NextGen. Computing and Storage at Scale

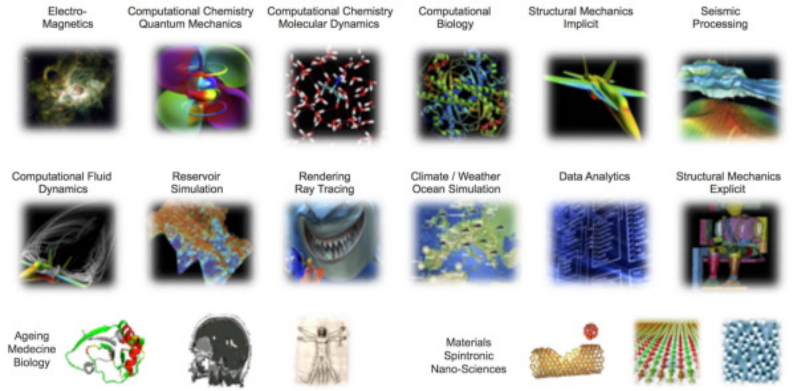
Overview and Implementation within the European HPC strategy

Dr. Sebastien Varrette

Workshop: "Accelerating Modelling and Simulation in the Data Deluge Era"

Fontainebleau, March 19th, 2018

Why HPC and BD ?



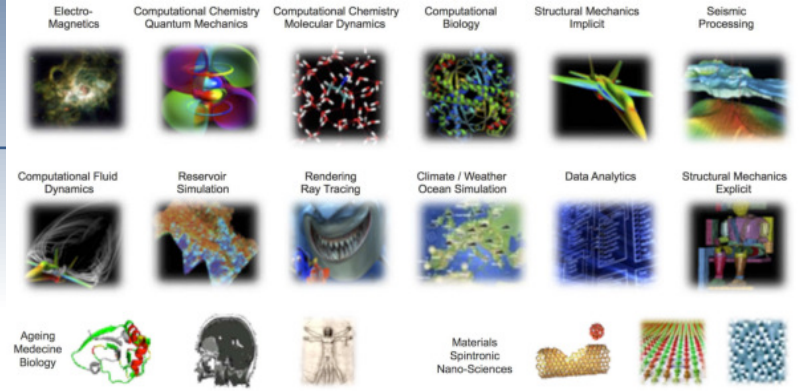
HPC: **H**igh **P**erformance **C**omputing
BD: **B**ig **D**ata

Andy Grant, Head of Big Data and HPC, Atos UK&I

To out-compete you must out-compute

Increasing competition, heightened customer expectations and shortening product development cycles are forcing the pace of acceleration across all industries.

Why HPC and BD ?



HPC: **H**igh **P**erformance **C**omputing
BD: **B**ig **D**ata

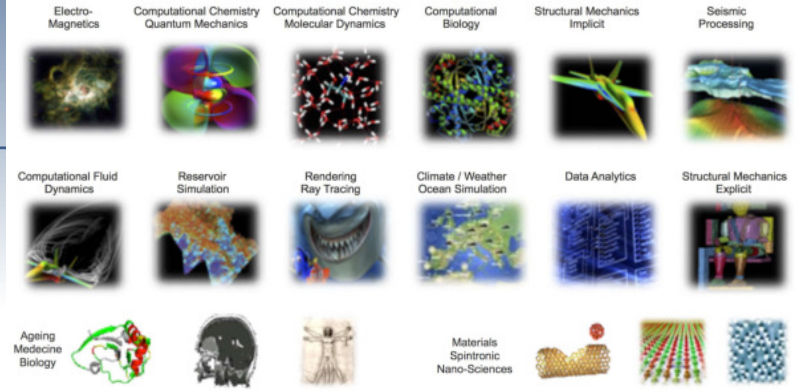
- Essential tools for **Science, Society and Industry**
  - ↪ All scientific disciplines are becoming computational today
    - ✓ requires very high computing power, handles **huge** volumes of data
- **Industry, SMEs** increasingly relying on HPC
  - ↪ to invent innovative solutions
  - ↪ . . . while reducing cost & decreasing time to market

Andy Grant, Head of Big Data and HPC, Atos UK&I

To out-compete you must out-compute

Increasing competition, heightened customer expectations and shortening product development cycles are forcing the pace of acceleration across all industries.

Why HPC and BD ?



HPC: **H**igh **P**erformance **C**omputing
BD: **B**ig **D**ata

- Essential tools for **Science, Society and Industry**
  - ↪ All scientific disciplines are becoming computational today
    - ✓ requires very high computing power, handles **huge** volumes of data
- **Industry, SMEs** increasingly relying on HPC
  - ↪ to invent innovative solutions
  - ↪ . . . while reducing cost & decreasing time to market

- HPC = **global race** (strategic priority) - EU takes up the challenge:
  - ↪ EuroHPC / IPCEI on HPC and Big Data (BD) Applications

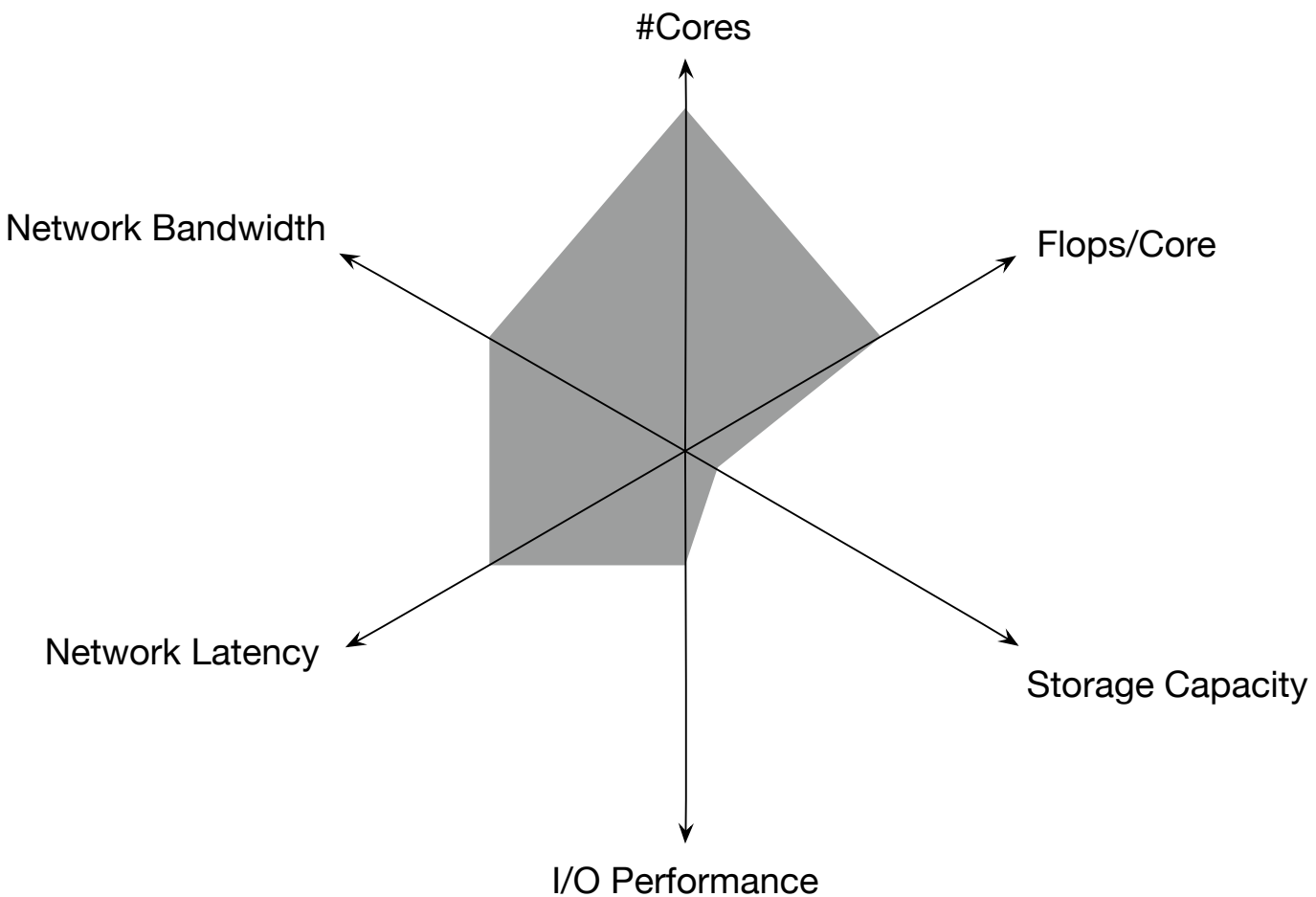
Andy Grant, Head of Big Data and HPC, Atos UK&I

