Evolución dell'Infrastruttura di Calcolo e Data Analytics per la ricerca

FERMI

High-end system, only for extremely scalable applications

Name: Fermi
Architecture: BlueGene/Q (10 racks)
Processor type: IBM PowerA2 @1.6 GHz
Computing Nodes: 10.240
Each node: 16 cores and 16GB of RAM
Computing Cores: 163.840
RAM: 1GByte / core (163 TByte total)
Internal Network: 5D Torus
Disk Space: 2PByte of scratch space
Peak Performance: 2PFlop/s
Power Consumption: 820 kWatts

N. 7 in Top 500 rank (June 2012)

National and PRACE Tier-0 calls

GALILEO

X86 based system for production of medium scalability applications

Name: Galileo
Model: IBM NeXtScale
Architecture: IBM NeXtScale
Processor type: Intel Xeon Haswell@ 2.4 GHz
Computing Nodes: 516
Each node: 16 cores, 128 GB of RAM
Computing Cores: 8.256
RAM: 66 TByte
Internal Network: Infiniband 4xQDR switches (40 Gb/s)
Accelerators: 768 Intel Phi 7120p (2 per node on 384 nodes
+ 80 Nvidia K80
Peak Performance: 1.2 PFlops

National and PRACE Tier-1 calls

PICO

Storage and processing of large volumes of data.
Data Analytics.

Name: Pico
Model: IBM NeXtScale
Architecture: Linux Infiniband cluster
Processor type: Intel Xeon E5 2670 v2 @2,5Ghz
Computing Nodes: 66+
Each node: 20 cores, 128 GB of RAM + 2 accelerators
Computing Cores: 1.320+
RAM: 6,4 GB/core

+2 Visualization nodes
+2 Big Mem nodes
+4 BigInsight nodes

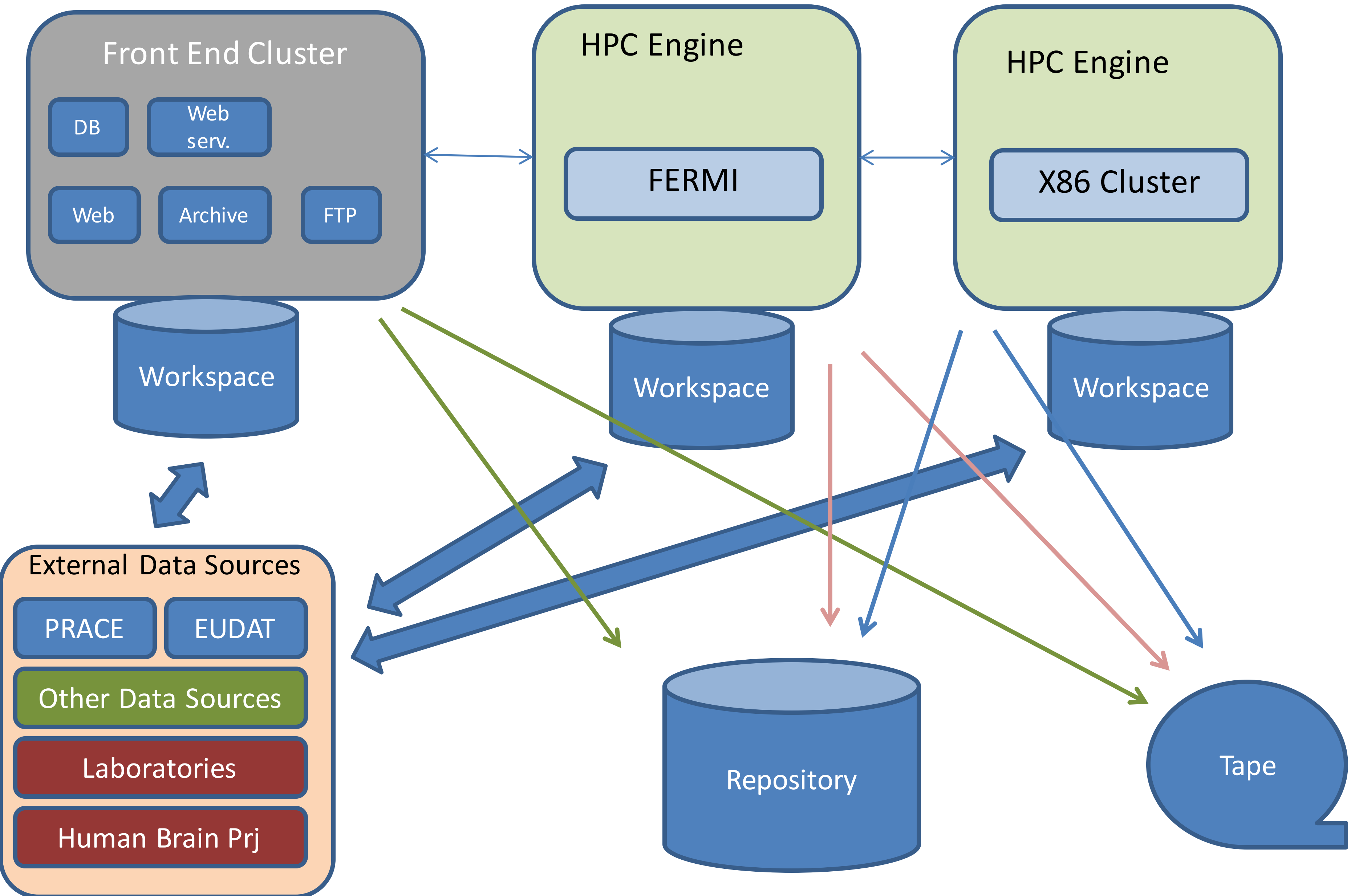
Integrated with a multi tier storage system:

