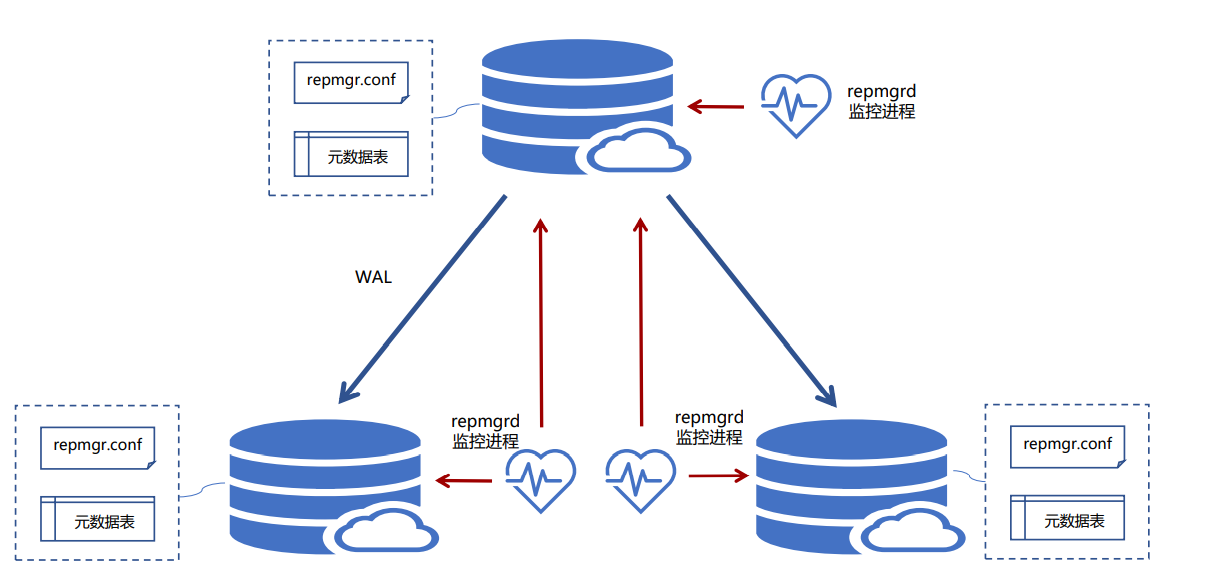
来源：http://blog.itpub.net/69990629/viewspace-2910063/

# 一、Repmgr

## 1.Repmgr特点

Repmgr是2010年由2ndQuadrant推出的一款PostgreSQL故障切换工具，repmgr是一个开源工具套件，用于管理PostgreSQL服务器集群中的复制和故障转移。它扩展了PostgresSQL内建的hot-standby能力，可以监控复制和执行管理任务。



### （1）Repmgr命令管理 (replication manager)

Repmgr是一个命令行工具，日常操作主要通过Repmgr进行操作，功能包括集群状态查看、switchover、克隆备库、失效节点重新加入等。

**用于执行管理任务的命令行工具，主要有以下方面作用：**

1. 设置备用服务器
2. 将备用服务器升级为主服务器
3. 切换主服务器和备用服务器
4. 显示流复制状态
5. clone恢复备机数据
6. 注册节点

### （2）Repmgrd守护进程 (replication manager daemon)

Repmgrd 是一个守护进程，支持故障检测、failover，监控和记录集群信息以及自定义脚本接受集群事件通知event\_notification\_command，它有一组预定义的事件，并将这些事件的每次发生都存储在 repmgr.events 表中。Repmgr 允许将事件通知传递给用户定义的程序或脚本，该程序或脚本可以采取进一步的行动，例如发送电子邮件或触发任何警报。

可以使用Repmgrd设置自动故障转移。Repmgrd 需要在启动 PostgreSQL 服务器时加载共享库“repmgr”。库名称应配置在 postgresql.conf 文件的shared\_preload\_libraries里。并且需要在 repmgr.conf 文件中设置  **failover=automatic** 参数。一旦设置了所有这些参数，Repmgrd 守护程序就会开始主动监控集群。如果主节点出现任何故障，它将尝试多次重新连接。当所有连接到主节点的尝试都失败时，Repmgrd 将通过选举选择最符合条件的备用节点作为新的主节点。

**它主动监视复制集群中的服务器并执行以下任务：**

1. 监控和记录集群复制性能
2. 通过检测主服务器故障并提升最合适的备用服务器来执行故障转移
3. 将有关群集中事件的通知提供给用户定义的脚本，该脚本可以执行诸如通过电子邮件发送警报等任务
4. repmgrd 根据本地数据库角色不同，其功能也不同：   
   主库：repmgrd仅监控本地数据库，负责自动恢复、同异步切换   
   备库：repmgrd监控本地数据库和主数据库，负责自动切换、复制槽删除

