**OFI WG Data Storage / Data Access Subteam Weekly telecom – 03/31/2015**

**OFIWG Download Site:** [www.openfabrics.org](http://www.openfabrics.org) 🡪OFED/OFA Resources 🡪 OpenFabrics Interfaces WG

**Agenda**

* roll call, agenda bashing
* Coverage for 4/7 meeting
* Use cases for completions - Bernard
* Continue with SNIA/RDMA summary – Doug Voigt
* Continue review of key objects and methods – Stan Smith

**Coverage for next week’s meeting**

Frank Yang

**Use Cases for Completions – see Bernard’s new table illustrating completion levels**

Described 4 completion levels:

- 1 payload handed to the ‘fabric - as before)

- 2 (as before)

- x payload visible to the remote consumer. Could be in cache or DRAM, but not necessarily persistent

- 3 payload is within the persistent domain

The real distinction between ‘x’ and ‘3’ is who is consuming the completion notice. In the case of ‘x’, the main consumer is the responder on the target side, whereas the main consumer of ‘3’ is the requester. If the requester receives such a completion, it knows that it can release its resources on the requester side.

Both case ‘x’ and case ‘3’ could be covered at the application layer. For example, the responder application could send back an ‘acknowledgement’ in the form of a SEND operation, as opposed to having the memory controller (persistent memory controller) interlocked with the network.

So the summary is that use case #3 has a strong use case.

**SNIA NVM Programming Model Discussion – Doug Voigt**

- four sessions at the Workshop last week, two on NVMe, one from Bernard, and one from Doug (SNIA)

- the NVMe ones are NOT part of SNIA

- not yet completely clear with which part of OFI SNIA should engage – DS/DA seems to make sense because of its storage orientation. Note that the SNIA NVM Prog Model has some user level components, which is consistent with other parts of the DS/DA work.

- Quick review of Doug’s slides from the workshop

- SNIA NVM Prog Model TWG (TWG) is suggesting the use of Memory Mapped Files. Essentially, open a file and give it space in the application virtual memory space. The buffers in the page cache are mapped and through a page fault mechanism, they appear in the users’ virtual memory space.

- Version 1 of SNIA NVM – similar in concept to verbs, but chose not to call it an API because different kernels will implement the API differently. This is virtually identical to ‘Verbs’ as described in the IB spec as a semantic description.

-NVDIMM – co-locate RAM and NVM on a single DIMM along with enough power to allow the RAM to be dumped to NVRAM.

- Focusing on NVM.PM

-NVM.PM.VOLUME mode

-NVM.PM.FILE mode

- Significant changes in NVMP Ver 1.1: atomicity and high availability:

- High availability becomes important for remote access.

- SNIA has published a white paper on this topic, including a set of requirements:

assurance of remote durability, efficient byte range transfers, efficient large transfers, atomicity of fundamental data types, resource recovery and hardware fencing after failure.

Want to make sure that these requirements can be boiled down and are included in our requirements document.

- NVM.PM.File.Map, Sync, Remote Optimized Flush: flushes CPU cache for indicated range.

Optimized flush: multiple ranges from user space (analogous to a disk write)

Optimized flush and verify: similar to write and verify read back with a disk drive.

Both assure that data is written to persistent media, but without the verify (you don’t’ have assurance that it was actually written).

- How this relates to OFI WG:

The semantic operations described above would be implemented as an API.

That API would be a consumer of APIs produced by OFI / DS/DA.

- Could be modeled as a local and remote filesystem in kernel space, such that when a write or flush occurs, it is reflected to the remote filesystem automagically.

Continuing from slide 15: Low Latency Remote OptimizedFlush

- The objective for optimized flush is to flush certain segments of the cache (or stuff left in the RNIC) into persistent memory. (This would be a really handy optimization for byte addressable memory.)

**Agenda for next meeting**

- Continue with Doug Voigt’s discussion of the SNIA NVM Programming model

- Review completion use cases, discuss feedback from OFI WG discussion.

- Continue with EP setup for the client side (Stan)

**Next regular telecom**

Next meeting: Tuesday, 4/7/15.

8am-9am Pacific daylight time

**NOTE:** We have switched over to using Webex (courtesy of Cisco). The URL for joining meetings is:

<https://cisco.webex.com/cisco/j.php?J=200935598&PW=67935ad6df07030d5f05044a5b0f>