

Администрирование Linux

Лекция 6.1

Диски и файловые системы

Иртегов Д.В.

Новосибирский гос. Университет

2022

«У меня один диск и весь C:»

- Зачем вообще нужны сложные конфигурации дисковой подсистемы?
 - Казалось бы, можно объединить все диски в один том LVM
 - Это упростит планирование
 - Исключит проблемы фрагментации «на /home место кончилось, а на /var еще полно»

Зачем нужно много разделов

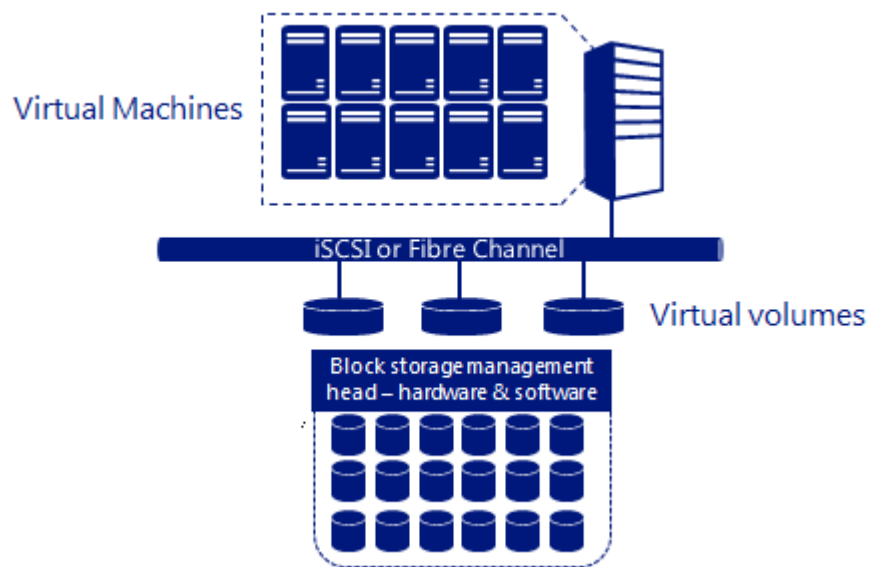
- Особенности физической конфигурации
 - Объединять диски разной производительности в один том [обычно] неэффективно
- Защита от переполнения
 - Юниксы плохо реагируют на переполнение / и /var,
 - Пользовательские данные имеют свойство разрастаться неограниченно
- Смягчение последствий аварий
 - Файловые системы не только смертны, но и внезапно смертны
 - Бэкап 20% диска восстановить быстрее, чем 100%
- Разные стратегии бэкапа
- Разные политики безопасности

Физические диски

- Типы дисков по производительности:
 - SSD
 - Локальные HDD
 - SAN
 - Сетевые ФС
- По гибкости при эксплуатации
 - Сетевые ФС
 - SAN
 - Локальные HDD/SDD

Что такое SAN

- Storage Access Network
 - Сеть доступа к дискам
 - Пул серверов обращается к разделяемому хранилищу



Filesystem Hierarchy Standard

- Разработан The Linux Foundation под влиянием Unix SVR3/4
 - / Корневая директория, содержащая всю файловую иерархию.
 - /bin Основные утилиты, необходимые в однопользовательском режиме
 - /boot Загрузочные файлы
 - /dev Специальные файлы устройств (сама по себе места не занимает)
 - /etc Общесистемные конфигурационные файлы (имя происходит от et cetera).
 - /home Содержит домашние директории пользователей
 - /lib Основные библиотеки, необходимые для работы программ из /bin и /sbin.
 - /proc Виртуальная ФС, представляющая состояние ядра операционной системы и запущенных процессов.
 - /sbin Основные программы для администрирования и настройки системы, например, init, iptables,
 - /tmp Временные файлы (см. также /var/tmp).
 - /usr Вторичная иерархия для программ; содержит большинство пользовательских приложений и утилит, используемых в многопользовательском режиме.
 - /usr/bin Дополнительные программы для всех пользователей, не являющиеся необходимыми в однопользовательском режиме.
 - /usr/lib Библиотеки для программ, находящихся в /usr/bin и /usr/sbin.
 - /usr/sbin Дополнительные системные программы (такие как демоны различных сетевых сервисов).
 - /usr/share Архитектурно-независимые общие данные.
 - /usr/local Третичная иерархия для данных, специфичных для данного хоста. share[2].
 - /var Изменяемые системные файлы, такие как логи,
 - /var/log Различные файлы регистрации.
 - /var/mail Почтовые ящики пользователей.
 - /var/tmp Временные файлы, которые должны быть сохранены между перезагрузками.

