Hewlett Packard
Enterprise

Leading the charge for ARM in HPC.

Perspective, opportunities and challenges 1 year after the last Research Summit.

Andy Warner
Distinguished Technologist
Advanced Technology Group

Comanche Collaboration

Hewlett Packard Enterprise

CAVIUM

redhat

Mellanox TECHNOLOGIES

arm

AMD

Develop, port, optimize, validate & support key hardware & software components

Provide real-world exposure. Select and port/develop workloads/projects

- *Hardware Platform*
- *Firmware & System management*
- *OS: RHEL 7.5*
- *Fabric: EDR InfiniBand*
- *Toolchains: gcc, Arm C/C++/Fortran)*
- *Libraries: ARM Performance Libraries & Open Source*
- *Runtime: HPE MPI, OpenMPI, MVAPICH, OpenSHMEM*
- *Profilers & debuggers: MAP, DDT*
- *GPU support*

- *Performance*
- *Stability*
- *Runtime environment enablement*
- *Lustre*
- *Math Libraries*
- *EDA + General Stack Environment*
- *HPC apps & mini-apps*
- *GPU enablement and machine learning*

Sandia National Laboratories

Los Alamos NATIONAL LABORATORY EST. 1943

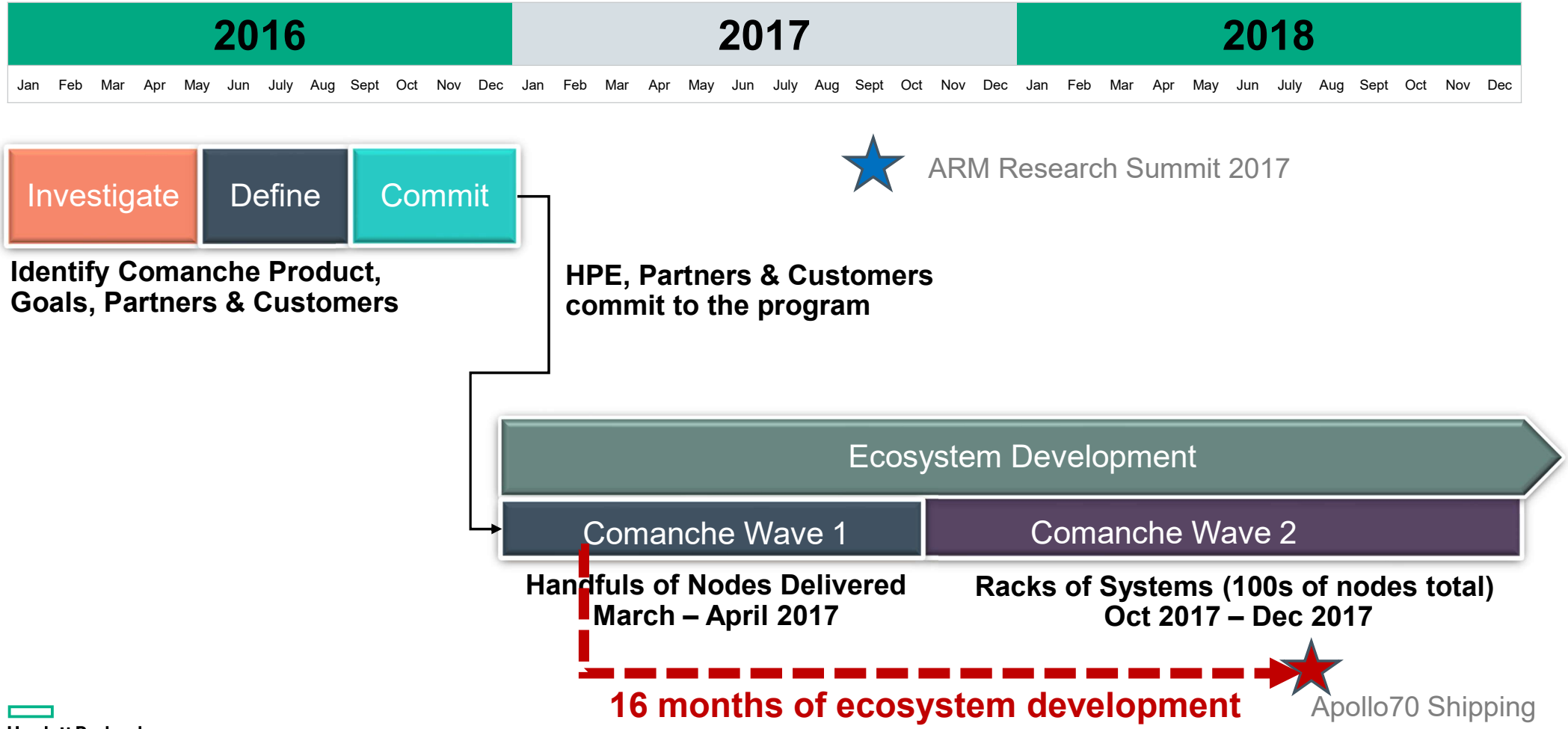
Lawrence Livermore National Laboratory

Argonne NATIONAL LABORATORY

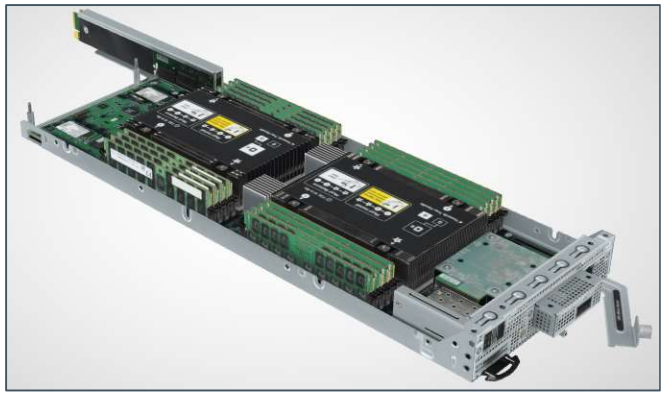
OAK RIDGE National Laboratory

arm

The Comanche Collaboration Timeline



