AHUG@ISC19 Workshop

FUJITSU

shaping tomorrow with you

The First SVE Enabled Arm Processor: A64FX and Building up Arm HPC Ecosystem

Shinji Sumimoto, Ph.D.
Next Generation Technical Computing Unit
FUJITSU LIMITED

Jun. 20th, 2019

Outline of This Talk

FUJITSU

- The First SVE Enabled Arm Processor: A64FX
  - A64FX: High Performance Arm CPU

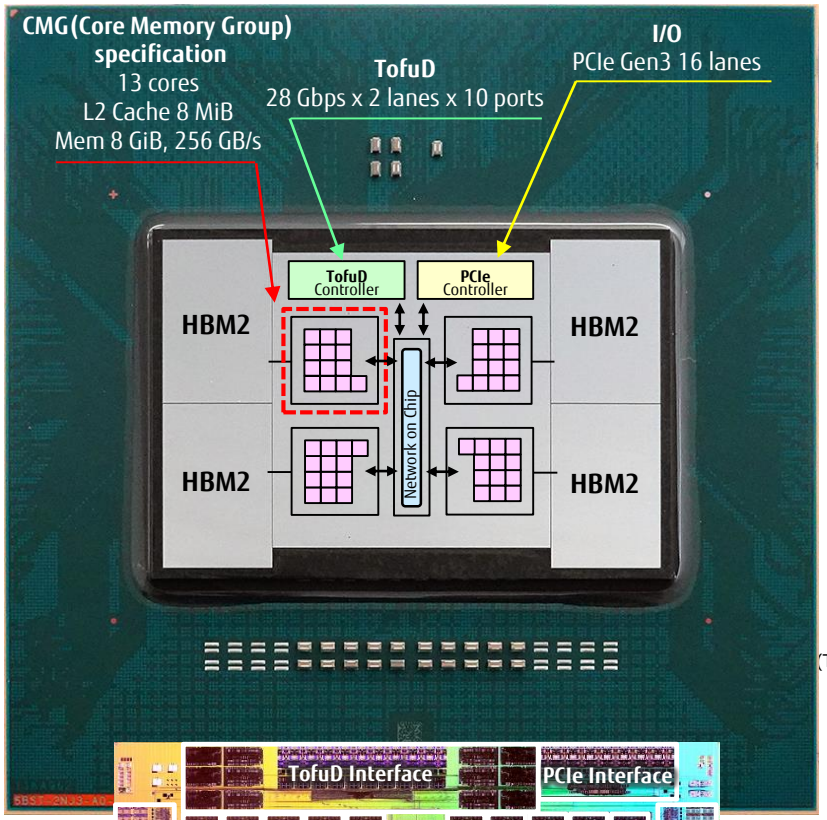
- Arm HPC Ecosystem Development
  - Arm HPC Software Topics
    - Activities with Arm, Linaro and OSS Community
    - OSS Application Porting Updates

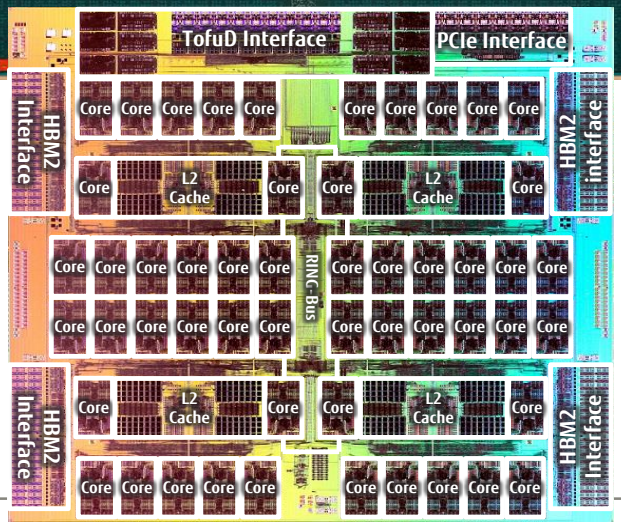
A64FX: High Performance Arm CPU

- Inheriting Fujitsu HPC CPU technologies with commodity standard ISA

High Performance Arm CPU "A64FX"

FUJITSU

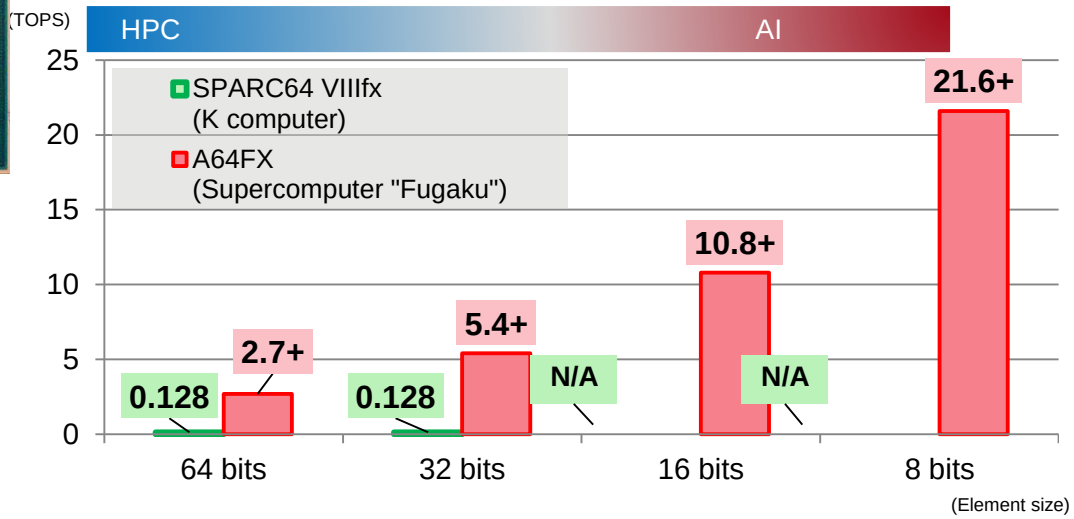




Architecture features

ISA	Armv8.2-A (AArch64 only) SVE (Scalable Vector Extension)
SIMD width	512-bit
Precision	FP64/32/16, INT64/32/16/8
# of cores	48 computing cores + 4 assistant cores (4 CMGs)
Memory	HBM2: Peak B/W 1024 GB/s
Interconnect	TofuD: 28 Gbps x 2 lanes x 10 ports

