Progress in Standardization of RDMA technology

Arkady Kanevsky, Ph.D
Chair of DAT Collaborative
Existing RDMA Transport Standards

Virtual Interface Architecture -
http://www.intel.com/design/servers/vi/developer/ia_imp_guide.htm
  • PRO: first standard RDMA architecture
  • CONS: no protocol, no APIs
  • FC-VI and VI/TCP protocols have been developed for it

InfiniBand - http://www.infinibandta.org
  • PRO: fully defined protocols on all levels, suitable for cluster interconnect, storage interconnect – single network for data center
  • CONS: no APIs, verbs only
Emerging RDMA Transport Standards

**RDMA Consortium** - [http://www.rdmaconsortium.org](http://www.rdmaconsortium.org)

- **PRO**: open protocol specs over existing IP infrastructure:
  - MPA (TCP framing), Direct Data Placement, RDMAP
  - upper layer protocols
    - verbs

- feeding to IETF

- **CONS**: early specifications, work in progress


- Work just starting on iWARP
RDMA API Standards

- Based on Intel API example (VIPL-1.0)
- On September 13, 2001 the VIDF ratified *VIDF Extensions to VIPL 1.0* Revision A.
- User level APIs only

- Direct Access Transport
- uDAPL
- kDAPL

**ICSC** - [http://www.opengroup.org/icsc/](http://www.opengroup.org/icsc/)
- Interconnect Software Consortium (under The Open Group)
- ITWG (user level)
- User level sockets
- Fabric management APIs
What is the DAT Collaborative?

Mission statement:

Define and standardize set of APIs that are:

- OS independent
- Transport independent
  - InfiniBand
  - IETF RDDP & RDMA Consortium
  - VI/TCP & FC-VI
- Fully exploit the capabilities of RDMA fabrics

Reference Implementation:  http://sourceforge.net/projects/dapl
Common High-Level Architecture & APIs

- **Mgmt Apps**
  - Mgmt APIs
  - Socket APIs
  - uDAPL (RDMA Transport)

- **User Apps**
  - IB Access

- **OS User Infrastructure**
  - IB Access
  - HCA Library

- **OS Kernel Infrastructure**
  - kDAPL
  - SRP, IPoIB

- **IB HCA**
  - HCA Driver

- **Open Group ICSC**
  - DAT API

- **DAT API**

- **Other Interconnects**

- **= supplied by HCA vendor**

- **= IB Consumers**
DAT Collaborative – Progress

Lightweight organization, low overhead
- No Promoters
- No Sponsors
- Members have all the rights of Promoters and Sponsors

Inaugural meeting at Veritas 6/27/2001
Web page www.datcollaborative.org
Reflector – dat-discussions @ yahoogroups.com

Major Milestones:
- kDAPL-1.0 API ratified (available on the web site)
- uDAPL-1.0 API ratified (available on the web site)
- Currently in Errata Phase (started in August 2002)

Weekly conference calls and monthly face to face meetings
The purpose of the Interconnect Software Consortium is to develop and publish software specifications, guidelines and compliance tests that enable the successful deployment of fast interconnects such as those defined by the InfiniBand™ specification. Software specifications include programming interfaces and protocols. The specifications to be developed include:

- Extensions to the UNIX Sockets API;
- An API that provides direct user application access to interconnect transport;
  - *uDAPL-1.0 is the starting point*
- APIs that provide application access to interconnect fabric management infrastructure.
RDMA ULPs

SDP – Socket Direct Protocol
- IBTA defines SDP (Annex A4)
- RDMAC working on changes needed for RDMA Protocol

- Direct Access File System (DAFS) Protocol - a new file-access protocol designed to take advantage of emerging RDMA (remote direct memory access) interconnect technologies such as InfiniBand, VI and iWARP.
  - Products shipping from Network Appliance
  - Demonstrated working systems by Fujitsu, Duke U., Harvard U., U. of BC, Broadband Storage

iSCSI extensions over RDMA
- RDMA Collaborative
Comparison of File Access Methods

**NFS**
- Application
- Buffers
- FS Switch
- NFS
- Buffer Cache
- TCP/IP
- NIC Driver
- NIC

**Local FS**
- Application
- Buffers
- FS Switch
- File System
- Buffer Cache
- SCSI Driver
- HBA Driver
- HBA

**DAFS**
- Application
- Buffers
- DAFS DAPL
- HBA Driver
- HCA Driver
- HCA

User Space
OS Kernel
H/W