To out-compete you must out-compute

Increasing competition, heightened customer expectations and shortening product development cycles are forcing the pace of acceleration across all industries.

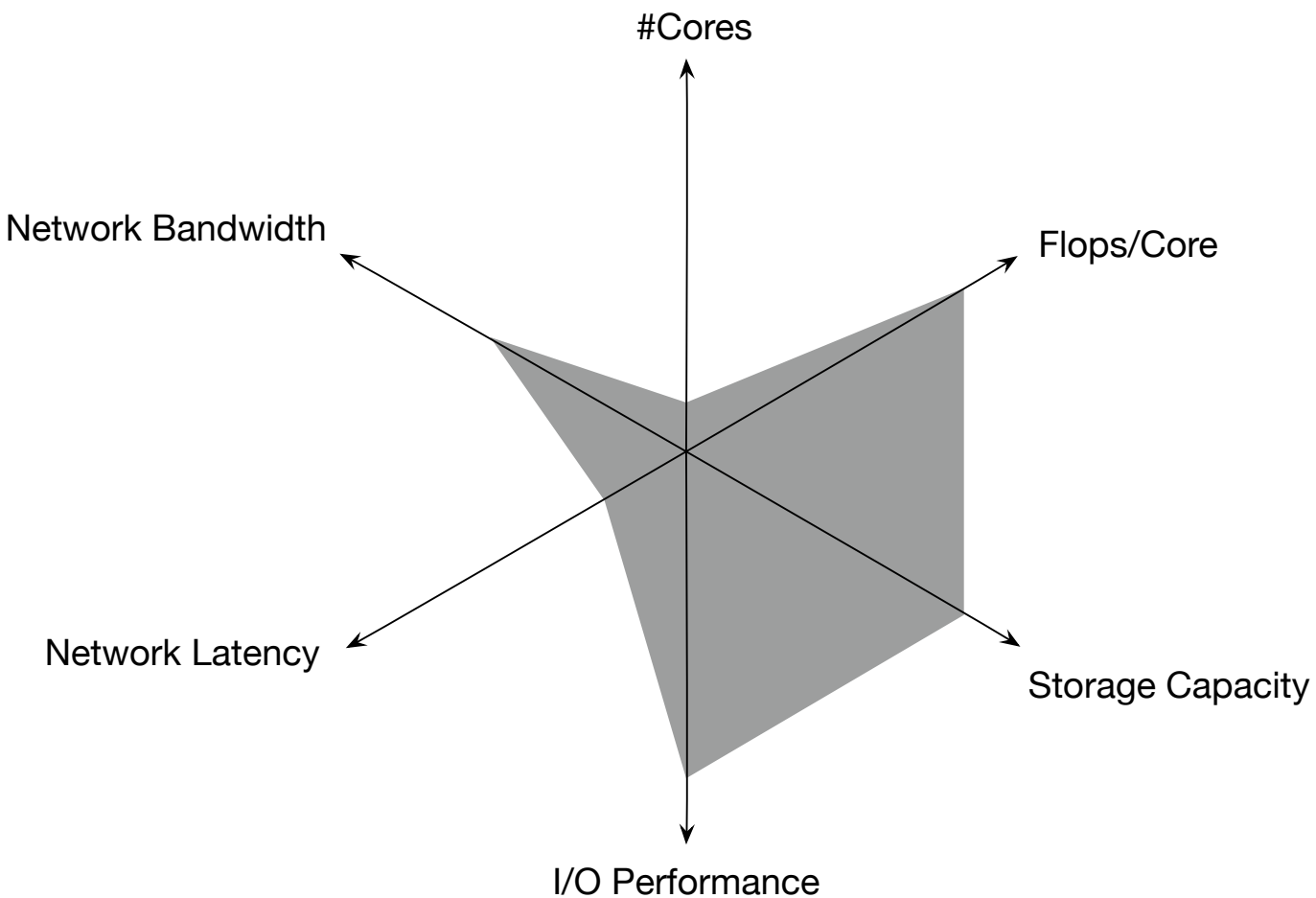
Different HPC Needs per Domains

Material Science & Engineering



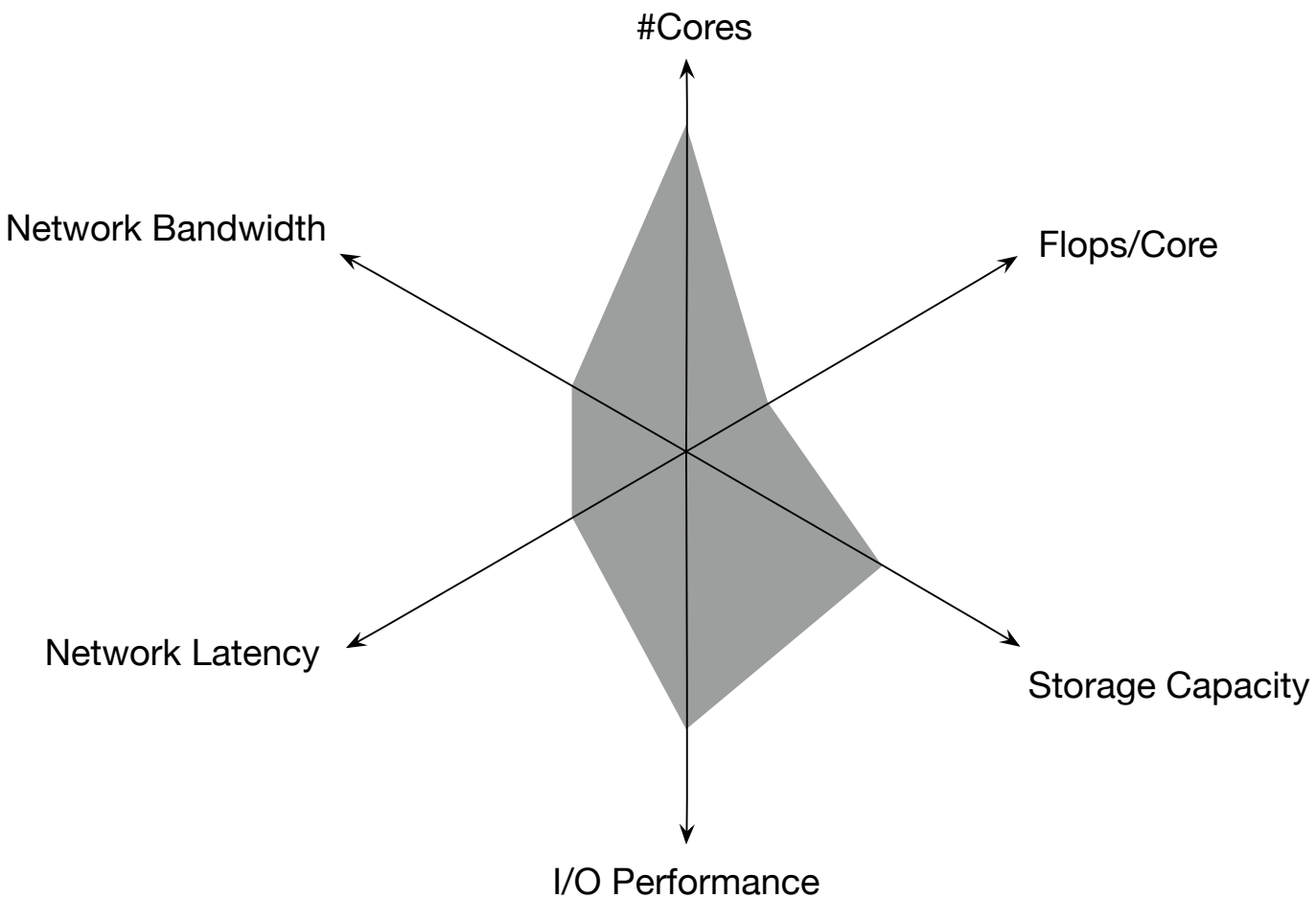
