NSF Workshop on Sustainable Data Centers

Jun 22-23, 2015

Palo Alto, CA

Workshop Co-Chairs
Prashant Shenoy, U. Massachusetts - Amherst
Thomas F. Wenisch, U. Michigan

Executive Summary

Data centers continue to be among the fastest growing users of electricity in the U.S. Due to aggressive adoption of cloud-based computing, the demands on data centers are growing exponentially, and both academia and industry will need to rethink how data centers are designed, built, and operated to be sustainable. Around the mid-2000's, the advent of mega-scale internet services and public cloud offerings led to a redesign of data center architectures, which addressed key inefficiencies, particularly in electrical and mechanical infrastructure. However, this first generation of improvement has plateaued. It is time for a second, holistic, clean-slate redesign of the data center, encompassing new server architectures, heterogeneous computing platforms, radical networking paradigms, new mechanical and electrical designs, intelligent cluster management, and radical rethinking of software architectures.

The NSF Sustainable Data Center (SDC) Workshop brought together industry practitioners, academic researchers, and government representatives to build the community and discuss the vision, challenges, and opportunities for SDC research for the next 5-10 years.

Attendees identified central challenges across three domains, the SDC software & hardware stack, SDC power & infrastructure, and resources & alliances to facilitate joint academic-industry research. The discussions identified four overarching research objectives: (1) Define the limits of sustainable data center efficiency and benchmark existing systems against these goals. (2) Scale software efficiency to match the rate of data growth. (3) Define algorithms, models, systems, and efficiency metrics that can incorporate smart power grid technologies into sustainable data centers. (4) Develop funding programs and incentivization mechanisms that encourage academic-industry alliances to facilitate knowledge transfer between cloud and data center operators and academic researchers. This report details specific challenges and recommendations to realizing each of these four objectives.

Introduction

Data centers are the core of modern business environments as computation has been rapidly moving into the cloud in the last decade. Data centers are among the fastest growing users of electricity in the U.S. consuming an estimated 91 billion kilowatt-hours of electricity in 2013, about 2% of the total electricity consumed by the US. When operating a data center of hundreds of thousands of servers, it is essential that they be operated effectively to improve energy efficiency and environmental sustainability. Due to aggressive adoption of cloud-based computing, the demands on data centers are growing exponentially, and both academia and industry will need to rethink how data centers are designed, built, and operated to be sustainable. Despite a decade of research and industrial innovation, a recent report from Natural Resources Defense Council (NRDC) indicates typical small and mid-size data centers hosting private clouds still hold many wasteful practices. Whereas best practices at mega-scale commercial cloud operators (e.g., Facebook, Microsoft, Google, and Amazon) have addressed the most egregious wastes (e.g., inefficient cooling), we nevertheless must find ways to transfer these best practices across the data center landscape and address the remaining performance and efficiency challenges that afflict even the largest installations.

At NSF, several core and crosscutting programs, including CSR, CyberSEES and CPS, have taken action and invested in numerous projects nationwide to address the fundamental issues of sustainable data centers (SDC). These projects can be broadly classified into three areas: individual servers, data center-level resource management, and energy supply. For individual servers, the ultimate goal is to design energy-proportional computing nodes by reducing both idle power (e.g., PowerNap) and dynamic power (e.g., DVFS). For data center-level resource management, the main goal is to utilize the resources (i.e., CPU, memory, bandwidth) in an effective and efficient way, without wasting resources. A better understanding of the workload is the key. Cloud providers can allocate right-sized server and network platforms to meet users' application requirements. At the energy supply, green power, such as solar and wind, are beginning to enter the data center power supply chain. By leveraging renewable power sources, local micro-grids may offset some or all of a data center's energy needs, particularly for small and mid-scale facilities. At the largest scale, incentives and mechanisms must be sought to encourage provisioning clean, reliable power in concert with the existing public grid.

Around the mid-2000's, the advent of mega-scale internet services and public cloud offerings led to a redesign of data center architectures, which addressed key inefficiencies, particularly in electrical and mechanical infrastructure. At the same time, accelerated need for efficient servers spurred a generation of research on CPU, memory, network, and storage power management techniques, which have led to a marked improvement in server efficiency and energy proportionality. However, this first generation of improvement has plateaued; further opportunity in the large-scale mechanical infrastructure is limited and no single server or network component stands out as the key source of inefficiency. Hence, it is time for a second, holistic, clean-slate redesign of the data center, encompassing new server architectures, heterogeneous computing platforms, radical networking paradigms, new mechanical and electrical designs, intelligent cluster management, and radical rethinking of software architectures while considering changing usage patterns (e.g., hybrid private/public clouds).

Although the need for broad input on sustainable data center design is acute, concerns about competitive advantage and user privacy have made open collaboration between academic researchers and cloud operators difficult. Academic researchers have limited access to production data center facilities and hence are not always aware of the real problems faced by practitioners. The immediate risk of this disconnect is that researchers might spend their time attacking imagined problems that are irrelevant to modern practice.

In addition to developing promising technologies to improve data center efficiency, we also need new metrics to assess the success of SDC research. Currently, power usage effectiveness (PUE) is a widely reported metric to assess the energy efficiency of a data center. The impact of renewables can be assessed via carbon usage effectiveness (CUE) to measure the combined impact of clean energy and energy efficiency on greenhouse gas emissions, and water usage effectiveness (WUE) can be used to assess the water usage of a data center. And yet, all three of these metrics fall short of describing the true efficiency of the data center. They fail to reflect waste at the enclosure/tray level (e.g., VRMs, server fans). Moreover, they do not assess the efficiency or value of the computation being performed and hence fail to reflect server hardware inefficiencies or software bloat.

The NSF SDC Workshop brought together industry practitioners, academic researchers, and government representatives to build the community and discuss the vision, challenges, and opportunities for SDC research for the next 5-10 years. More specifically, the objectives of the workshop include:

- Foster the SDC community, to increase interaction between academia and industry
- Set the vision and identify challenges and open problems, such as research reproducibility, benchmarks, experimental methods, and so on.
- Identify and exploit resource sharing mechanisms for workloads, traces, and so on.
- Seek opportunities to leverage the two recently funded NSFCloud testbeds to do SDC research (e.g., identify measurement and monitoring requirements)

Section 1: Recommendations for SDC Hardware & Software Stack

Sustainable data centers will require sustained improvement in the efficiency of systems software. Given the impending end of Moore's law, it is time to revise the state-of-the-practice in software design to eliminate endemic sources of inefficiency and accommodate next-generation, specialized hardware. To this end, there is a need for research that accomplishes two critical goals: (1) rigorously define the fundamental limits of efficiency for software and how far we are from them at the current time, and (2) improve energy-efficiency of software to match the exponential rate of data growth.

Objective #1: Define the limits of efficiency and benchmark existing systems against these goals.

Challenge #1: Develop metrics for quantifying system-level efficiency and sustainability.

It is not clear what are the right system-level metrics for efficiency and sustainability in datacenters, and whether efficiency (energy and/or computational) is the same as sustainability. Traditional measures, such as Power Usage Effectiveness (PUE) measure only certain sources of inefficiency (i.e., losses in voltage conversion or cooling) and fail to tie effective use of power/energy to the performance/value delivered by the system.

- Recommendation: Define nomenclature and metrics for a "Software PUE" -- an efficiency metric that quantifies the overhead of software components (analogous to the PUE metric used to quantify the overheads of electrical/mechanical infrastructure components).
- Recommendation: Take "common" operations, such as sorting, singular value decomposition, or deep neural networks, and derive benchmarks. Optimizing such benchmarks can improve a wide range of systems.
- Recommendation: Extend traditional asymptotic (big-O) analysis of software systems to rigorous study of the constants at scales of interest.
- Recommendation: Develop metrics for sustainability that consider resources and lifetime issues beyond energy consumption during operation. For example, consider the consumption of potable water, use of rare materials or unsustainable manufacturing/transportation processes, and the the reusability, recyclability, "up-cyclability" of system components.
- Recommendation: Academia needs input from industry on how sustainability is presently quantified by data center operators.

Challenge #2: Academics do not have access to large-scale data centers for research.

- Recommendation: The community should set up experimental clusters and make them available to researchers, for example, the NSF Cloud initiatives.
- Recommendation: Developing validated simulation frameworks that enable exploring optimization scenarios practically would help with datacenter research in academia. The challenge here is that this may be too hard of a problem. How far can we go with a simple framework/model? HPC has addressed a similar challenge in the past, but for a much more simple, predictable, and regular environment.
- Recommendation: Industry/academic research is currently focusing almost exclusively on the largest mega-data centers. There is a long tail of datacenters of different sizes and constraints, which, in the aggregate, represent a majority of the installed base and power consumption. It may make sense to increase research focus on micro-datacenters and edge servers that are more easily accessible. For example, container-based systems may be easier for researchers in academia to set up and manage.

Challenge #3: There are no representative benchmark suites for internet-scale distributed services.

There is currently no regularization of benchmarks that are real-world and relevant across companies for internet-scale systems. Longer term, interest is shifting towards what happens with Internet-of-Things sensors/machine-generated data. This shift needs to be reflected in the benchmarks used for sustainability research.

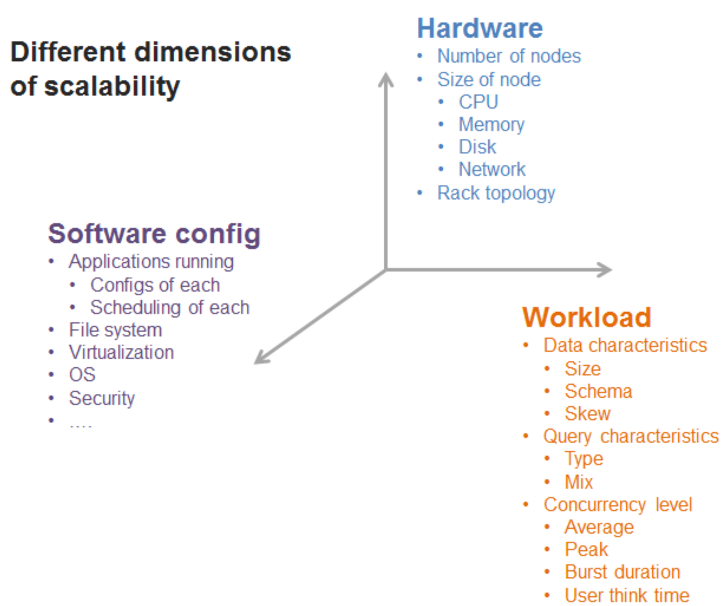
- Recommendation: The community can create an interface for researchers to formalize a short list of benchmark requests to industry affiliates, and for industry to express interest in sharing their applications/datasets/traces. It can also provide a platform for auditing the anonymization and/or incentivising the sharing of benchmarks. The two main classes of workloads that need to be included are user-interactive (websearch, email, key-value stores), and analytics (building on existing open-source frameworks). It is also critical to include application scenarios that arise in multi-tenant environments, since these elicit numerous challenges with respect to unpredictability and interference. In all cases, there is a need for representative input traffic and/or datasets to drive the applications.
- Recommendation: Machine-to-machine communication, smart cities, sensor networks, life-critical operations have strong geo-centric characteristics and mobility (data is not static) and all need real-time guarantees. There is a need to quantify how different such workloads are from current data center workloads, and what changes they would require in the hardware-software stack.

Challenge #4: There are no models for predicting how infrastructure components and workloads scale.

There is currently a lack of empirical or theoretical models that can predict scaling of a set of workloads on a system. A common question to ask is, given a workload, cluster hardware, and software/applications, what is the expected performance, and how does that scale across different meanings of scale? We currently lack any empirical or theoretical models to answer such scaling questions. Such questions aim to inform the community regarding what kind of problems can be scaled across what ranges. Subproblems:

- What is the meaning of performance as a function of [latency, throughput, 99%-tile tail, users supported, cost, energy, sustainability]?
- What is the meaning of scale across different hardware and workload dimensions?
- How do different meanings of performance at different scales change across different software application combinations?

- Recommendation: The general problem space is complex, and today both industry and academia talk about "scale" very haphazardly. Both should make more efforts to more thoroughly and systematically understand this space.



Objective #2: Scaling software efficiency to match data growth

New paradigms for communication, such as social media and video sharing, and new capabilities for acquiring and storing data, such as ubiquitous cameras and other sensors and wide availability of high-speed internet connections, are resulting in exponential growth in the production and consumption digital data. For example, Facebook reports more than 45 billion messages and 4 billion video views per day in Q1'2015 (see Appendix) with trends indicating continued growth.

Historically, Moore's Law (and, critically, supply voltage a.k.a Dennard scaling) has facilitated exponential improvements in computational capability such that data processing systems have been able to extract valuable information despite exponential growth in input data. As circuit advancements slow, new avenues must be found to continue to improve computational capability to keep pace with input data. A significant opportunity lies in improving the efficiency of legacy multi-layered system stacks that are orders of magnitude less efficient than "bare-metal" performance analysis suggests is possible.

Challenge #1: Layered software interactions introduce bloat.

Multiple layers of software and high-level languages improve programmer productivity, but over time generate software bloat. This bloat leads to wasted resources and unpredictable response times.

- Recommendation: Developing a suite of representative data center services can help with quantifying and breaking down bloat across the software stack and determining sources of unpredictability. Because a large fraction of inefficiency comes from the several levels of indirection in software, funding agencies can initiate programs to find the balance between the productivity/programmability of high level languages and the efficiency (in performance and energy) of low-level primitives.
- Recommendation: Investigate efficient memory usage, new memory technologies, and alternative memory/storage architectures as a way to reduce software bloat and improve energy efficiency.

Challenge #2: Systems lose efficiency due to variability at scale

Latency-sensitive interactive services like web search, ad serving, data/image retrieval, machine translation, or text to speech services, must process terabytes of data with sub-second latencies. Today's CPUs are highly effective at hiding the nanosecond-scale latency of memory accesses and operating systems are highly effective at hiding the millisecond-scale latency of disks.

However, modern high-performance networking and flash I/O frequently lead to situations where data are a few microseconds away. Neither hardware nor software offer effective mechanisms to hide these microsecond-scale stalls. Moreover, OLDI services typically rely on a strategy of sharding their data sets over hundreds or even thousands of servers to meet latency objectives. However, this strategy mandates that fully processing a request requires waiting for the slowest straggler among these servers. As a result, exceedingly rare events, such as transient network congestion, interrupts, OS background activity, or CPU power state changes, which have negligible impact on the throughput of a single server nevertheless come to dominate the latency distribution of the OLDI service. At 1000-node scale, the 5th '9 of the individual server's latency distribution becomes the 99% latency tail of the entire request. These two challenges cause OLDI operators to execute their workloads inefficiently at low utilization to avoid compounding stalls and tails with queueing delays. There is a pressing need for systems research to find ways to hide microsecond-scale stalls and track down and address the rare triggers of 99.999% tail performance anomalies that destroy application-level latency objectives.

- Recommendation: Develop new approaches and abstractions for I/O and CPU scheduling that are well-suited to tolerate microsecond-scale I/O latencies.
- Recommendation: Investigate root causes of performance variability and latency "tails" and design mechanisms that reduce the frequency and severity of rare events that cause delays.

Challenge #3: Systems must harness application-specific accelerators to gain power/energy efficiency.

The gains in performance and energy efficiency necessary for sustainable, scalable data centers cannot be achieved with general-purpose computing. Effectively harnessing hardware accelerators must therefore be at the forefront of the research agenda. GPUs and FPGAs are the most-commonly available accelerators today. Unfortunately, programming GPU accelerators is challenging, while programming FPGAs is nearly impossible for a typical software developer. Advancements in interfacing with accelerators and programming accelerators is therefore necessary to allow high performance and efficient integration of accelerators into future data centers.

- Recommendation: Develop methods for easier integration of specialized systems and accelerators into complex software stacks. Develop new languages, programming models, and frameworks. Develop methods for validating complex hardware/software systems. Find ways to accelerate hardware/software co-design to match the rapid pace of software and algorithmic innovation and enable frequent releases to production.

Challenge #4: Managing resources in increasingly heterogeneous environments.

Future data centers face critical resource management challenges. Several trends exacerbate these challenges, and have not received sufficient study, for example, the increased deployment of specialized hardware, multi-tenant environments, and applications that span mobile, edge network, and data center resources.

- Recommendation: Encourage research on the impact of real-time constraints and Quality-of-Service guarantees on performance and energy efficiency in these environments. Increase emphasis on automated techniques for monitoring and inferring system parameters for energy-aware resource control. Identify the role virtualization and compute/data migration can play in mobile settings where data ingress and egress locations are moving. Develop a fundamental understanding of variability and stalls across the system stack.

- Recommendation: Investigate auto-tuning and automated configuration management to reduce the complexity of mapping applications to suitable platforms. Find ways to specify performance objectives declaratively rather than specify conservative resource reservations and rely on automatic tuning to achieve these objectives.

- Recommendation: Develop mechanisms for resource isolation and management that do not introduce the overheads of existing virtualization schemes.

Challenge #5: Misalignment of financial incentives and sustainability goals

When a company is fully vertically integrated (e.g., large data center operator running first party apps), financial incentives are often well-aligned to reduce sustainability goals that reduce total cost of ownership. However, where boundaries exist either internally that limit optimization or between entities there is an optimization gap where major inefficiencies can occur. For example, while a large data center operator may provide very efficient compute, a small consumer of those resources may not utilize them efficiently. Independently, there is only a small sustainability gap, however many small users in aggregate can result to huge amounts of aggregate resources being used inefficiently.

- Recommendation: Encourage research on resource management schemes and incentives that align disparate financial interests with sustainability goals. For example, investigate mechanisms that facilitate cooperative resource optimization in multi-tenant data centers.

Section 2: SDC Energy and Power Infrastructure

Sustainable data centers will need to coordinate their energy use with the smart grid while incorporating smart grid technologies, such as demand response, energy storage and renewable energy. To this end, there is need to encourage research that achieves three goals: (i) rigorously define new efficiency metrics that go beyond traditional power efficiency; (ii) design of novel algorithms, models and system to exploit and incorporate smart grid technologies ranging from dynamic pricing, demand response, energy storage and renewables; (iii) explore the interactions between economics, policy and engineering issues.

Challenge #1: Design of novel algorithms, models and systems to exploit and incorporate smart grid technologies

The electric grid of the future will be smart from a number of perspectives: it will employ real-time automated demand response to modulate the load at peak periods, it will integrate an ever growing fraction of clean energy sources such as renewables, it will offer new economic structures such as dynamic pricing and real-time markets, and it will employ energy storage.

Data centers are well positioned to exploit many of these newer smart grid technologies. They can modulate their load in response to pricing signals---traditional power management schemes can be designed to not only respond to load but also pricing or demand-response signals. Data centers can enhance the manageability of the grid by helping the grid achieve its objectives. They can exploit their existing UPS systems to serve as energy storage and also employ thermal storage for cooling. However, realizing these advances will require new research into algorithms, models and systems to incorporate and optimize a broad range of future smart grid technologies.

- **Recommendation**: NSF should encourage systems research into a broad range of topics that lie at the intersection of data centers and the smart electric grid to spur the next generation of advances in this area.

Challenge #2: New Efficiency Metrics

Data centers have traditionally used a metric such as Power Usage Efficiency (PUE) as a measure of their efficiency. As sustainable data centers of the future incorporate a range of new technologies to enhance their efficiency and sustainability, new metrics and benchmarks become necessary to understand the benefits and costs of the methods. New metrics such as water usage efficiency (WUE) and carbon usage efficiency (CUE) that have been recently defined are a start, but more research is needed to define additional metrics. Such metrics need to consider the energy source, including renewables, capture the nature of the power load imposed on the grid (e.g., steady load, bursty loads etc), among others.

- **Recommendation**: Researchers should pursue novel metrics and benchmarks for future sustainable data centers from the perspective of sustainability and energy use. Such metrics and benchmarks will need to address (i) load profiles of data centers from the grid's perspective, (ii) elasticity of the load in terms of modulating its power profile and the energy agility of the system, and (iii) use of renewable and non-renewable energy sources. These metrics should enable researchers to rigorously compare the performance of new algorithms, models and systems and also encourage industry to benchmark and optimize their data centers along these dimensions.

- **Recommendation**: Industry involvement should be encouraged since they have insights into operational aspects of data centers, which is valuable for designing these new metrics and benchmarks.

Challenge #3: Interactions between economics, policy and engineering issues

Many research challenges in this area lie at the intersection of economics, policy and engineering/systems issues. Data centers may pay for their energy use based on variable or dynamic pricing models. They also have access to directly buy power from energy markets. They may choose to participate in capacity markets and demand-response that result in direct savings. Optimizing their energy costs is therefore a complex optimization challenge and require a cross-disciplinary effort in economics and computer science. Sustainability of data centers also touches upon key policy issues. Policy issues have driven the recent growth of renewable deployments such as solar and wind. While these policies directly impact the decisions made by data center operators on whether to use on-site or contracted renewable energy. However seamless integration intermittent sources of energy remains a challenge.

A key challenge is whether data centers can go beyond "merely" striving to be carbon-neutral. It is harder to be carbon neutral and also be a net positive for the manageability of the grid. Addressing this challenge requires work at economics, policy, engineering and computer science.

Finally, it is also important to involve the energy and utility industry, particularly those who are forefront of energy optimizations in the smart grid.

- **Recommendations**: Researchers should pursue cross-disciplinary investigations involving engineering, computer science economics, and public policy to facilitate novel advances in smart grid aspects of data center research. Funding agencies should encourage industry collaborations on such projects, especially from the energy and utility industry. Research projects should be encouraged to consider policy aspects of their research or work with policy experts as part of their broader impacts.