40 TByte of SSDs
5 PByte of Disks
16 PByte of Tapes

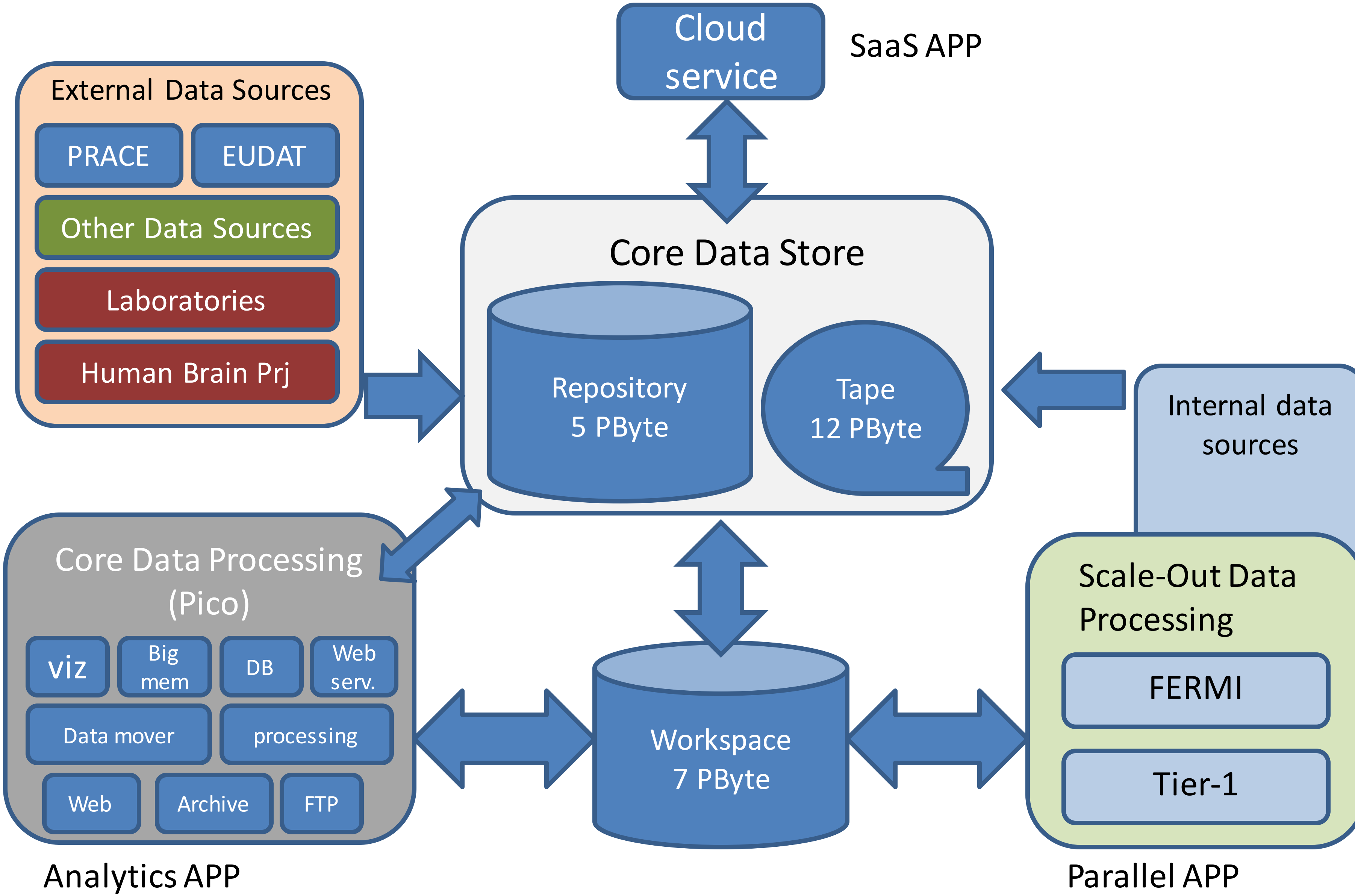
- New services to be defined on this system, taking advance from its peculiarities:
  - Low parallelism (less cores with respect to other systems, more cores/node) -> high throughput
  - Memory intensive (more memory/core and /node)
  - I/O intensive (SSD disk available)
  - DB based (a lot of storage space)
- New application environments:
  - Bioinformatics
  - Data anaysis
  - Engineerings
  - Quantum Chemistry
- General services
  - Remote visualisation
  - Web access to HPC
  - Cloud