Different HPC Needs per Domains

Biomedical Industry / Life Sciences



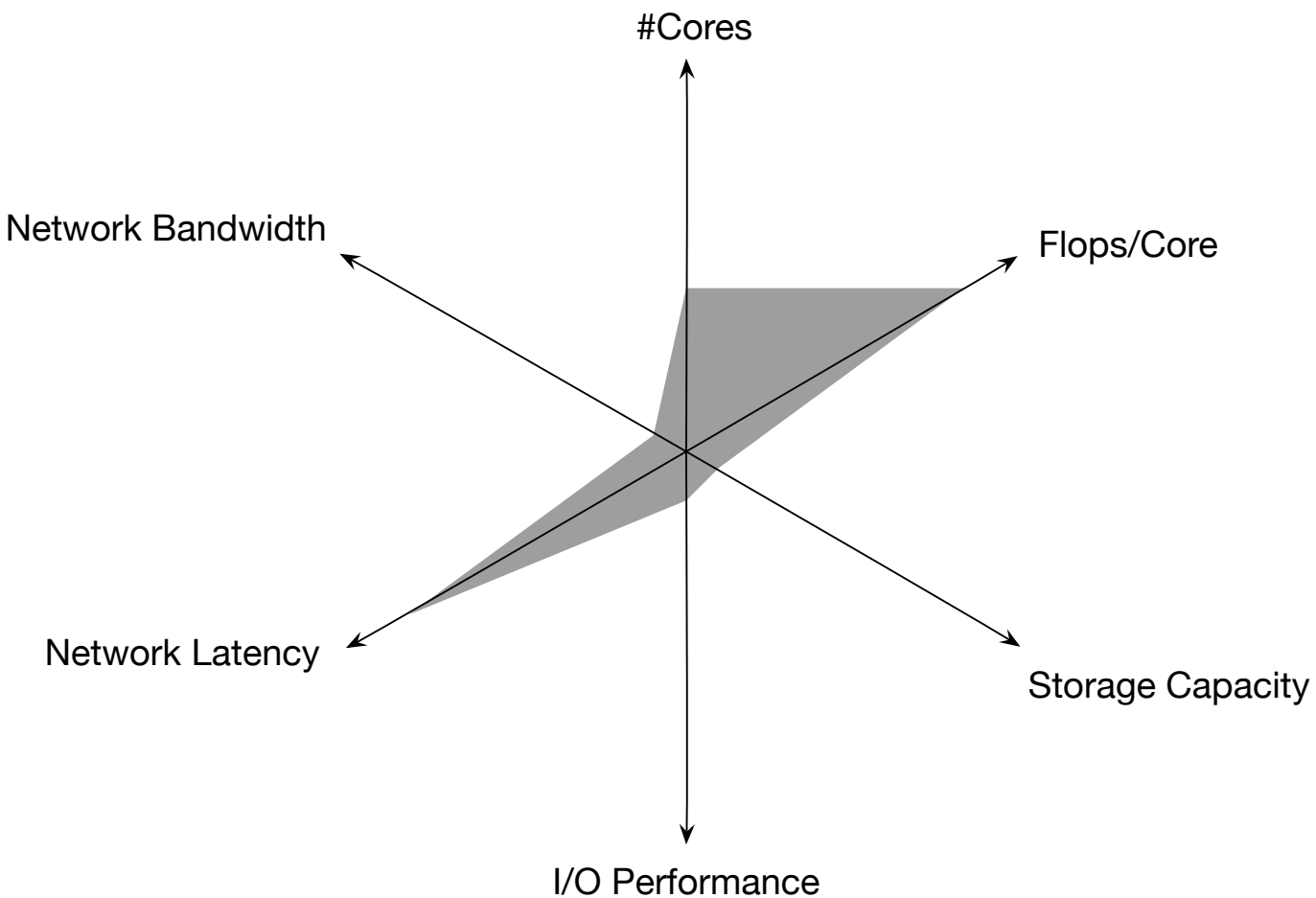
Different HPC Needs per Domains

Deep Learning / Cognitive Computing



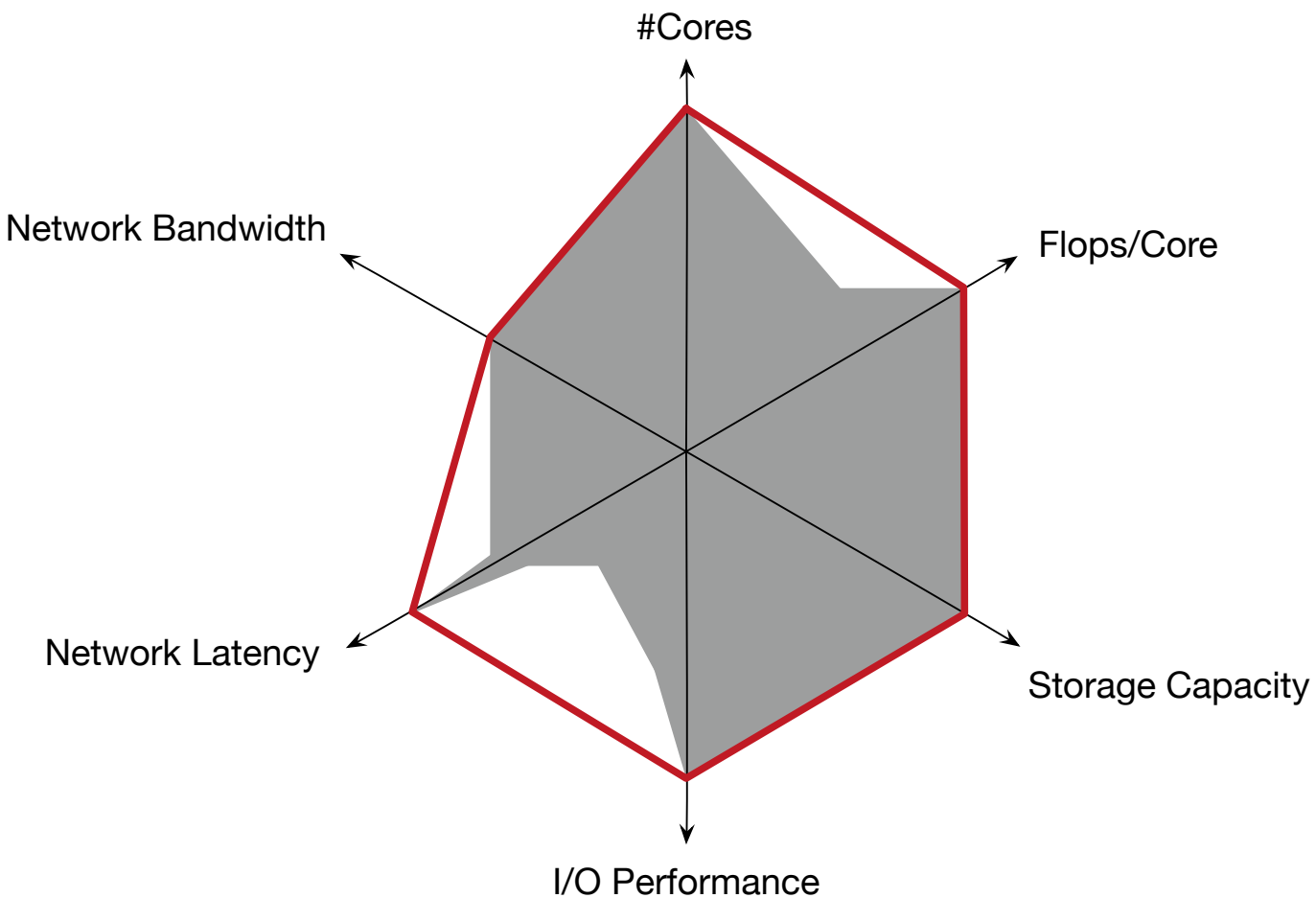
Different HPC Needs per Domains

IoT, FinTech



Different HPC Needs per Domains

ALL Research Computing Domains



Summary

Summary