HPE Apollo 70 System



Up to 4 servers in 2U

Specification	HPE Apollo 70 System
Processor	Cavium 64-bit Armv8-A ThunderX2™
CPU configurations	2 processors per node; up to 32 cores & 2.2 GHz
Memory	DDR4-2667 DIMM Support 16 DIMM slots; up to 512 GB per node
Drive Bays	8 LFF HDD/SSD 2 internal 2280 M.2 per node; up to 960GB each
I/O	Single-port OCP form factor Mellanox CX-5 100 Gb/s VPI Adaptor (InfiniBand or Ethernet) Dual-port SFP+ 10GbE Mellanox CX4 LOM Single Port RJ-45 1GbE Mgmt (BMC/IPMI) USB, VGA, UART
Expandability via 2U option	Up to 2 GPU or single additional PCIe (x16)
Infrastructure management	IPMI.2.0 Compliant BMC
Power Supply	Dual 1600W (hot plug)
Warranty	3 years (support, parts & labor)

Catalyst UK

Program Collaborators

UNIVERSITY OF LEICESTER

Program Partners

Hewlett Packard Enterprise

System Configuration

- 64 Apollo70 compute nodes:
  - Dual Cavium TX2 32core @ 2.2GHz
  - 128GB DRAM (16 x 8GB)
  - Mellanox ConnectX-5 EDR HCA
- 7 additional Apollo70 nodes:
  - Admin/head
  - Login/compile/service
  - Storage with 180TB of HDD/SSD
- Non-blocking EDR fat-tree
- Total of 4096 cores & 8TB RAM
- Software :
  - SLES12, HPC Module
  - SUSE Enterprise Storage
  - Mellanox OFED
  - ARM Allinea Studio