Section 3: Infrastructure, Resources and Industry Alliances

In this section, we consider issues related to infrastructure, resources and industry alliances such as: what infrastructure and resources are needed for SDC research? How should the community leverage existing investments from NSF and other funding agencies (e.g., NSF FutureCloud) for SDC research? What role should commercial cloud vendors play? How can funding agencies foster sharing of resources, traces/data? What role should industry play in supporting SDC research? How should funding agencies help foster strong academic - industry alliances in this area?

Challenge 1: Infrastructure and resources are needed for SDC research

Sustainable data centers is a nascent research area. Unlike more mature research areas where tools and workloads are available, this area lacks detailed workloads, workload characterization frameworks, metrics and simulation tools. While there are hardware testbeds such as Chameleon and others, they are not sufficient to address the needs in this areas---they lack realistic workload and lack environmental instrumentation (e.g., hardware or OS knobs to measure or control power usage).

- **Recommendations**: Funding agencies should support additional infrastructure projects such as ones that create datasets and workloads for the community or build system-wide simulation tools. Direct industry involvement in this type of work is highly desirable but also highly challenging due to confidentiality and privacy considerations. Nevertheless, where possible industry contributions should be sought in this area (the publicly available Google Cluster I/O dataset and Facebook's open compute design specs are examples where industry has made such contributions).

  For hardware testbeds, funding agencies should either support extending and enhancing existing testbeds to support energy and sustainability research. Smaller testbeds such as a self-contained container with computing infrastructure and instrumentation to measure and control power and cooling should also be considered.

Challenge 2: How to leverage existing investments from NSF (e.g., NSF FutureCloud) for SDC research?

There are many types of SDC research that require access to cloud testbeds. These include: load shifting between data centers, power supply/surplus conditions, reducing and increasing energy usage based on reliability, disaster recovery research, power capping, addressing the difference between peak to average provisioning, capacity planning, economization of security, and heterogenous computing with dedicated hardware for energy savings. As noted above, current cloud testbeds are necessary but they lack workloads and instrumentation to facilitate SDC research into such topics.

- **Recommendations**: NSF should consider additional investments in characterization frameworks for energy research. Straightforward hardware enhancements, when possible, to support SDC research should be considered (some enhancements are non-trivial and deserve their own project). Support for emulators for failure, power deficiency, power scarcity and surplus scenarios should be considered.

Challenge 3: What role should commercial cloud vendors play?

Commercial clouds offer several advantages that are not available in current research testbeds. Many researchers use cloud servers for running systems research experiments and these testbeds are also useful for SDC research.

- **Recommendations**: Use of commercial clouds incurs monthly usage costs. Funding agencies should consider partnering with commercial cloud providers where the cloud provider gives a grant to relevant funded projects to cover cloud usage costs.

Challenge 4: What role should industry play in supporting public (academic) SDC research?

Industry research project have often focused on near-term problems faced by industry, but it is Important for academics to focus on the long term and have a 5-10 year research horizon.

- **Recommendation**: Rather than ad-hoc one-to-one industry-academia project, consider establishing a consortium of academia and industry partners that will form a steering group to define and articulate longer-term grand challenge problems in sustainable data centers. The consortium could meet bi-annually with a changing focus---for instance, by defining a challenge/focus group, spinning it off (by finding resources, partners, specific objectives for the effort), and moving on to define another challenge.

Challenge 5: How should NSF help fostering strong academic - industry alliances in this area?

- **Recommendations**: SDC research is an area where fostering strong academic - industry alliances is especially important. To do so, NSF should consider supporting workshops such as this one where industry/academia can get together and discuss grand challenges on a particular thematic area. NSF could consider an industry advisory panel to seek inputs on research challenges faced by industry that could be addressed through academic research projects. NSF should sponsor collaborations between academics and industry, similar to the NSF computing innovation fellowship in terms of matchmaking support.

This could be supplemented with a match-making service that connects fellows with industry mentors. Finally NSF should help identify projects the public interest since these are ones that companies will have a strong interest for collaborations.

Appendix 1: Attendee List

Christos Kozyrakis, Stanford University
Adam Wierman, Caltech
Christopher Stewart, The Ohio State University
Sandeep Gupta, Arizona State University
Shaolei Ren, Florida International University
Hamed Mohsenian-Rad, UC Riverside
Ying Lu, University of Nebraska-Lincoln
Tarek Abdelzaher, UIUC
Xiaorui Wang, The Ohio State University
Peter Varman, Rice University
Chris Malone, Google
Amin Vahdat, Google
Karthick Rajamani, IBM
Weisong Shi, NSF
Ricardo Bianchini, Microsoft and Rutgers University
Amy Apon, NSF
Kate Keahey, University of Chicago and Argonne National Lab
Christina Delimitrou, Stanford University
Yanpei Chen, Cloudera
Michael Ferdman, Stony Brook University
Qiang Wu, Facebook
Cullen Bash, Hewlett-Packard Laboratories
Parthasarathy Ranganathan, Google
Raymond Parpart, University of Chicago
Dale Sartor, LBNL
Kim Hazelwood, Yahoo Labs
Qingyuan Deng, Facebook Inc.
Dahlia Malkhi, VMware Research
Jie Liu, Microsoft Research
Michael Wei, VMware
Chris Page, Yahoo! Inc.
Ravi Soundararajan, VMware, Inc.
Ali Saidi, ARM
Ravi Iyer, Intel
Thomas Wenisch, U. Michigan / Google
Prashant Shenoy, U. Mass-Amherst
Darleen Fischer, NSF
Charlie Manese, Facebook
Mike Marty, Google
Anand Sivasubramaniam, Penn State
Magnus Herrlin, LBNL
Zhenhua Liu, Stony Brook University
Bindu Madhavan, SAP & USC
Jishen Zhao, UCSC
Heiner Linz, Stanford University

Appendix 2: Attendee Position Statements

NSF Workshop on Sustainable Data Centers

Position Statement by:
Tarek Abdelzaher, *University of Illinois at Urbana Champaign*

Challenge: Hardware-software Co-design for Ultra-low-power Data Centers:

To attain the next order-of-magnitude cut in data center energy consumption, we must fundamentally re-think the underlying *hardware*. According to a recent report by the Natural Resources Defense Council (NRDC), US data centers consumed an estimated 91 billion kilowatt-hours of electricity in 2013. It is to be compared to 61 billion kilowatt-hours in 2006 according to the Environmental Protection Agency (EPA). The NRDC report further predicts that total data center energy consumption will reach 140 billion kilowatt-hours of electricity by 2020, emitting nearly 150 million metric tons of carbon pollution per year. This trend is roughly equivalent to a doubling of energy use every decade. In contrast, according to the Energy Information Administration, US residential energy consumption has risen by only 10% in the last 30 years (or only 3% per decade), and is presently declining. Similarly, the emissions of the transportation sector have increased by only 8% in the last decade and are presently declining as well.

Curbing data center consumption growth is challenging. While consumption trends of homes and cars are ultimately constrained by the overall population growth which is limited, data center consumption trends are more correlated with growth in digital data. According to the International Data Corporation (IDC), a premier global provider of market intelligence and data analytics, the digital universe will grow by a factor of 10, from 2013 to 2020, or from 4.4 trillion gigabytes to 44 trillion, thus more than doubling every two years. The need to derive personal and business value from the rapidly increasing volume of world data will continue to escalate demand on computing services and data centers.

In order to accommodate this trend, one possibility is to replace high-end servers by teams of low-power embedded boards. Recent boards on the market are more energy-efficient than their high-end computing counterparts. When they are pooled such that pool capacity collectively adds up to the original service capacity, their total energy demand remains an order of magnitude lower than that of the high-end servers. The approach has significant implications on data center software, which now has to be broken up into a lot more components that are individually a lot less powerful. The greatest impact is expected to be in the area of data center networking and data storage/caching protocols. Efficient utilization of network links shared by a larger number of smaller components requires novel communication protocol design. Similarly, data fragmentation into a larger number of smaller chunks requires novel data storage and retrieval protocols whose overhead remains manageable despite the significant increase in the number of storage elements. Finally, solutions are needed to maximize the end-to-end flow and reduce the end-to-end latency of workflows extending across the larger number of smaller components.

Position Statement

NSF Workshop on Sustainable Datacenters

Christina Delimitrou

Sustainability (efficiency) challenges. Much of the inefficiency in datacenters today comes from the lack of *performance predictability*. This is the result of several factors, including interference in shared resources, load spikes, and application-level bugs. Improving sustainability in datacenters requires a clean-slate approach in designing a system stack that provides end-to-end performance guarantees.

The following three directions are necessary:

1. **Hardware:** At the hardware level, there is need for isolation and partitioning mechanisms that enable predictable performance at high utilization. If isolation is the one promising approach in datacenter hardware, specialization is the second. FPGAs and special-purpose architectures are already making their way in today's datacenters, both for performance and efficiency. The question here becomes: how much specialization is needed to capture a significant fraction of a datacenter's compute cycles, given how frequently datacenter applications change? And, how does this affect the higher levels of the system stack?

2. **Operating Systems:** Strict hardware isolation eliminates a lot of OS complexity. Current OSes running user-driven applications spend the majority of their time in the OS scheduler or network stack as opposed to doing useful work. Simplifying the OS to handle protection and abstraction, not resource management, and offloading some operations to hardware (e.g., offload engines in recent NICs) can benefit both performance and efficiency. Specialization also allows the control plane to be tailored to the specific policies individual classes of applications care about, which conforms with the domain-driven design (DDD) of microservices.

3. **Application Design:** In datacenter applications the current first-order constraint is performance, with resource efficiency being an afterthought. This ends up hiding several orders of magnitude (or several generations of Moore's Law) of performance in the many levels of the software stack. Providing application designers with feedback on how to restructure applications to use fewer resources without performance penalties, e.g., through program synthesis has the potential to significantly improve efficiency.

Designing a new system stack for the datacenter also brings up the questions of how many levels of indirection we need to guarantee both predictability and elasticity, what APIs are needed across the stack, and whether design decisions used in traditional systems, e.g., 4K pages, virtual memory, etc. still make sense in warehouse-scale computers.

Finally, up until now this redesign improves predictability, albeit not in a formally provable way. Formal methods, such as proof assistants and languages like TLA+ are finding their way in distributed systems to provide correctness guarantees. These approaches should also be applied to formally reason about the performance characteristics and requirements of these systems.

How can NSF help: A major roadblock in datacenter research is the lack of representative applications, traces, and datasets. Most of it comes from the lack of incentive from the industry's standpoint, given the effort needed to anonymize data and software. It is also made worse by the fact that each academic group requests different datasets and/or traces. NSF can help in that direction by providing a systematic interface for research groups to express their requirements, and consolidate them to a small, manageable number of requests to industry that can be beneficial for the community as a whole.

Chameleon: A Large-Scale, Reconfigurable Experimental Environment for Cloud Research

Kate Keahey
Argonne National Laboratory

Cloud computing services have become critical to all major 21st century economic activities – yet, we are only beginning to understand this new important paradigm. Questions persist regarding applicability of the cloud platform to the emergent data-intensive and sensor-based applications, its suitability for high performance computing (HPC) applications, and its potential to leverage major emergent technologies such as Software Defined Networking (SDN) to name just a few. Answering those questions requires the ability to perform experiments at scale – in other words, an experimental testbed that would support experimentation with Big Data, Big Compute, and Big Instrument problems.

The Chameleon project has been funded under the NSFCloud initiative to provide such a large-scale platform to the open research community and allow them to explore transformative concepts in deeply programmable cloud services, design, and core technologies. Its reconfigurability will allow users to explore problems ranging from the creation of Software-as-a -Service to kernel support for virtualization. The broad range of supported research will include areas such as developing Platforms-as-a-Service, creating new and optimizing existing Infrastructure as a Service components, investigating software-defined networking, and optimizing virtualization technologies.

The Chameleon testbed, deployed at the University of Chicago (UC) and the Texas Advanced Computing Center (TACC), will consist of 650 multi-core cloud nodes (~14,500 cores total), over 5PB of total disk space, and leverage 100 Gbps connection between the sites. A large part of the testbed will consist of homogenous hardware to support large-scale experiments. This part of the testbed is composed of 12 racks, comprising 46 Xeon Haswell processors (42 compute and 4 storage servers) with OpenFlow-enabled switches; each rack will have 128 TB of storage and one of them will contain Infiniband network. In addition to distributed storage nodes, Chameleon will have 3.6 PB of central storage, for a persistent object store and shared filesystems. The testbed will also support heterogeneous units consisting of Atom microservers, ARM microservers, as well as a mix of servers with high RAM, FPGAs (Xilinx/Convey Wolverine), NVidia K40 GPUs, and Intel Xeon Phis to allow experimentation with high-memory, large-disk, low-power, GPU, and co-processor units. In its initial phase, the project leverages existing FutureGrid hardware at the University of Chicago and the Texas Advanced Computing Center with their FutureGrid configuration (i.e., as OpenStack clouds) to provide a transition period for the existing FutureGrid community of experimental users. This part of the testbed, called FutureGrid@Chameleon, hs been available to users since early December 2014.

To support a broad range of experiments, Chameleon will support a graduated configuration system allowing full user configurability of the software stack, from provisioning of bare metal and network interconnects to delivery of fully functioning cloud environments. In addition, to facilitate experiments and provide a “one stop shopping” for experimental artifacts, Chameleon will support a set of services designed to meet researchers needs, including support for experimental management, reproducibility, and repositories of trace and workload data of production cloud workloads based on both commercial and scientific clouds. The project will also provide innovative ways of integrating testbeds into the educational pipeline by designing and publishing new educational artifacts such as ready to deploy Chameleon appliances. We will encourage academic and commercial partners, as well as users, to submit and share artifacts, including traces and appliances that others can leverage and encourage discussion on the ways they can be represented.

The project is led by the Computation Institute at the University of Chicago and partners from the Texas Advanced Computing Center (TACC) at the University of Texas at Austin, the International Center for Advanced Internet Research (iCAIR) at Northwestern University, the Ohio State University, and University of Texas at San Antonio, comprising a highly qualified and experienced team. The team includes members from the NSF supported FutureGrid project and from the GENI community, both forerunners of the NSFCloud solicitation under which this project is funded. Chameleon will also form a set of partnerships with commercial and academic clouds, such as Rackspace, CERN and Open Science Data Cloud (OSDC), and will partner with other testbeds, notably GENI and INRIA's Grid'5000 testbed.

Sustainable Data Centers – A Position Paper

Bhuvan Urgaonkar, CSE, Penn State

It seems useful to classify data center sustainability related efforts into two categories: (a) ideas to reduce the energy consumption of existing or future data centers, and (b) ideas to align cost-efficacy – the primary objective of most commercial data centers – with sustainability-related goals. Much of existing research on sustainable data centers has been on (a). Numerous ideas related to making computer systems more "energy proportional" or improving the energy overheads of non-IT supporting infrastructure for reliable power delivery and cooling (improving the "PUE") fall into this category.

It is undeniable that efforts of class (a) will continue to be important for the evolution of sustainable data centers. One significantly novel set of options may arise if we focus (in addition to the more well-established way of looking at the problem as one of reducing the energy consumption of a given data center) on dampening the rate of growth of data centers. Briefly, this amounts to designing and operating data centers that operate at significantly higher levels of resource utilization than are prevalent today. This likely raises challenges for novel data center resource management, system software/middleware, and even novel ways of resource procurement for "tenant" applications of data centers. Most software today is designed with a view of machines and networks (virtualized) with relatively fixed capacity for most (if not all) resources. Can we design systems and (some subset of) application software that can deal with much larger uncertainty in the resource capacity of their underlying computers and networks? Interestingly, NSF's cloud systems, due to their much higher resource utilization than "real-world" data centers, might offer ideal platforms for testing the efficacy of such ideas.

It is our view that, efforts of type (b) have received relatively less attention from the research community and present us with particularly promising opportunities. Problems of this type would be concerned with the design of pricing and other incentive mechanisms. How can a data center, obviously interested in improving its profits, be motivated to operate in a sustainable way (e.g., in terms of being attentive to coincident peak occurrences on the grid or in terms of being sensitive to renewable vagaries)? Such issues, well-studied for more well-established consumers of energy, appear little studied for data centers. In fact, most data center power cost optimization/control literature seems to work with goals that do not incorporate intricacies of real-world utility tariff schemes (e.g., many variants of peak-pricing, tiered pricing, etc.) In addition to pricing/mechanism design between the utility and the data center, would it make sense for the data center itself to design pricing for its tenants to encourage behavior amenable to overall sustainability? These issues would require collaboration among experts from diverse disciplines. It seems that such collaboration could be fostered by the NSF through programs similar to CyberSEES. It also appears that if practitioners/researchers from industry (both data centers and electric utility) could share some information on the state-of-the-art in this area with their academic colleagues, this would serve as a very useful starting point for work on this exciting area.

Kim Hazelwood, Ph.D.
Principal Scientist / Director
Yahoo Labs – Systems Research Group

How Should Academia and Industry Collaborate?

First, let me say that collaboration between industry and academia is absolutely crucial to addressing the challenge of designing sustainable data centers. Academics have the freedom to embark on multi-year efforts toward developing long-term possibly risky solutions. Yet, academics cannot work alone on these endeavors. It's infeasible (and misguided, and wasteful) for any university to develop a 100K-1M node datacenter to allow academics to refine and explore their solutions. It's also infeasible to envision a university-based datacenter running actual production workloads.

At the same time, it is unrealistic to assume or expect that industry alone will sufficiently explore the design space. Companies are in the business of maximizing profits. This means getting things to work quickly and moving onto the next big challenge. It means that caring about energy is primarily a business case when it comes to power bills, and marginally as good press/advertising (but don't overestimate the importance of the latter!) The bar is set such that your goal is to (a) not break anything (b) get the system to perform better than it did before. (Note that I didn't say: get it to perform the best possible.) Those final tweaks are rarely worth the high cost of the engineer's time.

Clearly, neither domain is well suited to solve the problem alone. So how should the two collaborate? I think the current model that most faculty follow is to build a tight collaboration with one or more companies. The faculty members can send their students on summer internships at these companies, where aside from developing crucial engineering and design skills, the students can have access to the scale of challenges and breadth of data that is only present in industry. The companies agree to allow the interns to do the more risky or long-term research, and not to steal the students prior to graduation. In exchange, they get what amounts to a low-risk 3-month interview with a potential future hire. Good will results, coming from both sides. This internship-with-full-but-temporary-access-to-data model, combined with the unrestricted faculty research gifts that most companies provide to their close collaborators is a great model. I wouldn't change a thing, other than to encourage **everyone in academia** who works on data center research to ensure they build these close ties with industry.

(Note that this funding is completely complementary to the funding provided by NSF. The scale of industry gift money is typically an order of magnitude smaller than what the academics need, and what NSF typically provides.)

Kim Hazelwood, Ph.D.
Principal Scientist / Director
Yahoo Labs – Systems Research Group

How Can NSF Best Leverage Its Existing Investments in Cloud Testbeds?

I wanted to comment on this particular question mainly because I want to disagree with the premise of the question. First, I'll provide some background. While I was at Google, I did extensive analysis of actual data center and cloud workloads (which are quite different, by the way) and compared the characteristics to that of the standard benchmark suites available to academics. The differences were staggering. As an example, actual datacenter workloads consisted of applications whose binaries were several orders of magnitude larger than the largest of the SPEC benchmarks. This had significant implications for system design that spanned the cache and memory hierarchy as well as the energy-efficiency algorithms. This finding and the many other differences we uncovered between actual workloads and those of SPEC, CloudSuite, and other testbeds were published in ISCA 2015. Our takeaway message: we need new investments in workloads because none of the standard benchmarks are remotely close to reality.

The initial question asked was how NSF can leverage its existing investments in Cloud testbeds. For one thing, I do not know exactly what these investments are (which is already a bad sign). Sadly, I suspect that whatever investments NSF has made may be efforts to abandon, and a new plan should be developed looking forward.

Here's an example of something that will not work: Asking Google to release its internal benchmark suite to the public. This will never happen for several good reasons. The biggest reason is that the inputs to the benchmarks include actual user data, which legally cannot be released for obvious privacy reasons. A second, obvious reason is that Google has no incentive to release the secret sauce code and risk theft. Above all, this is not Google's problem to solve.

Whose problem is it then? I think that by leveraging a close collaboration between academics and industry, funded primarily by NSF, we can collectively solve the problem. Academics can create synthetic workloads that are provably similar to real production and cloud workloads by physically comparing these characteristics in-house at companies.

But of course, having access to the workloads is only half of the problem. Academics who wish to leverage these cloud workloads must also have access to data center-scale machines. Hence, we shouldn't see these public workloads as a way for academics to work in isolation from industry. Collaboration will always be key.

Architecture and Services to Enable Energy Reused Cloud

Jie Liu
Microsoft Research
One Microsoft Way, Redmond, WA 98052
jie.liu@microsoft.com

Energy reuse is a concept of using the same energy to achieve multiple goals, such as computing and heating, at the same time. One such example is *data furnace*, which suggests to use a cluster of cloud connected machines for home heating [1]. Overall, US households use 116 Billion kWh for home heating. By piggybacking 50% of it, the IT industry can double its size without increasing energy footprint. Recently, a few commercial companies, such as Nerdalize, CloudAndHeat, and Clise Property Management, are executing on this idea and have designed heating appliances with various form factors and capacities. They primarily operation on enterprise computing jobs and do not scale to the cloud.

