Linux Lustre client

Sanity prevails !!!
Altered reality

- Lustre is branching out
  - Ubuntu 16 client and server support
  - ARM and Power8 systems running newer kernels
  - Upstream client

- All are in bug squashing mode
  - Bugs are unique to these newer kernels and upstream
Upstream issues addressed on Intel branch

- Upstream client has limited scope testing
  - Only test run are racer and sanity
- Limited man power on upstream client
  - Developers
    - James Simmons
    - Neil Brown
    - Doug Ouchareck
  - Reviewers
    - Andreas Dilger
    - Patrick Farrell
- Seeing same bugs on Intel and upstream branches
Issues tackled on Intel branch for upstream

• LU-6245 / LU-6401 : UAPI header separation
  - Work completed for upstream and master
  - Packaging simplified
  - Sets the path forward to build lustre utilites against upstream client

• LU-10785 : xattrs and acl handling are broken
  - Lustre utilities heavily makes use of these. Big impact to users
  - Fixes for both upstream and master done

• Many IB fixes done to support newer kernels
  - Latest MOFED supported
  - Support for 64K pages done. Will land to 2.12
  - Working with Doug to port upstream
Issues tackled on Intel branch for upstream continued

- LU-9091: 64 bit time and tickless kernel support
  - No more jiffies sent over the wire
  - No more jiffies usage with /proc,/sys, or debug interface
  - Migrate away from jiffies to ktime_t and time64_t
    - Jiffies depend on HZ which is dependent on how kernel was built
    - Tickles kernel support

- LU-8066: sysfs and debugfs support
  - Very broken upstream. Working out issues on master
  - Debugfs is only available to root
  - Made lctl set_param -P almost functional
  - Udev event support

- LU-8980: tracepoint
  - Intel asked that tracepoint never be implemented for Intel or upstream
Impact on development for Intel branch

- Most kernel code style adopted.
  - If (ptr == NULL) should be if (ptr)
  - Mo more __u64 types please. Use u64, u32 etc.

- Handling 64 kernel time
  - Please use ktime_t or time64_t not int for time handling
  - Don’t use get_seconds(), time_after32() and other obsolete functions. Will go away in near future.
  - Don’t use jiffies except for schedule_timeout()

- No more proc entries

- New purposed features should be discussed with VFS maintainers
Impact on administration

- **New UAPI headers**
  - Located in `/usr/include/linux/lustre and /usr/include/linux/Inet`
  - Temporary wrappers in `/usr/include/lustre`. Build warnings.

- **For lctl set_param -P use need a udev rule**
  - `SUBSYSTEM=="lustre", ACTION=="change", ENV{PARAM}=="?*", RUN+="/usr/sbin/lctl set_param $env{PARAM}=$env{SETTING}"`

- **Can tune lustre with udev rules**
  - `SUBSYSTEM=="lustre", ACTION=="add", RUN="/usr/sbin/lctl set_param $attr{bulk_timeout}=200"
  - Udevadm info -a -p /sys/fs/lustre/...

- **Lustre reporting state changes like evictions, LNet health : LU-10756**
  - `SUBSYSTEM="lustre", ACTION=="change", ENV{STATE}=="RECOVER", RUN="/usr/sbin/lctl dk > /tmp/dump.log"`
Upstream client progress since last LUG

- More involvement
  - SuSE has express a strong interest
- Updated the TODO file for the roadmap
- All the checkpatch and style changes are nearly done
- Many bugs being resolved
- Technical debt elimination
  - Migration to workqueues – better scaling
  - Libcfs is almost gone.
  - Some Lustre specific code integrated into linux kernel
- Pre-2.9 version with many fixes from recent lustre
Remaining Upstream client task

- Resolve sysfs and debugfs issues
- Remove linked list use for kernel ↔ user land interface
  - Nodemap (Intel branch only) and LNet selftest
- ioctl cleanup and remove redirection
  - Move to netlink API which is very IO forward friendly
  - Remove many ioctls not used any more.
- Macro cleanup and no function pointer redirection
- Continue back porting fixes from master
Conclusion

- Strong interest in Upstream client
- Most major changes needed for upstream client completed
- Long standing bugs are being resolved
- Better integration with Linux kernel