### （3）相关元数据

表

repmgr.events：用来记录repmgr管理的事件信息

repmgr.nodes：复制群集中每个服务器的连接和状态信息

repmgr.monitoring\_history：repmgrd写入的历史备用监视信息

repmgr.voting\_term：【5.2新增】主要用来记录投票信息

视图

repmgr.show\_nodes：基于表repmgr.nodes，增加了显示上游节点的信息

repmgr.replication\_status：启用repmgrd的监视时，显示每个备用数据库的当前监视状态。

Repmgr元数据的schema可以存储在现有的数据库或在自己的专用数据库，repmgr元数据的schema不能驻留在不属于Repmgr管理的复制集群的数据库服务器上。

——   
**Repmgr本身不支持虚拟IP的功能** ，我们可以借助keepalived来实现虚拟IP的功能。另外在配置文件中，也可以设置promote\_command为一个自定义脚本

## 2.Repmgr如何仲裁哪台备机升主

* 每个备机检查到主机数据库故障后会进行重试，重试最后一次后，会去询问其他备用数据库。如果其它备用节点的最后一个复制的LSN或与主节点的最后一次通信的时间比当前节点的最后一个复制的LSN或最后一次通信的时间更近，则该节点不执行任何操作，并等待与主节点的通信恢复。
* 如果所有备机数据库节点都看不到主库，则它们将检查witness见证节点是否可用。如果也无法到达witness见证节点，则备机会假定主服务器端发生网络中断，因此不会继续选择新的主服务器。如果可以到达witness见证节点，则备机节点会假定主服务器已关闭，然后继续选择主节点。之后将升级配置为“首选”主节点的节点。每个备机数据库节点将重新初始化其复制，以跟随新的主数据库。

**Repmgr怎么选主**：  **当需要failover时，repmgr选举候选备节点会以以下顺序选举： LSN > Priority > Node\_ID。若LSN一样，会根据priority优先级进行比较，该优先级是在配置文件中进行参数配置，将priority设置为0会禁止参与选主。若优先级也一样，会比较节点的Node ID，小者会优先选举**

## 3.Repmgr如何处理脑裂场景（位置参数、witness见证节点）

### （1）位置参数

Repmgr使用location位置参数处理脑裂场景，其中每个节点应根据其所在的数据中心指定位置参数。在任何网络分裂的情况下，Repmgr将确保与主节点位于同一位置的节点的提升。如果它在该位置找不到任何节点，它将不会提升任何位置的任何节点。

location='location1'                    # 定义location

除此之外可以使用witness见证服务器的额外节点处理网络隔离，避免产生脑裂。

### （2）见证节点

witness见证节点重要用来处理集群主库和备库之间可能存在网络拥塞、延迟、路由等问题影响，导致主库还在正常工作，而备库无法联系主库的场景。通过设置witness节点可以针对主库与备库之间切换的检查完整性，即辅助备节点进行监控，避免因网络问题导致的脑裂现象。

见证节点主要的工作是帮助备库达到法定的数量。它是一个仅考虑多数票数的节点。该服务器上不需要安装PostgreSQL，因此在复制中没有任何作用。

当备机连不上主机了，就会连接witness见证节点，如果也连接不上见证节点，那判断自己网络故障了，如果能连上见证节点，则认为主机故障，见证节点的作用类似于一个信任的网关。

**witness必须配合Repmgrd**。 Remgrd启动后会作为常规服务运行并持续监视集群的运行状况。当达到与主机数据库失去联系的法定人数时，它将启动故障转移。它不仅可以自动升级备用数据库，还可以在多节点群集中重新启动其他备用数据库以跟随新的主数据库。

## 4.Repmgr优缺点

### （1）repmgr 优点

* Repmgr 提供了帮助设置主节点和备用节点以及配置复制的实用程序。
* 不使用任何额外的端口进行通信。如果想执行切换，那么它才需要配置无密码SSH。
* 通过调用已注册事件的用户脚本来提供通知。
* 在主服务器发生故障时可以执行自动故障转移。
* 对数据库侵入小，维护起来和正常主备流复制基本一致。
* 配置简单。

### （2）repmgr 缺点

* repmgr 不会检测备用库是否在恢复配置中使用未知或不存在的节点错误配置。即使节点在未连接到主/级联备用节点的情况下运行，节点也会显示为备用节点。
* 无法从 PostgreSQL 服务关闭的节点检索另一个节点的状态。因此，它不提供分布式控制解决方案。
* 它不能在备机单个节点down掉时，自动把其拉起。