Most current Internet and cloud services, such as search, advertisement, and e-commerce, are not suitable for this highly distributed architecture. They are usually data intensive applications, requiring large clusters connected by high performance data center network or fast response time to global users. The last-mile consumer network speed and the weak consistency provided on potentially unreliable servers, will cause poor quality of service. However, the explosion of Internet connected sensors and corresponding cognitive tasks is changing the nature of cloud computing. These are information extraction applications, such as video surveillance, scene understanding, and machine translations, which heavily rely on machine learning. Machine learning applications usually have two phases. The training phase is both data and computation intensive, requiring traditional highly integrated data center architecture. The second, running phase, on the other hand, needs little amount of input (e.g. sensor data) and output (e.g. labels), but intense computation on a single set of closely-knit machine. For example, classifying a single frame in a video stream requires 3GFPLOs. More complex tasks like object detection or semantic segmentation need three order of magnitude more companion [2]. Modern GPUs achieve 3TFLOPs per second at 235W. Heating a typical US household requires 0.9GJ per month, which converts to about processing 115 hours of video at 30 frames per second. Thus, the running phase is ideal for energy reuse scenarios like data furnaces.

I propose to investigate the architecture, services, and ecosystem to enable the energy-reused cloud (ERC). Some research problems include:

- Economical models, pricing, and incentive mechanisms to encourage energy reuse.
- Coordination between centralize and distributed clouds, including service architecture, provisioning, allocation, replication, and backup, machine learning model synchronization and updates, data security and privacy preservation, etc.
- Workload scheduling to match heating demand with computational tasks. This may include a global overlay network that can route source data to the right region and local schedulers that match demands with the right server.
- Server designs and control systems to optimize for heat transfer and low profile, as required by heating appliances.

[1] Jie Liu, Michel Goraczko, Sean James, Christian Belady, Jiakang Lu, Kamin Whitehouse. "The Data Furnace: Heating Up with Cloud Computing," HotPower 2011.

[2] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," CVPR 2014.

Position Statement for:

Dale Sartor, PE
Staff Scientist/Engineer
Lawrence Berkeley National Laboratory

May 18, 2015

*** What are the key research challenges for continued progress on sustainable data centers in the next 5-10 years?**

1. Funding

Some ideas (could use further thought/discussion):

1. Electrical distribution improvements
    a. Advanced UPS options
        i. Internal vs. external
        ii. Hi reliability auto bypass,
        iii. Examining the need for redundancy/Optimization of redundancy
    b. Direct Current power (eliminate conversions and improve reliability)
        i. Direct use of renewable DC generation
        ii. DC powering of IT equipment
        iii. DC powering of motors, lighting, communication
    c. On site generation and backup options
2. Cooling
    a. Address contamination/corrosion issues associated with outside air cooling
    b. Increasing operating temperatures and humidity range at low cost and with increased reliability
    c. Bringing advanced warm liquid cooling solutions to standard servers designed for multi generation (refresh)
    d. Novel passive heat removal (e.g. carbon nanotubes)
3. IT
    a. Develop productivity metrics
        i. Tools to aggregate utilization information
        ii. Standardize compute, storage, network energy efficiency metrics
    b. Redundancy in network not in data center
4. Human (and institutional) factors that inhibit change

*** What are the key operational challenges in building and managing sustainable data centers?**

1. Lack of knowledge/training/misinformation
2. Risk aversion
3. Poor communication between key players (e.g. facility and IT managers)

*** What mechanisms can NSF foster to enable researchers in academia to obtain design information, traces, measurements, etc. from operating data centers to facilitate continued research?**

1. Data center owner operators are more than happy to collaborate with researchers if they receive value and are not put at risk
2. Establish centers of expertise and provide industry access

*** How should academia and industry collaborate?**

1. Must be benefits to both sides. Typically benefits to industry are “one off,” so their out-of-pocket costs need to be covered

*** How can NSF best leverage its existing investments in cloud testbeds to these ends?**

1. Need description of cloud testbeds.
2. NSF could significantly influence data center sustainability via their normal grant making process (which typically does not encourage best practices in data centers).

NSF Workshop on Sustainable Data Centers

Cullen Bash
Hewlett-Packard Laboratories

Position Statement

The energy consumed by information technology continues to increase at virtually every level of the technology hierarchy, from the device to the cloud. The end of Dennard Scaling and the slowdown (and approaching end) of Moore's Law has resulted in significant increases in chip-level power consumption as over-clocking to improve performance is increasingly being seen as an alternative to scaling, even though the impact on power management is non-trivial. At the other end of the consumption spectrum from the device is the growing global appetite for IT. If the *public cloud* were a country, estimates indicate it would be one of the top five energy consumers globally. Although progress has been made over the past decade on incorporating sustainable practices in the design and operation of IT equipment, the results to date have been modest and primarily targeted towards thermal management, and power generation and delivery. There remain significant opportunities for academia, industry and government to partner in research and development related to sustainable data centers. Indeed, given that the IT industry is reaching a crossroads with the end of Moore's Law approaching, the era of Big Data emerging, revolutionary advances in computer architecture (particularly around non-volatile memory, photonic interconnects, and increasing levels of on-chip functionality integration), and increasing public concern over the environment, one could argue that the time for significant investment in research directed toward sustainability within information technology has arrived.

If sustainable data centers are defined as information technology installations that minimize resource consumption (environmental *and* computational) in light of geographic location, a number of key research and operational challenges worth pursuing within the next decade can be identified. The following is a non-exhaustive list:

Research Topics

- Local sustainable energy production;
- Temporal and spatial IT workload management;
- Comprehensive sustainability metrics for data centers;
- Energy proportionality;
- Energy reuse mechanisms.

Operational Challenges

- Integrated Management of facilities and IT systems;
- Availability and suitability of local environmental resources for power and cooling;
- Thermal and power management at 100+ kW/rack (also a potential Research Topic);
- Baseline local power generation from sustainable means (e.g. local microgrids);
- Cost effective energy reuse techniques;
- Installation cost for sustainable technology;
- Methods for reducing operational complexity in light of sustainable technology.

A number of these areas are large and complex and will take collaboration between industry, academia and government to realize success. Two of the more significant challenges to conducting research in this space are: 1) obtaining sufficiently anonymized workload data from a variety of sources and data types, and 2) obtaining access to data centers to conduct experiments. There may be an opportunity for an organization like the NSF to assist with creating a standard for the anonymization of data generated within data centers. Additionally, NSF has access to data center cloud testbeds that could possibly be leveraged as experimental test environments.

Mike Marty
Google
mikemarty@google.com

With the end of Dennard Scaling, increasing computation-per-watt in data centers is challenging and demands new cross-disciplinary approaches rather than riding silicon advances. While transistors are not becoming much more energy efficient, fortunately there exists opportunity to make better use of them.

First, the utilization of data centers is too low. Improving utilization can result in perhaps a one-time ~2-4x increase in computational efficiency. One reason for low utilization is that serving latency is impacted at higher utilizations, and these effects compound in distributed serving systems. With 10 subtasks, a one-in-a-thousand chance of suboptimal process scheduling will affect 1 percent of requests (recall that the request time is the maximum of all subrequests), but with 1,000 subtasks it will affect virtually all requests. Distributed applications can often take action to deal with stragglers if notified early enough. But our layers of abstraction lack sufficient feedback mechanisms-- requests can sit in various queues well past their deadlines: network queues, cpu scheduling queues, application work queues. Moreover machine-local techniques for energy reduction, such as processor power states, can actually be counterproductive in a distributed system.

Second, better efficiency can come from better application performance. However increasing performance and careful tuning takes significant software development cost, which can dominate the technical resources of a company. Improving application execution efficiency, while maintaining software development costs, can result in up to a ~5-10x (or more) improvement in overall computation efficiency. Performance is lower than it could be due to many reasons, many of which center around easy-to-use abstractions. A large fraction of software still pays performance penalties through the use of non-compiled runtimes (at Google, the heaviest lifting is often done using compiled languages like C++ and Go). Multicore threading models can also be extremely inefficient. As an example, applications often dispatch work to a large pool of worker threads because it does not know how many cores the kernel scheduler has allocated to it at any given point in time in a multiprogrammed mixed-use scenario (resulting in excessive context switching and loss of locality). Likewise kernel schedulers have the conundrum of not knowing how long a thread will execute before blocking-- should a runnable thread queue for the current CPU that has a hot cache or run immediately on a different core with a cold cache? Some of these problems are solved with space-sharing of cores with application-driven threading, rather than fine-grained time sharing with kernel-driven threading, but this presents its own set of challenges and is difficult to tune for.

Third, the premise of “wimpy cores” for greater energy efficiency cannot be unleashed without avoiding Amdahl’s Law bottlenecks that always seem to crop up. Unleashing wimpier cores can improve energy efficiency by ~2-4x. However replacing brawny cores with wimpy cores may require strong scaling (e.g., increased parallelism), and in warehouse-scale computers, this must be done across machines rather than constraining the parallelization problem to a single machine. Key to stronger scaling across machines is lower communication latency (particularly the tail latency). Yet low latency communication in conjunction with energy-efficient threading models is an unsolved problems (for the same “killer microsecond” reason discussed next).

Finally, general-purpose processors as we know them today are terribly inefficient compared to fixed-function hardware. Moving away from general-purpose processors towards fixed-function hardware can increase efficiency by 10x to 100x or more. With the premise of Dark Silicon, many researchers believe on-die accelerators is the key to unlocking significant increases in computational efficiency. The orchestration and use of accelerators will likely be handled by general-purpose processors, but there remains an open question as to how this will happen. On one hand, accelerators can be implemented as "complex" instructions that are invoked as part of the sequential execution of instructions. Techniques like simultaneous multithreading can use other, different functional units for different threads. On the other hand, accelerators can be treated as an asynchronous IO device. If the accelerator computation takes on the order of microseconds, computer systems lack good support for dealing with such "low latency" IO devices. I refer to this problem as the "killer microsecond" wherein computer system design has focused on the extremes (nanoseconds and milliseconds), and microsecond-scale IO latencies have poor solutions. Existing hardware threading mechanisms (such as SMT/hyperthreading) have sufficient contexts only to tolerate devices on the order of hundreds of nanoseconds, and software threading incurs too much overhead for microsecond-scale accelerator interactions.

Position Statement for NSF Workshop on Sustainable Data Centers

Qiang Wu, Facebook Inc.

I have been working on Facebook infrastructure since 2009 and have been focused on data center performance and power management in the past 3 years or so. The following statement reflects my personal view.

Facebook has over 1.4 billion users as of today. Large scale data centers are needed to support user applications. On one hand, we want to make sure we meet the user application requirements even under worst case scenarios (e.g., under a disaster recovery). On the other hand, for efficiency and sustainability, we want to minimize the total amount of resource used, including total number of servers, data center facility and space, and the mount of energy usage etc. However, to achieve this goal, there are several key challenges which need to be addressed.

(1) Intelligent power over-subscription: this allows us to aggressively and safely over-subscribe servers to fully utilize existing data centers and power infrastructure. So we can support user growth without building new unnecessary data centers. This would involve some smart workload/power modeling & prediction techniques so it will have just enough power in each data center under worst case scenarios (e.g., if one of the data centers is lost due to a natural disaster).

(2) Dynamic resource management for heterogeneous computing platforms: modern applications like Facebook have all kinds of workload (latency sensitive vs. latency tolerant, CPU bound vs. IO bound, etc). The computing platform can be very heterogeneous (different configuration, size, architecture, etc). It is challenging to manage such resources in a large scale in order to minimize the number of servers needed while meeting all user application requirements.

(3) Ensure reliability and safety: pursuing efficiency and sustainability should not be at a cost of reliability and safety. For example, it is not acceptable in practice if the optimization leads to increasing of hardware failure rate. For another example, the power safety should always be ensured (power outage is not an option). So, a real-time power monitoring and peak power management system (.e.g. power shifting, DVFS) will be needed to ensure power safety even in emergency cases.

(4) Operational complexity: all optimizations should be transparent and should not increase the operational complexity. Otherwise, the operational/human cost may offset any efficiency/sustainability benefits, and make it infeasible in practice.

Some of the key challenges are also what my colleagues and I have been trying to address at Facebook. We welcome collaborations with academic researchers to jointly attack these problems. Through our academic collaboration program, faculty and students can access Facebook data and system to conduct more advanced research. We hope NSF can help to facilitate more such kind of collaborations. This workshop, bringing together people from both academia and industry, is a great starting point.

Moving Sustainability Data Over the Last Mile
Dr. Christopher Stewart, The Ohio State University

An end user should be able to view and manage her energy and carbon footprint for the Internet services she patronizes. Armed with sustainability data, users can make sustainable choices in their web browsing activities. Competitions between users vying to “be greener” than each other will also improve the overall sustainability of the field (Carlson, 2010). Prior work has shown that green hosting sites (i.e., web hosts that expose their carbon footprint) reap sustainability-conscious end users while reducing carbon emissions (Stewart, 2013). Releasing sustainability data to end users consistutes the *last mile* in sustainable computing. In comparison, research on the first mile, i.e., improving sustainability within data centers, has been famously successfuly and continued investment on water efficiency, efficient parallelism (e.g., via FPGA) and renewable-aware algorithms are needed. However, the last mile presents a wide range of issues that thus far have prevented sustainability data from exiting the data center. Some of the challenges include:

Privacy violations: The term end user may describe a business using IAAS clouds or a human being accessing her social networking site. In both cases, sustainability data can be misused to infer private, sensitive data. On IAAS clouds, energy footprints correlate with cluster size---a metric occassionally used to measure the financial health of company. Similarly carbon footprint data observed over time significantly narrows the geographic regions where queries could have been processed. If Internet services process queries at sites nearest end users, carbon footprints can be used to track people. At the very least, these concerns demand careful authorization to access sustainability data, but Internet service managers may also rightfully worry about information released in when sustainability data is aggregated.

Introducing Security Holes: Energy footprints can reflect low-level cache access and sharing patterns (Shen, 2013). Prior work has shown such data can be exploited in IAAS platforms for network intrusion (Zhang, 2011). Internet managers need theoritical results on the safety of releasing footprint data.

Unexpected Market Effects: Energy and carbon footprints will be aggregated by social network, geographic region, SAT score, and shoe size. Undoubtedly, some feature space will reveal significant differences. In response, users will adjust their traffic patterns. Internet service providers should worry that these adjustments could hurt revenue. For example, if an IAAS provider reveals that its largest customer runs unsustainably high carbon footprints, the response from end users may hurt the IAAS providers profits.

We propose a two-pronged research agenda. First, researchers should create infrastructure tools (e.g., key-value stores, databases and data-parallel platforms) that manage carbon and energy footprint at request granularity. These tools will mitigate the impact of secondary market effects, allowing services to react quickly to changing user demands. A number of groups have developed tools in this space. Some examples include, Green Cassandra, Green Slot, Green Load Balancing, and CADRE. However, these tools have not been adopted widely in practice. One approach for industry impact is to adapt these toos to spot markets. Spot markets have volatility similar to carbon-offset markets but tools that exploit spot markets can translate directly to cost savings today.

For the second prong, researchers should devise black-box approaches to model energy and carbon footprints. The risks of unexpected market effects will continue to parallyze Internet service managers. However, users can benefit from reasonably accurate approximations of footprints. We are devising a non-invasive method to infer resource usage by combining response time data and publically available data (Deng, 2014). Our approach uses independent component analysis. However, it is one solution in a large space to explore. The general problem of wisely pruning the space of potential underlying resource usage patterns that could yield a sequence of response times is both broad and challenging. NSF can substantially strengthen research in this space by releasing an API for energy and carbon footprint measurement in its soon-to-be released cloud. Benchmarks tested on the NSF cloud will provide empirical validation for the proposed models.

In conclusion, sustainability data faces an array of challenges in the last mile related to privacy, security and market concerns. The research community must provide fine grained mechanisms for services to incrementally release data and adapt to their footprint. Also, the research community should serve as a catalyst, releasing best-effort and approximate data on sustainability to move the field forward. NSF can aid these research directions by supporting spot markets and by exposing sustainability metrics in its cloud offerings.

Bibliography:

To come (the references above are legit, trust me.)

Position Paper – Research Challenges in Sustainable Datacenters

Ricardo Bianchini

Over the last ~15 years, large datacenters have benefited from computer technology and physical infrastructure advances that substantially improved their energy efficiency (and, as a consequence, their sustainability). For example, large datacenter operators were able to lower Power Usage Effectiveness (PUE) scores from around 2 to the range of 1.1--1.2 in many geographical locations via smarter design of the physical infrastructure (power delivery and cooling). Operators were also able to ride Dennard's scaling and leverage increases in CPU performance without a corresponding increase in energy consumption. In addition, progress in fast low-power CPU states has enabled servers to become more power-proportional. Despite these developments, there is room for improvement, as the datacenters' computational resources are often poorly utilized and technology advances are either gone (Dennard's scaling) or are starting to falter (Moore's law). However, significant new energy efficiency and sustainability advances will likely be harder to achieve. Though we are already on the path to achieving some of them, others will require overcoming some research challenges. Next, I discuss some directions that I think the community could pursue.

First, I believe that we should start thinking about sustainability more broadly than just (operational) energy efficiency. In particular, it is important to realize that the very production (and eventual disposal/recycling) of IT, power delivery, and cooling equipment impacts the environment, i.e. we should target end-to-end sustainability. This means that any research that targets reducing the number of servers, the power infrastructure, or the cooling infrastructure used by datacenters will make them more sustainable. In some cases, the increased sustainability may even come at the cost of consuming more energy during datacenter operation. These end-to-end sustainability tradeoffs are not yet well understood. Nevertheless, researchers have done extensive work on reducing the number of servers using software techniques, e.g. by leveraging spare resources to run (useful) batch jobs on the same servers as interactive services without affecting their tail response times. Certain hardware accelerators may also achieve such reductions, while impacting the environment less significantly than full-blown servers. The key in acceleration is to ensure that servers remain fungible and benefit multiple workloads, so they can be maximally utilized. There are research challenges in doing so as well.

Second, the community should start regarding techniques that reduce tail response times as effective means of reducing the number of required servers (and potentially improving energy efficiency, as static energy is also reduced). For example, a recent paper shows that leveraging spare cores to increase the parallelism of long-running search requests increases the throughput that each server can achieve within a target tail response time. However, whether to use the spare resources for tail latency reductions or batch workloads depends on the business value that can be accrued from each. For example, a batch data analytics workload may substantially improve the quality of the interactive service and, thereby, deliver high business value. Thus, the challenges are in identifying these server-reducing techniques and applying them carefully to manage the overall performance, revenue, cost, and sustainability tradeoff. Hopefully, much more can be done in this space.

Third, the community should continue investing in approximation techniques, especially those that can provide bounds on the accuracy of the approximate results. Though some applications may not be amenable to such techniques, approximations can dramatically improve the performance and reduce the amount of resources needed by many applications (again, potentially translating into large reductions in the number of servers required and the energy consumed by datacenters). For example, a recent paper shows that certain MapReduce-based data analytics applications can be performed with orders of magnitude fewer resources at the cost of small and bounded inaccuracies, by using system-level techniques couched in Statistics. Such large savings will likely be difficult to achieve in other settings or at other layers of the stack. However, approximations open up additional challenges in business models, capacity planning, etc.

Fourth, much more challenging issues remain open for investigation. For example, (1) we still lack effective metrics for assessing software efficiency other than big-O notation, i.e. we have no good handle on the "constants"; (2) the techniques we use for fault tolerance, tail tolerance, and high availability, e.g. replication and request hedging, tend to be energy-hungry; and (3) long-term demand forecasting and capacity planning for cloud services remain extremely difficult to do accurately. Finally, whether/when alternative/renewable sources of energy (e.g., solar, fuel cells) will become cost-effective for large-scale datacenters and an overall gain for the environment remains an open question. Breakthroughs in these directions could have a tremendous impact on sustainability.

Key Challenge of Continued Progress in Sustainable Data Center Research

Mike Ferdman
Stony Brook University - Computer Science
http://www.cs.stonybrook.edu/~mferdman/

It is evident that the data center infrastructure is of critical importance for the continued evolution and expansion of day-to-day computing. It is also well known that the power consumption of today's data centers already makes up a noticeable fraction of the world's total generated power. When coupled with the observation that keeping up with demand requires bringing online new data centers and new servers continually, a question of sustainability emerges: at what point will data centers across the globe house so many machines that the cost of keeping them powered on will become intractable? Unfortunately, despite overwhelming evidence that the current data center growth trends are unsustainable in the long run, there exists no research roadmap to pursue in the community. Identifying a potential roadmap is the key challenge that stands in the way of progress.

The HPC community has long ago established a trend where concrete goals are set and achieved. *Petascale* has been reached and roadmaps are debated for reaching *exascale*. The HPC roadmaps identify the targets for performance and the challenges that stand in the way of meeting these targets. Data center research does not have such publicly acknowledged and accepted roadmaps. What's worse, it appears that researchers are primarily pursuing hot questions as dictated by the current data center operators rather than by a vision of the future.

Data center research requires well established roadmaps and goals. No doubt these are more difficult to devise than HPC-style goals. Data centers are more diverse, faster changing, and have a significantly shorter history to draw upon. More than likely, no single roadmap will ever be established, as perfect agreement among the various stake holders about what the future holds is unlikely to be achieved. However, even if there are multiple competing outlooks, having credible targets can enable researchers to focus on problems that actually address long-term data center sustainability. Until such goals are established, it is entirely unclear if the research being conducted is making significant strides toward sustainable data centers or simply shaving a few percentage points of energy costs for today's operators which, in the long run, will be largely irrelevant.

