Technology-Scalable Datacenters

Report from the Workshop on Next-Generation Datacenters

held in June 2011 in Brussels, Belgium

October 2011

Unit G3: Embedded Systems and Control
FP7 ICT Research Objective: Computing Systems
http://cordis.europa.eu/fp7/ict/computing

SEVENTH FRAMEWORK PROGRAMME

European Commission
Information Society and Media

Disclaimer: The views expressed in this report are those of the workshop participants and do not necessarily represent the official view of the European Commission on the subject.

Executive Summary

Information technology is now ever more than before an indispensable pillar of a modern day's society. Information is now at the core of the "supply-chain" for products and services in the modern world. Unfortunately, there are two phenomena that, unless mitigated, will dramatically slow down growth in information technology. First, the demand on information processing, storage and communication is growing faster than technology sustains. Second, the servers and datacenters forming the backbone of information technology are hitting fundamental technological barriers with energy, economic and environmental implications.

The report analyses the main research challenges in datacenter design:

- **Achieving mobile platforms' efficiency in server nodes**: digital platforms require drastic improvements in cost and energy efficiency - reminiscent of mobile platforms - far beyond "low-hanging fruits" of current datacenter optimisations. The target should be to decrease by at least two orders of magnitude the energy consumption at the server node over the next decade.
- **Taming the data deluge**: information technology is going through an inflection point from being "computation-" to "data-centric". Making sense of massive data requires a complete redesign of data management technologies and innovation in machine learning and probabilistic computing to reduce the size and dimensionality of data for services and applications.
- **Holistic integration**: future datacenter designs require a hand-in-hand collaboration of software and hardware to provide both energy (through specialization) and thermal management especially for future 3D-stacked server chips.
- **Federated datacenters**: despite the trend towards single or few giant datacenters, security, reliability/availability, affinity to data, proximity to energy sources, legal boundaries and the client base dictate the need for technologies to orchestrate operation and load balancing over a collection of physically distributed sites.

With a history of innovation and major breakthroughs in embedded and mobile platforms, energy-efficient technologies and a corresponding strong industrial backbone, Europe has the opportunity to seize the moment and lead in research and innovation to bring 100x improvements in datacentre efficiency and total cost of ownership by 2020 needed to sustain the demand on information technology growth.

The report, based on a workshop organised by the HiPEAC NoE, puts forward a number of concrete recommendations for European research priorities for datacenters both on the medium term (energy-proportional datacenters; tighter integration of servers having mobile efficiency with network fabric and infrastructure; and federated seamless services across distributed and potentially distant sites) and on the longer term (towards energy-neutral datacenters with specialised servers for datacenter workloads; probabilistic servers; computing with renewable energy; and global-scale federation of datacenters).

Table of Contents

Introduction

Today, we live in a digital world. Our daily needs are unimaginable without access to information. Communication, entertainment, social networking, and financial, transportation and health services are just a few examples of how our day-to-day interactions have transformed into data exchanges in the form of a stream of bits. Information technology, now ever more than before, is a necessity rather than a luxury to our existence and proper functioning as a citizen of the modern world.

At the center of this revolution is data. As individuals we need ubiquitous access, exchange and sharing of data with those we interact with. Similarly, businesses, governments and societies rely on collecting, analyzing and exchanging data to improve their products, services and ultimately enhance our lives. Data now also lies at the core of the supply-chain for both products and services in modern economies. In today's "knowledge era" of economics, there are a number of studies that indicate that 50% of the value in the developed world's economy is based on data [1].

Servers and datacenters form the backbone of this data revolution. Large organizations (e.g., enterprises, governmental agencies, research and academic institutions) have historically used machine rooms to host and operate a collection of computing, storage and networking servers. The machine rooms were designed to leverage the cost of building, hosting and operating the servers.

The exponential growth in IT made possible by advancements in semiconductor fabrication for the past several decades, however, has been met or surpassed by the growth in demand for computing. The net result is that computing platforms hosted in machine rooms have become denser (e.g., have a higher computational and storage capability per occupied unit volume), and machine rooms have been growing in size, increasingly hosting large collections of servers referred to as clusters or server farms. Because most of the activity within an IT department is centered around processing and storage of data belonging to an organization, machines rooms have evolved into entities that are increasingly referred to as datacenters.

