

Построение отказоустойчивого кластера PostgreSQL в среде Astra Linux 1.7 с использованием технологии racemaker&corosync

Константин Курочкин

Служба внедрения и сопровождения
Отделения разработки информационных систем

Параметры системы



ОС Astra Linux 1.7
PostgreSQL 11.10
Pacemaker 2.0.1
Corosync 3.0.1



Расетmaker

Расетmaker — мозг и по совместительству менеджер ресурсов кластера.

Главная задача — достижение максимальной доступности ресурсов, которыми он управляет, и защита их от сбоев



Сущности Pacemaker

Первая и самая важная сущность – это **узлы кластера**.

Узел (нода, node) кластера представляет собой физический сервер или виртуальную машину с установленным Pacemaker

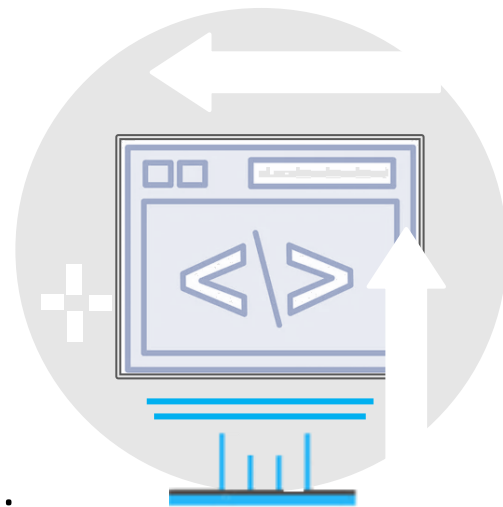
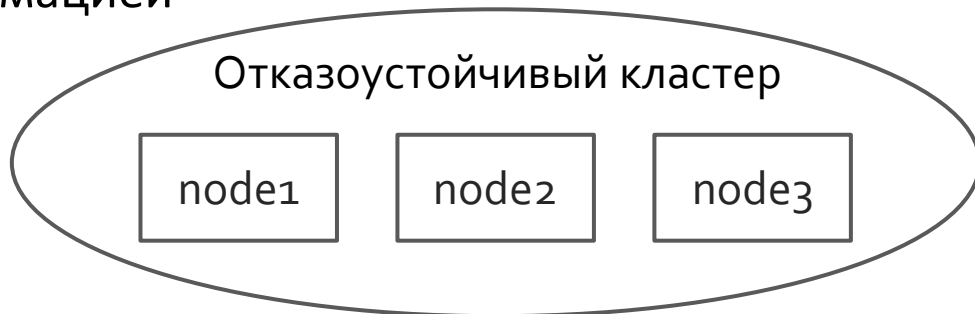


Сущности Pacemaker

Следующая важная группа сущностей Pacemaker – **ресурсы кластера**.

Ресурс для Pacemaker — это скрипт, написанный на любом языке.

Скрипт управляет сервисами в операционной системе. Главное требование к скриптам — уметь выполнять 3 действия: `start`, `stop`, `monitor` и делиться некоторой метаданной

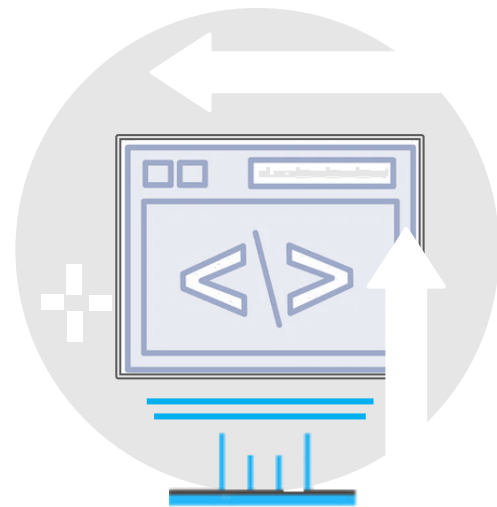


Сущности Расemaker

Ресурсы имеют множество атрибутов, которые хранятся в конфигурационном XML-файле Расemaker'a.

Наиболее интересные из них:

- Priority
- resource-stickiness
- migration-threshold
- failure-timeout
- multiple-active.