Position Statement - 2015 NSF Workshop on Sustainable Data Centers

Luiz André Barroso, Google Inc.

I would like to propose one research challenge for continued progress in sustainable data centers, but will start by removing some areas from our list. For the last decade a fair amount of effort has gone into reducing energy overheads of cooling, power distribution and conversion, as measured by the popular PUE metric. Modern facilities went from 2x overheads to just about 10%, and while improvements are still welcome it is clear that this is no longer an area with great opportunities for research impact. Likewise, research investments in job and server management schemes for migrating jobs in periods of low utilization in order to turn off servers are unlikely to pay off. We have learned that such schemes pay off only for a narrow set of jobs that tend to be stateless, and that are trivial to migrate or kill/restart, while being generally impractical for more complex jobs.

Interpreting sustainability as curbing the resource budget needed for accomplishing a given computing task, be it CPU cycles, storage bytes or joules, the biggest opportunity for large gains lies in building increasingly efficient distributed systems. A good software engineer can speed up almost any program by 20%, and that can translate 20% less compute equipment and energy consumption (particularly if the underlying equipment exhibits some degree of energy proportionality). A particularly difficult sustainability challenge is how to make better use of already deployed equipment, as evidenced by the rather depressingly low utilization levels reported by typical data centers. Facilities are often underutilized because it is hard to achieve high utilization while meeting service level guarantees to software services. In The Tail at Scale, Jeff Dean and I describe some aspects of this problem and present one example solution for distributed storage systems. Further research on software & hardware schemes that make it safer to share data center resources while respecting performance guarantees could greatly improve the effective computing capacity of large data centers.

Position Statement for 2015 NSF Workshop on Sustainable Data Center

Xiaorui Wang, ECE@The Ohio State University

1. Key research challenges for sustainable data centers in the next 5-10 years?

One of the key research challenges for sustainable data centers is how to maximize their performance under their stringent power/thermal constraints. We believe a methodology that can enable a data center to boost its computing performance for bursty workloads is to temporarily turn on more cores, which are supposed be off in the era of dark silicon due to thermal constraints. Recent studies have proposed Computational Sprinting, which allows a chip to temporarily exceed its power and thermal limits by turning on all its cores for a short time period, such that its computing performance is boosted for bursty computation demands. However, conducting sprinting in a data center faces new challenges due to power and thermal constraints at the data center level, which are exacerbated by recently proposed power infrastructure under-provisioning and reliance on renewable energy, as well as the increasing server density. We demonstrate the feasibility of this proposed methodology (called *Data Center Sprinting*) by analyzing the tripping characteristics of data center circuit breakers and the discharging characteristics of energy storage devices, in order to realize safe sprinting without causing undesired server overheating or shutdown. We have evaluated a prototype of Data Center Sprinting on a hardware testbed and in datacenter-level simulations. The experimental results show that our solution can improve the average computing performance of a data center by a factor of 1.62 to 2.45 for 5 to 30 minutes.

2. Key operational challenges in building and managing sustainable data centers?

One of the key operational challenges in managing sustainable data centers is how to minimize their cooling energy costs for sustainable cloud computing, while preventing undesired thermal emergencies. A major reason for data centers to have excessive energy consumption is the inefficient operation of their cooling systems (e.g., a set of Computer Room Air Conditionings (CRAC)), which can account for up to half of their energy consumption. Because of the lack of visibility in the operating conditions of a data center, the cooling systems often have to use unnecessarily low temperature set points in order to reduce the risk of creating any hot spot, resulting in excessive cooling energy consumption. To this end, we propose a holistic framework for multi-scale thermal monitoring, prediction, and management in data centers for sustainable cloud computing. In sharp contrast to existing work based on simplistic physical models, a key advantage of our solution is its reliance on rigorous *Computational Fluid Dynamics* (CFD) analysis as a theoretical foundation to capture the thermal dynamics of a data center in various overheating scenarios. We then design online detection and prediction algorithms, based on the offline CFD analysis, to monitor and predict in real time, the evolution of ambient temperatures in data centers. We believe the proposed framework can transform the way data centers are built and managed by helping break the long lasting barriers causing hot spots and high energy consumption. The outcome of this study can enable highly sustainable data centers with accurate thermal monitoring, prediction, and control.

3. How to enable researchers in academia to obtain data from operating data centers?

This is definitely an important topic for discussion. In our past ten years of doing research on data center energy efficiency, we often face the difficulties of getting real data from operational data centers. Although we managed to find some real traces online, the time scale and granularities of such traces are often insufficient for our studies. As a result, we often got criticized by reviewers during our paper submissions. The lack of data from real data centers has put many researchers in disadvantage, because some other researchers who have close connection with companies, like Google and Facebook, can get such data to publish in top-tier conferences. If we could have a standard way to get such data and share them among all the researchers, it might help make better research progress in this field.

4. How can NSF best leverage its existing investments in cloud testbeds to these ends?

NSF may put such testbeds online for all the researchers to share, like Emulab: https://www.emulab.net/

A Few Underexplored Areas of Sustainable Data Centers

Shaolei Ren
Florida International University

In the past few years, there have been numerous efforts dedicated to improving data center sustainability. While the progress to date is undeniably encouraging, the existing research has been primarily focused on owner-operated and dedicated data centers, like Google, which only represent a small fraction of the entire data center industry, in terms of both numbers and power demand. In this position statement, based on the ongoing research in my group, I will outline a few research areas along with their challenges that have been largely underexplored and, if still left unaddressed, would become a major hindrance for sustainable evolution of the data center industry.

- **Enabling Coordinated Power Management in Multi-Tenant Data Centers.** Multi-tenant data centers, like Equinix and Digital Realty, have been rarely studied, but they are very common in practice, consuming nearly as five times energy as Google-type data centers combined altogether. In a multi-tenant data center, the operator is mainly responsible for facility support (e.g., power and cooling) while individual tenants manage their own servers without coordination, thus creating many new challenges for the adoption of existing resource management-based sustainability practices. For example, how to coordinate tenants' power management to "follow" the intermittent on-site renewable energy for reducing data center's reliance on carbon-intensive electricity? How to optimize the operation of data center facility and tenants' IT equipment as a holistic system towards sustainability?

- **Greening Data Center Demand Response.** The traditional negative view that data centers are purely energy hogs is changing, as data centers have been playing an increasing role in emergency demand response (EDR) program, which is one of the widely-adopted demand response programs serving as the last line of protection against cascading failures in power networks. Nonetheless, data centers typically participate by turning on their on-site backup diesel generators, which are neither cost effective nor environmentally friendly. Fortunately though, data centers' huge yet highly flexible power demand, if successfully managed to reduce to a certain level as requested by grid operators, may become an ideal substitute of diesel generation for EDR. Key challenges, however, are the performance consideration. Furthermore, towards greening EDR for a multi-tenant data center, additional challenges will emerge, such as how to coordinate tenants' participation in EDR as they share the data center facility as an integrated system?

- **Achieving Water Sustainability in Data Centers.** Data centers are very "thirsty" and may each consume millions of gallons of water for cooling every day. Unfortunately, data center's huge water footprint has been neglected by the research community, despite the emergence of water sustainability as a critical concern. Furthermore, simply adopting the existing techniques for minimizing energy or carbon footprint may not reduce, and in some cases even increase, data center's water footprint. While advanced cooling systems, like air-side economizers, are useful for reducing data center's water footprint, most data centers are located in places where installation of such cooling systems is not feasible and instead chiller-based cooling systems combined with water-intensive cooling towers are the only option. In view of extended droughts that are becoming a norm in many parts of the world, how can data centers take the leadership to achieve water sustainability without compromising other important aspects such as performance and energy?

Position Statement

Hamed Mohsenian-Rad, Assistant Professor, ECE Department, University of California at Riverside

My interest in data center sustainability research is mainly at systems level with focus on understanding and shaping the interactions between data centers and power grid. I seek to answer questions such as: What are the unique aspects of load flexibility in data centers compared to other flexible loads in power systems? How can data centers benefit from direct interactions with distribution system operators and transmission system operators? What are the best of local energy resources (energy storage, renewable generators, etc.)? Accordingly, my approach to data center sustainability is rather “global”, to help data centers contribute towards improving the sustainability of the electricity grid as a whole, in order to meet data centers’ social responsibility as major energy consumers.

In my experience, one of the shortcomings in the current research on sustainable data centers is the gap between those who tackle the problem at systems / macro level and those who tackle the problem at device / micro level. As a result, the research efforts at systems / macro level often lack practicality by overlooking the complex nature of computing processes, while the research efforts at device / micro level often miss the diverse opportunities that are available to data centers to achieve sustainability. NSF can play a leading role in addressing this shortcoming by encouraging more cross-disciplinary this field with teams of computer scientists, computer engineers, and power engineers.

Sustainable Data Centers – Economic models capturing business cost and benefit

Karthick Rajamani, IBM Research

Datacenters are often just a portion of businesses with various factors dictating their operations besides efficiency and sustainability. Many decisions on IT equipment, cooling and power distribution infrastructure are made based on the application(s) to be run and the impact to the business from availability, responsiveness etc. and not based on efficiency considerations. But these decisions impact the efficiency of the datacenter. The most effective method to get datacenter operators to address efficiency/sustainability is to establish a direct, economic relationship from efficiency/sustainability to business value. Generic models for this would be a starting point but the models will be able to influence practice only when made readily consumable/customizable for the specific datacenter's business and operations. Such models would need to capture the increase in (usually acquisition) cost, if any, for any efficiency optimization or sustainability-enhancing measure as well as the reduction in (usually operational) cost from their adoption. As these can change with time, maintaining a living model of their relation to the total cost of acquisition (TCA) and total cost of ownership (TCO) is necessary.

Many datacenter operations are fragmented among different organizations or units within the same business or even among completely different businesses leveraging each other for different services/revenue and together make up the totality of a functioning datacenter. As an example, consider company A providing the real-estate, cooling and power provisioning infrastructure to company B renting them to operate a server hosting service. Company B's inclination to choose energy-efficient IT equipment and management practices is directly dependent on what fraction of its operating costs the efficiency aspects impact. Depending on the terms surfaced to company B (whether it pays for some, a portion or none of the energy consumed by the IT equipment) it may or may not care about the efficiency aspects of the IT equipment. Optimizing the overall efficiency of the datacenter would require cooperative behavior from both A and B for which each has to be able to appreciate the economics of efficiency/sustainability for their individual businesses or units.

As datacenters grow larger with 100K or more servers and 10s-100s of MW of operating power they need to better understand the complete flow of materials and energy into/from the datacenter. The models to drive efficiency and sustainability would need to factor in anticipated demand for computing translated to materials and energy, when and at what threshold major investments are needed, when smaller investments for operational optimizations would suffice and when larger investments for efficiency improvements makes for better economics and when/where/what broader integration with other operating entities or factoring of other considerations would make economic sense for sustainability, e.g., use of co-generation, exploitation of local weather conditions, use of renewables – energy and materials.

To summarize, advancing sustainable datacenter mission needs strong focus on development of
1) Continuous, living, customizable model for impact of operational efficiency on the overall economics of running a datacenter within the business that it is a part of.
2) Efficiency model that articulates the cost and benefit to each participant in the datacenter operation.
3) Complete economic model for the flow of materials and energy into/from the datacenter that can factor in anticipated demands in computing, materials, energy and support for its integration into corresponding models for extra-datacenter entities/operations such as co-generation, power generation and distribution, and materials recycling.

A joint, collaborative approach between academic institutions, industry and relevant governmental/non-commercial organizations is essential for the creation of such a model/framework for datacenter efficiency-and-sustainability economics.

In the past ten years there have been huge improvements in data center efficiency. Large amounts of waste in cooling, air handling, and power conversion has been addressed and continued process scaling has made data centers much more efficient. However, the next ten years of improvements will be substantially harder to achieve.

Moore's Law continues to provide increased silicon area for processors; however there has been a substantial slowing in the scaling of power and performance of DRAM. Ideally, as DRAM capacity was increased so too was the bandwidth to access it, but the reverse is actually true today. To guarantee signal integrity, as more capacity is added to a system, the DRAM interface has to run at lower frequencies and thus bandwidth. In addition, the power required to access these DRAMs is remaining largely constant, thus increasing their contribution to total system power.

These trends will limit the efficiencies that can be extracted from future server systems. While process geometry scaling will allow for denser processing, providing enough capacity and bandwidth to sustain that scaling will be a large challenge.

Traditional DRAM scaling has focused solely on the cost per bit. While, novel DRAM interfaces (e.g. HMC, HBM) have addressed half the issue, they have limited capacity at substantial costs. There are a number of areas where improvements could be made to provide increased capacity and sufficient bandwidth which range from new memory technologies that improve density and cell-size, energy-proportional buses that don't have large static power components, and better utilization of memory capacity and bandwidth by the processor that reduce the external bandwidth and capacity needs (caching, smarter software, compression).

Like many other areas worthy of research, one of the difficulties in developing these technologies is a lack of representative benchmarks and methodologies to make the evaluation tractable. Much of this work must be done holistically at the system level where components are optimized together and cannot be effectively done in isolation. Without representative benchmarks, any research into more effective caching, tiered memory, and similar areas will be difficult to evaluate. Similarly, methodologies must exist to make the evaluation of these technologies tractable. There is a huge disparity between the typical memory capacity of real systems (100s of GB) and that of the tractable systems that are evaluated today (1s of GB).

Energy Storage: The Keystone for Sustainable Datacenter Power Management

Anand Sivasubramaniam
The Pennsylvania State University
Email: anand@cse.psu.edu

The mismatch between a datacenter's power demands and the supply/distribution side restrictions, is a key challenge to the sustainability of datacenters from the energy, power, carbon emissions and cost perspectives. Given the large power draws, a datacenter in a given geography has a tremendous consequence on the number/capacities of power plants that need to be provisioned, especially if its peak demand periods coincide with the peak demands of other consumers in that region. Though deterministic in their generation capacity to meet this demand, thermal power plants can have a high carbon footprint. Opting for renewable power, on the other hand, makes datacenters very susceptible to associated vagaries of generation capacities. The mismatch also impacts the capital and operating expenditures of datacenters. The datacenter has to often support a peak load that also coincides with peak tariff periods of the day. Further, the capital cost of the power infrastructure is proportional to the peak power draw, that is rarely (and possible never) drawn.

Traditional computing knobs – server and device shutdown, dynamic voltage-frequency scaling, workload migration within and across datacenters – have been extensively explored to control the demand for these purposes. Only recently, has energy storage been studied for demand-response within a datacenter. By storing energy during excess supply capacities, it can be harnessed during demand excesses for addressing the mismatch. To date, much of this work has tried to leverage lead-acid batteries of existing UPS devices for such demand response. Such work suggests re-purposing these devices, originally intended to be called upon for rare power outages, for demand-response. Rather than engineer a solution with existing devices in the datacenter, we need to investigate more fundamental issues related to provisioning energy storage for datacenter Demand-Response:

What/Where? Instead of trying to re-purpose what is available, can we customize energy storage for datacenter demands, given that it is becoming a large enough industry? UPS units have simply adopted the economic sweet-spot of a technology (lead-acid battery) available in the market based on power/energy needs. They are sized more for power, since they need to handle only a few minutes of load before generators kick-in upon an outage. Decoupling the energy and power needs, and decoupling the usage (backup vs. demand-response) can open possibilities for more efficient energy storage provisioning. Further, there are technologies (e.g. flow batteries) that even allow independent sizing for the power and energy needs. It is essential to figure out which technologies to use (including combinations), and where to use them (server, rack, PDU or datacenter layers) specifically for datacenter demand-response. Traces of real power demands at each layer is needed to undertake such a study.

When/How? Even if present, energy storage in the datacenter is treated as a ``non-computational'' entity, and is managed by its own control loop. This can again make its usage sub-optimal. Making energy storage as one more resource (just as CPUs, memory and disks) that is explicitly managed introduces several new challenges and opportunities. Software could leverage rich APIs to bank a joule or select which watt to tap into (from utility power or power from the energy storage device including specifying which such device). Apportioning and managing the watts/joules, including those in the energy storage devices, amongst different co-existing applications can offer much more efficient ways of meeting performance SLAs within the constraints posed by the electricity supply/distribution.

Notes on Sustainable Data-Centers

Dahlia Malkhi, Ravi Soundararajan, Michael Wei
VMware

VMware's software runs in data-centers across the world, and not surprisingly we see virtualization infrastructure as key in data-center sustainability. Moving forward, we identify need and opportunity for improvement along three dimensions.

One dimension includes the continued enhancement of virtualization technology and tools surrounding it. Our company makes on-going effort towards improving resource utilization in a number of directions which bear direct relevance to sustainability.

The second dimension is democratizing the virtual OS infrastructure. This may sound contradictory to the commercial viability of companies like VMware, but in fact, we live in a world where software interacts across administrative boundaries and across organizational domains, and VMware strives to operate collaboratively in the IT space.

The third is an emerging eco system around hybrid private/public data-centers, federation of data-centers, and federation of clouds.

Existing virtualization challenges/hurdles

Existing big data-center technologies are key to making computing infrastructure dramatically more sustainable in very effective ways. Virtualization makes much **less wasteful utilization of all infrastructure resources** -- CPU, network, storage. Virtualization drives a key tenet of sustainability, **dematerialization**, which means to do more with less. This has been VMware's role in the industry since the beginning, and it continues leading and driving the virtualization world to a great extent.

Many challenges and opportunities lie ahead:

1. Reduced footprint VMs that still support legacy applications: provide the performance of containers with the security of VMs. This also shrinks the resource needs of an average user. As a result, consolidation improves. This means that more end users can be accommodated in the same hardware footprint, reducing energy needs and the carbon footprint of each user.

2. Full HW virtualization still under research and development, including NICs, GPGPUs, disaggregate memories, NV memory devices, and others.

3. Virtualization can help **prolong hardware life** at various levels of the software stack.

Software running within a virtual machine can be agnostic of the hardware it is run on. Enhanced support of legacy hardware can prolong the lifecycle of old hardware. Furthermore, **fault tolerance** can mask faulty hardware from software, enabling software to continue running on faulty hardware which would otherwise be discarded.

4. Advanced cluster resource management techniques are critical for effective utilization, driving **power consumption and hardware waste down**.
    a. It is important to continue research in migrating and auto-scaling VMs in order to efficiently utilize hardware and provide best performance. With more efficient hardware and applications that run more quickly, we can consolidate more applications on less hardware.
    b. Improve implementation of management operations. While it is important to improve resource management, this should not be done at the expense of inefficiently using hardware for other purposes. For example, each VM migration or resizing takes power (whether it be CPU cycles or network bandwidth). Arriving at optimal solutions without regard for the time or space complexity for such solutions does not help with improving the power budget, and unduly utilizing network traffic for management operations vs. application needs is just an example of 'robbing Peter to pay Paul.'

We refer to the collection of software technologies which provide infrastructures for data-center cluster as *the cluster OS* .

Democratizing the data-center OS

To date, much of the cluster OS technology is being developed by and deployed at a handful of companies (Amazon, Google, ..). There needs to be a **commoditization** of cluster-OS for data-centers (DCs). This will drive the same benefits of reduced energy and prolonged HW usage outside the “5 big computers”, and will help push the envelope of possibilities in this space.

Some topics:

- Increasing virtualization throughout storage and networking. By continuing research in storage and network virtualization, we can help drive the use of commodity components throughout the stack. These components are typically cheaper and lower-power than specialty hardware.
- Commoditizing fault-tolerance: today, the big cloud providers (Amazon, Google, Microsoft…) are able to stretch the life of their hardware because of the amount of data and insight they have about hardware failures. As a result, they are able to build cluster-OSes which can run in spite of failures. We believe that this level of fault-tolerance should be commoditized, which would not only provide more reliable private data centers, but greatly reduce the amount of hardware unnecessarily thrown

away just because these companies don't have the same level of insight that the big cloud providers do. Understanding how to exchange information about failures between software and hardware will be at the center of enabling this commoditization.
- Increasing efficiency of virtualization for special-purpose hardware. As an example, consider virtualizing the GPGPU such that high-performance applications can be migrated from a private testbed to a public cloud. While a lot has been done in this space, efficient interposition of the hypervisor with guest VMs GPGPU work is still an unsolved challenge. As a result, the cost for a high-performance cluster can be large, and better virtualization can drive down such costs.

Federating the cloud

An innovative approach for making private data-centers more sustainable is to create a market for taking advantage of spare computing resources. Solutions in this space could be leveraged in several forms:
- a federation of cloud providers to provide cooperative services
- hybrid private-cloud clouds
- federation of private clouds open for leasing in ``airBnB''-like mode

There are many exciting technical challenges.

The first and foremost is security. We envision a novel *hosting-mode* which delegates platform control to a remote manager and disallows even the local admin to control the machine. There are clearly challenges with managing hosting, maintaining keys associated with hosting hypervisors and hosted VMs. Likewise, there would be need to work with secure data encryption and encrypted communication.

Additional issues that need to be addressed are:

- a way for users to monetize their spare capacity and sell this capacity to other people
- support for multi tenancy across the board
- support for dynamic VM migration

Position statement: Adam Wierman

I will focus my position statement on an area of sustainable data centers that I think is often overlooked, but has a huge potential for impact on the adoption of renewable energy in the coming decade.

The typical story surrounding data centers and energy is very myopic one: Data centers are energy hogs that need to be made more sustainable. This message is pervasive in the media, and it certainly rings true. However, while data centers make up a significant and growing percentage of electricity usage, it is still small (~3%). Thus, making data centers sustainable locally is certainly valuable, but I want to argue that **data centers can have a bigger impact on sustainability if one takes a wider view and uses them as a tool to make the whole power system more sustainable**.