Что стоит выносить в отдельные разделы

- /boot, swap – технически необходимо
- /home – если у вас много пользователей
 - На рабочих станциях /home вообще монтируют по NFS
- /var – многие приложения (Apache, почтовые сервера) хранят там данные
 - Можно вынести отдельные подкаталоги /var, например, /var/www, /var/spool/mail
 - Стоит задуматься про /var/tmp
- /opt, /srv – некоторые прикладные пакеты предпочитают жить там
- Некоторые приложения умеют хранить данные в сыром разделе, без ФС

Рекомендации от RedHat

- Swap
 - $\leq 2\text{GB}$ 2x RAM
 - 2GB – 8GB 1x RAM
 - 8GB – 64GB 0.5x RAM
 - $> 64\text{GB}$ 4GB
- /boot 250Mb
- / (root, /usr) 3-5Gb
- /var 384Mb
 - минимальная потребность самого RHEL

Файловые системы

- Поддерживаемые ФС
 - ext2/ext3
 - ext4 – в CentOS6 ставится по умолчанию
 - xfs, jfs, reiserfs
 - tmpfs – теряется при перезагрузке
 - NFS
- Экспериментальные ФС
 - btrfs, zfs
- ФС с ограниченной функциональностью
 - fat16/32, NTFS, ISOFS (cdfs, iso9660), UDF
 - smbfs

Восстановление после сбоя

- `/sbin/fsck` – аналог `CHKDSK.EXE` в Windows
 - На самом деле, wrapper для fs-specific модулей
 - Для журнальных ФС, первым делом пытается накатить журнал
 - Может провести полную проверку (mark'n'sweeper всех структур данных)
 - Может провести тест поверхности диска

Что такое сбой

- Сбои, которые могут повредить ФС
 - Выключение питания
 - Kernel panic/зависание сервера до reset
 - У SAN – сбой сети
 - У hot-plug дисков – выдергивание без размонтирования
- Может привести к неконсистентности метаданных
 - Потерянные блоки (экстенты)
 - Блоки (экстенты) посчитанные дважды
 - Потерянные файлы и каталоги («сироты»)

Что делает fsck

- Традиционные ФС
 - Dirty flag в суперблоке
 - Mark'n'sweeper: ищем все валидные файлы, все остальное объявляем «мусором»
 - Потерянные файлы (inode) размещаем в каталоге lost+found
- Журнальные ФС
 - Операции над метаданными проводятся как транзакции
 - Сначала описание транзакции пишется в журнал
 - Потом транзакция накатывается на рабочие данные
 - Потом транзакция помечается как завершенная
 - После сбоя, читается журнал и незавершенные транзакции откатываются
- Журнал не панацея
 - При аппаратных сбоях или ошибках в драйвере ФС накат может привести к потере данных
 - Но зато накат журнала гораздо быстрее традиционного чека

«Вытаскивание себя за волосы»

- Как происходит чек корневой ФС после сбоя?
 - Откуда берется /sbin/fsck?
 - Чекать смонтированную ФС нельзя!
 - Но если смонтировать ее R/O, то можно
 - Поэтому / сначала монтируется R/O, потом (если нужно) чекается, потом перемонтируется на R/W

Еще полезные опции при МОНТИРОВАНИИ

- noexec – файлы на этой ФС не будут исполняемыми
 - Сильно затрудняет жизнь вирусам
 - Стоит ставить на /tmp, /var/tmp
 - Стоит задуматься об этом для /home, /var
 - Зависит от того, как система и /home будут использоваться
- nosuid – не работает setuid бит
 - Стоит ставить на всех удаляемых носителях
 - Стоит задуматься для сетевых ФС

Маленькие хитрости

- Unionfs/aufs – каскадно-объединенное монтирование
 - Объединяет несколько ФС в одну, например, CDFS и tmpfs
 - Используется на liveCD/liveFlash
- Loopback device – позволяет смонтировать образ ФС, размещенный в файле

Управление разделами

- `fdisk` – MSDOS & BSD disklabel only
 - `fdisk /dev/sda` – командная строка
 - `p` – распечатать
 - `n` – новый раздел
 - `w` – записать и выйти
 - `q` – выйти без сохранения
- `parted` – MSDOS, GUID disklabel
- `mkfs` – создать файловую систему

parted

```
[root@vzhn ~]# parted /dev/sda
GNU Parted 1.8.1
Using /dev/sda
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) help
  check NUMBER                do a simple check on the file system
  cp [FROM-DEVICE] FROM-NUMBER TO-NUMBER  copy file system to another partition
  help [COMMAND]              prints general help, or help on COMMAND
  mklabel,mktable LABEL-TYPE  create a new disklabel (partition table)
  mkfs NUMBER FS-TYPE         make a FS-TYPE file system on partition NUMBER
  mkpart PART-TYPE [FS-TYPE] START END     make a partition
  mkpartfs PART-TYPE FS-TYPE START END     make a partition with a file system
  move NUMBER START END       move partition NUMBER
  name NUMBER NAME            name partition NUMBER as NAME
  print [free|NUMBER|all]     display the partition table, a partition, or all devices
  quit                        exit program
  rescue START END           rescue a lost partition near START and END
  resize NUMBER START END     resize partition NUMBER and its file system
  rm NUMBER                   delete partition NUMBER
  select DEVICE               choose the device to edit
  set NUMBER FLAG STATE       change the FLAG on partition NUMBER
  toggle [NUMBER [FLAG]]     toggle the state of FLAG on partition NUMBER
  unit UNIT                   set the default unit to UNIT
  version                     displays the current version of GNU Parted and copyright
  information
```

Упражнение

- Создать средствами VBox новый VDI файл объемом 1Гб
- (.iso можно стереть, он нам больше не понадобится)
- Подключить его к вашей виртуалке как второй диск (должен быть виден как /dev/sdb)
- Создать при помощи fdisk раздел Linux ext4
- Создать файловую систему на весь раздел, настроить ее монтирование при загрузке в /opt

Упражнение

- Скопировать образ .iso с расширениями гостевой ОС в /opt (он небольшой)
- Создать loopback device (losetup)
- Смонтировать полученный образ