**OFI Data Storage / Data Access Subteam Weekly telecom – 05/02/2017**

**DS/DA Shared Documents:** <http://downloads.openfabrics.org/WorkGroups/ofiwg/>

**Agenda**

* roll call, agenda bashing
* libfabric PM discussion

**Opens**

Mark Carlson – call for sessions for SNIA Storage Developers’ Conference is now open. Mark suggests that a session from OFA would be very welcome. Technical details, different from last years’ session. Could use it to recruit people to this group – what’s the group working on, etc.

**libfabric PM discussion – Chet Douglas (Intel)**

Chet has a libfabric proposal, but has been awaiting progress in the IBTA (LWG).

* what can we do today without any changes?
  + ‘appliance method’ – send a number of writes to the target machine bypassing L3 cache, meaning that data goes directly to the persistence domain. A write followed by a read by the initiator ensures ordering. The extra read amounts to 2uS, so many customers are finding it useful. This is an ‘upper level protocol’ implementation. May require BIOS to enable non-allocating write transactions (to avoid L3 cache).
  + ‘general purpose server method’ – takes advantage of DDIO (direct data I/O – still uses allocating writes which means data is written to L3, and still requires a following send that includes a list of cache lines to flush. Software on the target side uses the list to force a flush. Uses an SFENCE, which is an ISA construct, at the target end to force a flush.
  + Both mechanisms have been implemented today by Intel. Uses Intel’s NVML library to do all this, with libfabric under the covers.
  + All require additional messages to make this happen.
* What’s going on in the industry:
  + Example use case: the ‘tail of the pointer’ case. Want to update the log tail pointer after all data has been made persistent.
  + Includes the concept of a cache flush and a non-posted write to force cache flushes.
  + RDMA Writes 🡪 RDMA Flush 🡪 Non-posted write (to update the pointer)
  + RDMA Flush and RDMA NP WRITE are new wire opcodes
  + Intel proposed an optimization of combining the Flush and NP Write into a thing called RDMA Fenced Write w Commit. Net result is to eliminate one round trip.
  + Tom Talpey proposed an optimization combining the above together with the RDMA writes, into one operation.
* What’s going on in SNIA
  + Doug Voigt published the HA white paper some time ago.
  + Ordering should be based on a memory region basis.
* Will soon have performance data for 100G (current numbers are for 40G).

**Next meeting**

* NVDIMM ACPI and UEFI spec review.
* Review latest performance numbers and use cases (100G).

**Webex Recording:**

|  |  |
| --- | --- |
| [**Play recording**](https://cisco.webex.com/cisco/ldr.php?RCID=4f3fe451970fe6060a975d92a2a7f08b) (1 hr 3 min) | |
| Recording password: tGQsTWd7 |  |

**Next regular telecom:**

Next meeting: Tuesday, 05/16/17

8am-9am Pacific daylight time

**Logistics:**

See the OFA’s central calendar (<https://openfabrics.org/index.php/ofa-calendar.html>) for current meeting logistics.