Fujitsu High Performance Computing Ecosystem

Dr. Pierre Lagier
Chief Technology Officer
Fujitsu Systems Europe

Human Centric Innovation in Action

Innovation
Flexibility
Simplicity
Three challenges for HPC centres

- Compute Power
- Cooling
- Integration factor
Compute Power: which engine?

SPARC64 XIfx
- Proprietary implementation of SPARC processor
- 32 cores, 24MB L2 cache
- ~1 TFlops peak performance*

Intel XEON E5 2600
- The most popular processors on earth, largest number of available applications
- Up to 22 cores (EP series), 55MB L3 Cache
- ~0.6TFlops peak performance*

Intel Knights-Landing
- Will it be the new HPC engine?
- 72 advanced Silvermont cores on one die with 8~16GB on-package multi-purpose memory
- >3TFlops peak performance*

* 64 Bits double precision floating point operations
Fujitsu HPC Platforms

PRIMEHPC FX100
- focusing on Japanese academic customers and national projects
- Proprietary technologies

PRIMERGY CX-400 (Xeon based)
PRIMERGY CX-xxx (KNL based)
- widely used Intel technology, x86 and KNL, with worldwide support over the 5 continents
- Standards compliant
How many compute nodes per rack?

Peak Performance

New CX-xxx
- General purpose and hybrid parallel applications
- Integrated mid-range to high end

FX100 High End HPC
- Proprietary platform for hybrid parallel applications
- Integrated high end HPC

CX-400 Flexible HPC solution
- General purpose and hybrid parallel applications
- Scales in storage memory and number of cores

Compute nodes

>500 TFlops
40 KW

235 TFlops
35 KW

96 TFlops
34 KW

160*

216**

80*

*assuming 40U rack fully populated
**specific proprietary rack form factor

Peak Performance

>500 TFlops
40 KW

235 TFlops
35 KW

96 TFlops
34 KW

160*

216**

80*
Let's cool down!

ALMA - Atacama desert, Chili

The world's largest radio telescope, ALMA, located 5,000 meters above sea level in Chile.

The ACA Correlator, a ultra-high-speed data processing system Fujitsu deployed solution is based on PRIMERGY servers and FPGA.

King Abdulaziz University - Jeddah, KSA.

The largest university in the kingdom, using PRIMERGY supercomputer.

Annual mean temperature is 28.4 degrees Celsius (83.2 degrees Fahrenheit), with an maximum recorded average of 40° C (104F) in summer.

From the warmest to the coolest, Fujitsu has a solution which fits your requirements.
Cool-safe® LCT: How it Works

Facility water requirement
- Maximum water temperature
  - Facility supply: 40 °C
  - Facility Return: 59 °C
- Water pressure
  - Max: 100 PSI
  - Min: 10 PSI

Heat exchanger
Liquid-to-liquid heat exchanger between facilities liquid loop and server liquid loops. Facilities and server liquids are kept separate and never mix.

Pump/cold plate units
atop CPUs, GPUs and RAM circulate liquid through server and RackCDU, collecting heat and returning to RackCDU for exchange with facilities liquid.

Coolant reservoir and control

Pipes
move cooling liquid to and from RackCDU to servers

Leak Detection Panel
(not shown)
The future of HPC
Building an Exa-Scale* machine

*1,000,000,000,000,000,000 operations/s

Number Cruncher
AMD / Nvidia
Intel
ARM

Memory
3D
Memristor

Interconnect
IB
TOFU
ARIES

Power plant
ExaScale Workload

- Remote Visualization
- Traditional HPC workload

Virtual Network

Sanbox VM

Heterogeneous / Hybrid Parallel Applications

Private “On Demand” Parallel Storage

Large Hybrid Parallel Applications

Global Permanent Parallel Storage

Hybrid Parallel Applications

Dedicated “On Demand” Parallel Storage

Global Interconnect

VM Provisioning
Traditional HPC Workload

- An ExaScale Machine is nothing but a “Super Massively Parallel System”
- TCO and ROI will be critical due to the scale of investments and operations
- Reverse trend from Departmental or project oriented solution to globally centralized IT

Normal end users will not change dramatically because of ExaScale
- Any existing application will run on it, even single core ones
- Integrated legacy HPC environment will remain almost the same
- Remote visualization will play an important role
Hybrid Parallel Workload

- Large hybrid parallel applications will take advantage of ExaScale system
- Global permanent parallel storage will stay as general purpose data storage and sharing solution

- Specific parallel applications will ensure the transition to mature new parallel programming paradigms
- Dynamic “on demand” parallel storage will provide better flexibility as well as higher performance
Extended usage

- The “sandbox VM” allows end user to work in highly secured private environment
- Virtual network and interconnect integration will provide full secured parallel environment support

VM provisioning will introduce more flexibility for end users, extending the capabilities of today Linux clusters

This will include both classic VM provisioning as well as remote visualization
Important facts to take in account when moving toward future massively parallel machine.
Where power consumption really sits?

<table>
<thead>
<tr>
<th>Operation</th>
<th>Power Cost (picoJoule)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiply &amp; Add</td>
<td>50</td>
</tr>
<tr>
<td>On package DRAM read</td>
<td>500</td>
</tr>
<tr>
<td>DRAM read</td>
<td>5000</td>
</tr>
<tr>
<td>Interconnect send</td>
<td>20000</td>
</tr>
<tr>
<td>Network send</td>
<td>40000</td>
</tr>
<tr>
<td>Parallel storage write</td>
<td>100000</td>
</tr>
</tbody>
</table>

An Exascale system will be viable down to interconnect level with minimal data exchange between the compute nodes.

Input/Output will need to be revisited to avoid performance showstopper.

With current technologies Cloud HPC remains questionable in front of these numbers.
Where data will sit?

Static legacy parallel data storage
- LUSTRE and similar proven technologies extended with HSM for long time archiving
- Basis for providing applications and data to the compute nodes

Dynamic temporary data storage
- Ultra high speed dedicated parallel storage for single application, created on demand
- Likely NVRAM resident

Private temporary data storage
- Privately owned by a set of VMs
- Accessed through virtual network/interconnect
ExaScale will only work if software is ready

- Combining vectorization / multi-threading / multi-processing for a single application will increase its intrinsic complexity
- Existing MPI based application might need a rewrite of data distribution for better hybrid parallelism
- Scientists will need to work with scientific computing experts to exploit the potential of new architectures
What matters to end users?

- Democratization
- Team organisation
- Application-centric
- Work management
- Coherent interface
- Expert methods

Software Stack
Appliances

Industry ready solutions

• Validated architecture
• Preconfigured and tested systems
• “Intel Cluster Ready” compliance

On going work

• System oriented appliance unifying HPC and Data Analytic
• Application oriented appliances for solution deployment like Life Science
Fujitsu HPC Gateway – Broadening HPC access

Simplify HPC

- No more scripts – Job preparation in seconds rather than hours
- HPC on the Desktop – Intuitive collaborative workplace for newcomers through to practised users
- Productive at first login – More users can work with HPC even with little/no IT skill

Build in Expertise
Our Values

Innovation

Flexibility

Simplicity

We integrate the best available technologies to Fujitsu High Performance Computing Solution.

Anywhere, anytime, always available, our experts have the answers to your requirements.

We provide you with a comprehensive set of tools to make easier the use of a supercomputer.