Peak performance (Chip-level)

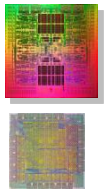


1 Peta FLOPS System: K computer vs. Post-K

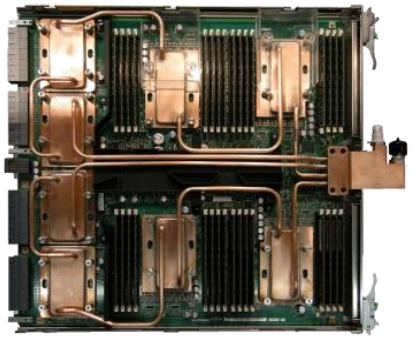
FUJITSU

- K computer
  - 80x compute racks & 20x disk racks
- Post-K (Now Fugaku)
  - 1x rack w/ SSDs

SPARC64 VIIIfx



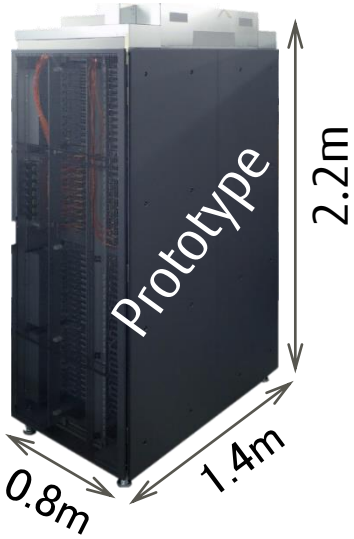
+ ICC

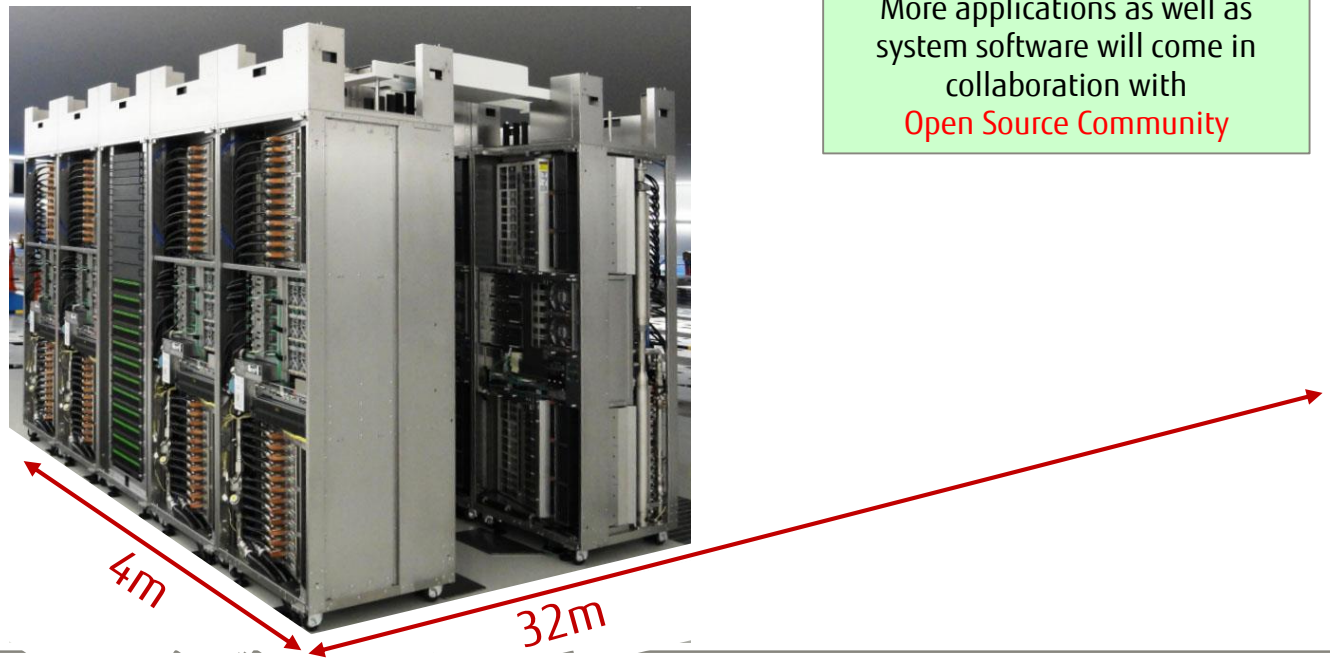


	K computer	Post-K
Compute nodes	7,680(=96x80)	384
IO nodes	4,80(=6x80)	
Footprint (m^2)	128(=4x32)	1.1
	SPARC Linux	Arm Linux

More applications as well as system software will come in collaboration with Open Source Community