In particular, a powerful alternative view is that data centers could become a crucial resource for easing the integration of renewable energy into the grid system wide. That is, a key consequence of energy efficiency improvements in data centers is that their electricity demands are now very flexible. They can shed 10%, 20%, even 30% of their electricity usage in as little as 10 minutes by doing things such as pre-cooling, adjusting the temperature, demand shifting, quality degradation, geographical load balancing, etc. This flexibility, combined with the size of energy usage, makes data centers an amazing target for demand response.

This view of data centers as demand response candidates is important because it is exactly the lack of demand response and/or large scale storage that is the biggest hurdle for the integration of renewable energy into the grid. Renewable sources tend to be intermittent, unpredictable, and uncontrollable, but if demand response and/or large scale storage is available, then it can be used to smooth the fluctuations and make them much easier to incorporate. The problem is that large-scale storage is extremely expensive and demand response is typically quite difficult to obtain. In contrast, a 20 MW data center with 20% flexibility can serve the same role as hundreds of thousands of houses participating in a demand response program or nearly 1 MW of energy storage.

Unfortunately, despite the potential of data centers for demand response, the current reality is that data centers perform little, if any, demand response. And, worse, if they do provide demand response it is through use of behind-the-meter generation, which tends to be extremely dirty. There are many reasons for this, but perhaps the biggest is simply that the demand response programs that exist today are not suited for the load profile and risk tolerance of data centers, for which availability and performance are crucial concerns.

Consequently, there is much work to be done before the true potential of data center demand response can be realized. The research ahead is highly challenging and interdisciplinary, e.g., requiring work on the management of data center participation in demand response programs and the design of new demand response markets, as well as providing tools for the integration of data centers into power system modeling.

Currently, since much of the research in these directions falls between the cracks of the computer science, economics, and power systems communities, it is a difficult place for researchers to position themselves. Thus, NSF can have a big impact in fostering research in this direction through broadening the scope of research targeted in areas like NETS, CPS, EPAS, etc. Additionally, work in this area requires not just industrial interaction with tech companies, but also interaction with utilities and ISOs. Thus, the types of traces, measurements, etc. necessary to facilitate research are complementary to what is typically discussed in the sustainable data center community.

Workshop on Sustainable Data Centers

Peter Varman

Datacenter operational efficiencies will need to be increased to decrease power and energy consumption, while maintaining performance demands of applications.

Two directions of potentially fruitful investigation in this context are:

(i) How do we use NVM technology in the context of sustainable data centers?

(ii) Develop aggressive resource management techniques to optimize utilization (or reduce energy/power directly). Optimization of multiple resources in an integrated manner has promise to improve the datacenter resource utilization significantly compared to the independent resource-at-a-time optimizations currently employed.

Topics for NSF Meeting – Yahoo Suggestions

Chris Page, cpage@yahoo-inc.com

1. *"Prosumer" Approach to Electricity Production, Consumption & Management.*

The size and load profile of data centers have traditionally made them attractive but relatively passive customers for utilities. Data centers could take a more active role in both what they ask of the electricity grid & what they can provide back to the grid to answer next-generation utility challenges.

Specific topics of interest:

- How can data center electricity demand, storage capacity, and other factors facilitate the integration of more intermittent power sources on to the grid (eg wind & solar)? Example: Rutgers University GreenSlot program.

- Are there more efficient and/or more cost effective strategies for frequency regulation of electricity that could benefit both utilities and data centers? Could this include sharing infrastructure for FR, both on and offsite, or aggregating FR infrastructure at the data center level to benefit the grid?

- Distributed Energy Resource optimization: what are the efficiency and cost savings opportunities (eg, reduced congestion, deferred substation builds) for DER sharing/partnerships between data centers & utilities?

- Follow the Sun/Follow the moon optimization: What's the actual potential for energy savings, load reduction & carbon reduction of long-range time-of-day load migration between regional data centers? What sort of incentives would need to be in place to ensure optimizing for use of clean energy rather than just chasing cheapest (and possibly dirtiest) possible electricity? What other benefits/unintended consequences might result?

2. *Efficiency through right sizing and design for end-use.*

The Tier system was a useful framework in the early development of data center design but has become an outdated model that can hinder innovation and easily lead to overbuild and over design. Likewise, PUE was and remains a good performance metric to start with, but is limited in terms of what sort of efficiency gains it can measure. Reexamining needs and purposes for data centers, then designing and right-sizing accordingly, is a key research challenge.

Specific topics of interest

- Fitting level of redundancy to what's needed for the job, including "flex-tier"
- Providing temperature and humidity requirements that fit the equipment & job, rather than over-cooling to accommodate the most sensitive/expensive equipment
- Tackling economic/institutional barriers, like overengineered SLA requirements for colo providers or "split incentives" between colo customers who don't pay the utility bill & colo providers who run the facility
- Large-scale measurement of actual equipment mortality rates at different T and rH and duty cycles, including servers, network equipment, etc.

Position Statement for NSF Workshop on Sustainable Data Centers

Ying Lu, Associate Professor, University of Nebraska-Lincoln

In the next 5-10 years, to continue the progress on sustainable data centers, I think that the research community should focus on addressing the following key problems and challenges.

We should continue accelerating the cloud computing adoption in both industry and academia. As many previous studies have pointed out, smaller data centers often lack the incentives, resources, and expertise to investigate and adopt energy efficiency measures. Thus, instead of building and maintaining many small energy-inefficient data centers to support the increasing demands for IT services, scientific computations, and data processing, we should build energy-efficient public clouds and continue accelerating the cloud computing adoption in both industry and academia.

In terms of building and managing energy-efficient public clouds, the research community should focus on the following aspects. We should better leverage renewable energy in data centers. How to intelligently manage workloads and available energy sources is one of the key operational challenges in future energy-efficient data centers. In addition, since demand response has the potential to significantly ease the adoption of renewable energy in smart electricity grids, we should build data centers that adopt demand response schemes to dynamically manage their electricity loads in response to power supply conditions.

To accelerate the cloud computing adoption, the following key research challenges must be addressed. First, we need to investigate and remove the obstacles of cloud migration. Second, efficient, dynamic, and automatic provisioning schemes must be developed to support the smooth execution of cloud applications. Third, we must have secured clouds.

To enable researchers in academia to obtain design information, traces, measurements, etc. from operating data centers to facilitate continued research, NSF may consider funding researchers in academia to spend summers as visiting faculty members in industry.

Qingyuan Deng

* What are the key research challenges for continued progress on sustainable data centers in the next 5-10 years?
- increase the over all data center utilization (computing and power capacity) while not sacrificing performance / reliability or increasing operational complex – energy is more efficiently used while servers are running within higher utilization - how to reach that without introduce performance interference / reliability issues / and operational complex.

* What are the key operational challenges in building and managing sustainable data centers?
- the goal to build sustainable data center sounds to me kindly conflicting with the operational simplicity, so simplicity / less managing involvement is an importance factor to make it doable.

* How can NSF best leverage its existing investments in cloud testbeds to these ends?
- Don't quite know the existing investments.

* How should academia and industry collaborate?
- I think the most efficient way is to let faculty to take 1~2 years sabbatical / on-leave to work in the industry
1. Especially for data centers research, it's almost impossible to build even nearly similar scale datacenter at school, so many important problems, requirements, initiatives are originated from industry;
2. Many aspects / data regarding data centers are confidential and it's hard to access outside companies, so from the legal perspective it's much easier;
3. Comparing to sending students for internship, other than the NDA problems, faculties have better overall understanding and experiences discovering potential research opportunities

* What mechanisms can NSF foster to enable researchers in academia to obtain design information, traces, measurements, etc. from operating data centers to facilitate continued research?
- NSF could provide some incentives encouraging faculties spending sometime working in the industry.

Position statement for NSF Workshop on Sustainable Data Centers

Yanpei Chen, Software Engineer, Cloudera

Cloudera has visibility into use cases across a range of important industry verticals. We are a leading big data vendor. In 2014 we had over $100M in revenue, and in 2013 over half of the Fortune 500 companies are our customers. Our market position gives us a good understanding of big data needs across verticals such as finance, telecommunications, healthcare, government, retail, and others.

Energy costs is a concern for the largest (roughly 5%) of our customers. These are the largest customers by cluster size and the size of their business. The remaining customers are rapidly growing their clusters, sometimes with more than one expansion in a year. Thus, we anticipate energy efficiency will be a concern for an increasing number of our customers.

Cloudera is in a good position to influence how energy efficiency is measured across the industry. Cloudera is a member of industry standard benchmark consortia including the Transaction Processing Council (TPC), the Securities Technology Analysis Center (STAC), and others. Within these consortia, Cloudera is jointly defining several draft big data benchmarks. Cloudera also has a large ecosystem with over 1500 partners. Any measurement mechanism that makes its way into our partner certification program becomes a de-facto standard run by our partners.

The following is my personal position as an individual member of the technical community.

My work on big data energy efficiency started in 2009, when the term "big data" has not yet become popular [1]. This and contemporary work in other areas led us to realize we need to design large scale systems based on real-life workloads. Thus, we built the SWIM workload replay tool for MapReduce [2], which is now a part of Cloudera's partner certification suite. A side finding from the work is that there are promising energy efficiency techniques, provided we are willing to accept some constraints in how we manage the workload [3]. This sequence of work leaves unaddressed the following questions important both for academia research and for real-life cluster planning and operations in the industry:

- **How is energy efficiency different from regular computational efficiency?** Even without considering energy, the more efficient system in the traditional sense will lead to the same hardware being able to serve a larger workload. Hence for the same workload increase, slower capacity increase, hence less energy spent. PUE approaches 1 means computational efficiency and data center efficiency converge. So do we need to be concerned with energy efficiency at all, or is it already fully incorporated within traditional measures of computational efficiency?
- **How can energy efficiency measurement methods capture realistic behavior?** Many hard design challenges arises from the dynamics of the serviced workload over time. Efficiency and energy efficiency measurement methods can distort technical merit when they measure an unrealistic good case or corner case.
-
- **How do we design for common workloads given that we do not know what is common?** Companies such as Google, Facebook, Twitter have visibility into their own workload as a single case study. Vendors such as Cloudera sees many customer workloads but do not and should not have access to their proprietary information. How do we progress and avoid a multitude of point solutions that do not generalize? This is a concern beyond energy efficiency.

- **How do we design for energy efficiency without distorting cluster operations in a way that vastly decreases business value?** Many energy efficiency techniques, including the ones in my own work [3], places constraints on cluster operations that vastly decreases the business value. Such constraints relate to how data is placed, when do the jobs/queries get executed, and the performance of the jobs/queries. Energy efficiency must be achieved while still serving the cluster's businesses needs.
- **How can we design for energy efficiency at large scale?** Many design challenges are challenging only at scale. Presently, customers ask for proof-points on clusters of 100s of nodes. The demand for scale and the cost of proof-of-concept at scale increases continuously. How do we proceed? This is a concern beyond energy efficiency.
-

[1]. Y. Chen, L. Keys, and R. H. Katz. Towards Energy Efficient MapReduce. UC Berkeley EECS Tech Report EECS-2009-109. 2009.

[2]. Y. Chen, A. Ganapathi, R.Griffith, and R. Katz. The Case for Evaluating MapReduce Performance Using Workload Suites. MASCOTS 2011.

[3]. Y. Chen, S. Alspaugh, D. Borthakur, and R. Katz. Energy Efficiency for Large-Scale MapReduce Workloads with Significant Interactive Analysis. EuroSys 2012.

Performance Variability and Sustainable Computing
Amin Vahdat
May 2015

Computing infrastructure accounts for a substantial and growing fraction of the world's energy consumption. At the same time, an increasing fraction of compute is moving into data centers. Hence, improvements to computing efficiency within the data center can lead to outsize improvements in overall computing efficiency and the sustainability of planetary computing infrastructure.

In this brief position statement, we highlight *Performance Variability* as a key issue to improving compute efficiency in the datacenter.

Conquering performance variability

Distributed systems often run at the speed of the slowest component [e.g., Tail at Scale]. However, developers do not know which components will be slowest ahead of time, in large part because the slowest component will shift as a result of deployment conditions, failures, etc. This means that services often run integer factors slower than what would otherwise be possible if we could more tightly control system behavior. Sources of variability are growing and rampant:

- CPU performance depends on the number of tasks scheduled on the same machine and the memory/cache access behavior of those tasks. Consider issues of voltage scaling and cache pollution. Cluster schedulers consider all available cores to be equal, not accounting for antagonist behavior or overall machine load.
- Storage behavior is highly variable depending on access patterns and device behavior. Consider the differences in behavior of block access when: i) the block requires a disk seek, ii) the block can be retrieved as part of a streaming read, iii) the block is accessed in flash, iv) the block is written to flash but invokes wear leveling algorithms, v) the block is accessed from a remote DRAM cache, and vi) the block is accessed from a local DRAM cache. The sources of variability will only grow as we introduce next generation NVRAM as another source of interesting tradeoffs.
- Network behavior is highly dependent on software stack behavior and concurrent access across shared links. Datacenter network latency can vary by two orders of magnitude in going from an uncongested clear path with 10us latency to a 1ms congested path that involves just 1MB of router buffer queueing in a sample 10Gb/s path. TCP behavior is a large unknown even under good conditions and can become a huge bottleneck in the case of incast. While cluster schedulers consider datacenter networks to be flat for the purposes of scheduling, available bandwidth and its

variability actually depends on the particular network paths and the behavior of ToR and server co-resident tasks.
- The CPU scheduler can introduce unpredictable and seemingly random delays in the time a process can react to external events, such as RPC communication. These delays can range from zero to multiple milliseconds.

The above sources of performance variability compound, often multiplicatively. The end result is that distributed services must be sized and provisioned for performance at the 99% or even 99.9% of accesses to the service. Batch-oriented services that involve substantial distributed data access must often join data across hundreds or thousands of services. Once again, the critical path here often depends on the speed of the slowest component in such bulk-synchronous operations. Such provisioning means that computation can proceed 10-100x slower than the ideal case in the common case.

The infrastructure naturally consumes substantial additional power as a result. While improving performance is a useful independent goal, slower performance would be much easier to accept if the power consumed were proportional to the performance delivered. Unfortunately, this is not the case.

There are some interesting reasons for the lack of energy proportionality. First, the best way to address latency variability is often to poll for event completion in distributed systems. There is no efficient, low-latency mechanism to wake processes. Second, the transitions from low to medium to high energy states is at best rough and paradoxically a huge source of performance variability. Third, servers and CPU schedulers are built for throughput so very good at filling in "background work" during periods of, e.g., IO blocking. But it's not clear whether the switching back and forth to the background work, shifting to lower power states, etc. is actually worthwhile from the perspective of reducing power consumption in large-scale distributed systems.

Unfortunately, the problem is only getting worse. With the slowing of CPU performance gains, we are turning to domain-specific accelerators like GPUs and FPGAs. Basic architectural operations may themselves become probabilistic in the future, as will access to storage. The basis for this position statement is to consider the energy consumption of large-scale datacenter computing in the context of large-scale distributed systems, rather than as piecemeal local optimizations that may actually be taking us away from the end goal of sustainable computing.

Characterizing and optimizing server power consumption must be in the context of large-scale distributed systems with substantial sources of variability, both internal to the computation under measurement, but also external based on antagonistic failure conditions and co-resident services. Contrast this with current mechanisms for characterizing server power consumption as a function of synthetic singer-server workloads. The first step in this direction may be to develop realistic benchmarks capable of highlighting the sources of performance variability and capturing their impact on both performance and, more importantly, power consumption.