HPC Computing Hardware

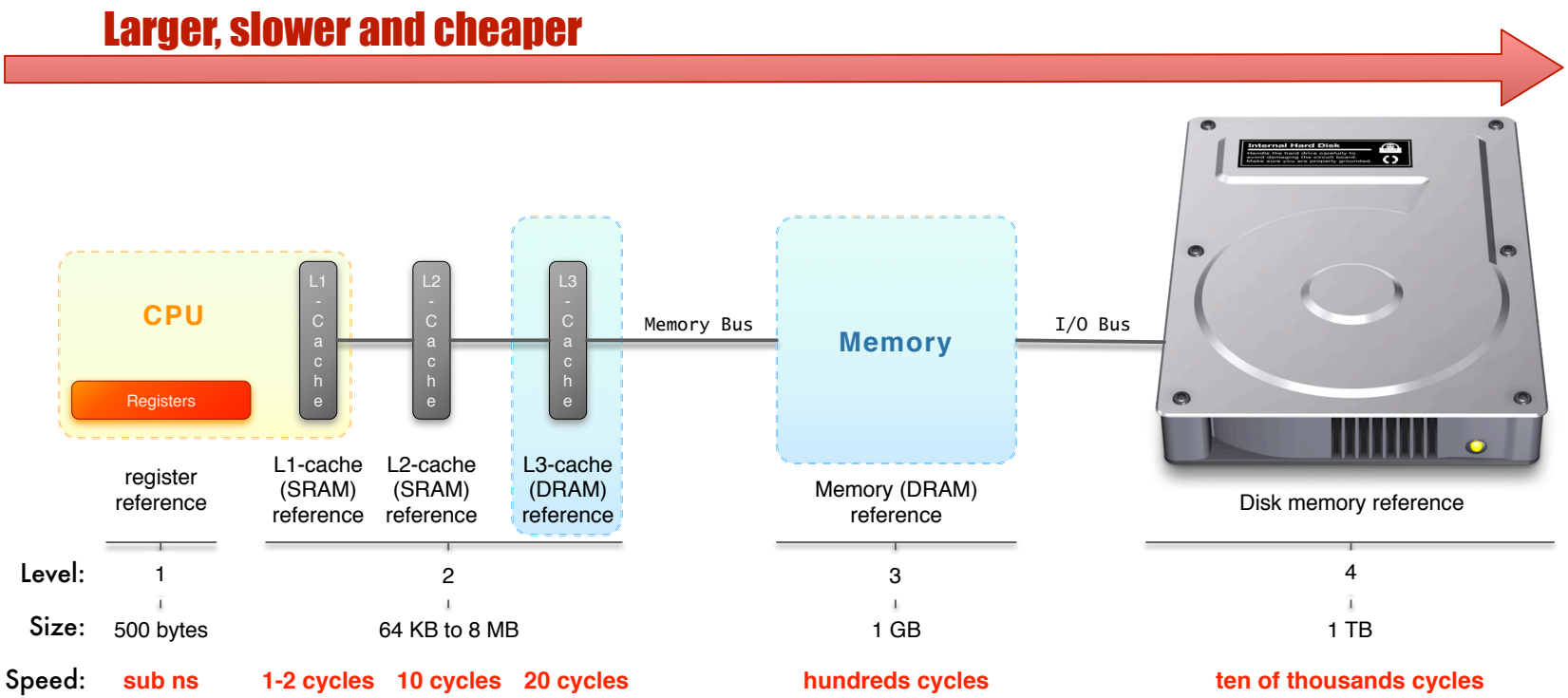
- **CPU** (Central Processing Unit)
  - ↪ highest software flexibility
  - ↪ high performance across all computational domains
  - ↪ Ex: Intel Core i7-7700K (Jan 2017) $R_{peak} \simeq 268.8$ GFlops (DP)
    - ✓ 4 cores @ 4.2GHz (14nm, 91W, 1.75 billion transistors) + integrated graphics