Infrastructure Evolution

HPC "island" Infrastructure



(data centric) Infrastructure



Computing Infrastructure

Tier0: Fermi
Tier1: Galileo
BigData: Pico

Tier0: new
(HPC Top10)
BigData: Galileo/Pico

Tier0 BigData:
50PFlops
50PByte

today

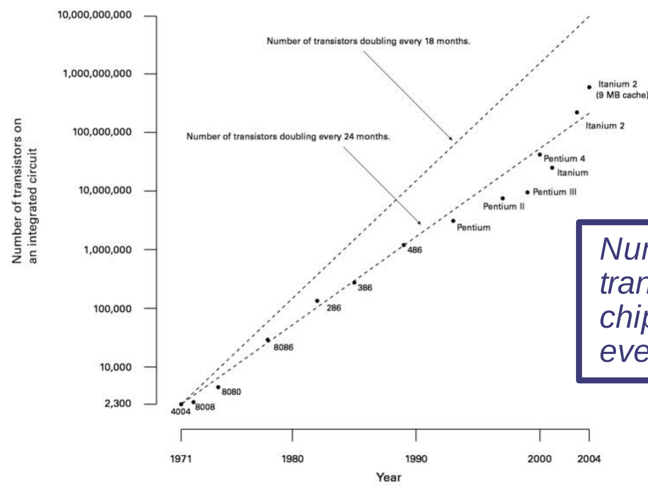
1Q 2016

2018

Challenges & Opportunity

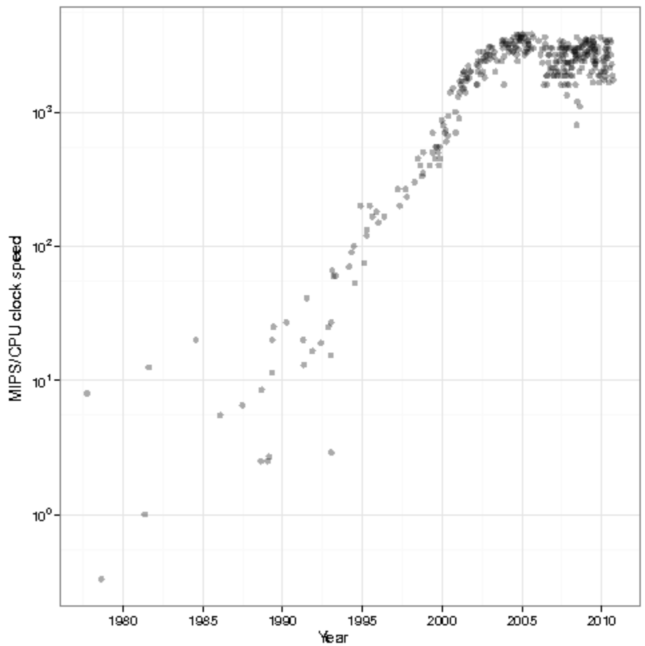
HPC Trends

Moore's Law



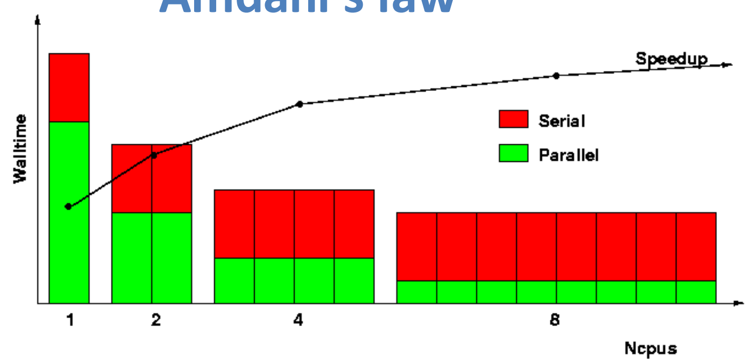
Number of transistors per chip double every 24 month

Dennard scaling law (downscaling)



The core frequency and performance do not grow following the Moore's law any longer	*Increase the number of cores to maintain the architectures evolution on the Moore's law*