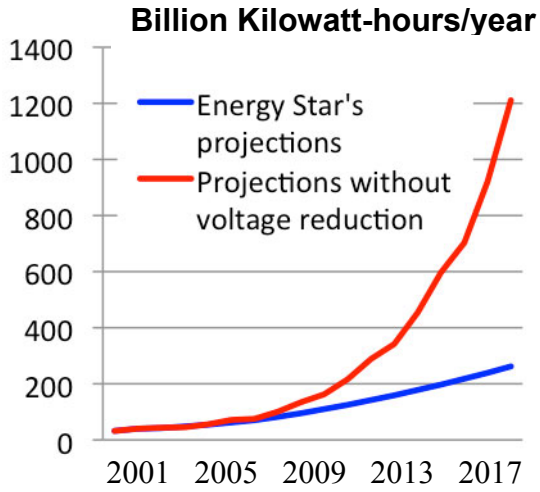
In the future, the main scalability impediment to computing in general and datacenters specifically will be energy, with both economic and environmental implications. Since the industrial revolution energy has been (and is projected to remain for the next several decades) the number one cause of humanity's concern (Lectures by Richard Smiley [14]). Apart from a continued growth in demands for scaling IT platforms, energy prices also grow due to higher demands and higher costs of extracting conventional forms of

energy (e.g., fossil fuel) because of both scarcity and environmental concerns, increasing the cost of IT.

More importantly, the semiconductor fabrication technologies and circuits that have fuelled the IT revolution for the past four decades have hit fundamental efficiency limits. While projections indicate a continued scaling of CMOS-based platforms in density and cost for another decade, CMOS has hit an energy-scalability wall resulting in diminishing returns in transistor energy efficiency with further increases in density. The net result is tokened "the economic meltdown of Moore's law" for servers and datacenters by the Kenneth Brill of the Uptime Institute [2].

Datacenters' electricity bill is skyrocketing. Market estimates [McKinsey] report that in 2008 the world's 44 million servers consumed 0.5 percent of all electricity and produced 0.2 percent of all carbon dioxide emissions, or 80 megatons a year, approaching the emissions of entire countries like Argentina or the Netherlands. A recent high-end ("Internet-scale") datacenter is reported to draw 48 MW, equivalent to 40000 homes. Typical high-end centers draw 20-30 MW (equivalent to 15000 - 20000 homes). With current technology, investments for building and equipment of high-end datacenters are now over a few billion dollars. Even worse, estimates indicate that in the long term if current trends continue, the companies operating datacenters will spend more money per year on energy than on equipment [3].

In 2007, datacenters in Western Europe consumed a whopping 56 terawatt-hours (TWh) of power per year. According to the EU, this figure is likely to almost double to 104 TWh by 2020 [15].



The figure to the right plots Energy Star's [3] electricity usage measured and projected up to 2011. We extrapolate these values linearly up to 2017 and re-calculate them adjusting for the error in voltages if voltages stop scaling down given the current ITRS projections. Not surprisingly the electricity usage increases exponentially if not mitigated. The projected increase in energy prices (not shown) would only exacerbate the rate at which the total cost of ownership of datacenters would increase.

The European Ecosystem

Given that datacenters form the backbone of information processing, storage and communication in the modern world, and information is a key component of supply-chain for both products and services, the European economy is highly dependent on scalability and efficiency of datacenters. Innovation in datacenter design is critical for both established enterprises with large investment and stake in IT as well as for smaller businesses and startups who could otherwise not be able to afford ever-increasing costs of owning IT. A slowdown of IT scalability in terms of cost and performance will undoubtedly negatively impact enterprises in Europe.

Europe has an excellent opportunity to lead in innovating novel technologies for datacenters with an established and favourable ecosystem of enterprises. Europe has long been a leader in mobile and portable digital platform designs, much of which will likely lay the foundation for future servers and datacenters. Server blades based on ARM are appearing in the market. The extended use of open-source software in datacenters makes easier the migration of certain datacenter tasks to new architectures with activities like Linaro bringing additional momentum. ST Microelectronics is one of the lead manufacturers of Longsoon processors which lay the foundation for China's future datacenters, as well as the SPEAr line of low-power SoCs that emerge in datacenter server racks. SAP and Ericsson are top five (in sales) global providers of software and enterprise services with Ericsson also the lead telecom equipment provider. Bull has long been a pioneer of server clusters and OVH and 1&1 are among the big players of server and datacenter capital owners worldwide. Companies like Nokia and Alcatel-Lucent are also involved in operation or design of datacenters. Xyratex, is a leading storage vendor. Gnodal is developing high-performance datacenter Ethernet.