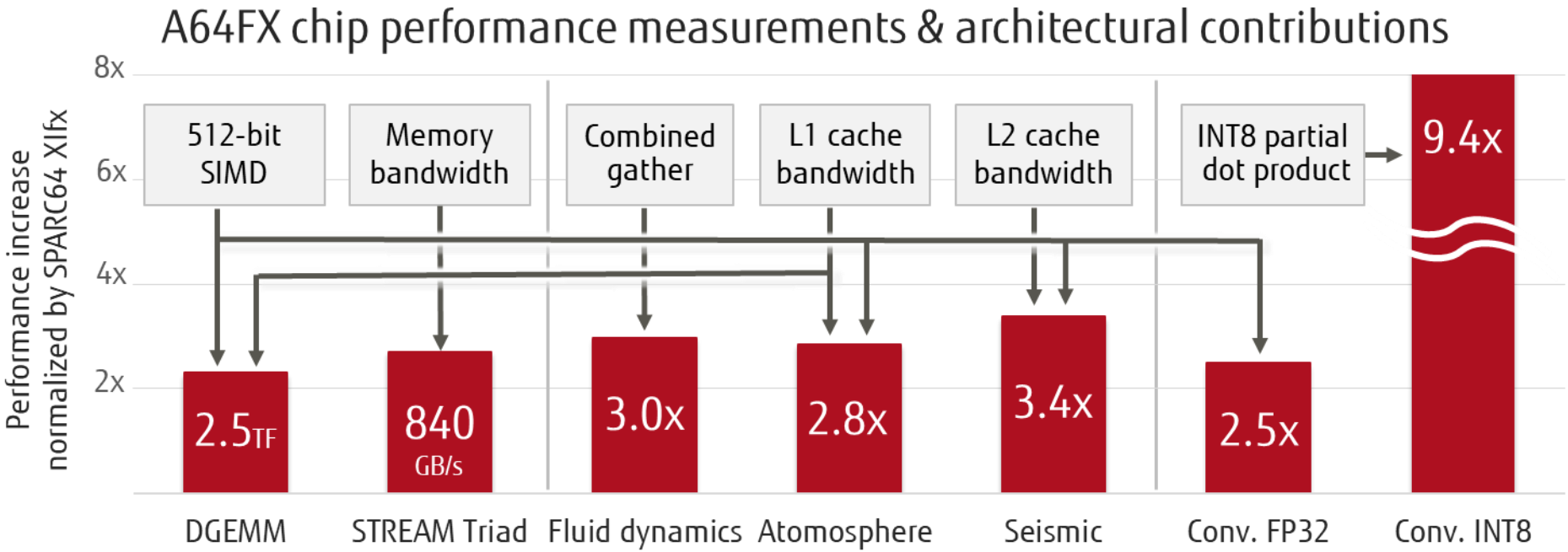
A64FX





A64FX CPU Performance Evaluation

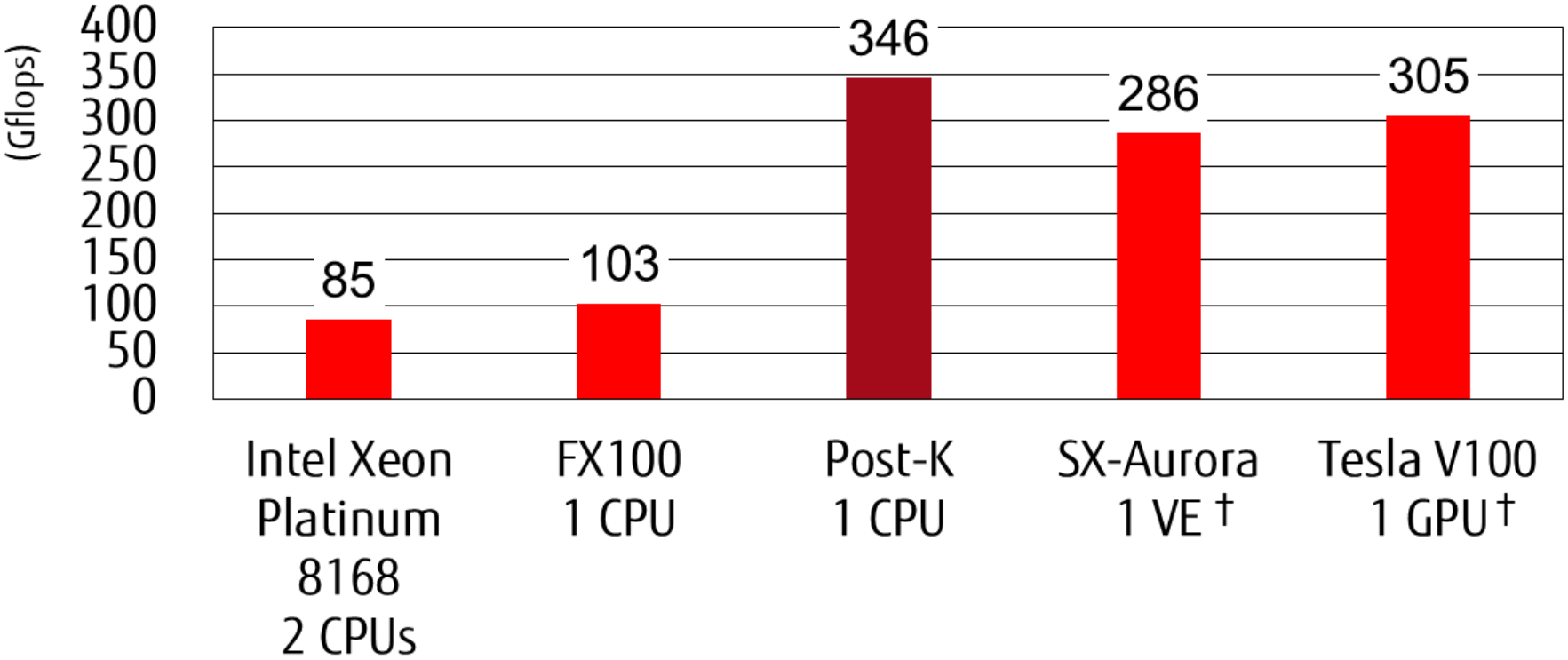
- Over 2.5x faster in HPC & AI benchmarks than SPARC64 XIfx



A64FX Performance Comparison(1/2)

FUJITSU

Himeno Benchmark (Fortran90)



