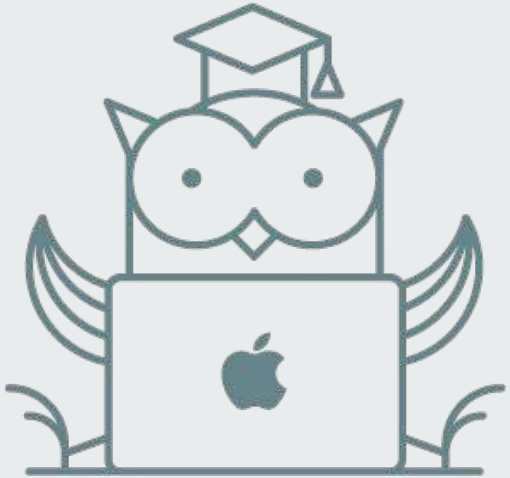
Курс «Администратор Linux»

Дисковая подсистема: LVM, Файловые системы.

Занятие # 3

Дмитрий Молчанов
Григорий Ожегов



Содержание

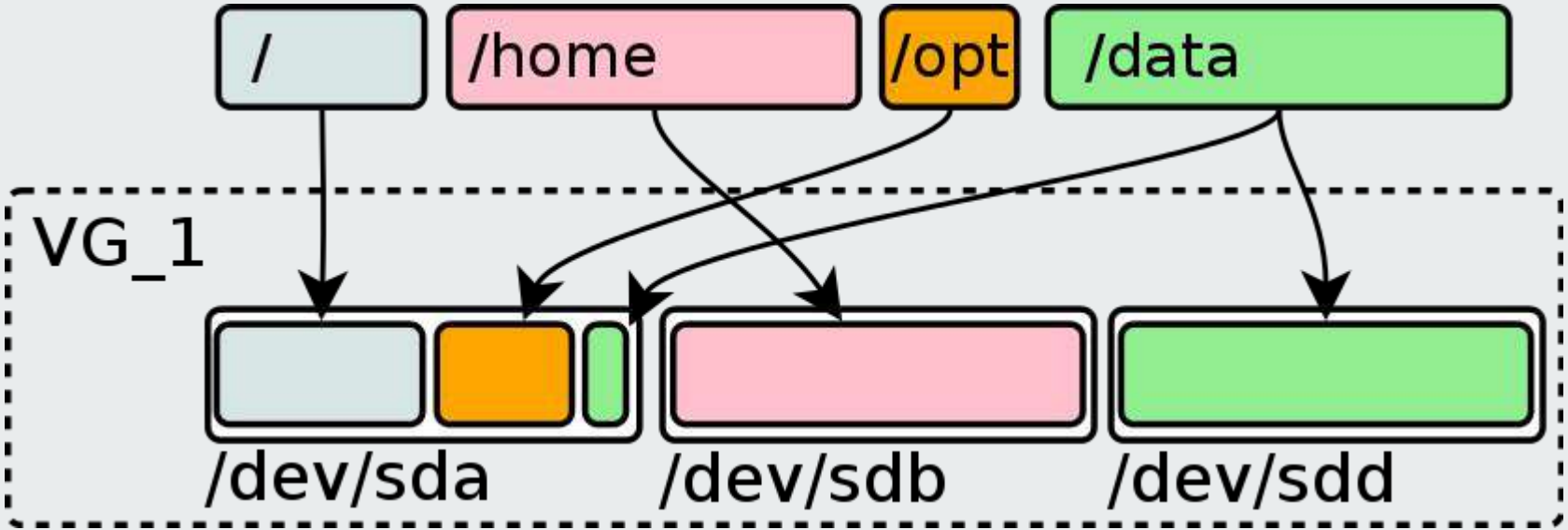
- LVM
- Организация файловых систем
- Файловые системы:
  - ext2/3/4
  - xfs
  - reiser
  - tmpfs/ramfs

Дисковая подсистема: LVM

LVM - Logical Volume Manager. Специальная подсистема ядра, которая добавляет дополнительный уровень абстракции от “железа”, позволяя управлять дисковым пространством и решать разнообразные задачи по управлению дисковой подсистемой:

- группировка физических томов
- создание и изменение размеров логических томов
- Snapshot’ы
- Thin provisioning (NEW)
- Cache Volumes (NEW)
- LVM RAID

Дисковая подсистема: LVM



LVM: Physical Volumes/Физические тома

На самом нижнем уровне подсистемы LVM лежат физические тома (physical volume), но на самом деле это любое блочное устройство:

- Диск
- Партиция на диске
- loop device (файл как блочное устройство)

Команды:

- `pvcreate`
- `pvremove`
- `fdisk/sfdisk`
- `losetup`

LVM: Volume Groups/Группировка дисков

Группировка дисков в LVM позволяет объединить доступное пространство нескольких Physical Volumes.

Команды:

- `vgcreate`
- `vgextend`
- `vgremove`

LVM: Logical Volumes/тома

Внутри групп блочных устройств можно создавать логические тома, которые являются самым верхним слоем “пирога” lvm и являются блочными устройствами, внутри которых можно создавать файловые системы или даже другие экземпляры LVM.

logical volume может быть распределен по нескольким физическим устройствам - вынужденно, утилизируя “остатки” свободного пространства, или намеренно - через lvmraid/striped (raid0) тип тома. Во втором случае мы получим бенефиты raid0 при чтении/записи.