Amdahl's law



maximum speedup tends to $1 / (1 - P)$
P= parallel fraction

The upper limit for the scalability of parallel applications is determined by the fraction of the overall execution time spent in non-parallel operations.

Peak Performance	Exaflops 10^18	Moore law
FPU Performance	Gigaflops 10^9	Dennard law
Number of FPUs	10^9	Moore + Dennard
App. Parallelism	serial fraction 1/10^9	Amdahl's law
System TCO	Power >10^6 Watt	Moore + Dennard

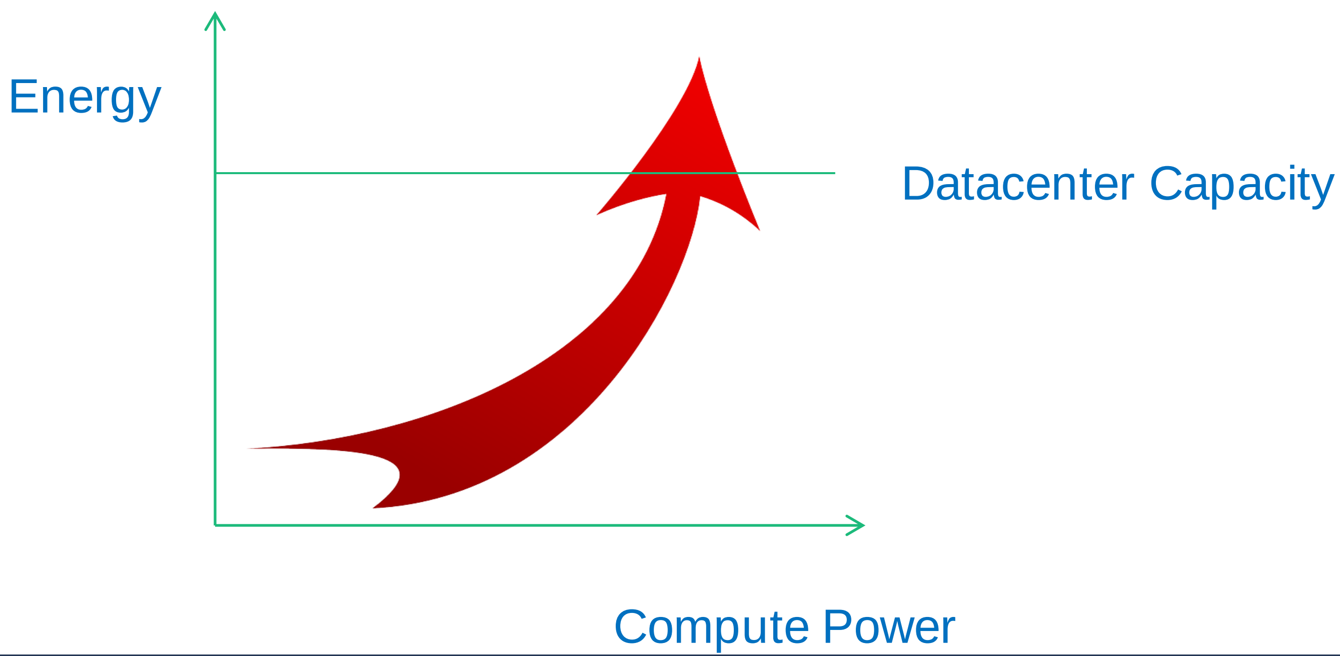
opportunity

challenge

Energy trends

“traditional” RISK and CISC chips are designed for maximum performance for all possible workloads

