-------------------------------------+--------------+ |
b5ffe505-4ab6-11e6-b04b-28b2bd168d07 | ONLINE       | | b60907e7-4ab6-11e6-afb7-
28b2bd168d07 | ONLINE       | +-------------------------------------+----------
----+
```

强制执行新的集群成员资格配置时，请确保确实要停止将要从集群中强制退出的所有节点。在上述情况下，如果s3，s4和s5并非真正不可访问而是online状态，则它们可能已经形成了自己的仲裁策略（5分之3，因此占大多数）。在这种情况下，将集群节点列表强制设置为s1和s2可能会造成人为的裂脑情况。因此，在强制执行新的成员资格配置之前，确保要剔除出集群的节点确实已关闭是非常重要的；如果没有关闭，请在继续操作之前将其关闭。

使用group_replication_force_members系统变量成功强制新的组成员身份并取消该集群的阻塞后，请确保清空系统变量。 group_replication_force_members必须为空才能在集群内成功执行START GROUP_REPLICATION语句。