Команды:

- `lvcreate`
- `lvchange/lvextend`
- `lvremove`

LVM: Logical Volumes/тома

Размер логических томов можно изменять на лету, если это позволяет место в группе устройств. Соответственно, если позволяет файловая система, расположенная на этом томе, можно менять ее размер, наращивая его на лету, без перезагрузки.

Команды:

- `lvcreate`
- `lvchange/lvextend`
- `lvremove`

LVM: lvm snapshots

Размер логических томов можно изменять на лету, если это позволяет место в группе устройств. Соответственно, если позволяет файловая система, расположенная на этом томе, можно менять ее размер, наращивая его на лету, без перезагрузки. Снепшоты позволяют получить “замороженный” список тома, в то время как основной том будет доступен для записи.

Для COW (Copy On Write) снепшотов необходимо наличие свободного места в VG - оно будет использоваться для хранения изменений вносимых в основной том.

Чаще всего это используется для бэкапов.

LVM: Thin Provisioning

Функция Thin Provisioning позволяет еще более тонко управлять дисковым пространством не стесняя себя, выделяя места больше чем есть на самом деле. В комбинации с тем, что соответствующие снепшоты можно использовать в RW. Эта возможность позволяет эффективно использовать место для виртуалок разделяя один том с базовой системой и храня в снепшотах изменения относительно базовой версии.

LVM: Cache Volumes

LVM Cache volume позволяет использовать, например, ssd в качестве кэша для “ускорения” медленных томов в VG за счет хранения копий наиболее часто используемых блоков на быстром накопителе, это делается прозрачно для пользователя.

Файловые системы

- Иерархическая организация
- Точки монтирования
- Файловые системы

Файловые системы: Иерархия

Файловая система в linux (да и unix в целом) имеет иерархическую/древовидную организацию. Принято минимизировать количество разделов в корне и размещать там стартовые ветки иерархий. В корне мы обычно видим следующие каталоги:

- /boot - информация необходимая для загрузки
- /bin,/sbin - системные исполняемые файлы.
- /etc - файлы конфигурации системы и приложений
- /home - домашние каталоги пользователей
- /var - динамически изменяемая информация (БД, Кэши, логи)
- /lib[64] - системные библиотеки
- /usr - пользовательские (или системные) программы.

Так же есть ветвь /usr/local предназначенная для локального ПО.

Файловые системы: inode

Файловые системы которые чаще всего встречаются на серверах (ext*, xfs) используют inode для хранения мета-информации о файле:

- размер
- Права
- Времена модификации данных, метаданных, доступа
- владелец/группа
- Тип
- selinux context

Файловые системы: Проблемы

Во время администрирования вы можете с высокой вероятностью столкнуться со следующими проблемами:

- No space left on device:
  - Закончилось место
  - Закончились inode
- Невозможно создать директорию - уперлись в лимит 32000 поддиректорий в директории.

Файловые системы: Точки монтирования

Любой каталог в linux может быть точкой монтирования другой файловой системы, и переходя в этот каталог мы попадаем на другую файловую систему.

Файловые системы: журналирование

Журналирование файловой системы - предварительная запись информации об изменении, а иногда и данных для изменения в специальную область - журнал. После выполнения операции запись удаляется из журнала. Это помогает восстанавливать систему в случае сбоев.

Файловые системы: ext2/3/4

- ext2 - исторически “стандартная” для Linux. файловая система решавшая много ограничений своих предтечей - ext и minix. Считается эталоном производительности. Поддерживается online resize
- ext3 - Логическое продолжение ext2, расширены ограничения на размер файлов и тома, добавлена возможность журналирования:
  - writeback - записывается только информация об изменении (мета-данные)
  - ordered - то же самое, что и wb, только информация в журнал записывается ДО изменения.
  - journal - записывается все - и информация об изменении и данные, сильно влияет на производительностью.
- ext4 - Логическое продолжение ext3. Номинально сильно увеличены ограничения на размер тома, по факту из коробки все еще 4Тб на том, возможность хранить ext. attributes в Inode, увеличение inode (128->256b), решен вопрос со вложенными каталогами (>32000)

Файловые системы: xfs

XFS - высокопроизводительная журналируемыея файловая система родом из SGI (Silicon Graphics).

+ Динамическая аллокация inode + Дефрагментация на лету + Потенциально лучшая производительность + Встроенные средства для резервного копирования/снепшотов (xfsdump/xfsrestore) + “отсутствие” жестких ограничений на размер файловой системы	- Малая ремонтопригодность - выше вероятность сбоя из-за хранения большого количества данных в памяти.

Файловые системы: tmpfs

Файловая система хранящаяся в памяти - очень быстро, но не сохраняется между перезагрузками. Можно использовать только для кэшей.

```
mount -t tmpfs none /mnt -o size=100M
```

Если не указать size - будет выделено половину памяти.

Полезные ссылки:

- http://xgu.ru/wiki/LVM

Домашнее задание:

- to be added. (Эксперименты с RAID и LVM)

Спасибо за внимание

Дмитрий Молчанов
Григорий Ожегов