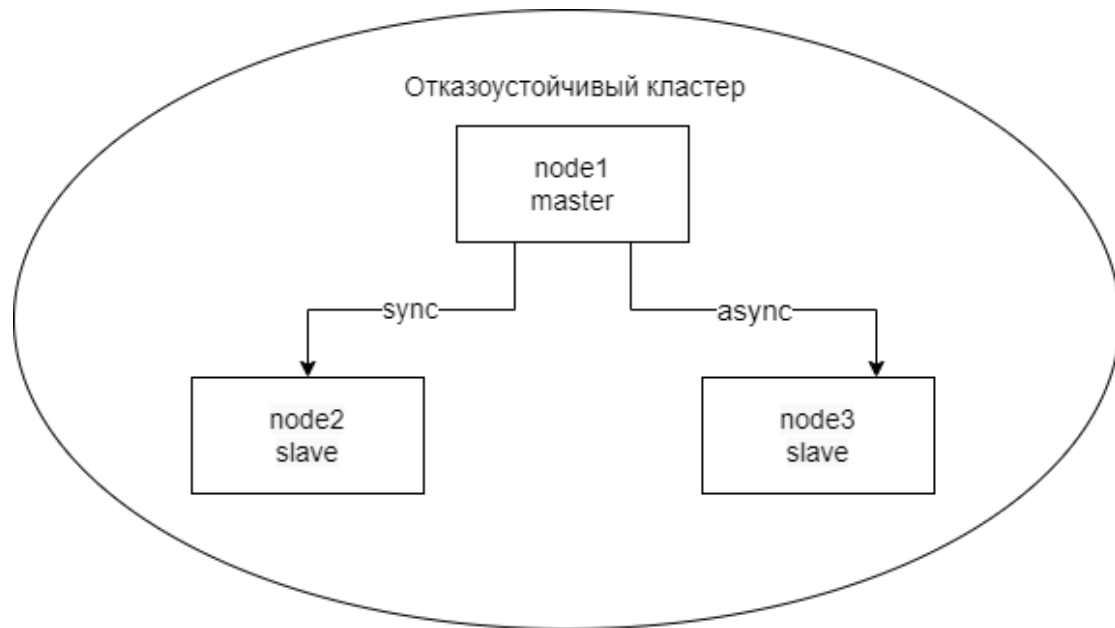
Corosync

Это сервис системы группового общения.

Обеспечивает сетевое взаимодействие узлов, т.е. передача сервисных команд (запуск/остановка ресурсов, узлов и т.д.), обмен информацией о полноте состава кластера (quorum) и т.д.



Кворум



Кластер с кворумом – это когда количество «живых» узлов больше половины общего количества узлов в кластер

Условия для работы Расетmaker

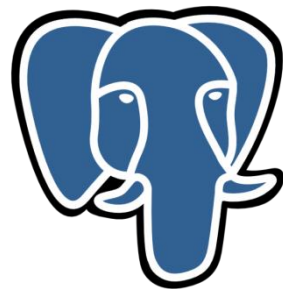
- Синхронизация времени между узлами в кластере
- Разрешение имен узлов в кластере
- Стабильность сетевых подключений
- Наличие у узлов кластера функции управления питанием/перезагрузкой с помощью IPMI(ILO) для организации «фенсинга» (fencing – изоляция) узла.
- Разрешение прохождения трафика по протоколам и портам



Особенности использования PostgreSQL

При построении отказоустойчивого кластера PostgreSQL, необходимо учитывать ряд особенностей:

1. Pacemaker при запуске кластера с PostgreSQL размещает файл блокировки LOCK.PSQL на узле с мастером СУБД;
2. Архивы WAL-ов рекомендуется размещать на общих (shared) устройствах хранения, доступных как Мастеру, так и Реплике. Это позволит упростить процесс восстановления Мастера после сбоя и перевода его в режим Slave.

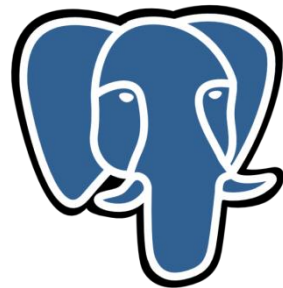


PostgreSQL

Особенности использования PostgreSQL

3. Во избежание возникновения ситуации со split-brain необходимо выполнение двух важных условий при построении отказоустойчивых кластеров:

- Наличие кворума в ОУК. Это означает, что в кластере должно быть не менее 3-х узлов;
- Наличие устройств фенсинга на узлах с СУБД. При возникновении сбоя устройства «фенсинга» изолируют сбойнувший узел – посылают команду на выключение питания или перезагрузку (poweroff или hard-reset).



PostgreSQL

Виды сбоев на узлах кластера



Сбой по питанию на текущем мастере или на реплике



Сбой процесса PostgreSQL



Потеря сетевой связности между каким-либо из узлов и остальными узлами



Сбой процесса Pacemaker/Corosync

Виды планового обслуживания кластера



Замена вышедшего из строя оборудования (не приведшего к сбою)



Апгрейд оборудования



Обновление софта



Смена ролей Мастера и Реплики

Спасибо за внимание!

Служба внедрения и сопровождения
Отделения разработки информационных систем



Москва,
Графский пер., д.14, к.2



+7 (495) 747-7040
(# 101, # 122)



develop@fors.ru