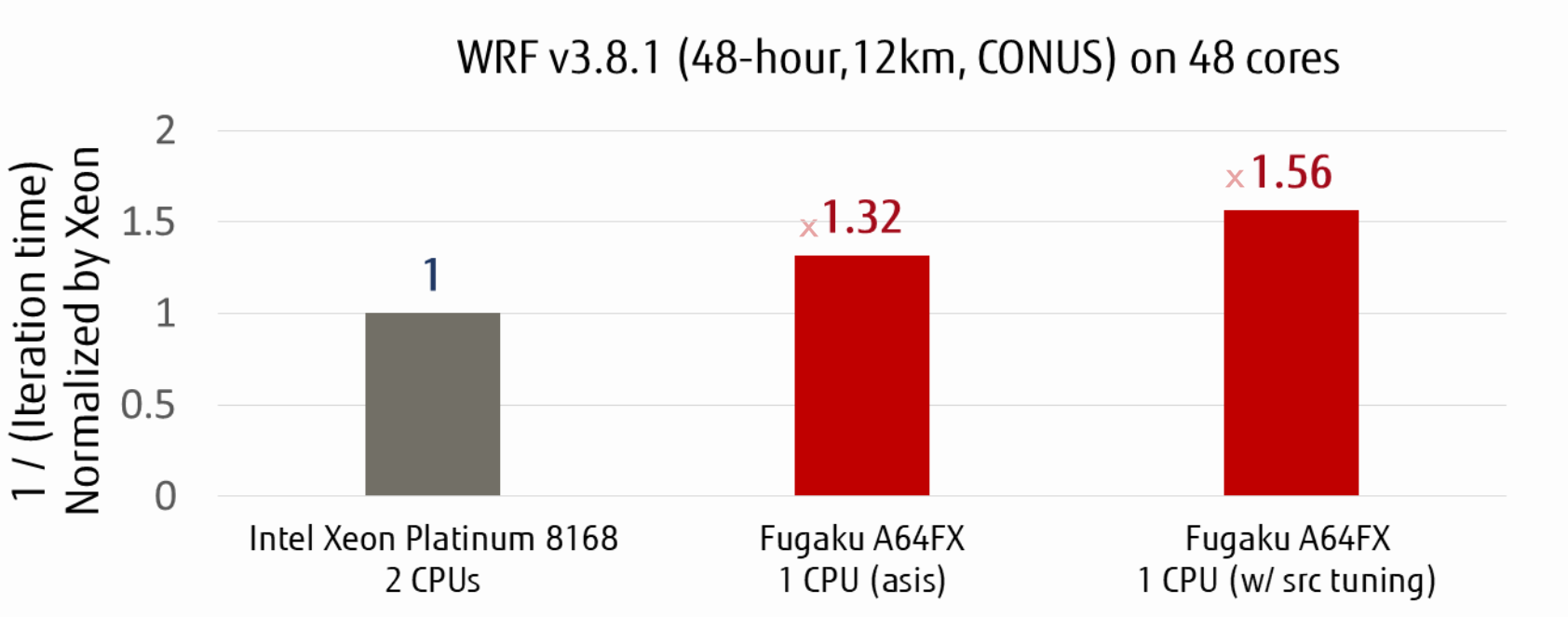
† "Performance evaluation of a vector supercomputer SX-aurora TSUBASA", SC18, https://dl.acm.org/citation.cfm?id=3291728

A64FX Performance Comparison(2/2)

FUJITSU

WRF: Weather Research and Forecasting model

- Vectorizing loops including IF-constructs is key optimization
- Source code tuning using directives promotes compiler optimizations



Arm HPC Software Topics: Activities with Community

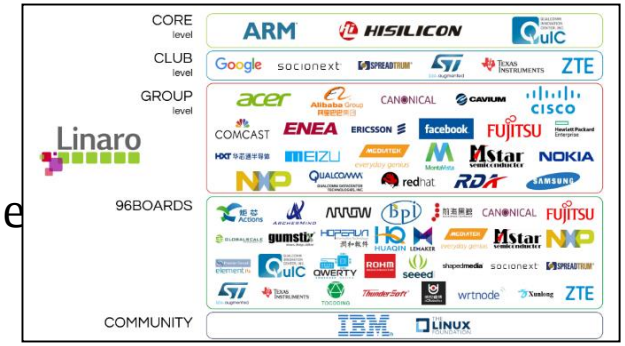
- With Arm and Linaro
- With OSS Community: SPACK, Open MPI and Lustre

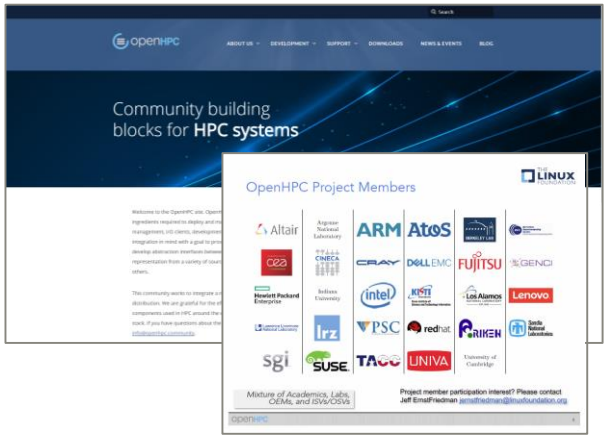
Strong Relationship with Arm HPC community

FUJITSU

- Arm
  - Great Establishment and Contribution to Arm HPC base such as SVE Support of Linux GCC and OpenHPC
    https://developer.arm.com/hpc
- Linaro
  - Building binary portability on Arm HPC
    - Standardization of Arm Basic System Software (Linux Kerne glibc, GCC etc.) and Upstreaming to OSS community
    - Developing and upstreaming SVE software to OSS community
      https://www.linaro.org/sig/hpc/
- OpenHPC
  - Developing Standard IA and Arm HPC software portability
  - Distribution
    - 2017/11: v1.3.3 the first Arm version distributed
    - 2019/6: v1.3.8 for Arm distributed

    http://www.openhpc.community/





Activities with Arm and Linaro

FUJITSU

- LLVM SVE upstreaming and OSS porting with Arm
  - Variable Vector Length Support for LLVM Community in cooperation with Arm
- OpenHPC with Linaro:
  - Mr. Okamoto(Fujitsu) has been selected a 2018-2019 TSC(Technical Steering Committee) member
- Development Status with Linaro
  - LLVM/Clang for aarch64 Improvement: now ongoing
    - Register allocation, Software pipelining support, Vectorization/SIMDization
    - Pushing SVE support to the LLVM community in cooperation with Arm, Variable Vector Length Support is critical issue to introduce to LLVM tree.
  - QEMU/SVE Development: for building SVE software development
    - V4.0.0 released: https://www.qemu.org/

QEMU/SVE Development with Linaro

https://www.qemu.org/

FUJITSU

2019/4/23: Version 4.0.0 Released

DOWNLOAD CONTRIBUTE DOCUMENTATION BLOG

Run operating systems for any machine, on any supported architecture

Run programs for another Linux/BSD target, on any supported architecture

Run KVM and Xen virtual machines with near native performance

QEMU is a member of Software Freedom Conservancy.

Latest releases

4.0.0
Apr 23rd 2019
signature – changes

3.1.0
Dec 11th 2018
signature – changes

3.0.1
Apr 12th 2019
signature – changes

2.12.1
Aug 2nd 2018
signature – changes

Full list of releases

Home
Download
QEMU advent calendar

Contribute
Reporting a bug
New contributors
Submitting a patch
Contributor FAQ

Documentation
Manual
Developer docs
Wiki

This site is made available under the terms of a number of licenses.