HPC Computing Hardware

- **CPU** (Central Processing Unit)
  - ↪ highest software flexibility
  - ↪ high performance across all computational domains
  - ↪ Ex: Intel Core i7-7700K (Jan 2017) $R_{peak} \simeq 268.8$ GFlops (DP)
    - ✓ 4 cores @ 4.2GHz (14nm, 91W, 1.75 billion transistors) + integrated graphics

- **Accelerators** (from *less* to *least* software flexibility)
  - ↪ **GPU** (Graphics Processing Unit) Accelerator
    - ✓ Ex: Nvidia Tesla V100 (Jun 2017) $R_{peak} \simeq 7$ TFlops (DP)
    - ✓ 5120 cores @ 1.3GHz (12nm, 250W, 21 billion transistors)
    - ✓ Ideal for Machine Learning workloads
  - ↪ **Intel MIC** (Many Integrated Core) Accelerator
  - ↪ **ASIC** (Application-Specific Integrated Circuits)
  - ↪ **FPGA** (Field Programmable Gate Array)

HPC Components: Local Memory

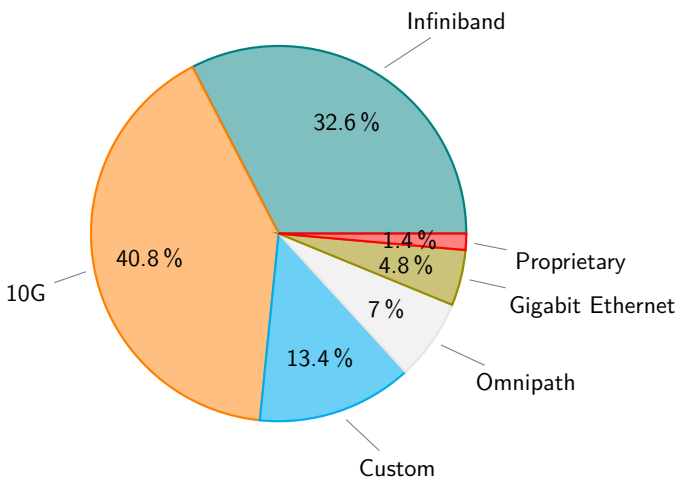


- SSD (SATA3) R/W: 550 MB/s; 100000 IOPS **450 €/TB**
- HDD (SATA3 @ 7,2 krpm) R/W: 227 MB/s; 85 IOPS **54 €/TB**

HPC Components: Interconnect

- **latency**: time to send a minimal (0 byte) message from A to B
- **bandwidth**: max amount of data communicated per unit of time

Technology	Effective Bandwidth		Latency
Gigabit Ethernet	1 Gb/s	125 MB/s	40μs to 300μs
10 Gigabit Ethernet	10 Gb/s	1.25 GB/s	4μs to 5μs
Infiniband QDR	40 Gb/s	5 GB/s	1.29μs to 2.6μs
Infiniband EDR	100 Gb/s	12.5 GB/s	0.61μs to 1.3μs
100 Gigabit Ethernet	100 Gb/s	1.25 GB/s	30μs
Intel Omnipath	100 Gb/s	12.5 GB/s	0.9μs

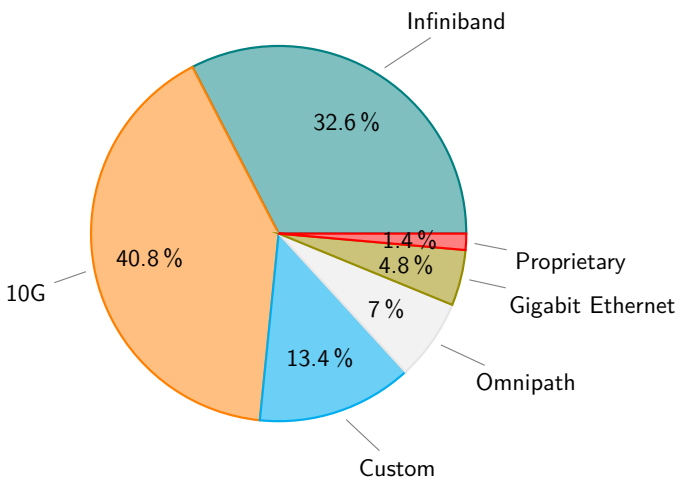


[Source : www.top500.org, Nov. 2017]

HPC Components: Interconnect

- **latency**: time to send a minimal (0 byte) message from A to B
- **bandwidth**: max amount of data communicated per unit of time

Technology	Effective Bandwidth		Latency
Gigabit Ethernet	1 Gb/s	125 MB/s	40μs to 300μs
10 Gigabit Ethernet	10 Gb/s	1.25 GB/s	4μs to 5μs
Infiniband QDR	40 Gb/s	5 GB/s	1.29μs to 2.6μs
Infiniband EDR	100 Gb/s	12.5 GB/s	0.61μs to 1.3μs
100 Gigabit Ethernet	100 Gb/s	1.25 GB/s	30μs
Intel Omnipath	100 Gb/s	12.5 GB/s	0.9μs



[Source : www.top500.org, Nov. 2017]

Network Topologies

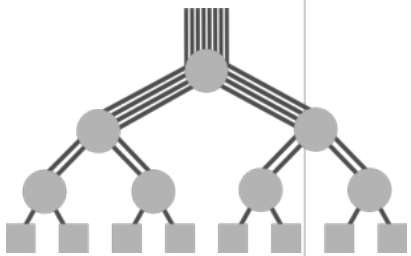
- **Direct** vs. **Indirect** interconnect
  - ↪ *direct*: each network node attaches to at least one compute node
  - ↪ *indirect*: compute nodes attached at the edge of the network only
    - ✓ many routers only connect to other routers.

Network Topologies

- **Direct** vs. **Indirect** interconnect
  - ↪ *direct*: each network node attaches to at least one compute node
  - ↪ *indirect*: compute nodes attached at the edge of the network only
    - ✓ many routers only connect to other routers.

