

# Территориально распределенная конфигурация среды виртуализации

Платформа виртуализации R-Платформа поддерживает создание распределенных сред виртуализации и хранилища в рамках двух разнесенных площадках в рамках расстояния синхронной репликации. Данная конфигурация так-же известна под названием “растянутый кластер”.

## Основные требования к площадкам

Распределенный кластер предъявляет следующие требования к сетевой инфраструктуре:

- Задержки отправки пакетов между двумя площадками по сети хранения не должны превышать 5мс.
- В качестве интерфейсов сети хранения должны использоваться два канала со скоростью не менее 10Гбит/сек каждый.
- В качестве сети управления должен использоваться два выделенных интерфейса со скоростью работы не менее 1 Гбит/сек.

## Параметры настройки машины "Свидетеля" для распределенной среды виртуализации

При создании территориально распределенной конфигурации на платформе виртуализации R-Платформа суммарное количество узлов на двух площадках должно быть четным а также должна быть развернута дополнительная машина с сервисом метаданных хранилища называемая “свидетель” на третьей площадке. Свидетель может быть выполнен как в виде физического сервера так и в виде виртуальной машины.

Требования к виртуальной машине свидетеля:

- 2 vCPU;
- Виртуальный диск 64 ГБ, предпочтительно расположенный на SSD;
- Объем оперативной памяти - 4 ГБ.
- Все указанные объемы ресурсов должны быть выделены со 100% резервированием (процессор, память, дисковая емкость).
- 1 интерфейс сети управления;
- 1 интерфейс сети хранения;
- Тип виртуальной машины - Centos 7;

[domestic:rosplatforma:guide:advanced:metrocluster:whitness](#)

## Создание территориально-распределенной конфигурации

Установка платформы виртуализации в формате метро-кластера практически не отличается от развертывания системы в рамках одной площадки. Дополнительно к базовым задачам по

установке и настройке узлов виртуализации необходимо выполнить следующие процедуры.

1. Перед созданием кластера хранения необходимо указать расположение узлов и создать таким образом домены отказа;
2. Создать кластер хранения расположив диски с ролью метаданных равномерно по доменам отказа.
3. Задать уровень отказоустойчивости с учетом настроенного расположения;
4. Добавление узлов в Р-Управление;
5. Запуск службы отказоустойчивости shaman;

## Настройка домена отказа на сервере

Каждый компонент Р-Хранилище обладает адресом, который состоит из 5 идентификаторов: `room.row.rack.host_ID.cs_ID` Первые три адреса пути `room.row.rack` можно настроить в файлах конфигурации `/etc/vstorage/location` на каждом физическом сервере <sup>1)</sup>. Последние два компонента `host_ID.cs_ID` генерируются автоматически:

- `host_ID` является уникальным, случайно генерируемым идентификатором хоста, который создается во время установки и хранится в `/etc/vstorage/host_id`;
- `cs_ID` является уникальным идентификатором службы, который генерируется при создании сервера фрагментов.

Для задания расположения сервера между двумя центрами обработки данных необходимо на каждом сервере создать конфигурационный файл, описывающий его расположение

[/etc/vstorage/location](#)

```
room.row.rack
```

где:

- `room` - номер виртуального помещения в формате числа от 0 до 9. ;
- `row` - номер ряда;
- `rack` - номер стойки в ряду;

Данные номера не обязательно должны совпадать с номерами реального расположения оборудования а используются для создания логических сущностей в рамках домена отказа.

Серверы, которые находятся в одном домене отказа должны иметь одинаковые значения файла `location`.

Например для создания распределенной конфигурации с пятью доменами отказа файл `location` может принимать следующие значения:

1. `0.0.1` - для первого домена отказа на первой площадке;
2. `0.0.2` - для второго домена отказа на первой площадке;
3. `0.0.3` - для третьего домена отказа на второй площадке;
4. `0.0.4` - для четвертого домена отказа на второй площадке;