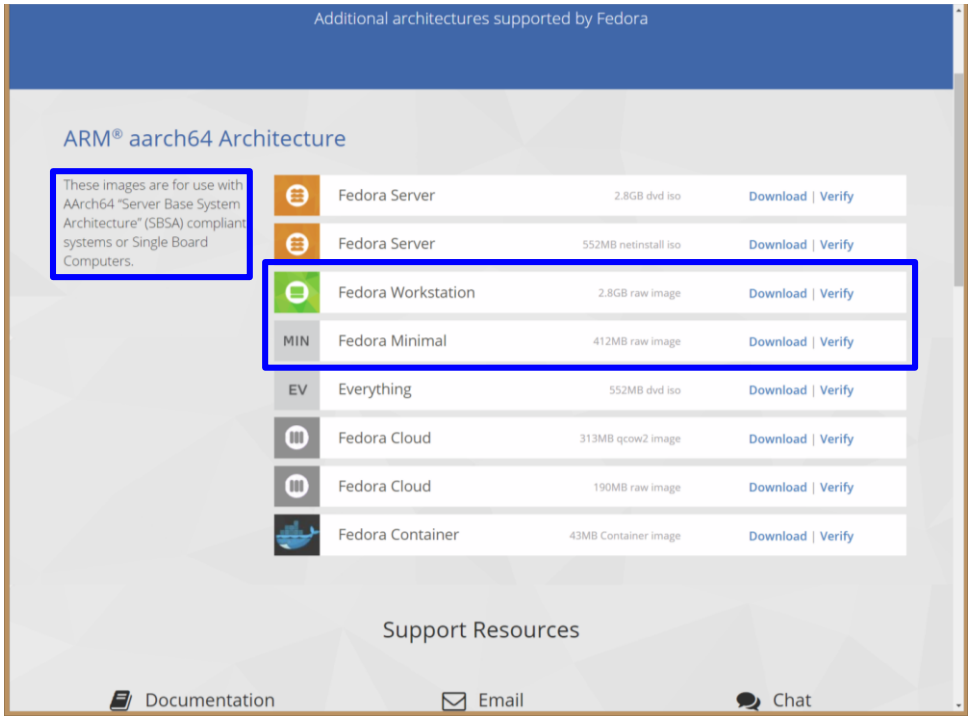
Arm

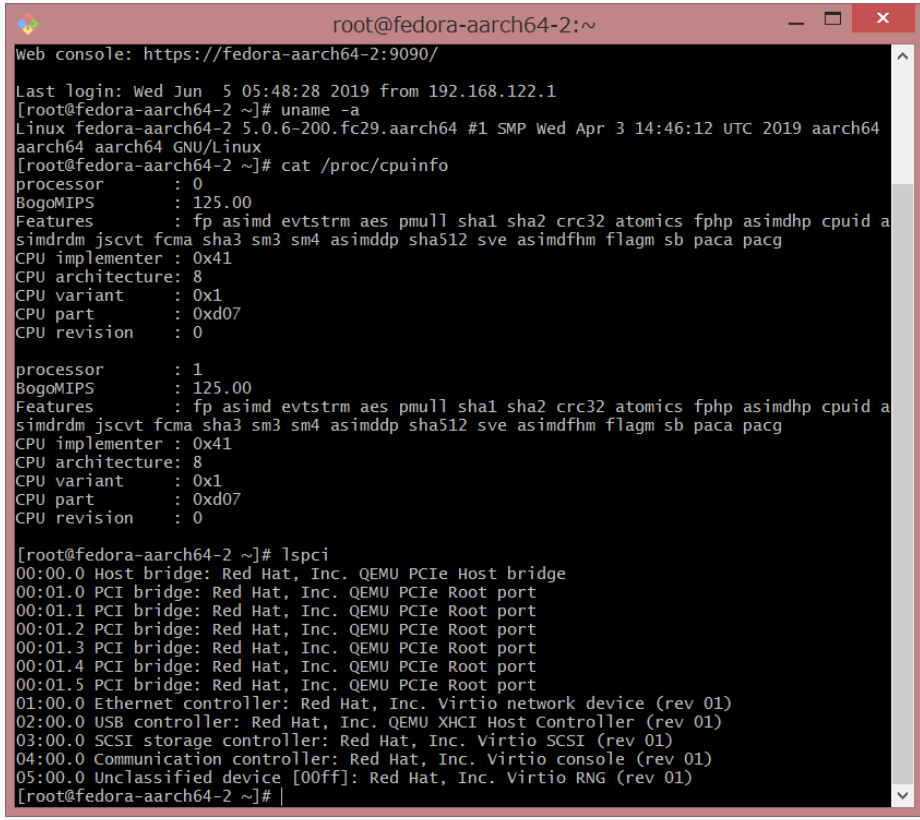
- Implement the ARMv8.0-SB extension
- Implement the ARMv8.0-PredInv extension
- Implement the ARMv8.1-HPD extension
- Implement the ARMv8.1-LOR extension (as the trivial "no limited ordering regions provided" minimum)
- Implement the ARMv8.2-FHM extension
- Implement the ARMv8.2-AA32HPD extension
- Implement the ARMv8.3-PAuth extension
- Implement the ARMv8.3-JSConv extension
- Implement the ARMv8.4-CondM extension
- Implement the ARMv8.5-CondM extension
- Implement the ARMv8.5-FRINT extension
- Implement the Armv8.5-BTI extension for system emulation mode
- New machines "musca-a" and "musca-b1" -- these model the Arm "Musca" development boards
- New machine "mps2-an521" -- this is a model of the AN521 FPGA image for the MPS2 devboard
- Support TBI (top-byte-ignore) properly for linux-user mode
- The micro:bit board now boots simple MicroPython programs since device emulation for the timer, GPIO, NVMC and RNG has been added
- The cubieboard model now implements the 'A' SRAM
- AArch64 processors can now boot from a kernel placed over 4GB into RAM
- The stellaris boards ("lm3s6965evb", "lm3s811evb") now implement the watchdog timer device
- The BLK_MAX register in the TZ MPC device now reports the correct value
- The u-boot "noload" image type is now supported for the Arm virt board
- The Arm virt board now permits more than 255GB of RAM
- stm32f2xx_usart: Do not update data register when device is disabled
- virt board ACPI tables: COHACC override flag now correctly set in IORT SMMUv3 node
- AArch32 exception return is fixed to permit a switch from Mon->Hyp mode
- ftgmac100: implement the new MDIO interface on Aspeed SoC
- Emulation of the ARM PMU has been improved
- Fix various places where we failed to UNDEF invalid A64 instructions
- Don't UNDEF a valid FCMLA on 32-bit inputs