From the research funding point of view, there is already a preliminary set of projects to build on. The FP7 research objective "Computing Systems"[1] is funding a number of projects that address issues related to the next generation of datacenters. The Eurocloud project addresses the design of server chips based on designs coming from energy-efficient mobile platforms, while the IOLANES project addresses issues related to storage for datacenters. New projects like RELEASE address the programmability of large-scale datacenters.

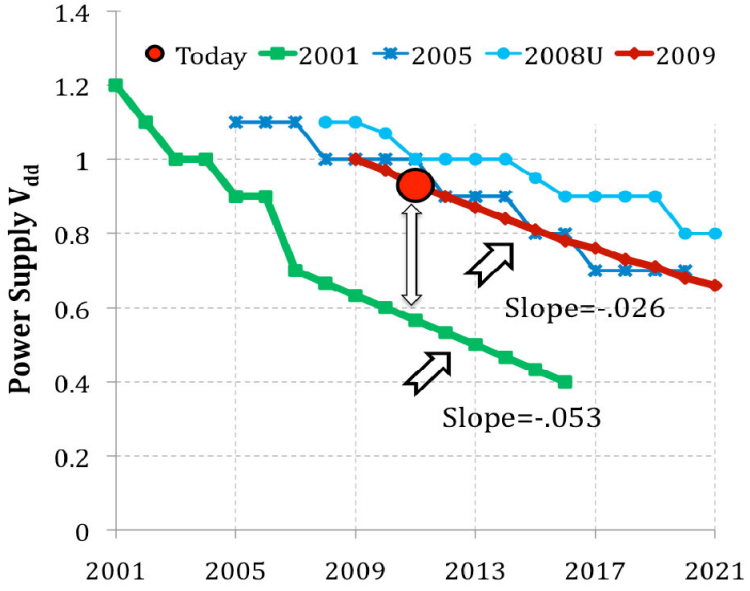
Affordable IT and datacenter ownership is not just limited to warehouse scale installations (e.g., Google, Facebook, Amazon). Technology and usage trends require dramatic improvements in server and datacenter efficiency both at a single enterprise scale (e.g., private clouds) and in installations that specialize in providing IT as a service (e.g., public clouds). It's this type of datacenter usage that Europe will not only mostly benefit from but also can pioneer.

[1] http://cordis.europa.eu/fp7/ict/computing/home_en.html

Challenge 1: End of Dennard Scaling

Improvements in semiconductor fabrication processes along with architectural and circuit innovation have led to exponential increases in performance and a reduction in cost of digital platforms since the inception of microprocessors. This phenomenon has been referred to as Moore's Law. In conjunction with Moore's law, the key enabler in truly realizing this exponential proliferation of digital platforms has been a commensurate improvement in energy efficiency of transistors described by Dennard.

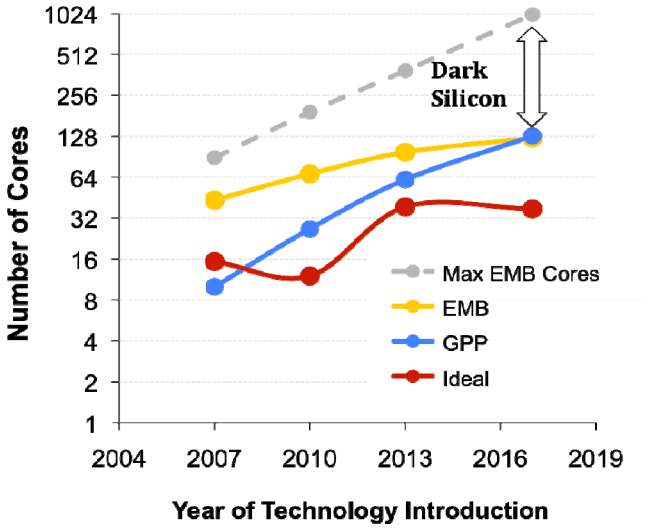
Dennard scaling describes a reduction in chip voltage levels that allow for maintaining a constant chip energy/power envelope while doubling the transistor count every fabrication process generation. Because power is quadratic in transistor supply voltages, a small reduction in voltages would allow for a larger reduction in overall energy allowing for doubling the transistors and operating them at higher frequencies.