Main HPC Topologies

- **CLOS Network / Fat-Trees** [Indirect]
  - ↪ can be fully non-blocking (1:1) or blocking (x:1)
  - ↪ typically enables **best performance**
    - ✓ Non blocking bandwidth, lowest network latency

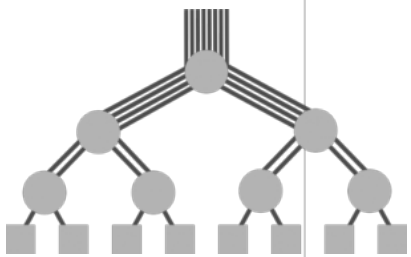


Network Topologies

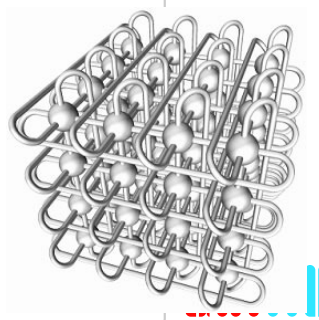
- **Direct** vs. **Indirect** interconnect
  - ↪ *direct*: each network node attaches to at least one compute node
  - ↪ *indirect*: compute nodes attached at the edge of the network only
    - ✓ many routers only connect to other routers.

Main HPC Topologies

- **CLOS Network / Fat-Trees** [Indirect]
  - ↪ can be fully non-blocking (1:1) or blocking (x:1)
  - ↪ typically enables **best performance**
    - ✓ Non blocking bandwidth, lowest network latency



- **Mesh or 3D-torus** [Direct]
  - ↪ Blocking network, cost-effective for systems at scale
  - ↪ Great performance solutions for applications with locality
  - ↪ Simple expansion for future growth



[Big]Data Management: Disk Encl.

- $\simeq$ **120 K€** - enclosure - 48-60 disks (4U)
  - $\hookrightarrow$ incl. redundant (**i.e.** 2) RAID controllers (master/slave)

[Big]Data Management: FS Summary

- **File System (FS)**: Logical manner to *store, organize & access* data
  - ↪ (local) **Disk FS** : `FAT32`, `NTFS`, `HFS+`, `ext4`, `{x,z,btr}fs...`
  - ↪ **Networked FS**: `NFS`, `CIFS/SMB`, `AFP`
  - ↪ **Parallel/Distributed FS**: `SpectrumScale/GPFS`, `Lustre`
    - ✓ typical FS for HPC / HTC (High Throughput Computing)

[Big]Data Management: FS Summary

- **File System (FS)**: Logical manner to *store, organize & access* data
  - ↪ (local) **Disk FS** : `FAT32`, `NTFS`, `HFS+`, `ext4`, `{x,z,btr}fs...`
  - ↪ **Networked FS**: `NFS`, `CIFS/SMB`, `AFP`
  - ↪ **Parallel/Distributed FS**: `SpectrumScale/GPFS`, `Lustre`
    - ✓ typical FS for HPC / HTC (High Throughput Computing)

Main Characteristic of Parallel/Distributed File Systems

Capacity and Performance increase with #servers

[Big]Data Management: FS Summary

- **File System (FS)**: Logical manner to *store, organize & access* data
  - ↪ (local) **Disk FS** : `FAT32`, `NTFS`, `HFS+`, `ext4`, `{x,z,btr}fs...`
  - ↪ **Networked FS**: `NFS`, `CIFS/SMB`, `AFP`
  - ↪ **Parallel/Distributed FS**: `SpectrumScale/GPFS`, `Lustre`
    - ✓ typical FS for HPC / HTC (High Throughput Computing)

Main Characteristic of Parallel/Distributed File Systems

Capacity and Performance increase with #servers

Name	Type	Read* [GB/s]	Write* [GB/s]
`ext4`	Disk FS	0.426	0.212
`nfs`	Networked FS	0.381	0.090
`gpfs (iris)`	Parallel/Distributed FS	**10.14**	**8,41**
`gpfs (gaia)`	Parallel/Distributed FS	7.74	6.524
`lustre`	Parallel/Distributed FS	4.5	2.956

* maximum **random** read/write, per IOZone or IOR measures, using 15 concurrent nodes for networked FS.

HPC Components: Data Center

Definition (Data Center)

- Facility to house computer systems and associated components
  - ↪ Basic storage component: **rack** (height: 42 RU)

HPC Components: Data Center

Definition (Data Center)

- Facility to house computer systems and associated components
  - ↪ Basic storage component: **rack** (height: 42 RU)

Challenges: Power (UPS, battery), Cooling, Fire protection, Security

- Power/Heat dissipation per rack:
  - ↪ HPC **computing** racks: **30-120 kW**
  - ↪ **Storage** racks: **15 kW**
  - ↪ **Interconnect** racks: **5 kW**
- Various **Cooling** Technology
  - ↪ Airflow
  - ↪ Direct-Liquid Cooling, Immersion...

Power Usage Effectiveness

$$PUE = \frac{Total\ facility\ power}{IT\ equipment\ power}$$

New Trends in HPC

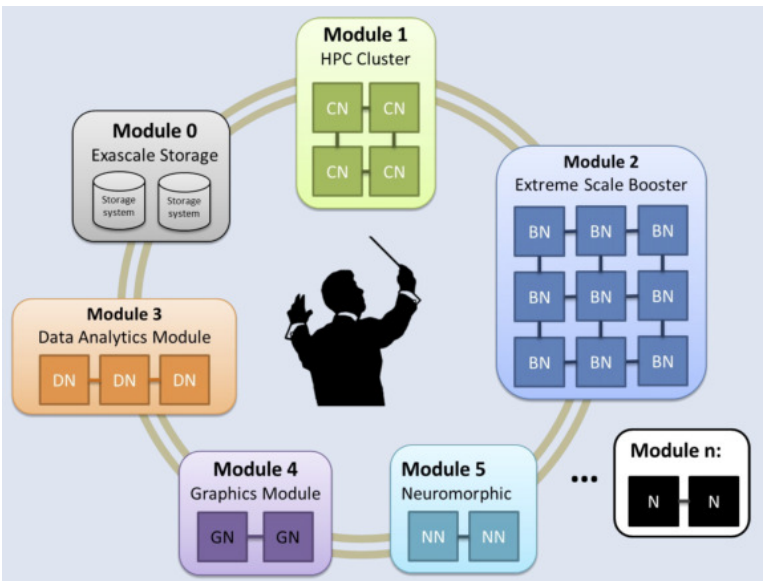
- **Continued scaling** of scientific, industrial & financial applications
  - ↪ . . . well beyond Exascale
- New trends changing the landscape for HPC
  - ↪ Emergence of **Big Data analytics**
  - ↪ Emergence of (**Hyperscale**) **Cloud Computing**
  - ↪ **Data intensive Internet of Things (IoT)** applications
  - ↪ **Deep learning & cognitive computing** paradigms

[Source : EuroLab-4-HPC]

[Source : IDC RIKEN report, 2016]

Toward Modular Computing

- Aiming at **scalable**, **flexible HPC infrastructures**
  - ↪ *Primary* processing on CPUs and accelerators
    - ✓ **HPC & Extreme Scale Booster** modules
  - ↪ *Specialized modules* for:
    - ✓ **HTC & I/O intensive** workloads;
    - ✓ **[Big] Data Analytics & AI**



[Source : "Towards Modular Supercomputing: The DEEP and DEEP-ER projects", 2016]

Summary

HPC International State of Affairs

Global race toward Exascale Technology

IDC-Projected Exascale Investment Levels (In Addition to System Purchases)

U.S.