A lot of silicon to maximize single thread performace

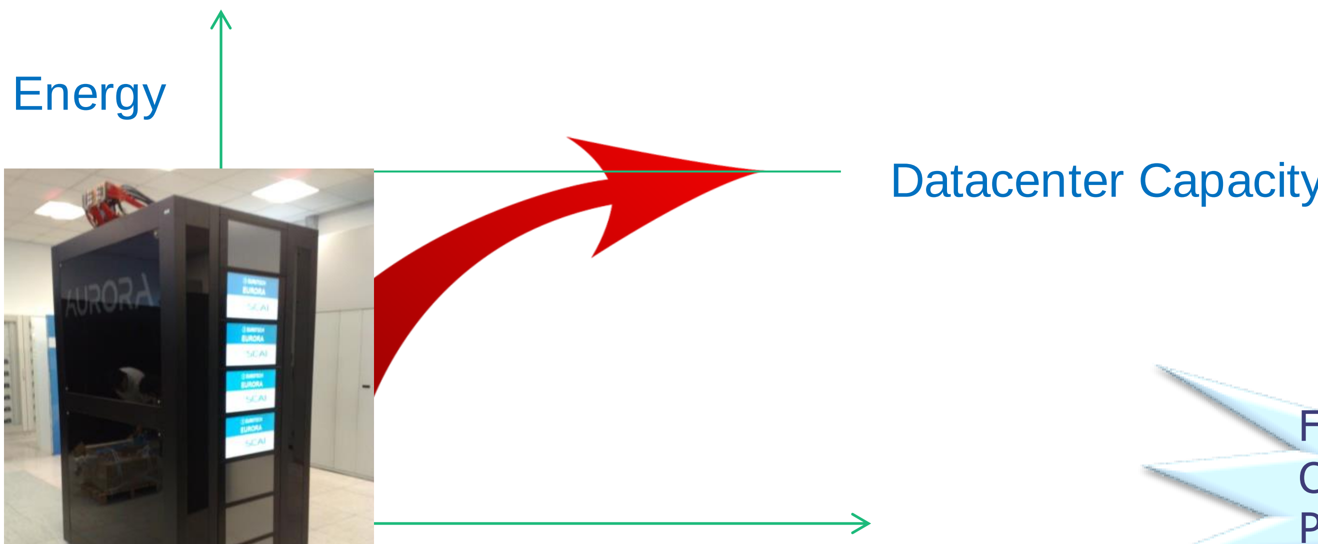


Change of paradigm

New chips designed for maximum performance in a small set of workloads

→

Simple functional units, poor single thread performance, but maximum throughput

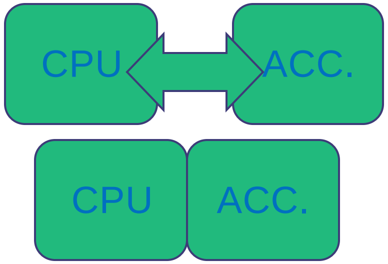


EURORA
#1 in The Green500 List June 2013

FETHPC H2020 Project,
Coordinated by
Politecnico di Milano
Prof. Cristina Silvano

Converging Exascale architecture model

Hybrid



Nvidia GPU

AMD APU

Homogeneus

ARM

Intel

System attributes	2001	2010	"2015"		"2018"	
System peak	10 Tera	2 Peta	200 Petaflop/sec		1 Exaflop/sec	
Power	~0.8 MW	6 MW	15 MW		20 MW	
System memory	0.006 PB	0.3 PB	5 PB		32-64 PB	
Node performance	0.024 TF	0.125 TF	0.5 TF	7 TF	1 TF	10 TF
Node memory BW		25 GB/s	0.1 TB/sec	1 TB/sec	0.4 TB/sec	4 TB/sec
Node concurrency	16	12	O(100)	O(1,000)	O(1,000)	O(10,000)
System size (nodes)	416	18,700	50,000	5,000	1,000,000	100,000
Total Node Interconnect BW		1.5 GB/s	150 GB/sec	1 TB/sec	250 GB/sec	2 TB/sec
MTTI		day	O(1 day)		O(1 day)	

Applications Challenges

-Programming model

-Scalability

-I/O, Resiliency/Fault tolerance

-Numerical stability

-Algorithms

-Energy Awareness/Efficiency