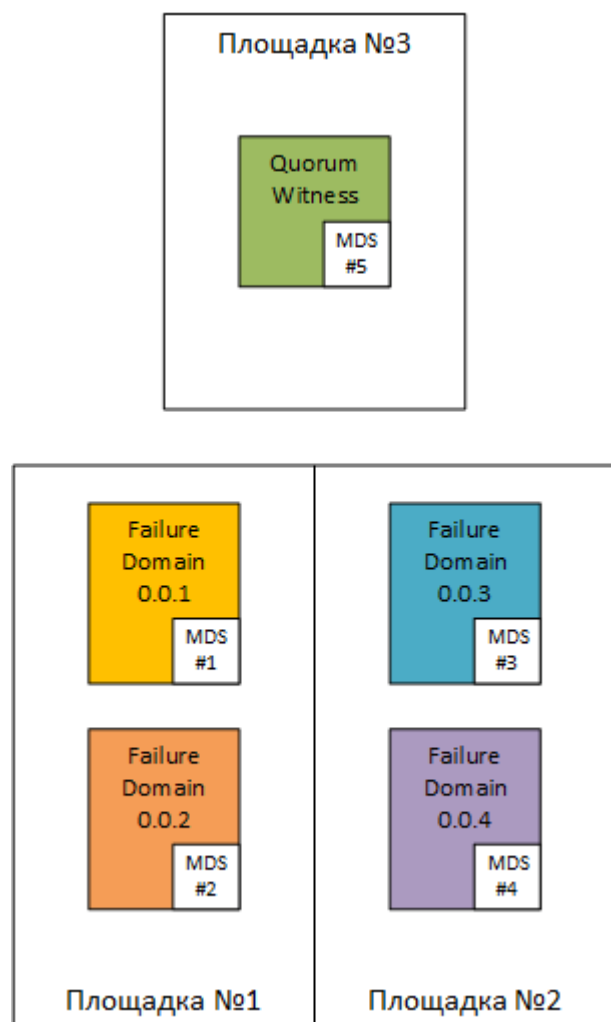
## Создание кластера и задание расположения дисков метаданных в доменах отказа

Создание кластера должно производиться только после установки корректного значения в файле `location`.

Рекомендуемое количество сервисов метаданных в одном кластере хранения равно пяти. Поэтому для конфигурации с четырьмя доменами отказа (по 2 на каждую площадку) необходимо назначить по одному сервису метаданных в каждом домене отказа. Пятый домен отказа и диск с ролью метаданных будет расположен на "свидетеле".

Машина свидетеля должна иметь только диски двух следующих ролей или один диск совмещающий две эти роли:

1. Система;
2. Метаданные.



После создания кластера хранения необходимо создать хранилище для среды виртуализации штатным способом.

## Задание уровня отказоустойчивости

Для конфигурации метро кластера необходимо задать уровень отказоустойчивости для создаваемых хранилищ равным `gf=4`. Данное требование обусловено тем что в случае выхода из строя одной площадки система будет сохранять как минимум две копии каждого блока данных.

Во избежание возникновения ситуации когда все 4 копии блока могут оказаться на одной площадке необходимо указать домен отказа равный `rack`. Так как объединение серверов в 4 домена отказа путем правки файла `/etc/vstorage/location` проводилось именно на третьем уровне расположения - `rack`. Для этого, после создания хранилища для среды виртуализации или iSCSI или S3 необходимо установить атрибут распределения блоков командой:

```
vstorage -c %CLUSTER_NAME% set-attr -R  
/mnt/vstorage/vols/datastores/%DATASTORE_NAME% failure-domain=rack
```

, где:

- `%CLUSTER_NAME%` - имя кластера хранения;
- `%DATASTORE_NAME%` - имя хранилища виртуальных машин.

Для хранилищ другого типа настройка производится аналогичным образом.

## Добавление узлов в Р-Управление

После создания кластера хранения необходимо добавить все вычислительные узлы в систему управления виртуализацией Р-Управление. Сервер “свидетеля” не добавляется, потому что запуск виртуальных машин на данном сервере производиться не должен.

## Запуск службы отказоустойчивости shaman

Служба `shaman` запускается на всех серверах которые предназначены для запуска и выполнения виртуальных машин. На свидетеле команда `hastart` не выполняется и служба `shaman` не должна запускаться на данном сервере во избежание распределения на данный сервер виртуальных машин в случае сбоя.

Необходимо настроить максимальное количество узлов при котором служба `shaman` продолжает работать. При превышении указанного числа сбой считается критичным и автоматическое восстановление работоспособности виртуальных машин не производится.

Параметр задается на одном из узлов, который подключен к кластерной службе.

```
shaman set-config RELOCATION_SKIP_THRESHOLD=n
```

, где число узлов `n` необходимо установить равным суммарному числу узлов площадки или больше на один.

1)

для дополнительной информации см. [man](#) страницу по `vstorage-config-files`

From:

<https://micronode.ru/> - **micronode.ru**

Permanent link:

<https://micronode.ru/domestic/rosplatforma/guide/advanced/metrocluster>

Last update: **2023/01/17 13:15**

