OPENFABRICS ALLIANCE

Windows RDMA File Storage

10TH ANNUAL INTERNATIONAL DEVELOPER WORKSHOP

Tom Talpey, Microsoft

Filesystems Panel track

SMB3

- The primary Windows remote file protocol
- Long SMB history, since 1980's
  - SMB1 since Windows 2000 ("CIFS" before that)
  - SMB2.0 with Vista and Windows Server 2008, 2.1 in 7/2008R2
- Now at dialect 3
  - SMB3.0 with Windows 8/Server 2012, SMB3.02 in 8.1/WS2012R2
- Supported:
  - Full Windows File API
  - Enterprise applications
  - Hyper-V Virtual Hard Disks
  - SQL Server
  - New in Windows Server 2012 R2:
    - Hyper-V Live Migration
    - Shared VHDX - Remote Shared Virtual Disk MS-RSVD

SMB3 Features

- Connection management
  - Dialect negotiation, validation
- Authentication
  - Integrity (signing) and/or privacy (encryption)
- Multichannel
  - Provides both trunking/bandwidth and availability
- Resilience and recovery to network failure
- RDMA (in Windows Server)
- File I/O semantics (Win32, and others)
  - With control and extension semantics
  - Filesystem and IOCTL passthrough
- Remote access
  - NTFS, VHD, Named Pipes, RPC, Memory, …

SMB Direct (RDMA)

- Transport layer protocol adapting SMB3 to RDMA
- Fabric agnostic
  - iWARP, InfiniBand, RoCE
  - IP addressing
    - IANA registered (smbdirect 5445)
- Minimal provider requirements
  - Enables greatest compatibility and future adoption
  - Only send/receive/RDMA Write/RDMA Read
  - RC-style, no atomics, no immediate, etc.
- Supported inbox in WS2012 and WS2012R2:
  - iWARP (Intel and Chelsio RNICs at 10 and 40GbE)
  - RoCE (Mellanox HCAs at 10 and 40GbE)
  - InfiniBand (Mellanox HCAs at up to FDR 54Gb)

SMB Multichannel

Full Throughput

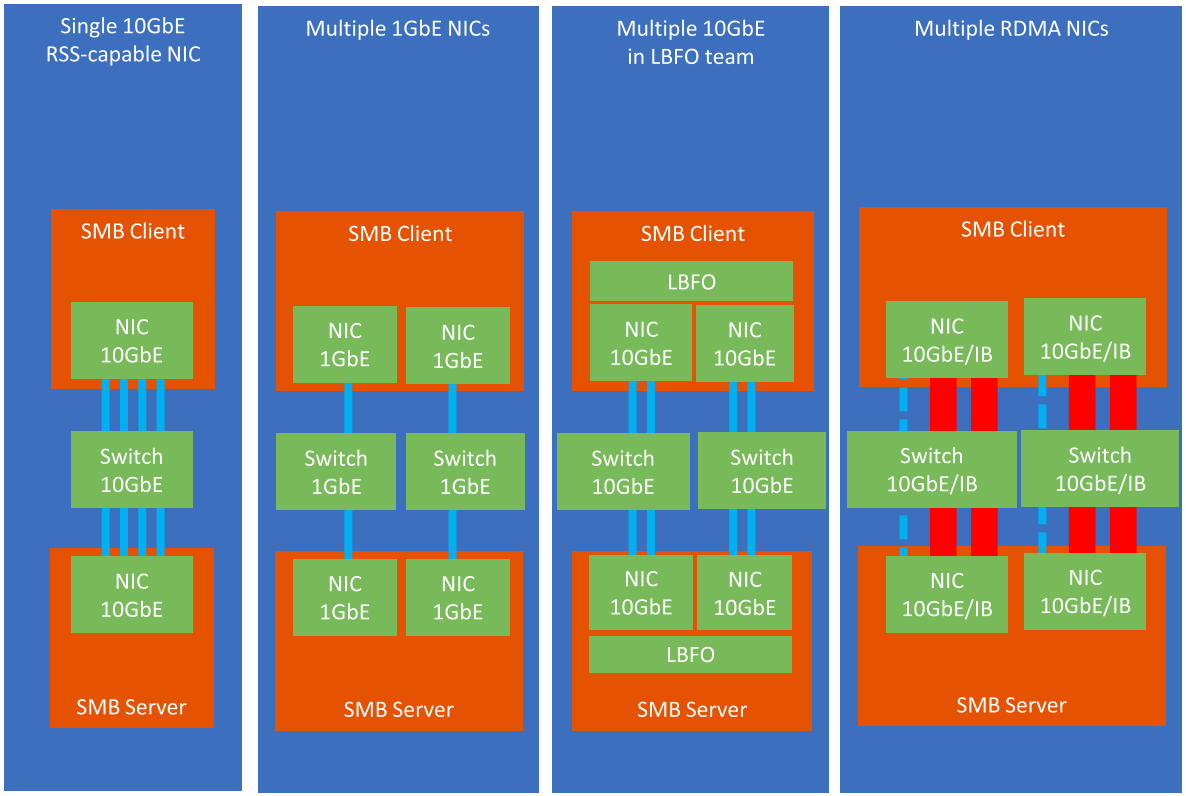
- Bandwidth aggregation with multiple NICs
- Multiple CPUs cores engaged when NIC offers Receive Side Scaling (RSS) or RDMA used – NUMA-aware