Appendix 3: Lightning Round & Presentation Slides

facebook

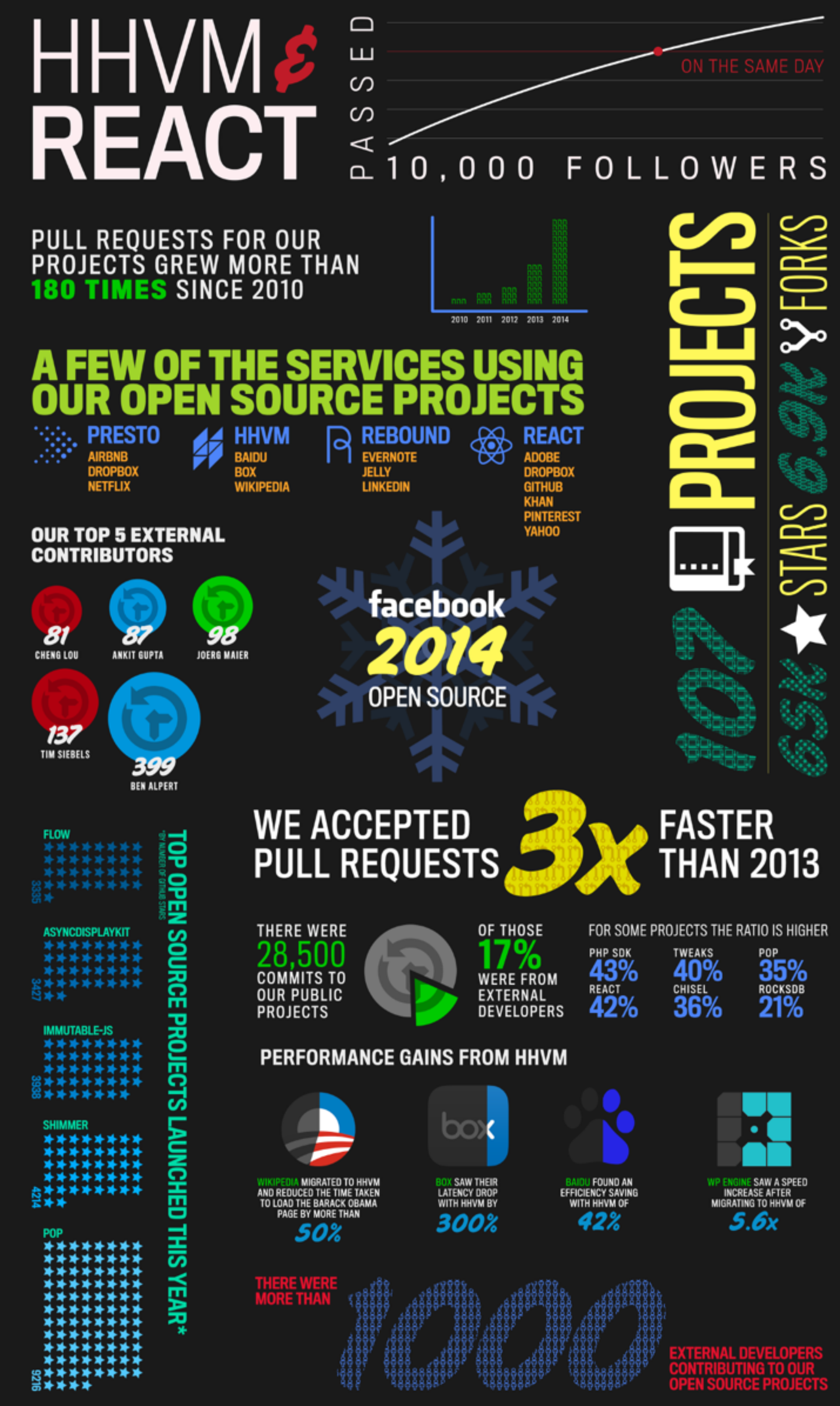
Facebook and the Open Compute Project

Charlie Manese
Infrastructure
NSF SDC - June 22, 2015

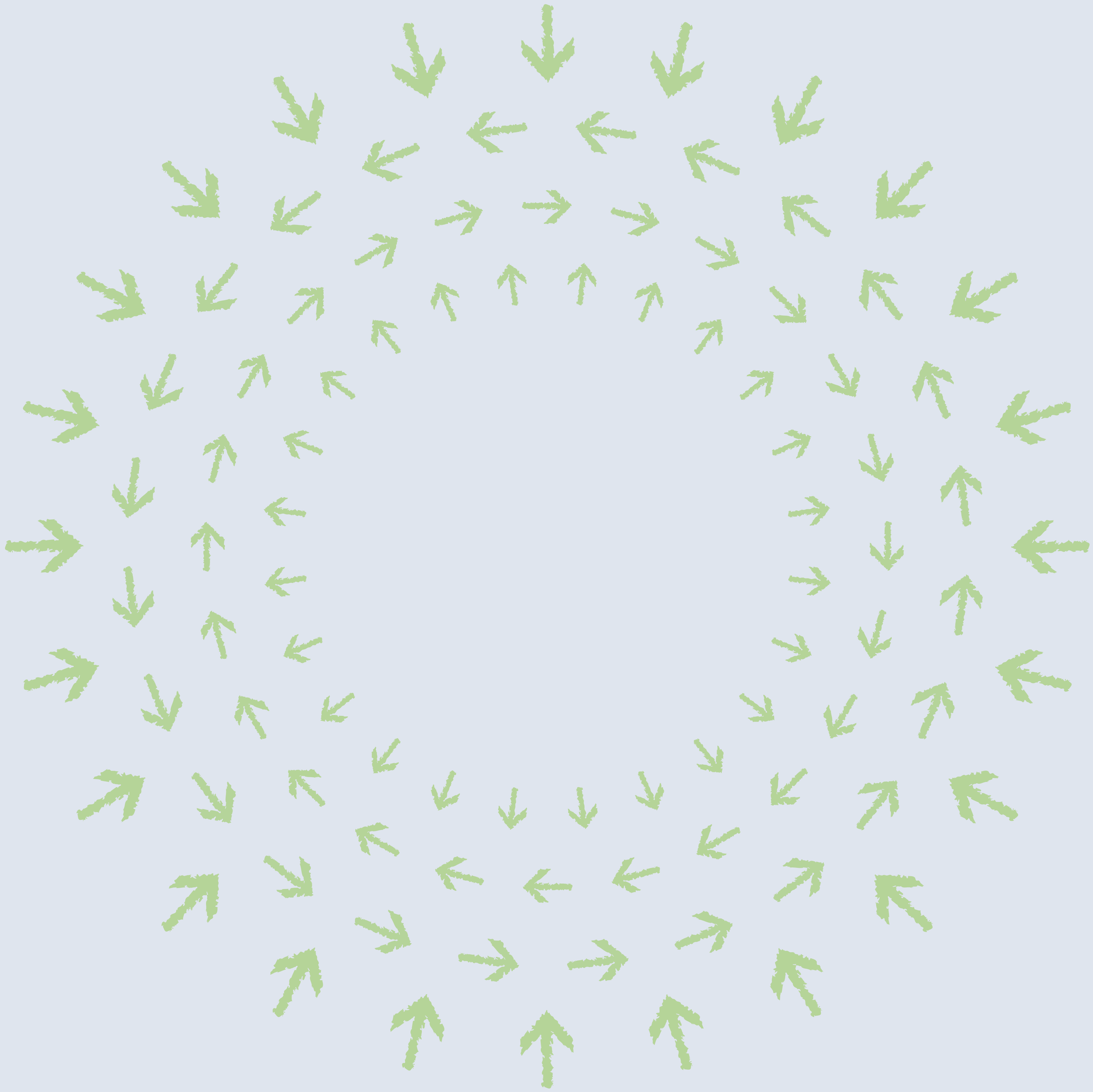

facebook
December 2010

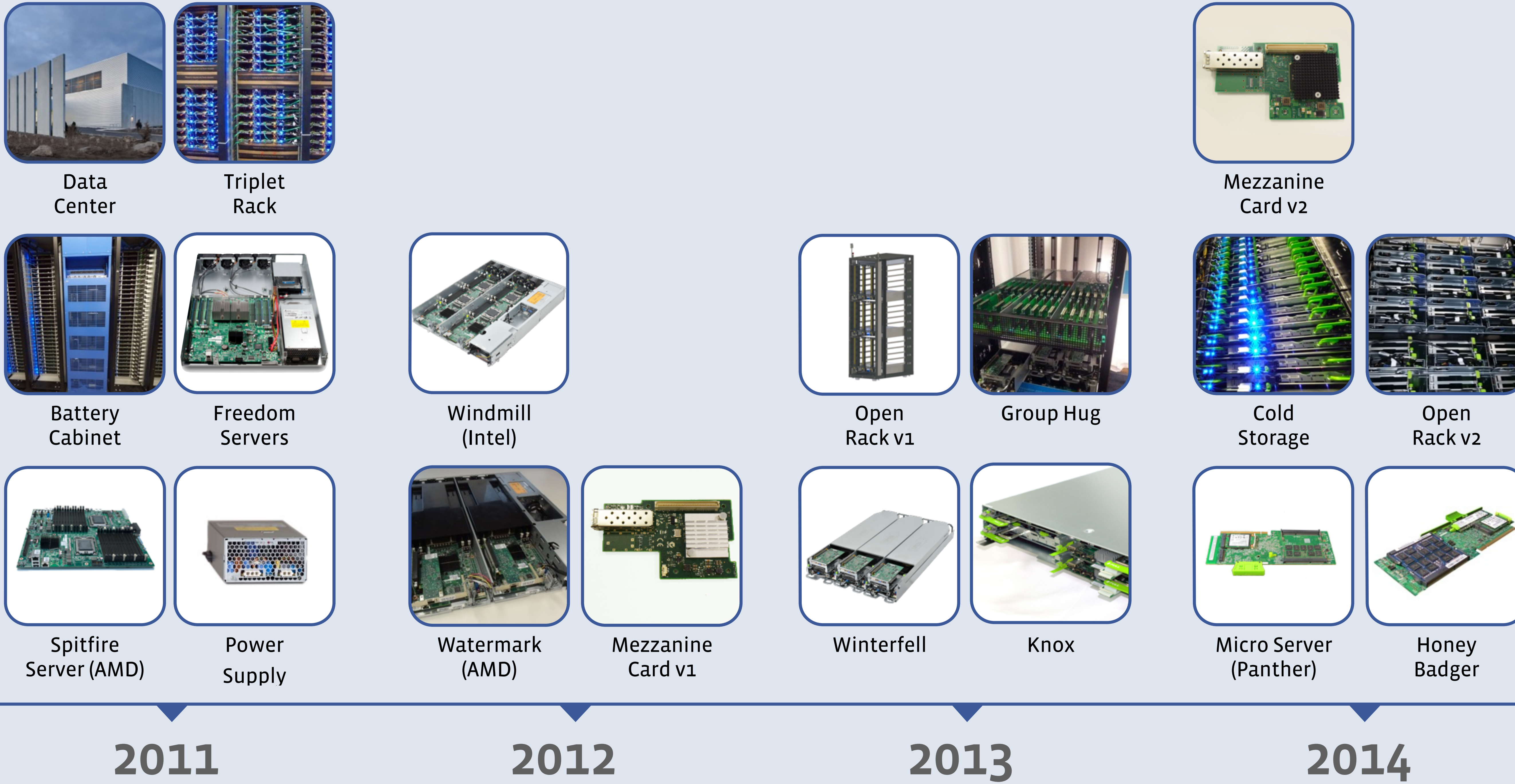
January 2015


facebook
Our Facebook Community
1.44 BILLION
people on Facebook each month
800 MILLION
people on WhatsApp each month
700 MILLION
people on Groups each month
600 MILLION
people on Messenger each month
300 MILLION
people on Instagram each month
45 BILLION
messages sent daily
4 BILLION
video views daily
800 MILLION
people offered access through Internet.org
650 MILLION
people connected to a sports Page
2 MILLION
advertisers
Q1'15 community highlights



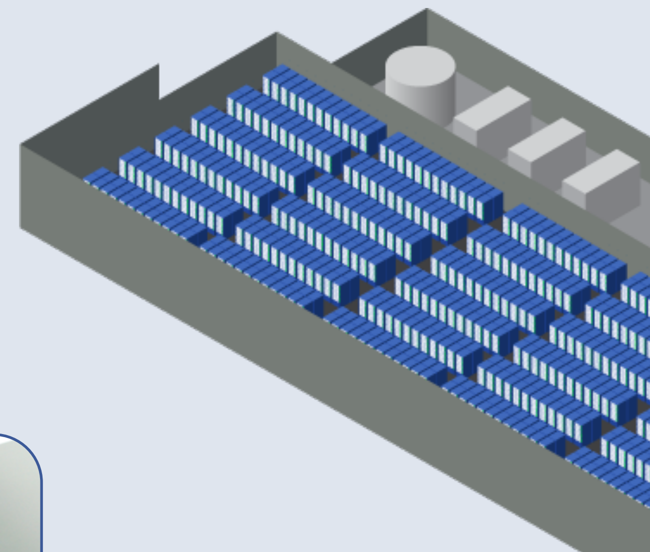
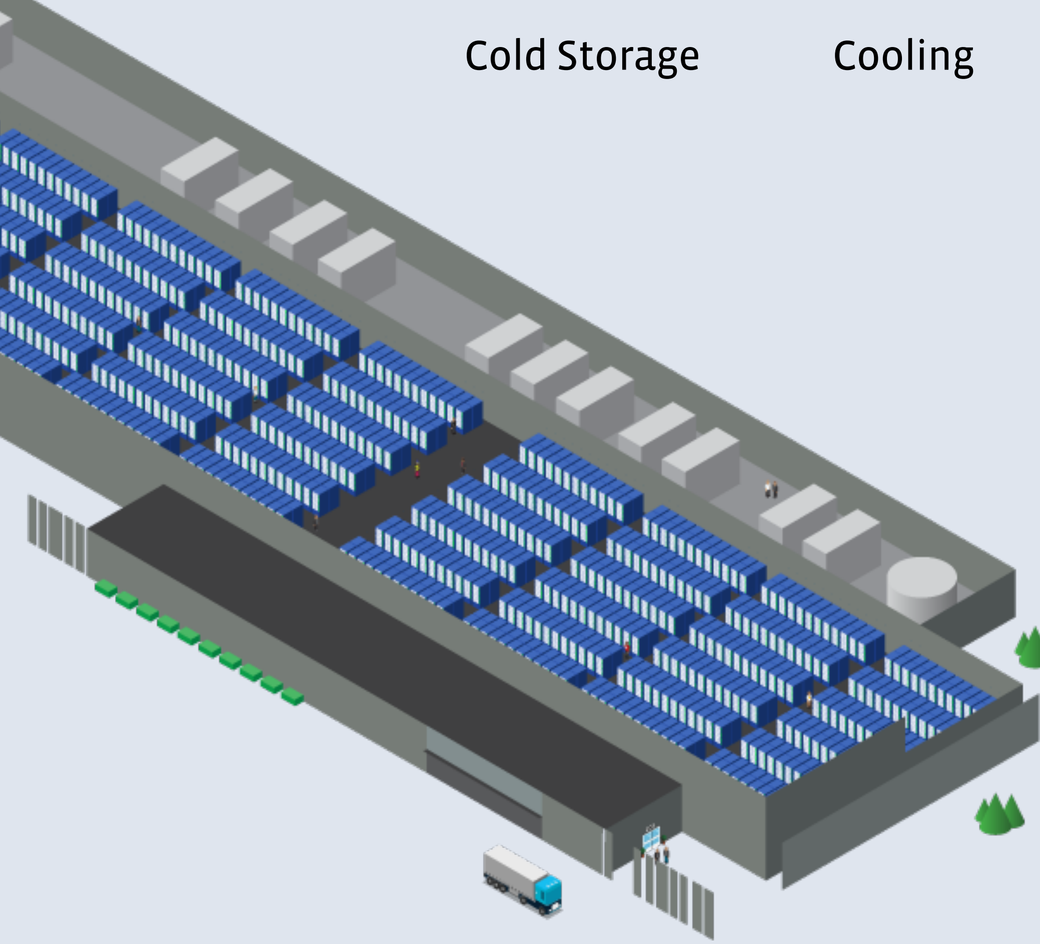
HHVM & REACT
PASSED 10,000 FOLLOWERS
ON THE SAME DAY
PULL REQUESTS FOR OUR PROJECTS GREW MORE THAN 180 TIMES SINCE 2010
2010 2011 2012 2013 2014
107 PROJECTS
65K STARS
6.9K FORKS
A FEW OF THE SERVICES USING OUR OPEN SOURCE PROJECTS
PRESTO
AIRBNB
DROPBOX
NETFLIX
HHVM
BAIDU
BOX
WIKIPEDIA
REBOUND
EVERNOTE
JELLY
LINKEDIN
REACT
ADOBE
DROPBOX
GITHUB
KHAN
PINTEREST
YAHOO
OUR TOP 5 EXTERNAL CONTRIBUTORS
81 CHENG LOU
87 ANKIT GUPTA
98 JOERG MAIER
137 TIM SIEBELS
399 BEN ALPERT
facebook 2014 OPEN SOURCE
WE ACCEPTED PULL REQUESTS 3x FASTER THAN 2013
THERE WERE 28,500 COMMITS TO OUR PUBLIC PROJECTS
OF THOSE 17% WERE FROM EXTERNAL DEVELOPERS
FOR SOME PROJECTS THE RATIO IS HIGHER
PHP SDK 43%
TWEAKS 40%
POP 35%
REACT 42%
CHISEL 36%
ROCKSDB 21%
TOP OPEN SOURCE PROJECTS LAUNCHED THIS YEAR*
*BY NUMBER OF GITHUB STARS
FLOW 3335
ASYNCDISPLAYKIT 3427
IMMUTABLE-JS 3938
SHIMMER 4214
POP 9216
PERFORMANCE GAINS FROM HHVM
WIKIPEDIA MIGRATED TO HHVM AND REDUCED THE TIME TAKEN TO LOAD THE BARACK OBAMA PAGE BY MORE THAN 50%
BOX SAW THEIR LATENCY DROP WITH HHVM BY 300%
BAIDU FOUND AN EFFICIENCY SAVING WITH HHVM OF 42%
WP ENGINE SAW A SPEED INCREASE AFTER MIGRATING TO HHVM OF 5.6x
THERE WERE MORE THAN 1000 EXTERNAL DEVELOPERS CONTRIBUTING TO OUR OPEN SOURCE PROJECTS





Data Center
Triplet Rack
Battery Cabinet
Freedom Servers
Spitfire Server (AMD)
Power Supply
Windmill (Intel)
Watermark (AMD)
Mezzanine Card v1
Open Rack v1
Group Hug
Winterfell
Knox
Mezzanine Card v2
Cold Storage
Open Rack v2
Micro Server (Panther)
Honey Badger
2011
2012
2013
2014