- $1 to $2 billion a year in R&D (including NRE)
- Investments by both governments & vendors
- Plans are to purchase multiple exascale systems

EU

- About 5 billion euros in total
- Investments in multiple exascale and pre-exascale systems
- Investments mostly by country governments with a little from the EU

China

- Over $1billion a year in R&D
- Investments by both governments & vendors
- Plans are to purchases multiple exascale systems each year
- Already investing in 3 pre-exascale systems by 2017/18

Japan

- Planned investment of just over $1billion* (over 5 years) for both the R&D and purchase of 1 exascale system
- To be followed by a number of smaller systems ~$100M to $150M each
- Creating a new processor and a new software environment

©Hyperion Research 2017

** Note that this includes both the system and R&D*

HPC International State of Affairs

Global race toward Exascale Technology

IDC-Projected Exascale Dates and Suppliers

U.S.
- Sustained ES: 2023
- Peak ES: 2021
- Vendors: U.S.
- Processors: U.S.
- Initiatives: NSCI/ECP
- Cost: $300-500M per system, plus heavy R&D investments

EU
- Sustained ES: 2023-24
- Peak ES: 2021
- Vendors: U.S., Europe
- Processors: U.S., ARM
- Initiatives: PRACE, ETP4HPC
- Cost: $300-$350 per system, plus heavy R&D investments

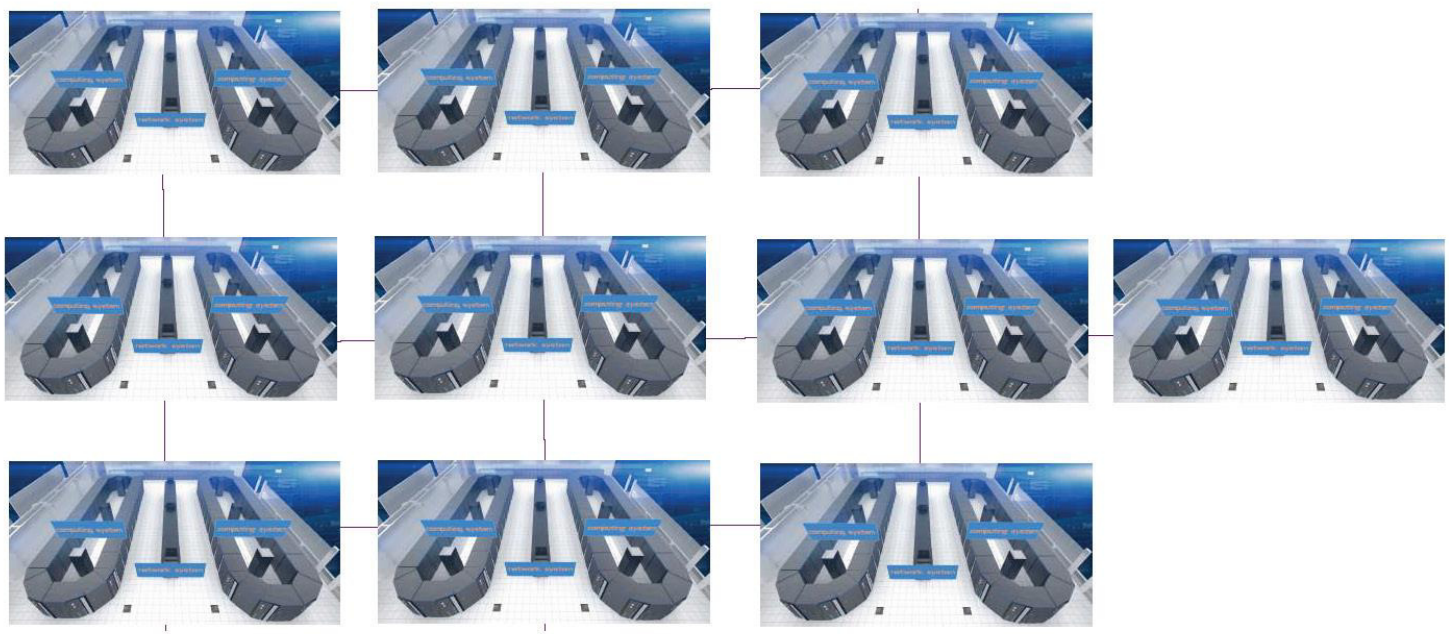
China
- Sustained ES: 2023
- Peak ES: ~~2020~~ 2019…
- Vendors: Chinese
- Processors: Chinese (plus ~~U.S.~~?)
- 13th 5-Year Plan
- Cost: $350-500M per system, plus heavy R&D

Japan
- Sustained ES: 2023-24
- Peak ES: Not planned
- Vendors: Japanese
- Processors: Japanese
- Cost: $600-850M, this includes both 1 system and the R&D costs…will also do many smaller size systems

Exascale Feasibility

We Can Build an Exascale System Today?

Connect together 10 Sunway TaihuLight systems



Require 150 MW of power, programming for 100 M threads, and $2.7B price tag

22

European HPC strategy

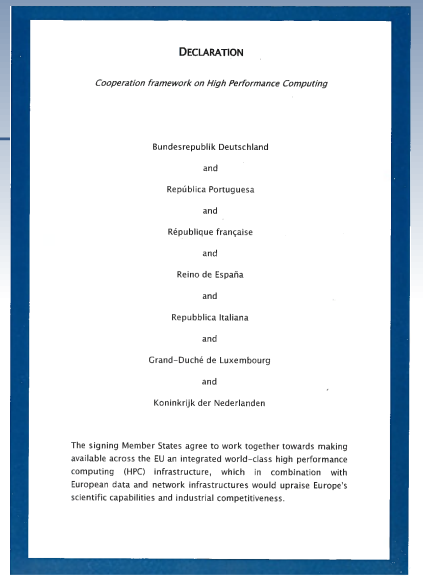
- EU HPC strategy initiated in 2012
  - ↪ implementation within H2020 program

European HPC strategy

- EU HPC strategy initiated in 2012
  - ↪ implementation within H2020 program

- More recently:
  - ↪ IPCEI on HPC and Big Data (BD) Applications (Nov. 2015)
    - ✓ Luxembourg (leader), France, Italy & Spain
    - ✓ Testbed around Personalized Medicine, Smart Space, Industry 4.0, Smart Manufacturing, New Materials, FinTech, Smart City...

