

Odin

OS:DAY

Семь проблем Linux контейнеров

Сергей Бронников
OpenVZ community manager

Проблема 1: Эффективная виртуализация

- Виртуализация - это разделение ресурсов
- 30 лет назад: дорогие мейнфреймы
- Наше время: виртуальные машины
- Частичное решение: поддержка виртуализации в ЦП
- Накладные расходы
- **Решение:** изоляция процессов
 - все видят всех (файлы, процессы, сеть, информация о процессе)
 - древнее решение - chroot()
 - namespaces (mnt, UTS, IPC, PID, net, user)

Проблема 2: Общие ресурсы

- Все контейнеры используют одни и те же ресурсы: ЦП, ОЗУ, НЖМД и т.д.
- Необходимо **честное** разделение ресурсов
- Нужна приоритизация
 - «Все животные равны. Но некоторые животные равны более, чем другие» — Джордж Оруэлл
- **Решение 1:** контроль ресурсов в OpenVZ
- **Решение 2:** cgroups

Проблема 3: простое управление ограничениями

- Использование **User Beancounters** нетривиально:
 - пользователь должен настроить все эти параметры
 - некоторые параметры взаимозависимы
- Мы сделали набор рабочих конфигураций
- **Решение 1:** vSwar
- **Решение 2:** vctmd

Проблема 4: быстрая “живая” миграция

- Мы можем мигрировать контейнер OpenVZ с одного физического сервера на другой без выключения контейнера
- Мы хотим быструю миграцию даже для “толстых” контейнеров:
 - большой диск: используем shared storage
 - много памяти: ???
- как проходит процесс миграции
- **Решение 1:** сетевой swar
- **Решение 2:** итеративная миграция памяти

Проблема 5: основная ветка

- OpenVZ разрабатывается в стороне
- Затем мы пытаемся добавить наши изменения в upstream (ванильное Linux ядро)
- **Проблема:** разработчики ванильного ядра не принимают наши изменения
- **Решение 1:** переписать с чистого листа
- **Решение 2:** CRIU (Checkpoint and Restore In Userspace)

Проблема 6: общая файловая система

- Контейнер это директория на НЖМД, все контейнеры используют одну и ту же файловую систему
- Журнал файловой системы - узкое место
- “Живая миграция”: rsync меняет inodes
- Нельзя сделать снапшот для отдельного контейнера
- Тип файловой системы и её свойства фиксированы
- **Решение 1: LVM**
- **Решение 2: loop**
- **Решение 3: ploop**

Проблема 7: неэффективное использование устройств хранения

- Большинство пользователей - хостинг провайдеры
- Средняя утилизация дискового пространства - 37%
- **Проблема:** Неиспользованное место
- **Решение 1:** SAN
- **Решение 2:** Virtuozzo Storage

Спасибо за внимание, вопросы?

Сергей Бронников sergeyb@odin.com, @estet