# **RDMA Flush Extension**

Tom Talpey, Microsoft NFSv4 WG virtual f2f April 29, 2020

# Flush Extension

- A proposed extension to RDMAP and DDP ("iWARP")
- Support for:
  - FLUSH placement guarantee for remote visibility and persistence
  - ATOMIC WRITE transactional 64-bit write, ordered to FLUSH
  - VERIFY hash calculation to verify contents of remote region
- Similar effort under way in InfiniBand Trade Association
  - For IB and RoCEv1/v2, as an Annex
  - Compatible semantics for IB, RoCE and iWARP
  - IBTA extension does \*not\* currently include VERIFY

# Flush Extension IETF Document

- First published in February 2016 (twenty-sixteen)
  - As "draft-talpey-rdma-commit"
    - https://datatracker.ietf.org/doc/draft-talpey-rdma-commit/
  - Described fundamental requirements and concepts
  - Broad consensus led to significant work offline
- Updated in March 2020
  - Authorship from multiple companies
  - Updated requirements and concepts, added specific protocol
  - Still to be written: ordering rules, local interface and local processing
- To discuss today:
  - Whether it's appropriate to adopt as NFSv4 WG work item

#### IETF WG Adoption

- RDDP Work Group no longer active
- Candidates for adoption:
  - TSVWG: "catch-all" transport area, but not specifically RDMA-focused
  - Independent: Possible, but undesirable to update an existing IETF spec this way
  - NFSv4: Relevant, and RDMA expertise exists here
- Recommendation: NFSv4 WG adoption

# Related NFSv4 WG Activity

- RPC-over-RDMA
  - Existing effort, currently being extended as rpcrdmav2
  - No inter-dependence (in either direction), but strong relation
- pNFS NVMEoF use, possibly including PCIe peer-to-peer
  - Presentation earlier in this meeting
- pNFS "Push Mode" layout
  - <u>https://datatracker.ietf.org/doc/draft-hellwig-nfsv4-rdma-layout/</u>
  - Expired, no activity at present
- RDMA expertise and RDMA development activity
- Storage relevance to Persistent Memory-equipped servers
  - Also relevant to "remote shared memory" models

# Current NFSv4 WG Activity

- Informal discussion, with broad support
- Comments received on-list (thanks Chuck!)
  - Deferring response until this discussion
- Final decision will require
  - NFSv4 WG consensus and proposed charter addition
  - WG chairs, Area Director, and other IETF agreement
  - Updated draft as draft-nfsv4-rdma-placement-extensions (e.g.)

# Discussion