European HPC strategy


DECLARATION

Cooperation framework on High Performance Computing

Bundesrepublik Deutschland
and
República Portuguesa
and
République française
and
Reino de España
and
Repubblica Italiana
and
Grand-Duché de Luxembourg
and
Koninkrijk der Nederlanden

The signing Member States agree to work together towards making available across the EU an integrated world-class high performance computing (HPC) infrastructure, which in combination with European data and network infrastructures would upraise Europe's scientific capabilities and industrial competitiveness.

- EU HPC strategy initiated in 2012
  - ↪ implementation within H2020 program

- More recently:
  - ↪ IPCEI on HPC and Big Data (BD) Applications
    - ✓ Luxembourg (leader), France, Italy & Spain
    - ✓ Testbed around Personalized Medicine, Smart Space, Industry 4.0, Smart Manufacturing, New Materials, FinTech, Smart City...

- **Latest advances**:
  - ↪ EU Member States sign EuroHPC (Mar. 2017)
    - ✓ common effort to create/grow European supercomputing ecosystem
    - ✓ Federation of national/regional HPC centers (see also PRACE2)
  - ↪ EU Objective with EuroHPC:
    - ✓ 2-3 **Pre-exascale** systems by 2019, **2 exascale** systems by 2021

EU HPC Strategy Implementation

- **European Technology Platform (ETP) for HPC**
  - ↪ Industry-led forum feat. HPC stakeholders
  - ↪ Providing EU framework to define HPC research priorities/actions
    - ✓ UL (P. Bouvry, S. Varrette, V.Plugaru) part of ETP4HPC (2016-)
    - ✓ See Strategic Research Agenda, 2017 European HPC Handbook...

EU HPC Strategy Implementation

- **European Technology Platform (ETP) for HPC**
  - ↪ Industry-led forum feat. HPC stakeholders
  - ↪ Providing EU framework to define HPC research priorities/actions
    - ✓ UL (P. Bouvry, S. Varrette, V.Plugaru) part of ETP4HPC (2016-)
    - ✓ See Strategic Research Agenda, 2017 European HPC Handbook...

- EU **COST Actions**, for instance:
  - ↪ NESUS: Network for Sustainable Ultrascale Computing
  - ↪ cHiPSet: High-Performance Modelling and Simulation for BDA

EU HPC Strategy Implementation

- **European Technology Platform (ETP) for HPC**
  - ↪ Industry-led forum feat. HPC stakeholders
  - ↪ Providing EU framework to define HPC research priorities/actions
    - ✓ UL (P. Bouvry, S. Varrette, V.Plugaru) part of ETP4HPC (2016-)
    - ✓ See Strategic Research Agenda, 2017 European HPC Handbook...

- EU **COST Actions**, for instance:
  - ↪ NESUS: Network for Sustainable Ultrascale Computing
  - ↪ cHiPSet: High-Performance Modelling and Simulation for BDA

- **PRACE** - Partnership for Advanced Computing in Europe
  - ↪ Non-profit association, 25 member countries, now entering PRACE2
  - ↪ Providing access to **Five EU Tier-0** compute & data resources
  - ↪ Luxembourg 25th country to join (Oct. 17th, 2017)
    - ✓ Official Delegate/Advisor (P. Bouvry/S. Varrette) from UL

EU HPC Strategy Implementation

- **European High-Performance Computing Joint Undertaking**
  - ↪ EuroHPC JU effectively operational starting **Jan 1st, 2019**
    - ✓ administrative management from Luxembourg
  - ↪ Public and private members
    - ✓ EC, 14 MS, representatives from supercomputing/BD stakeholders
    - ✓ Governing Board (public members)
    - ✓ Industrial & Scientific Advisory Board (private members)
  - ↪ EU Objective with EuroHPC:
    - ✓ 2-3 **Pre-exascale** systems by 2020, **2 exascale** systems by 2022
    - ✓ Pending decision on hosting countries

EU HPC Strategy Implementation

- **European High-Performance Computing Joint Undertaking**
  - ↪ EuroHPC JU effectively operational starting **Jan 1st, 2019**
    - ✓ administrative management from Luxembourg
  - ↪ Public and private members
    - ✓ EC, 14 MS, representatives from supercomputing/BD stakeholders
    - ✓ Governing Board (public members)
    - ✓ Industrial & Scientific Advisory Board (private members)
  - ↪ EU Objective with EuroHPC:
    - ✓ 2-3 **Pre-exascale** systems by 2020, **2 exascale** systems by 2022
    - ✓ Pending decision on hosting countries

EuroHPC Budget: 2×486 M€

EU HPC Strategy Implementation

- **European High-Performance Computing Joint Undertaking**
  - ↪ EuroHPC JU effectively operational starting **Jan 1st, 2019**
    - ✓ administrative management from Luxembourg
  - ↪ Public and private members
    - ✓ EC, 14 MS, representatives from supercomputing/BD stakeholders
    - ✓ Governing Board (public members)
    - ✓ Industrial & Scientific Advisory Board (private members)
  - ↪ EU Objective with EuroHPC:
    - ✓ 2-3 **Pre-exascale** systems by 2020, **2 exascale** systems by 2022
    - ✓ Pending decision on hosting countries

EuroHPC Budget: 2×486 M€

- **European Processor Initiative (EPI)**
  - ↪ Initial plan vs current plan...
  - ↪ **120 M€** via Framework Partnership Agreement (FPA)

Summary

Conclusion

- New trends changing the landscape for HPC, with convergence of
  - ↪ **Big Data analytics** and (**Hyperscale**) **Cloud Computing**
  - ↪ Data intensive **Internet of Things (IoT)**
  - ↪ **Deep learning & cognitive computing** paradigms

Conclusion

- New trends changing the landscape for HPC, with convergence of
  - ↪ **Big Data analytics** and (**Hyperscale**) **Cloud Computing**
  - ↪ Data intensive **Internet of Things (IoT)**
  - ↪ **Deep learning & cognitive computing** paradigms

- All **new deployments are** (normally) **planned modular**
  - ↪ the **right module for the right application**
  - ↪ assumes data-locality aware schedulers and execution
  - ↪ **improved network backbone** between sites: **GEANT upgrade**

Conclusion

- New trends changing the landscape for HPC, with convergence of
  - ↪ **Big Data analytics** and (**Hyperscale**) **Cloud Computing**
  - ↪ Data intensive **Internet of Things (IoT)**
  - ↪ **Deep learning & cognitive computing** paradigms

- All **new deployments are** (normally) **planned modular**
  - ↪ the **right module for the right application**
  - ↪ assumes data-locality aware schedulers and execution
  - ↪ **improved network backbone** between sites: **GEANT upgrade**

Several On-going Strategic HPC efforts in Europe...

- ETP4HPC, EU COST Actions etc.
- PRACE, now entering PRACE2
- IPCEI on HPC and Big Data (BD) Applications
- EuroHPC for concrete Exascale deployment.

Questions?

http://hpc.uni.lu

Dr. Sebastien Varrette
University of Luxembourg, Belval Campus:
Maison du Nombre, 4th floor
2, avenue de l'Université
L-4365 Esch-sur-Alzette
mail: sebastien.varrette@uni.lu