Open data center stack

Cold Storage

Cooling

Open Rack

Leopard

Knox

Wedge

Battery

Power

6-Pack

Software

Network

Servers
& Storage

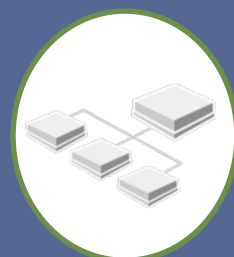
Data
Center

HipHop Virtual Machine

5x – 6x faster than Zend

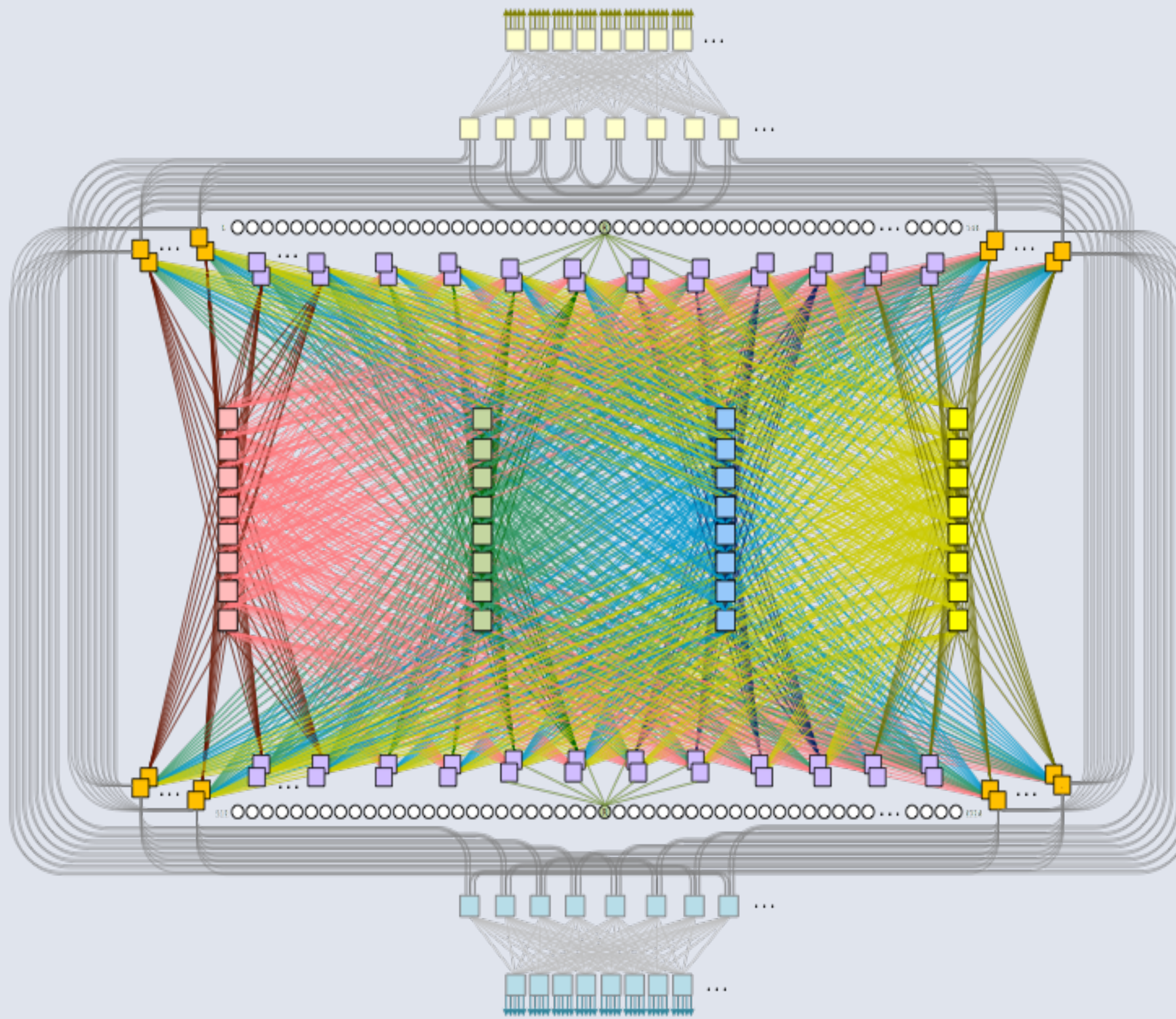
Software

Network



Servers
& Storage

Data
Center



Software

Network

Servers & Storage

Data Center

Original OCP designs

38 %

Energy Efficiency

Cost

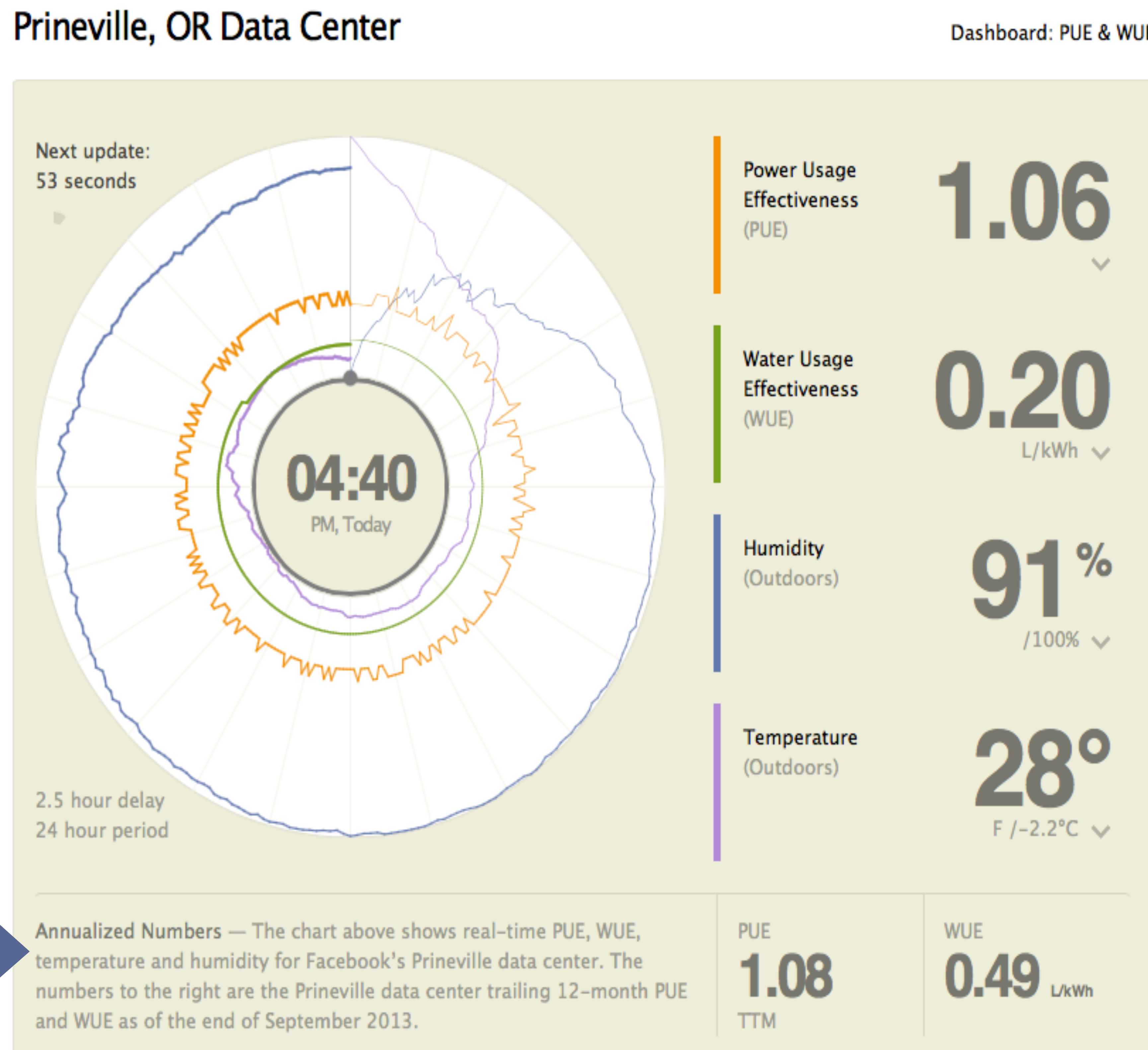
24 %

Software

Network

Servers
& Storage

Data
Center



Efficiency gains with OCP

$2 Billion

Efficiency gains with OCP

Annual Energy Savings

Annual Carbon Savings

Design principles

- Efficiency
- Scale
- Simplicity
- Vanity Free
- Easy to Operate

DATA CENTER

Facebook greenfield datacenter

Goal

- Design and build the most efficient datacenter eco-system possible

Control

- Application
- Server configuration
- Datacenter design


Prineville, OR


Forest City, NC


Luleå, Sweden

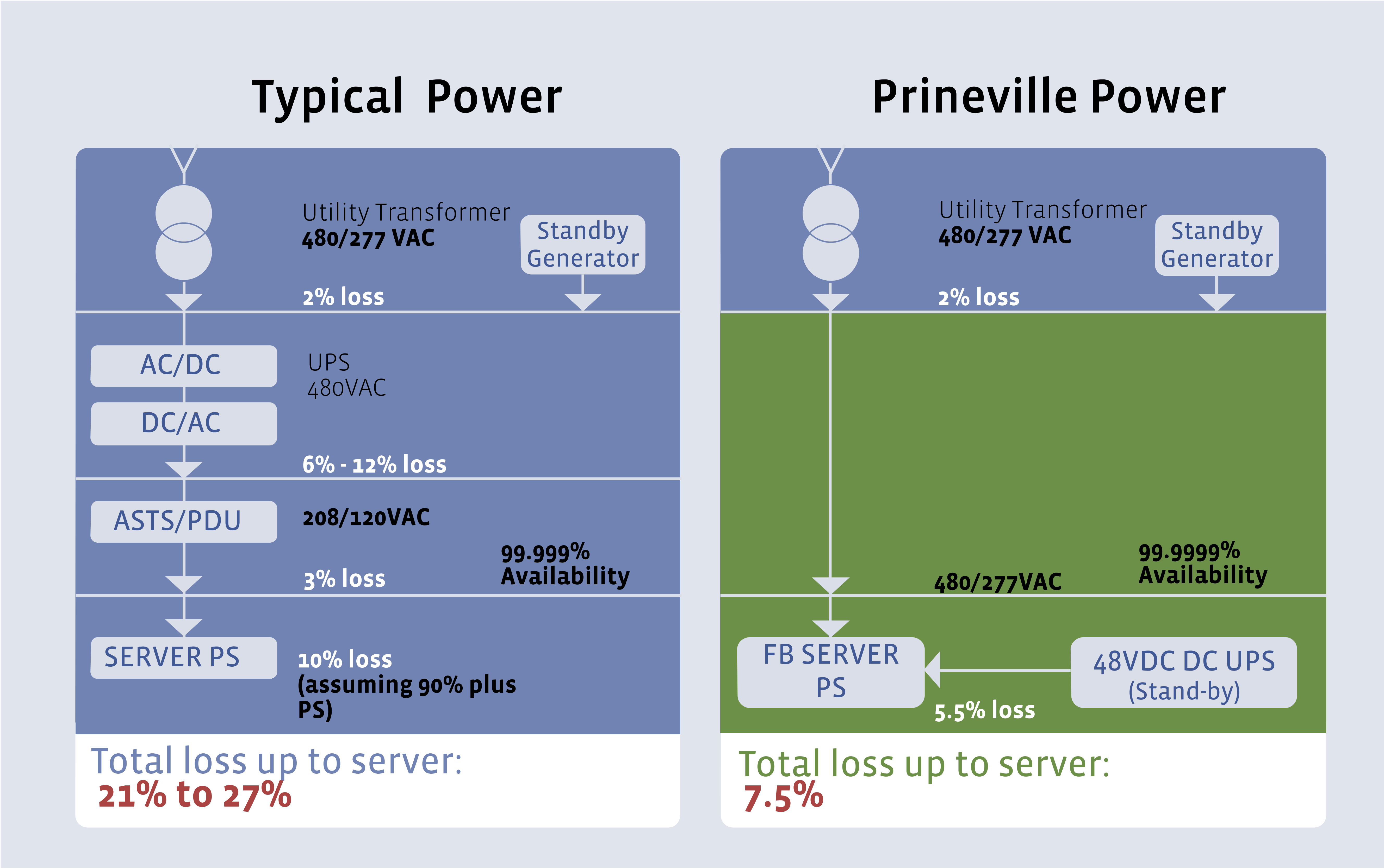

Altoona, IA
Hodnett
Photographics LLC

Electrical overview

- **Eliminate** 480V to 208V transformation
  - Used 480/277VAC distribution to IT equipment

- **Remove** centralized UPS
  - Implemented 48VDC UPS System

- **Result** a highly efficient electrical system and small failure domain


Typical Power
Utility Transformer
480/277 VAC
Standby Generator
2% loss
AC/DC
UPS
480VAC
DC/AC
6% - 12% loss
ASTS/PDU
208/120VAC
99.999% Availability
3% loss
SERVER PS
10% loss (assuming 90% plus PS)
Total loss up to server:
21% to 27%
Prineville Power
Utility Transformer
480/277 VAC
Standby Generator
2% loss
99.9999% Availability
480/277VAC
FB SERVER PS
48VDC DC UPS (Stand-by)
5.5% loss
Total loss up to server:
7.5%

Reactor power panel

- Custom Fabricated RPP
  - Delivers 165kW, 480/277V, 3-phase to CAB level
  - Contains Cam-Lock connector for maintenance wrap around
- Line Reactor
  - Reduces short circuit current < 10kA
  - Corrects leading power factor towards unity (3% improvement)
  - Reduces THD for improved electrical system performance (iTHD 2% improvement)
  - Power consumption = 360 Watt

Battery cabinet

- Custom DC UPS
- 56kW or 85kW
- 480VAC, 3-phase input
- 45 second back-up
- 20 sealed VRLA batteries
- Battery Validation System
- Six 48VDC Output
- Two 50A 48VDC aux outputs

Mechanical overview

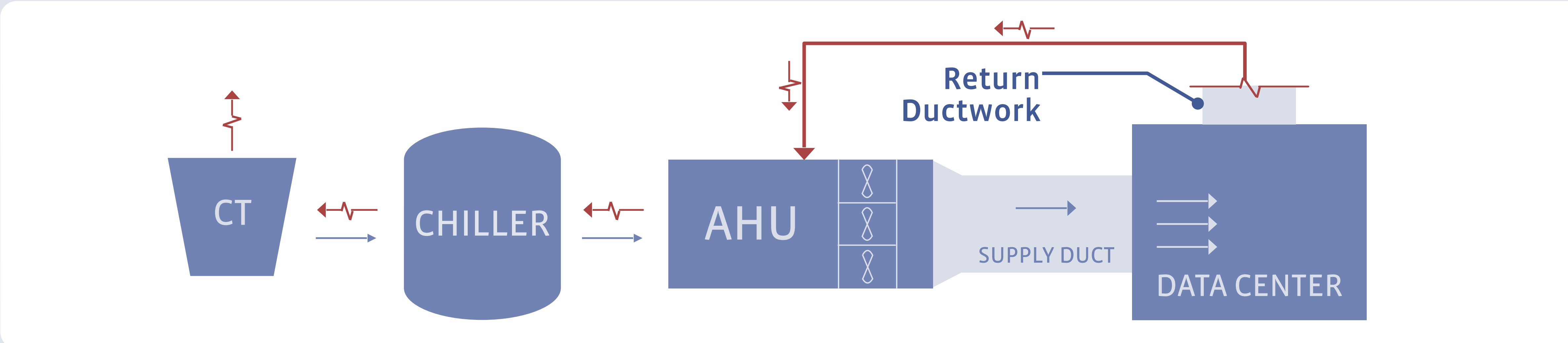
- **Removed**
  - **Centralized** chiller plant
  - HVAC **ductwork**

- **System Basis of Design**
  - ASHRAE Weather Data: N=50 years
  - TC9.9 2008: Recommended Envelopes

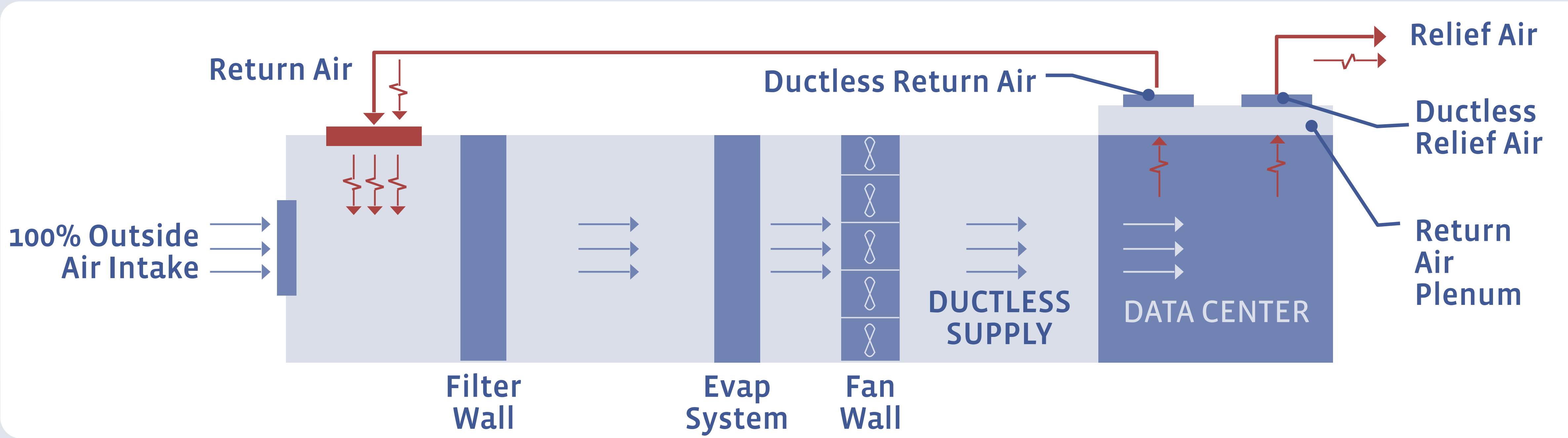
- **Built-up penthouse air handling system**

- **Server waste heat is used for office space heating**

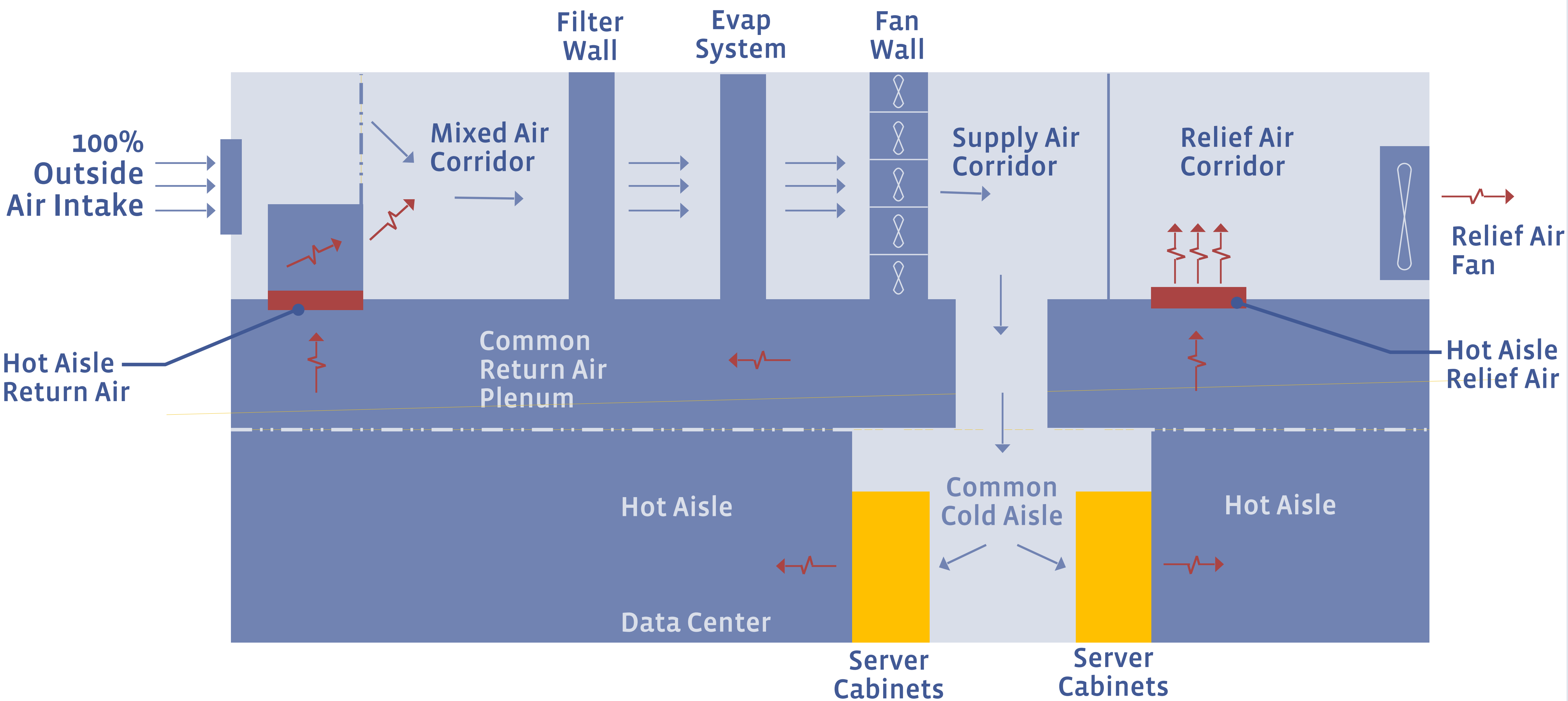
Typical datacenter cooling

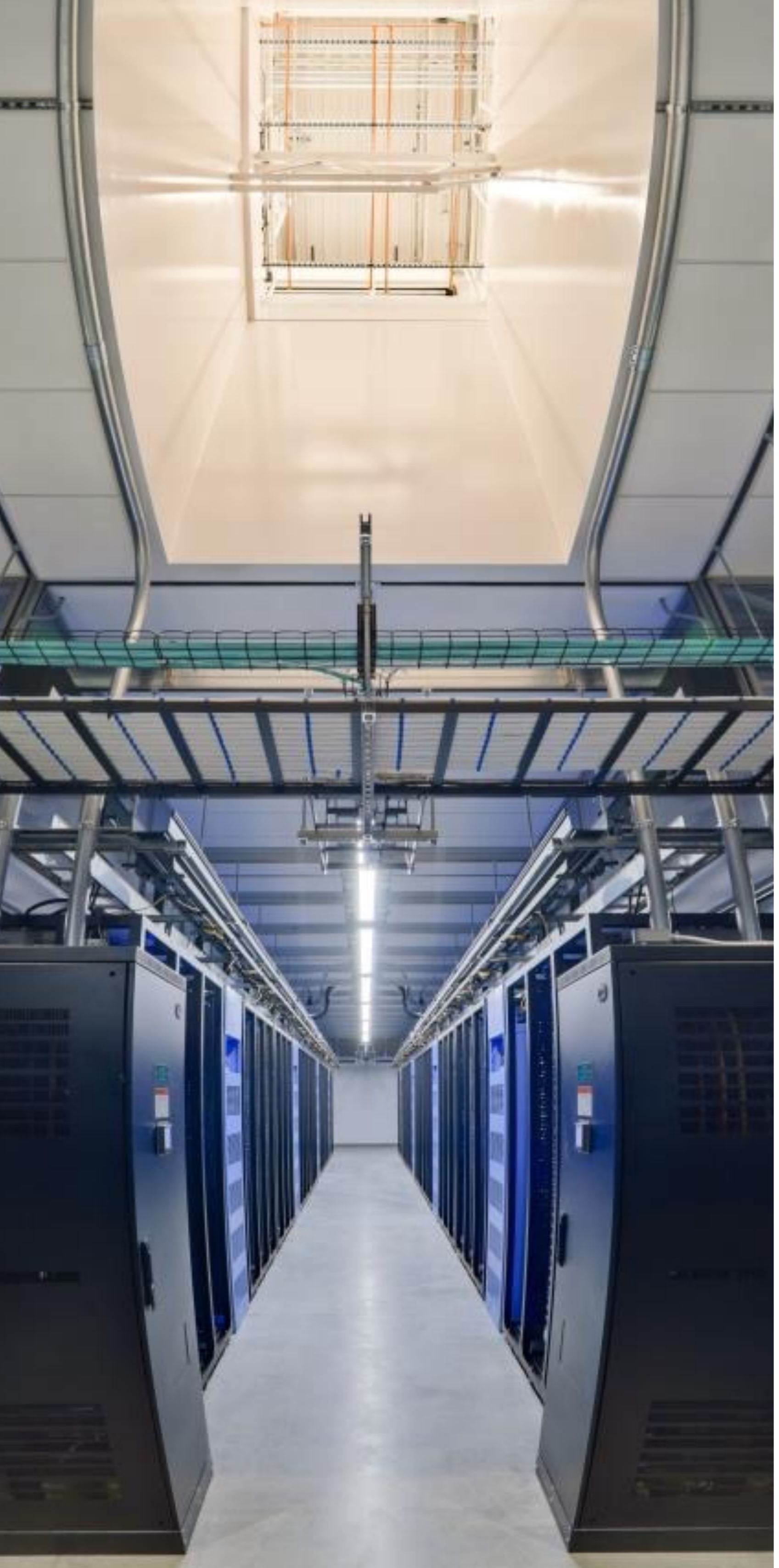


Prineville datacenter cooling



PRN datacenter cooling






Cold aisle pressurization – ductless
supply

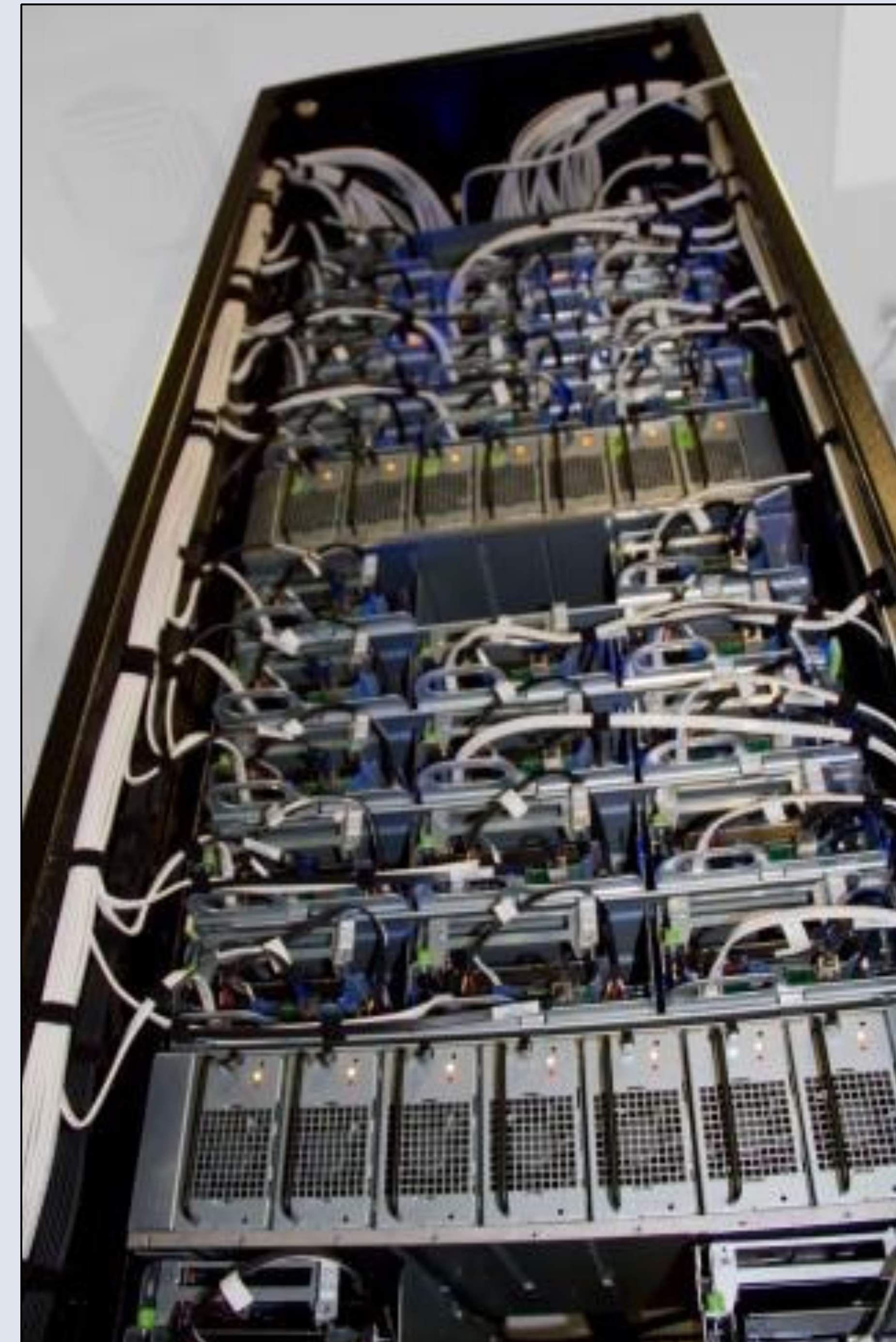
Basis of design comparison

PRN1A1B	PRN1C1D	FRC1A1B	LLA1A1B
80°F/27°C inlet	85°F/30°C inlet	85°F/30°C inlet	85°F/30°C inlet
65% humidity	80% humidity	90% humidity	80% humidity
20°F/11°C ΔT	22°F/11°C ΔT	22°F/11°C ΔT	22°F/11°C ΔT