The figure above plots ITRS [4] projections from 2001 until 2009 for transistor voltage levels. The figure indicates that indeed Dennard scaling has dramatically slowed down. The net result has been a paradigm shift in chip design to multicore architectures with a levelling in chip frequencies. In a nutshell, digital platforms have hit a technologically-induced energy barrier that is forcing designers towards more efficient use of transistors and simpler cores.
In the short-term there will be a push towards leaner hardware to maximize data access rate and processing for a given a power budget. Portable processor

technologies as in the FP7-funded EuroCloud server [5] project, emerging memory (e.g., 3D-stacking, PCM) and storage (e.g., SSD) technologies will help dramatically improving densities and efficiencies in servers and datacenters.

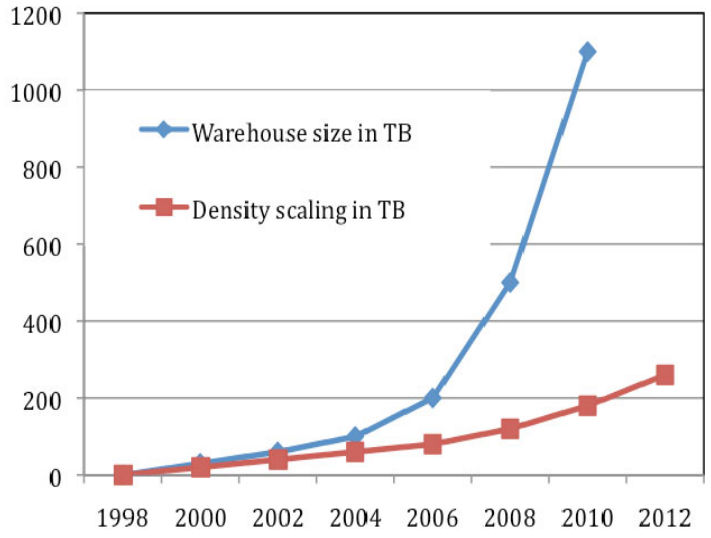
Unfortunately, in the long term, multicore designs are not a panacea. Despite the inherent scalability of threaded server workloads, increasing core counts can't directly translate into performance improvements because chips are physically constrained in power. Similarly, there is a limit to how simple and power-efficient individual cores can get. The figure below (from [6]) shows the limits of scaling and the inevitable emergence of dark silicon. The figure plots the optimal throughput design (for all frequency/voltage levels provided) for a given technology node based on ITRS projections and performance/power models for OLTP workloads on IBM DB2 for general-purpose cores (Niagara, GPP), embedded cores (ARM 11, EMB), and fully-customized cores (Ideal). These results assume that chip bandwidth limitations are mitigated through emerging wide-IO and 3D technologies and as such power is the only limiter for all core designs, leading to dark silicon.



Much as specialization laid the foundation for energy-efficient mobile platforms since the 90's, specialization and tighter integration of applications with hardware will drive the research and innovation in servers and datacenters. The growing demand on datacenter utility (e.g., see Challenge 2) on the one hand and the slowdown of Dennard scaling mean that a reduction of two orders of magnitude in energy is necessary to enable scalability of datacenters into the next decade.

Challenge 2: Taming the Data Deluge

Our life is increasingly data-driven. Digital platforms have made it possible to collect and analyze data to improve every aspect of our life be it health, entertainment, commerce, government or research. Science is now entering a "Fourth Paradigm" [7] where data mining along with theoretical, empirical and simulation-based frameworks form the backbone of modern science. We are entering an era of "Data Deluge" and are only starting to tap into its vast potentials. There are clear examples now where data deluge has resulted in dramatic improvements in technologies that were otherwise considered ineffective. Language translation and augmented reality are two examples where data-driven approaches have shown major strides in applicability and effectiveness. Exploiting the data deluge can be seen as the second wave of the IT revolution where data (and not computation) takes center stage.