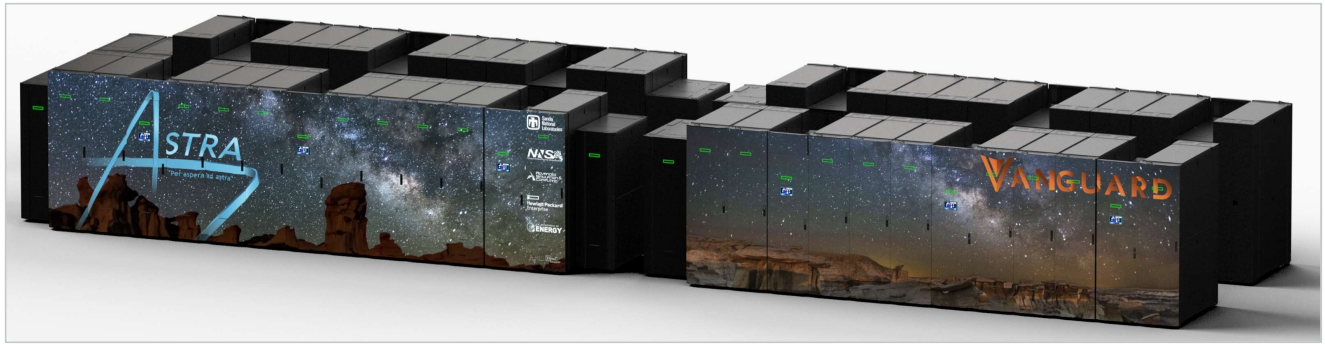
Vanguard Astra

WORLD'S MOST POWERFUL ARM SUPERCOMPUTER

- 2,592 HPE Apollo 70 compute nodes
  - 5,184 CPUs, 145,152 cores
  - 2.3 PFLOPS system peak
- Cavium Thunder-X2 ARM SoC, 28 core, 2.0 GHz
- Memory per node: 128 GB (16 x 8 GB DR DIMMs)
  - Aggregate capacity: 324 TB
  - Aggregate bandwidth: 608 TB/s (stream triad)
- Fabric: InfiniBand EDR, Fat-Tree, Mellanox ConnectX-5
  - 112 x leaf switches, 3 x 648-port spine switches
- Storage: HPE Apollo 4520 All–Flash Lustre
  - Capacity: 403 TB (usable)
  - Bandwidth: 240 GB/s
- Liquid cooled
  - Total 1.2 MW
  - Compute racks are cooled by 12 MCS300 in-row coolers

Sandia National Laboratories

VANGUARD



Hewlett Packard Enterprise

Vanguard Astra – Leadership Class Performance

Bringing Balance Back to DOE Systems

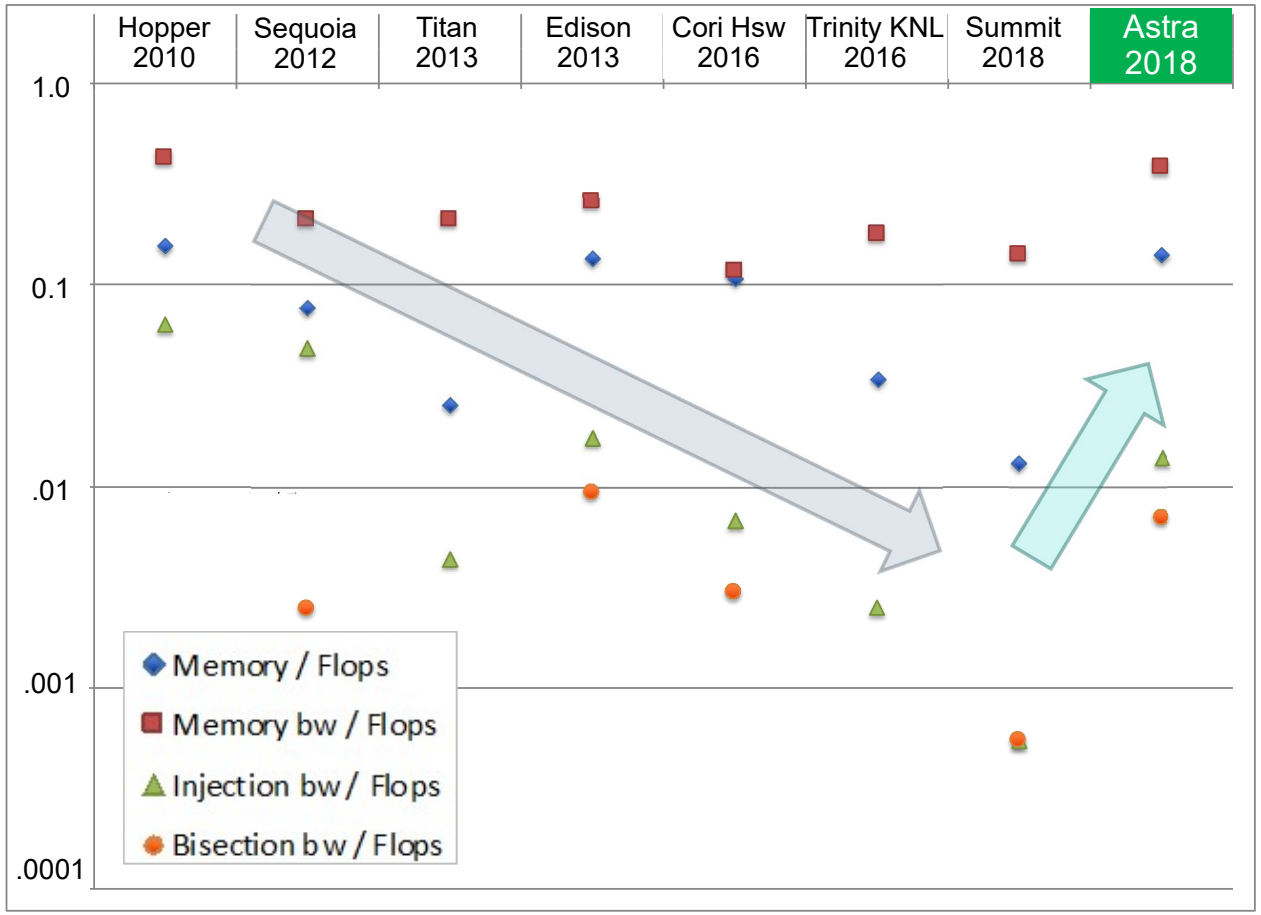
Vanguard Astra ratios versus ORNL's Summit:

