# New Filesystem Freeze API

Fernando Vazquez <fernando@oss.ntt.co.jp>
Linux Plumbers Conference, September 8th 2011

# Filesystem freeze

- A simple definition
  - Capability of suspending writes to a filesystem which is usually coupled with a thawing option to resume writes
- Use case
  - Allows backup systems to snapshot a consistent state

# Filesystem freeze in Linux

- Behavior
  - Suspend writes
    - Including all vfs I/O submission APIs and mmap I/O (the latter only since Linux 3.0)
    - We need to get the metadata and journal out so there are several special cases for them
  - Sync filesystem
  - Call filesystem specific freeze function
- Used to be a xfs specific ioctl but is now a VFS interface

# Current API

- VFS API
  - Freeze: ioctl(fd, FIFREEZE, 0)
  - Thaw: ioctl(fd, FITHAW, 0)
- Warts
  - It is possible to umount a frozen filesystem (as of Linux 3.1-rc5)
    - However it is not possible to thaw by block device
  - No check API
    - There is no reliable way to know whether a filesystem is frozen or not

# Filesystem freeze and virtualization

- Use case: hypervisor initiated live-snapshots
  - A live-snapshot is a snapshot taken while a virtual machine is running
- Live-snapshot is a multi-step process
  - Request consistent filesystem state
  - Put storage driver backend (emulator) in a quiesced state
  - Create snapshot (possibly leveraging storage backend specific snapshotting capabilities)
  - Update virtual disk image if needed
  - Release consistent filesystem state
- The first and the last operation require collaboration from the guest

# Hypervisor initiated live-snapshots

- Guest collaboration is achieved through agents running inside the guest
  - Linux (KVM) guests: virtagent
  - Windows: VSS
- Virtagent uses the Linux filesystem freeze API, which comes with some risks
  - If the agent exits or is killed while the filesystem is frozen who is going to thaw it?
    - We may not be able to restart the agent and the guest's root user is likely to be unaware of what is going on
  - Even if we manage to restart the agent when it goes away accidentally it is not possible to establish what the state of the file system is reliably
    - There is no check API!

# New filesystem freeze API

- FIISFROZEN: vfs ioctl to check freeze state
- BLKISFROZEN: add block device ioctl to check freeze state
  - Useful to thaw unmounted frozen filesystems
  - Might get rid of this if returning EBUSY when trying to unmount a frozen filesystem is acceptable
- FIGETFREEZEFD: freezes the indicated filesystem and returns a file descriptor; as long as that file descriptor is held open, the filesystem remains open
  - Since the filesytem is automatically thawed when the file descriptor is closed, if the agent goes away the filesystem will be automagically thawed by kernel
- Filesystem freeze fs ioctls: FS_FREEZE_FD, FS_THAW_FD, FS_ISFROZEN_FD
  - Added new parameter to FIGETFREEZEFD ioctl to indicate whether the filesystem should be frozen on fd open
  - Useful when you are trying to restart the agent

# Future work

- Try to get it merged upstream
- VSS (Volume Shadow Copy Service)-like API?

# Questions?