Unfortunately, data is growing faster than technology can sustain. The Economist [8] reports that by one estimate all of earth produced about 150 exabytes of data in 2005. The same estimate for 2010 is 1200 exabytes. Li & Fung, a supply-chain managing firm saw a tenfold increase in data through the networks in just 18 months. The US video surveillance technologies improvements have resulted a 30-fold increase in data in just one upgrade.
The figure above plots the results of a survey by Wintercorp [9], plotting the maximum data warehouse size among enterprises. The results indicate that indeed data warehouses are growing on average faster than the increase in storage densities in digital platforms [10].

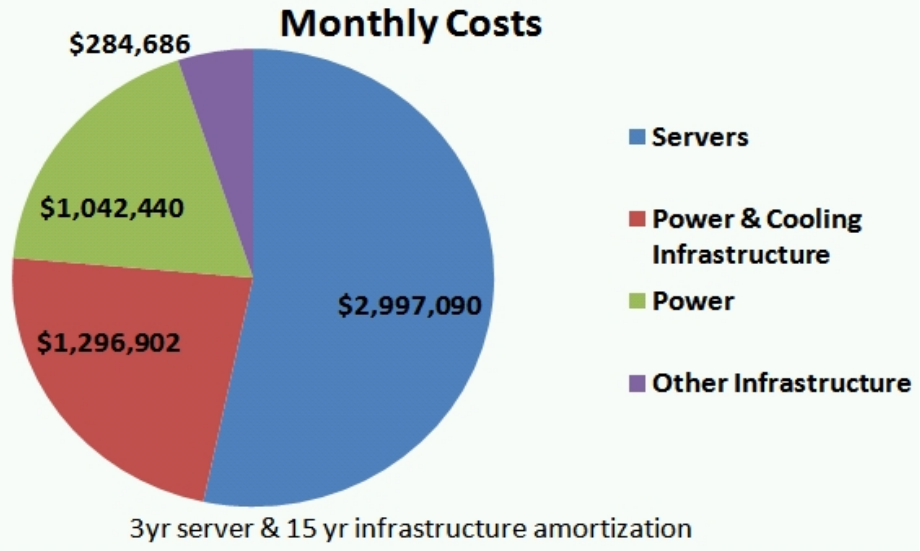
The data deluge brings about a major paradigm shift in computing, along with a number of accompanying challenges and opportunities in designing datacenters. First off, technologies to enable reductions in energy must scale faster than Moore's law to keep up with data deluge. The latter argues for opportunistic optimizations in the entire system stack from algorithms all the way down to circuits to enable closer integration of processing and storage of massive data. Plotting through and making sense of massive data will require a complete redesign of data management technologies (beyond conventional database

management systems) to locate data, and innovation in machine learning to reduce the size and dimensionality of data for services and applications. Data deluge also opens up the possibility of probabilistic computing to enhance scaling because not all processing, storage and communication of data requires the same guarantees in quality [11].

Secondly, the data deluge gives rise to concerns about privacy, security and reliability. Large enterprises (e.g., Google, Amazon, Microsoft, Apple and Facebook) that are big datacenter capital owners are all betting on collecting an analyzing data to personalize IT. Unfortunately, personalized computing also has the disadvantage that information that is otherwise private can end up in a database either through sensors (e.g., cell phones) or just given voluntarily based on trust. The net result is that future datacenters must rely on technologies that guarantee security and reliability for a given set of data based on a negotiated contract (i.e., an SLA).

Challenge 3: Holistic Integration

Energy management and minimization in today's datacenters is directly related to cooling, power delivery and eventually energy sources. The figure below [12] plots a breakdown of cost in a modern datacenter. In the long term, unless mitigated, energy costs (e.g., power delivery and cooling) will become a substantial if not dominant fraction of the total cost of ownership [3].