- 0.14 DRAM Bytes / Flops (10X Summit)
- 0.38 DRAM Bytes per second / Flops (3X Summit)
- 0.014 Injection bw / FLops (26X Summit)
- 0.007 Bisection bw / FLops (13X Summit)

Versus Edison:

- Similar peak Flops and total memory capacity with half the CPUs
- Increased memory bandwidth
- 50% of the power

Leadership System ratios – aggregate memory tiers



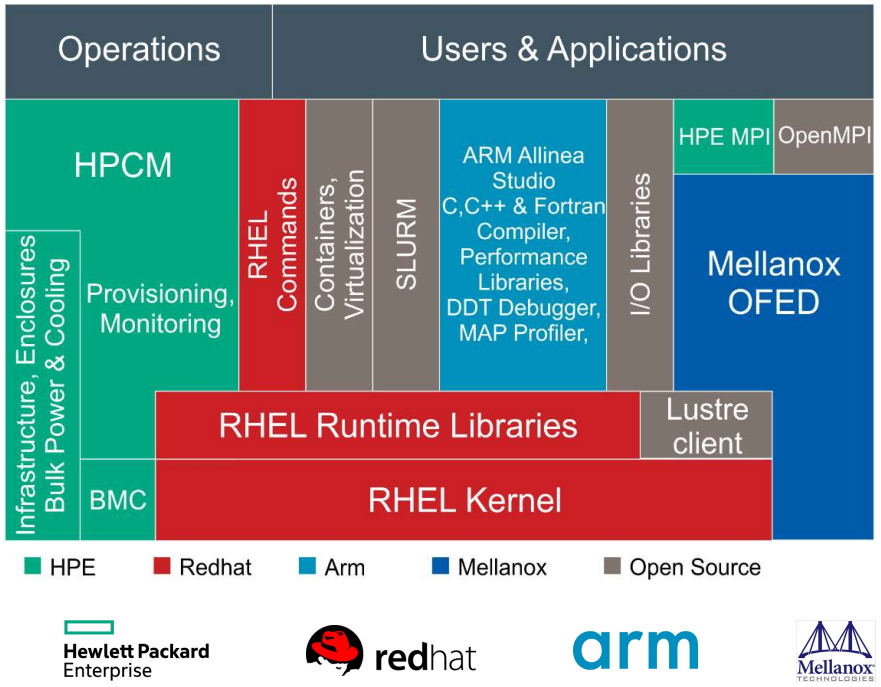
Hewlett Packard Enterprise

Vanguard Astra – Leadership Class Open Source Software

Support Vanguard program mission to develop ecosystems and validate applications on new architectures at scale, reduce risk, and increase choice.

Comprehensive software solution that combines HPE, open source, and third party software:

- HPE:
  - HPE Message Passing Interface
  - HPE Performance Cluster Management
- Redhat:
  - RHEL 7.5
- Close collaboration between HPE Open Leadership Software Stack (OLSS) initiative and Sandia Advanced Tri-Labs Software Environment (ATSE) to provide leadership class open source runtime.
- ARM:
  - Allinea Studio: Compiler, Libraries, DDT & MAP
- Mellanox:
  - OFED, UCX, HPC-X, SHArP



Software ecosystem advancements in the last 12 months

- RHEL for Arm announced at SC'17
- HPE Apollo 70 first Arm-based server to be certified with SLES 12
- RHEL certification of Apollo 70 is complete, formal announcement waiting on some logistics
- Mellanox OFED and HPC-X for ARM
- LLNL Tri-Lab Operating System Stack (TOSS) added aarch64 target
- Lustre client
- Arm Allinea Studio
- OpenHPC (Arm Allinea Studio builds)
- HPE:
  - HPE MPI
  - HPCM HPE system management software (arm compute, arm or x86 admin nodes)

Opportunities & Challenges for the HPC community

– SVE
  - How to deliver the potential performance:
    - Open toolchain support: gcc & llvm; specifically fortran
    - Broad library support
    - Using SVE to deliver memory bandwidth

– Drive and simplify application/workflow availability:
  - Build on the “make things boring” achievements
  - Collaboration and openness are crucial

Hewlett Packard
Enterprise

Thank you

Andy Warner
Distinguished Technologist
Advanced Technologies & Exascale
andy.warner@hpe.com