QEMU on Armv8+SVE with Fedora

FUJITSU

- Fedora/aarch64 : https://alt.fedoraproject.org/en/alt/
  - Fedora 29 supports SVE Enabled Kernel and GCC Compilers
  - Easy to Use: Downloading raw-image and Running with virt-manager
    - CPU type: max (not cortex-xx)





Our goals of SPACK world and testing status

- In collaboration with LLNL and R-CCS

About SPACK package manager

FUJITSU

Spack is a flexible package manager for HPC

- How to install Spack:

```
$ git clone https://github.com/spack/spack
$ . spack/share/spack/setup-env.sh
```

- How to install a package:

```
$ spack install hdf5
```

- HDF5 and its dependencies are installed within the Spack directory.
- Unlike typical package managers, Spack can also install many variants of the same build.
  - Different compilers
  - Different MPI implementations
  - Different build options

Get Spack!

http://github.com/spack/spack

@spackpm

LLNL-PRES-806064

github.com/spack/spack

Lawrence Livermore National Laboratory ECP

6

Software Installation with SPACK on QEMU/Aarch64 w/ SVE

- Installing hdf5

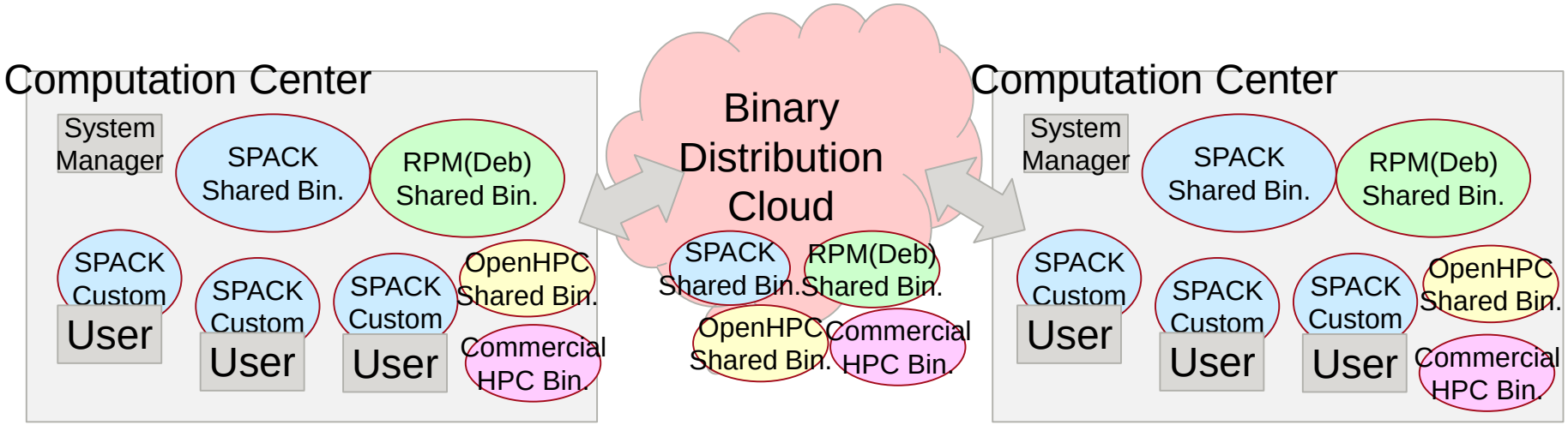
root@fedora-aarch64-2:~/rpmbuild/BUILD/spack

```
9-aarch64/gcc-8.3.1/libsigsegv-2.11-tlfhlq6u5uoz3i6qsxgmod7ab6jjsmzg
==> m4 is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-aarch6
4/gcc-8.3.1/m4-1.4.18-2roqylljdtmbcfmroa4s6nsz72vmshpl
==> libtool is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-a
arch64/gcc-8.3.1/libtool-2.4.6-44nmtplzmyybub7jvebbpvhxiq3lyuw4
==> pkgconf is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-a
arch64/gcc-8.3.1/pkgconf-1.4.2-7yrvgat7h22mf4t36q6fwxev6kc77eab
==> util-macros is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora
29-aarch64/gcc-8.3.1/util-macros-1.19.1-x54p26zd733pf5ofh74utyut4ceklakx
==> libpciaccess is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedor
a29-aarch64/gcc-8.3.1/libpciaccess-0.13.5-ni5zzvuihac33dotbl4poe4jlz4zrgoj
==> xz is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-aarch6
4/gcc-8.3.1/xz-5.2.4-kznfjcl5vf7rhbtfhfbnq66p2marowe6
==> zlib is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-aarc
h64/gcc-8.3.1/zlib-1.2.11-wvyqelnuvl4ojvgndiv6w3aqmod46e7c
==> libxml2 is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-a
arch64/gcc-8.3.1/libxml2-2.9.8-per5r3nnbgy7z7bwn534k4f2vpwq7pz2
==> ncurses is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-a
arch64/gcc-8.3.1/ncurses-6.1-42y2rth5vtgmbwooheru6ns2yzvbvlx7
==> readline is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-
aarch64/gcc-8.3.1/readline-7.0-2pshwgmvetbwsyoirdzpyzhhmf5ymzvj
==> gdbm is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-aarc
h64/gcc-8.3.1/gdbm-1.14.1-urny6skistmlybsajq2kxuykua67whpk
==> perl is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-aarc
h64/gcc-8.3.1/perl-5.26.2-3m4qn32mjr7qpcqlb37or6yiitsaoovm
==> autoconf is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-
aarch64/gcc-8.3.1/autoconf-2.69-jvzdjiyicuebnpkko6ubbwyplsbei73n
==> automake is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-
aarch64/gcc-8.3.1/automake-1.16.1-7avqrgia7j3ktcd4vnr54o5k4npw4qbn
==> numactl is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-a
arch64/gcc-8.3.1/numactl-2.0.11-3nzbl2sozjdrpz7oy7orl5r5lfsowhq5
==> hwloc is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-aar
ch64/gcc-8.3.1/hwloc-1.11.9-wksu5y2gfqzpki76pnlbtsu5uynbvr4u
==> openmpi is already installed in /root/rpmbuild/BUILD/spack/opt/spack/linux-fedora29-a
arch64/gcc-8.3.1/openmpi-3.1.3-zh4lpbw4yp7nr65rn23yqybxsgod2ydo
==> Installing hdf5
==> Searching for binary cache of hdf5
==> Warning: No Spack mirrors are currently configured
==> No binary for hdf5 found: installing from source
==> Fetching https://support.hdfgroup.org/ftp/HDF5/releases/hdf5-1.10/hdf5-1.10.4/src/hdf
5-1.10.4.tar.gz
################################################################################# 100.0%
==> Staging archive: /root/rpmbuild/BUILD/spack/var/spack/stage/hdf5-1.10.4-hfsnkhwtahzbl
ihrcuu7zflo7xxoflps/hdf5-1.10.4.tar.gz
==> Created stage in /root/rpmbuild/BUILD/spack/var/spack/stage/hdf5-1.10.4-hfsnkhwtahzbl
ihrcuu7zflo7xxoflps
==> No patches needed for hdf5
==> Building hdf5 [AutotoolsPackage]
==> Executing phase: 'autoreconf'
==> Executing phase: 'configure'
==> Executing phase: 'build'
```