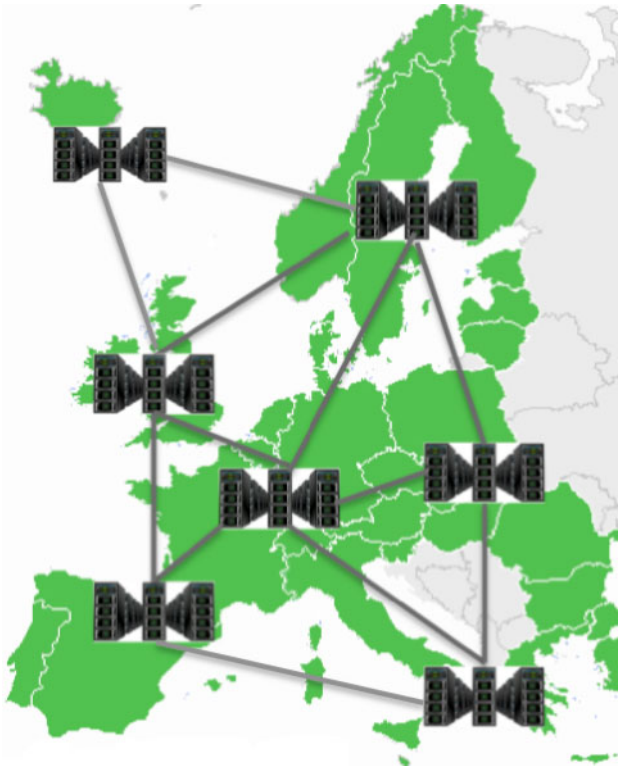


These trends have a number of implications on the design of future datacenters. In the short term, datacenter owners will push the temperatures to much higher levels to reduce cooling costs and handle server faults in software. For instance, Microsoft's fourth generation datacenters are already running at many degrees higher temperatures than prior generations and are mostly cooled with exterior air at ambient temperatures. As servers get cheaper (relative to the operation cost), software takes center stage in detecting, isolating and dealing with faulty hardware. Scalable and efficient software fault tolerance with minimal hardware support will likely emerge as the dominant avenue to provide reliability in datacenters.

In the medium to long term, a tighter integration of infrastructure (e.g., cooling, power delivery and energy source) along with holistic power minimization and management will drive datacenter design. Future designs require a hand-in-hand collaboration of software and hardware to provide both energy (through specialization) and thermal management. Thermal management requires hardware support for instrumentation (e.g., distributed sensors across the datacenter) and software support to mine the data and drive both power management (e.g., through hardware knobs or load balancing) and cooling. Future datacenters are likely to benefit from liquid-cooling technologies that have dramatically higher heat removal efficiencies than today's air-cooled systems both at the chip, blade and rack level [13].

Challenge 4: Federated Datacenters

Although single giant datacenters are today dominating the headlines, efficiency and reliability also dictate that an IT ecosystem exists in the form of a federated set of datacenters as opposed to a single physical entity. There are a number of reasons dictating the need for a federated organization. Data has affinity to its source of origin (e.g., sensor data) or must reside in a specific geographical (e.g., proximity to usage or infrastructure) or legal boundary. Many datacenter owners today opt for geographical locations with physical proximity to energy source (e.g., hydro) and low ambient temperatures. Moreover, communicating data may either be impractical (e.g., in the case of massive data) or have legal implications. Finally, to enhance reliability, sensitive data must be kept (replicated and) redundantly in multiple physical sites. As such, depending on the application and usage, data and datacenters may be inherently distributed and must co-operate in a federated manner.



Federated datacenters, including geo-distributed mini/micro datacenters, lead to a number of critical research challenges related to networking, and distributed system and data management technologies. Technologies to bridge across multiple datacenters must enable seamless integration of services across a heterogeneous collection of datacenters each with specialized servers, operating systems, service interfaces, intranetworking and storage technologies. The federated datacenters must collectively collaborate to provide both location-dependent services and load balancing as well as reliability and availability in case of disaster recovery and failover.

Recommendations for a Research Program on Datacenter Design Innovation in Europe

In this report, we include a list of recommendations for medium-term and long-term research. The medium-term priorities include:

- **Achieving mobile platforms efficiencies in server nodes**: Innovation in processor system-on-chip, memory, and persistent storage hierarchies to enable data management at an order of magnitude less energy. Traffic-proportional interconnects and switching at all levels (from intra-core to inter-rack) as well as I/O subsystems.

- **Energy-proportional distributed systems**: Designs for network fabrics that dissipate energy commensurate to usage, traffic and network load. Designs for data management across multiple installations to provide reliability and availability for a negotiated service-level agreement.