Automatic Failover

- SMB Multichannel implements end-to-end failure detection
- Leverages NIC teaming (LBFO) if present, but does not require it

Automatic Configuration

- SMB detects and uses multiple paths
- Zero-config – simply install adapters

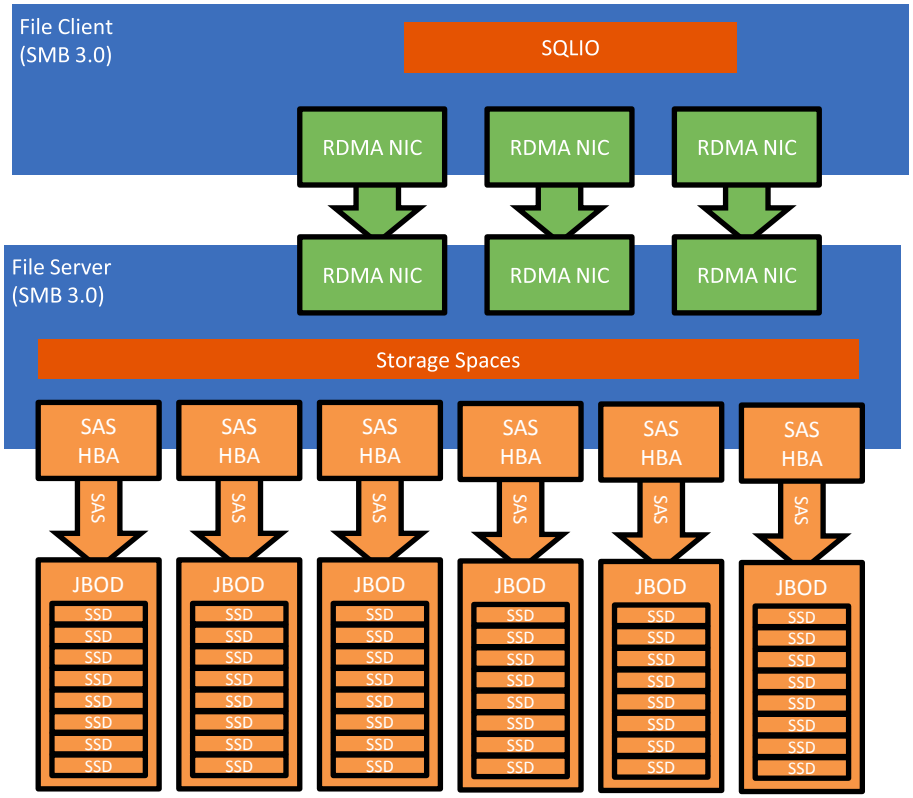


Performance

- Key file I/O workloads
  - Small/random – 8KB, IOPS sensitive
  - Large/sequential – 512KB, bandwidth sensitive
  - Smaller, and larger (up to 8MB), are also relevant
- Multichannel achieves higher scaling
  - Bandwidth and IOPS from additional network interfaces
  - Affinity and parallelism in endnodes
- Unbuffered i/o allows zero-touch to user buffer
- Strict memory register/invalidate per-I/O
  - Key to enterprise application integrity
  - Performance maintained with remote invalidate and careful local behaviors

Performance

Single client, single server, 3xIB FDR multichannel connection, to storage and to RAM



Workload	IOPs
8KB reads, mirrored space (disk)	~600,000
8KB reads, from cache (RAM)	**~1,000,000**
32KB reads, mirrored space (disk)	~500,000

Throughput **>16 Gbytes/second**
45% better than Windows Server 2012

Larger I/Os (>32KB) – similar results, i.e. larger i/o is not needed for achieving full performance!

Link to full demo in “Resources” slide below

Resources

- Jose Barreto's blog
  - http://smb3.info/
  - The Rosetta Stone: **Updated Links on Windows Server 2012 R2 File Server and SMB 3.02**
    - http://blogs.technet.com/b/josebda/archive/2014/03/30/updated-links-on-windows-server-2012-r2-file-server-and-smb-3-0.aspx
  - Performance Demo
    - http://blogs.technet.com/b/josebda/archive/2014/03/09/smb-direct-and-rdma-performance-demo-from-teched-includes-summary-powershell-scripts-and-links.aspx
- SNIA Storage Developer's Conference
  - SDC presentations (multiple years, SMB track)
    - http://www.snia.org/events/storage-developer/archive
- Protocol documentation
  - Microsoft Open Specifications
    - http://www.Microsoft.com/protocols
  - **[MS-SMB2]: Server Message Block (SMB) Protocol Versions 2 and 3**
    - http://msdn.microsoft.com/en-us/library/cc246482.aspx
  - **[MS-SMBD]: SMB2 Remote Direct Memory Access (RDMA) Transport Protocol**
    - http://msdn.microsoft.com/en-us/library/hh536346.aspx
- Microsoft Technet
  - Improve Performance of a File Server with SMB Direct
    - http://technet.microsoft.com/en-us/library/jj134210.aspx
- Windows Kernel RDMA interface
  - NDKPI Reference (provider)
    - http://msdn.microsoft.com/en-us/library/windows/hardware/jj206456.aspx

Thank You

OPENFABRICS ALLIANCE

10TH ANNUAL INTERNATIONAL DEVELOPER WORKSHOP