Our goals of Arm HPC Ecosystem with SPACK

FUJITSU

- Pushing forward building up Arm HPC w/SVE Ecosystem
  - Building tools and applications for Arm HPC w/SVE
  - Distributing and Sharing execution binaries for not only in computational centers but also in Arm HPC users in the world.
  - Final Target is binary distribution including OpenHPC distribution
- Sharing community build binary packages in the world
  - Not only private package building environment but also sharing binaries in centers, countries, and world
  - Not limited open source packages, but also commercial binaries.



Testing Status

FUJITSU

- Testing Environments:
  - Platforms: Thunder X2 system and Qemu 3.1 for Aarch64 w/ SVE
  - OS: CentOS 7.x, Fedora29
  - Networking environment: Proxy Internet Access w/ User Authentication

- First Impression:
  - Working fine very easily on Aarch64 environment
  - Good for custom binary building tool for each execution environment
    - Rpm packages on the Internet are unified binaries.
  - Sometimes timeout on downloading because of heavily congestion internet
    - Curl gave-up and build error occurred
  - Some applications/tools fail to compile not implemented for Aarch64
    - Overall test results are as follows

Result of package installation using Spack

FUJITSU

- Total 3,225 packages are registered (releases/v0.12.1 6th June 2019)
- Confirmation condition
  - Compiling without modifying anything except URL and checksum
  - Using gcc 4.8.5
  - No confirmation for execution
- # of Success

	NOW (Jun. 2019)	Jan. 2019
X86(gcc)	2,386/3225(73%)	2,284/2,907(78%)
Arm(gcc)	2,199/3225(68%)	1,336/2,907(45%)

- Evaluation Status
  - Around 70% of packages are success on Aarch64
  - Existing Several failure patterns
    - Some of them can be fixed by modification of configuration files

Post-K(Now Fugaku) Software Stack

FUJITSU

- Post-K system supports SBSA/SBBR
  - Keeping binary compatibility with the other Aarch64 based systems.

Post-K Applications

FUJITSU Technical Computing Suite / RIKEN Advanced System Software

Management Software	Hierarchical File I/O Software	Programming Environment
System management for highly available & power saving operation	Application-oriented file I/O middleware	XcalableMP
Job management for higher system utilization & power efficiency	Lustre-based distributed file system FEFS	MPI (Open MPI, MPICH)
		OpenMP, COARRAY, Math Libs.
		Compilers (C, C++, Fortran)
		Debugging and tuning tools

arm

Linux OS / McKernel (Lightweight Kernel)

Post-K System Hardware

Post-K
Under Development
w/ RIKEN

Open MPI: from SC18 BoF Slides

https://www.open-mpi.org/papers/sc-2018

FUJITSU

Open MPI: from SC18 BoF Slides

https://www.open-mpi.org/papers/sc-2018

FUJITSU

MPI for the Post-K Computer

- Post-K MPI based on Open MPI
  - Support A64FX (Armv8.2-A+SVE) and TofuD
  - Plan to use Open MPI 4.0 and PMIx 2.1
- Contribution to Open MPI from post-K MPI
  - Persistent collectives [*see next page*]
  - Datatype for half-precision floating point [*early 2019*]
  - Thread parallelization of pack/unpack [*early 2019*]

Half-precision(FP16)datatype development started in cooperation with ANL and Mellanox

FUJITSU

MPI-4.0 or MPI-3.2 ?

Persistent Collectives in MPI-4.0

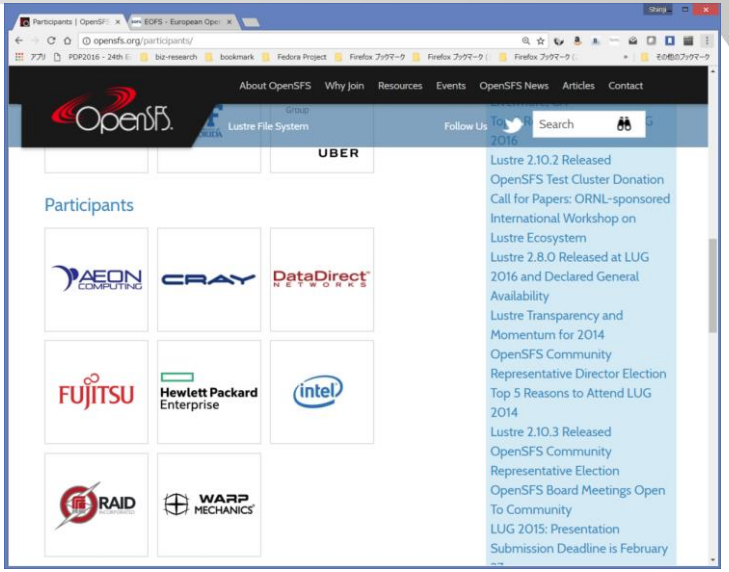
- Persistent collectives are in Open MPI 4.0.x
- Overlap computation & communication and reduce communication initialization cost
- Use MPIX_ prefix because standardization is not complete

```
MPIX_Bcast_init(
  buf, count, ..., &req);
for (...) {
  MPI_Start(&req);
  // ... your computation
  MPI_Wait(&req, &stat);
}
MPI_Request_free(&req);
```