- **Tighter integration with infrastructure**: Designs for synergistic energy management in both servers and infrastructure including power distribution and cooling. Innovation would require online models and technologies for sensing and throttling energy to provide (real-time) performance and cost guarantees.

- **Federated datacenters / Multi-Datacenter architecture**: Designs for collaborative distributed datacenters to implement both location- and workload-dependent services, redundancy for disaster recovery and failover. Innovation in networking fabrics and distributed systems to extend services across multiple heterogeneous datacenters. Architecture of systems based on multiple datacenters, including scheduling, power management and data/workload placement

Longer term priorities for Horizon 2020 (2014-2020) that would enable improvements in efficiency and reductions in electricity by orders of magnitude include:

- **Specialized hardware for servers**: Designs for the end of the Dennard Scaling and Dark Silicon require identifying opportunities for specialization in emerging data-intensive workloads. Programming paradigms for application and system software targeting heterogeneous hardware that enables exposing energy as a first-class constraint. Design of hardware and software that facilitate verification. Optical interconnects for all communication and new persistent storage architectures leading to new design trade-offs.
- **Probabilistic servers**: Server designs for computing paradigms that trade off fidelity in computation and data access for improved efficiency. Machine learning technologies to reduce the size and dimensionality of

datasets for analytics. Probabilistic paradigms for error-resilient computation and data access to enable pushing circuits beyond guardbands.

- **Computing with renewable energy**: Renewable sources of energy are highly variable in availability. Innovation in technologies that help improve energy capture, and help throttle server energy usage with energy storage and availability.

- **Datacenter everywhere**: global scale federation of compute nodes to deliver workload-optimized compute and network solutions.

European industry and academia should now join forces in executing an ambitious research and innovation programme in datacenters. Otherwise, the danger is clear and present: Europe will simply not be part of the 21st century data-centric economy.

References

[1] OECD, “The Knowledge-Based Economy”, Paris, GENERAL DISTRIBUTION OCDE/GD(96)102.

[2] Kenneth G. Brille, “The Invisible Crisis in the Data Center: The Economic Meltdown of Moore’s Law”, Uptime Institute white paper, 2007.

[3] ENERGY STAR Program, “Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431”, 2007.

[4] Data from ITRS, www.itrs.net, July 2011.

[5] EuroCloud Server, www.eurocloudserver.com.

[6] Nikos Hardavellas, Mike Ferdman, Babak Falsafi, Anastasia Ailamaki, “Toward Dark Silicon in Servers”, IEEE Micro special issue on Big Chips, July/August 2011.

[7] Tony Hey, Stewart Tansley, and Kristin Tolle, “The Fourth Paradigm: Data-Intensive Scientific Discovery”, Microsoft.

[8] The Economist, “The Data Deluge”, February 25th, 2010.

[9] Richard Winter, “Scaling The Data Warehouse“, Information Week, 2008, www.intelligententerprise.com.

[10] EMC, “The Digital Universe”, http://www.emc.com/leadership/programs/digital-universe.htm.

[11] Babak Falsafi, “Reliability in the Dark Silicon Era”, IOLTS Keynote, 2011.

[12] James Hamilton’s Blog, www.mvdirona.com, 2008.

[13] J. B. Marcinichen, J. Olivier, and J. R. Thome, “Reasons to Use Two-phase Refrigerant Cooling”, *ElectronicsCooling*, 2011.

[14] Richard Smalley, “Richard Smalley on Energy”, University Professor Lecture Series, 2003.

[15] Dennis Bouley, Estimating a Data Center’s Electrical Carboon Footprint, http://www.apcmedia.com/salestools/DBOY-7EVHLH_R0_EN.pdf

Participants and Input

Babak Falsafi (Editor)	EPFL & EcoCloud
Angelos Bilas	FORTH & University of Crete
Koen De Bosschere	Ghent University
Kristof De Spiegeleer	Dacentec
Philippe Dobbelaere	Alcatel-Lucent
Luigi Grasso	BULL
Elio Guidetti	ST Microelectronics
Andrew Moore	University of Cambridge
Emre Ozer	ARM
Yanos Sazeides	University of Cyprus
Daniel Scheibli	SAP
András Vajda	Ericsson
European Commission:	Max Lemke
	Rolf Riemenschneider
	Panagiotis Tsarchopoulos

Technology-Scalable Datacenters

European Commission
Information Society and Media