Lustre Community: OpenSFS and EOFS

FUJITSU

- OpenSFS: US Based Non-profit Organization
  - President: Stephen Simms (Indiana Univers*ity*)
- EOFS: EU Based Non-profit Organization
  - President: Frank Baetke (HPE)



- Lustre for Arm
  - Fujitsu is member of OpenSFS and will support Lustre based products.
- Two Major Events
  - Lustre User Group(LUG)
    - LUG 19@Houston, 2019/5/15-17 http://opensfs.org/events/
  - Lustre Admins and Devs workshop(LAD)
    - LAD 19@Paris, 2019/9/24-25 https://www.eofs.eu/events/lad19

Slides Archives are on each site

2018/11: Whamcloud has started Lustre client support on Arm based platforms

https://www.ddn.com/press-releases/ddn-unveils-professional-support-lustre-arm-based-rm-platforms/

DDN Unveils Professional Sup

https://www.ddn.com/press-releases/ddn-unveils-professional-support-lustre-arm-based-rm-platf···

アプリ PDP2016 - 24th Eu biz-research bookmark Fedora Project Firefox ブックマーク Firefox ブックマーク (1 Firefox ブックマーク (2 その他のブックマーク

DDN UNVEILS PROFESSIONAL SUPPORT FOR LUSTRE CLIENTS ON ARM-BASED PLATFORMS

Allows HPC and AI Users to Confidently Deploy Arm Architectures for Mission Critical Applications

SANTA CLARA, Calif. – November 12, 2018 – DataDirect Networks (DDN®) today announced that its Whamcloud division, the foremost Lustre support provider and driving force behind Lustre innovation, is delivering professional support for Lustre clients on Arm® architectures. With this support offering, organizations can confidently use Lustre in production environments, introduce new clients into existing Lustre infrastructures, and deploy Arm-based clusters of any size within test, development or production environments.

As the use of Lustre continues to expand across HPC, artificial intelligence (AI) and data-intensive, performance-driven applications, the deployment of alternative architectures is on the rise.

"With DDN's Whamcloud division now fully supporting the Lustre client on Arm-based systems, users have more choice and can now introduce Arm mission-critical Lustre infrastructures with confidence," said Robert Triendl, senior vice president of global sales, marketing, and field services at DDN "Whamcloud's support is timely and aligns with market demand as customers seek an expanded range of alternative architectures such as Arm-bas

QUESTIONS? Contact a Storage Specialist!

Arm's advanced, energy-efficient processor designs are enabling the intelligence in more than 130 billion silicon chips and securely powering products from the edge to the hyperscale. Arm's expanding momentum in the high-performance computing market is evidenced with recent announcements of deployments on Marvell® ThunderX2® 64-bit Armv8-A processor by leading research and scientific customers, such as Sandia National Laboratories with the Astra Supercomputer.

"The adoption of Arm-based systems in HPC is accelerating to support users at all stages of their application development and providing more options for the diverse range of organizations deploying HPC systems," said Brent Gorda, senior director of HPC, Infrastructure Line of Business, Arm. "Whamcloud's support for Lustre on Arm is a key enabler for broader adoption of Arm in HPC and delivers more architecture options for the HPC community."

Tweet this: Great news for #HPC! The @DDN_Limitless @Whamcloud division announces professional support for Lustre clients on Arm-based platforms – http://bit.ly/2F9KkrT

"Sandia has deployed Lustre-based parallel file systems for many years to support its high-performance computing enterprise needs. Astra, the world's fastest Arm-based platform, will use a flash-based Lustre file system that we expect will maximize the end-to-end efficiency of our mission workload," said James Laros, Vanguard-Astra program lead at Sandia National Laboratories.

Long recognized as a staple technology for those with the most demanding data requirements, Lustre is deployed in thousands of data centers in healthcare, energy, manufacturing, financial services, academia, research and HPC labs, and consistently is selected by top 100 HPC sites as the file system of choice for the world's fastest computers.

"Lustre is critical to scalable, high performance system solutions, and deployments continue to expand across the globe," said Mike Vildibill, vice president, Advanced Technologies Group, HPE. "Collaborating with Whamcloud enables us to advance Lustre adoption for all of our customers, spanning a diversity of system platforms."

"The software ecosystem for Arm based servers continues to gain momentum, and Whamcloud's commitment to enablement, optimization and support of the Lustre file system adds another key component for our partners and customers," said Larry Wikelius, vice president, ecosystem and partner enabling at Marvell Semiconductor, Inc. "Marvell's ThunderX2 processor delivers the compute and memory performance that addresses the demands for Lustre performance at scale

Arm HPC Software Topics: OSS Application Porting Updates

OSS apps porting at Arm HPC Users Group

(http://arm-hpc.gitlab.io/)

- Twelve primary OSS applications are listed and being tested in the Users Group for each compilers, collaboratively w/ Arm

Application	Lang.	GCC	LLVM	Arm	Fujitsu
LAMMPS	C++	Modified	Modified	Modified	Modified
GROMACS	C	Modified	Modified	Modified	Modified
GAMESS*	Fortran	Modified	Modified	Modified	Modified
OpenFOAM	C++	Modified	Modified	Modified	Modified
Siesta*	Fortran	Ok in as is	Issues found	Issues found	Modified
NAMD	C++	Modified	Modified	Modified	Modified
WRF	Fortran	Modified	Modified	Modified	Modified
Quantum ESPRESSO	Fortran	Ok in as is	Ok in as is	Ok in as is	Modified
NWChem	Fortran	Ok in as is	Modified	Modified	Modified
ABINIT	Fortran	Modified	Modified	Modified	Modified
CP2K	Fortran	Ok in as is	Issues found	Issues found	Modified
NEST*	C++	Ok in as is	Modified	Modified	Modified
USQCD (MILC)	C	Ok in as is	Modified	Modified	Modified
BLAST*	C++	Ok in as is	Modified	Modified	Modified

* Registered by Fujitsu

FUJITSU

shaping tomorrow with you