RACK, SERVERS, AND STORAGE

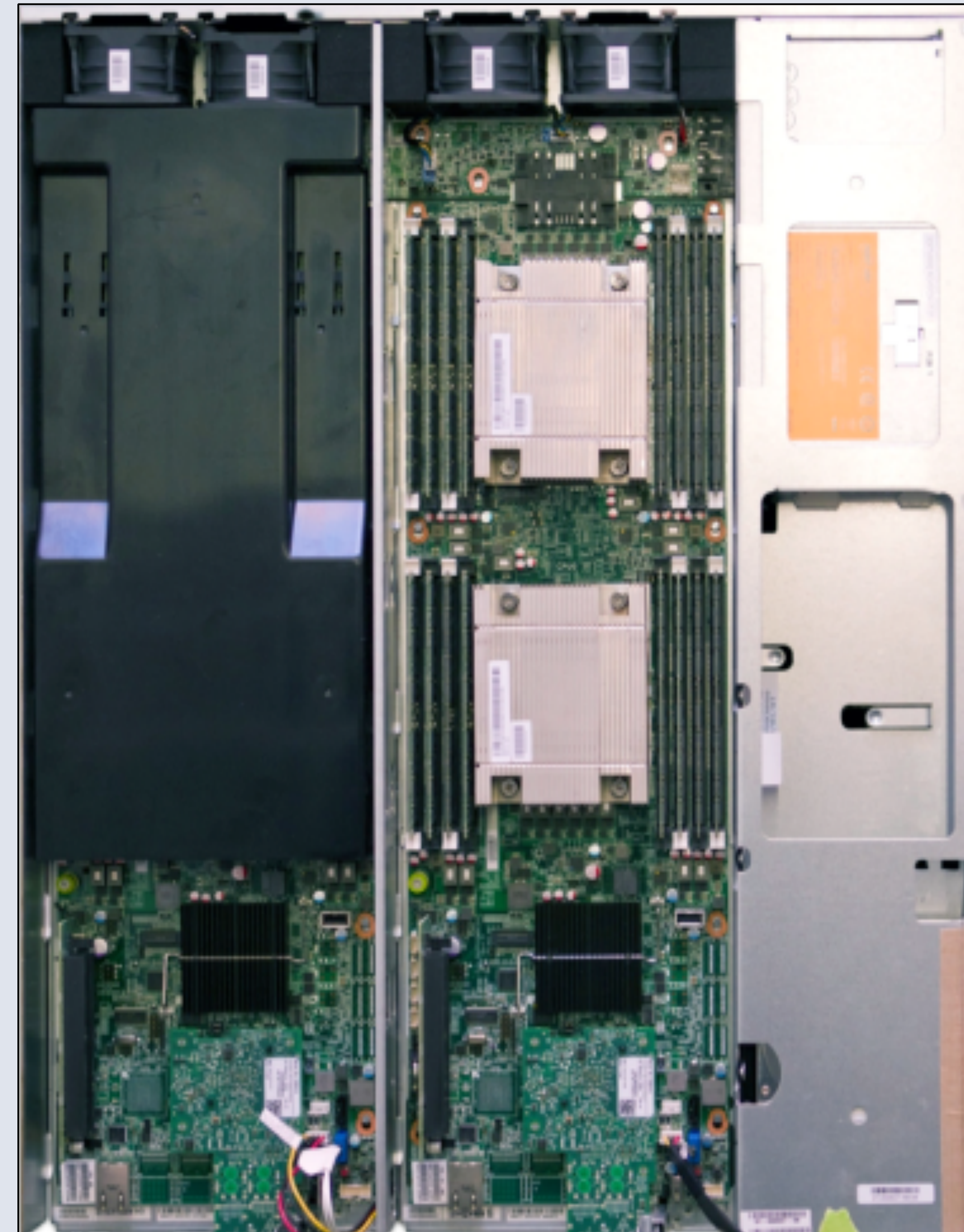
Open Compute Rack: Open Rack

- Well-defined "Mechanical API" between the server and the rack
- Accepts any size equipment 1U – 10U
- Wide 21" equipment bay for maximum space efficiency
- Shared 12v DC power system



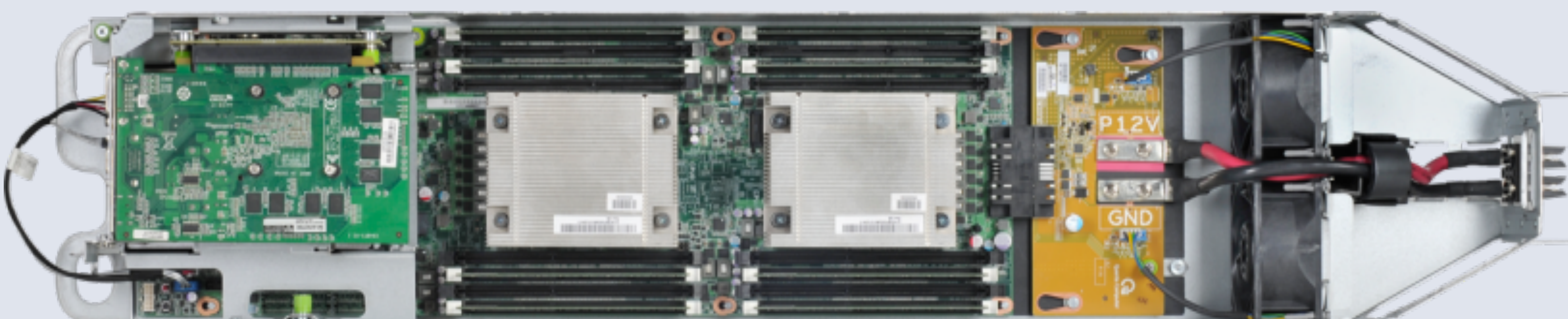
Open Compute Server v2

- First step with shared components by reusing PSU and fans between two servers
- Increased rack density without sacrificing efficiency or cost
- All new Facebook deployments in 2012 were “v2” servers



Open Compute Server v3

- Reuses the “v2” half-width motherboards
- Self-contained sled for Open Rack
- 3-across 2U form factor enables 80mm fans with 45 servers per rack

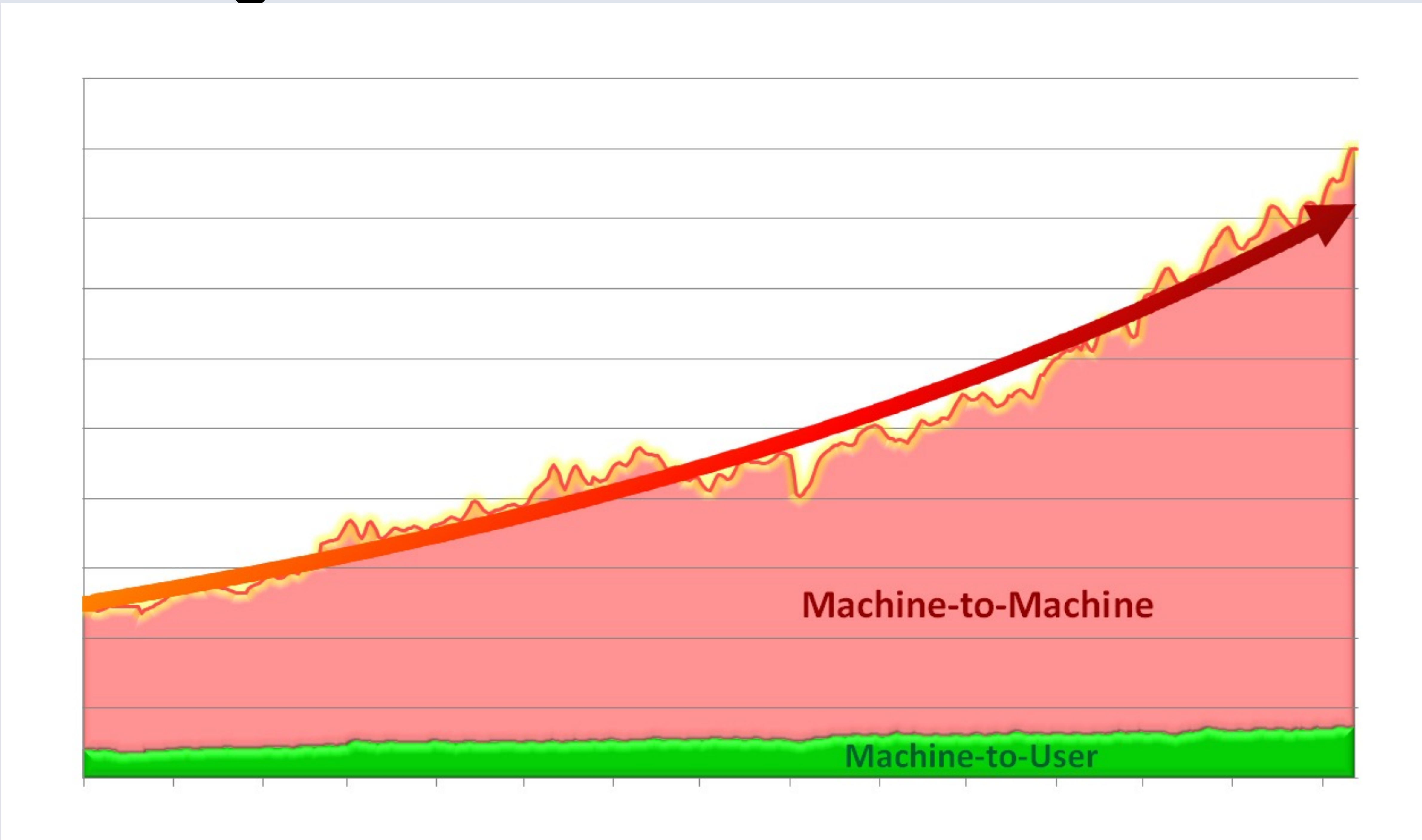


Open Vault

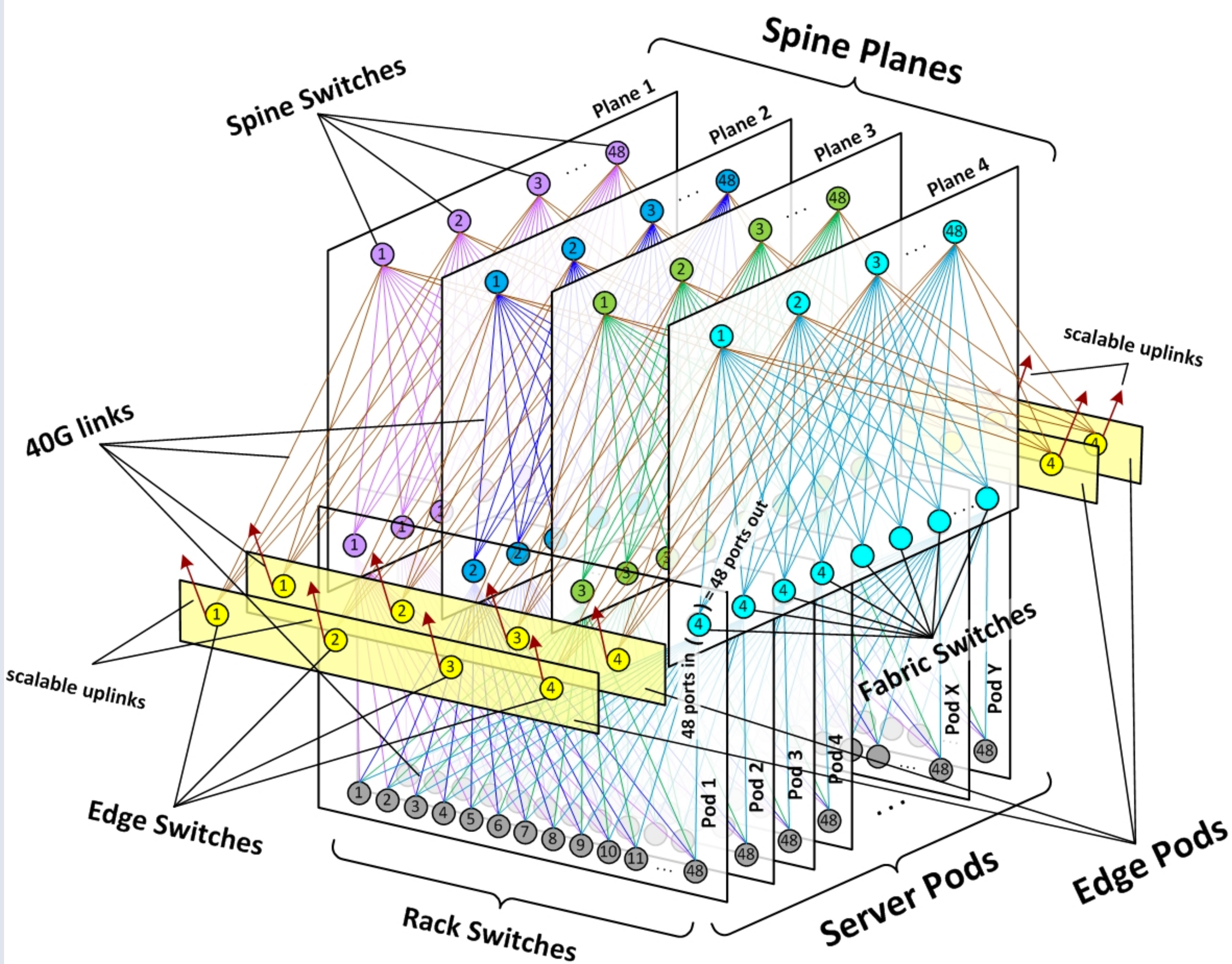
- Storage JBOD for Open Rack
- Fills the volume of the rack without sacrificing hot-swap

NETWORK

Traffic growth

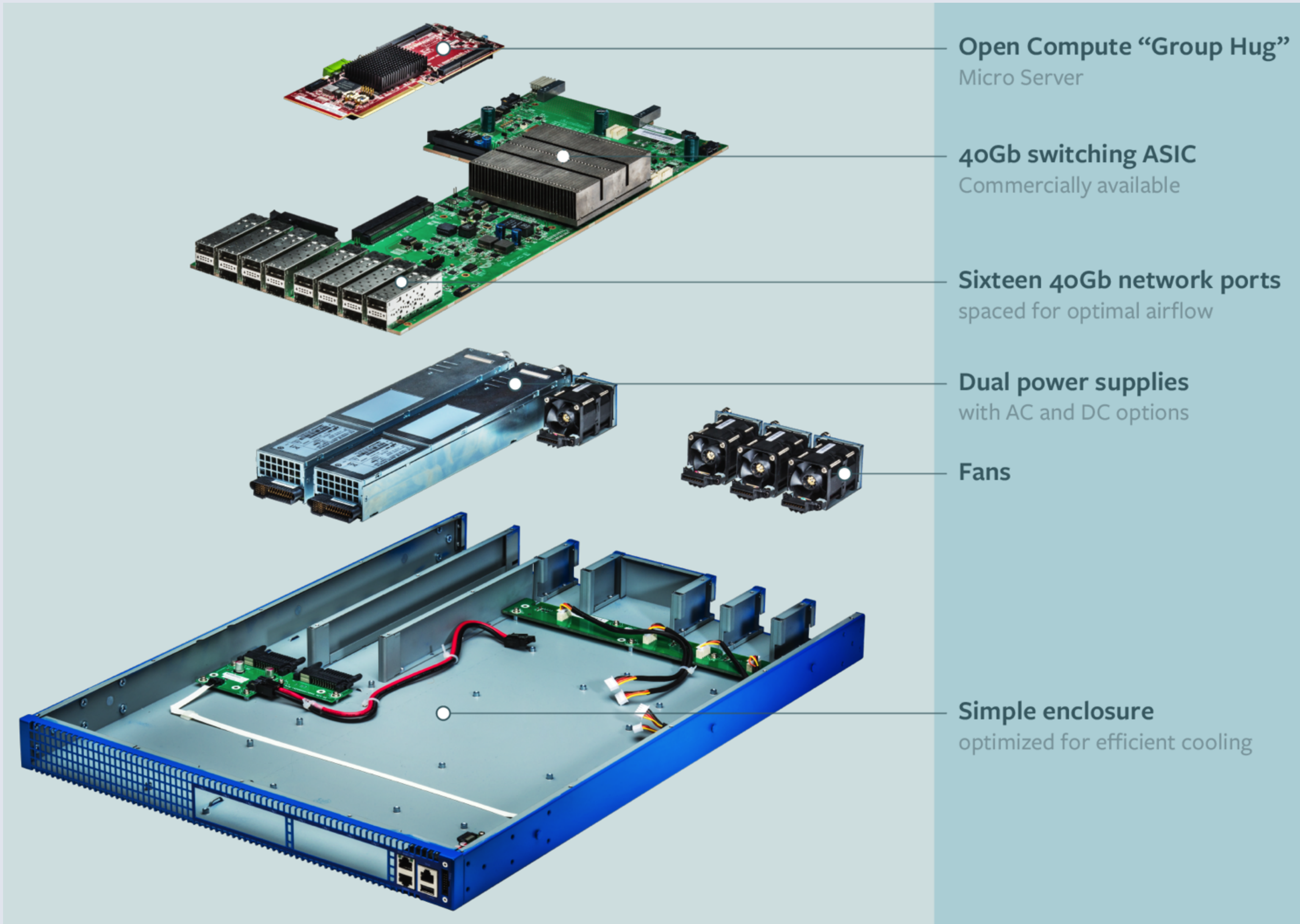


Fabric

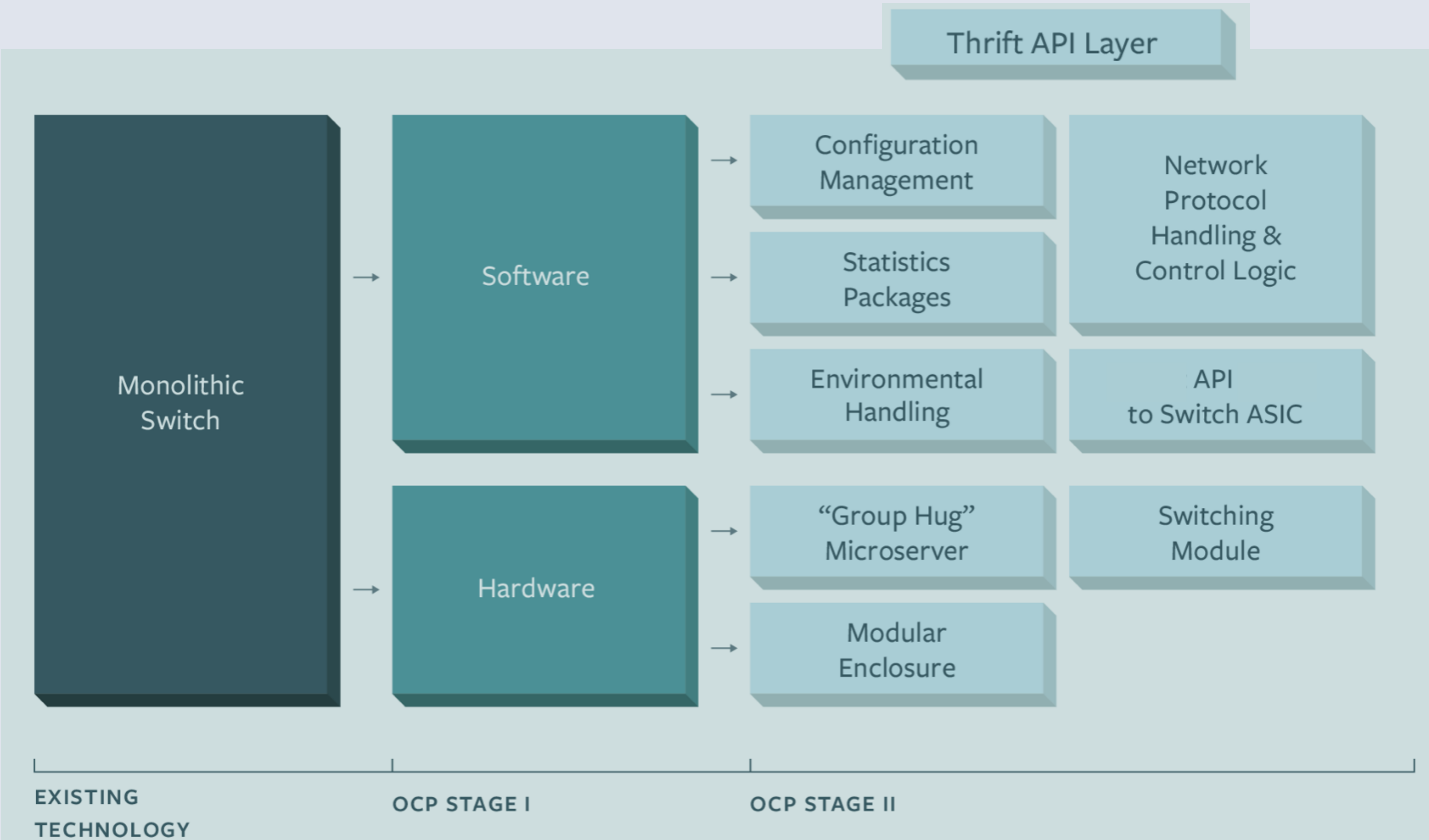


Spine Planes
Spine Switches
Plane 1
Plane 2
Plane 3
Plane 4
scalable uplinks
40G links
48 ports in ( ) = 48 ports out
Fabric Switches
scalable uplinks
Pod 1
Pod 2
Pod 3
Pod 4
Pod X
Pod Y
Edge Switches
Rack Switches
Server Pods
Edge Pods

Wedge


Open Compute "Group Hug"
Micro Server
40Gb switching ASIC
Commercially available
Sixteen 40Gb network ports
spaced for optimal airflow
Dual power supplies
with AC and DC options
Fans
Simple enclosure
optimized for efficient cooling

FBOSS

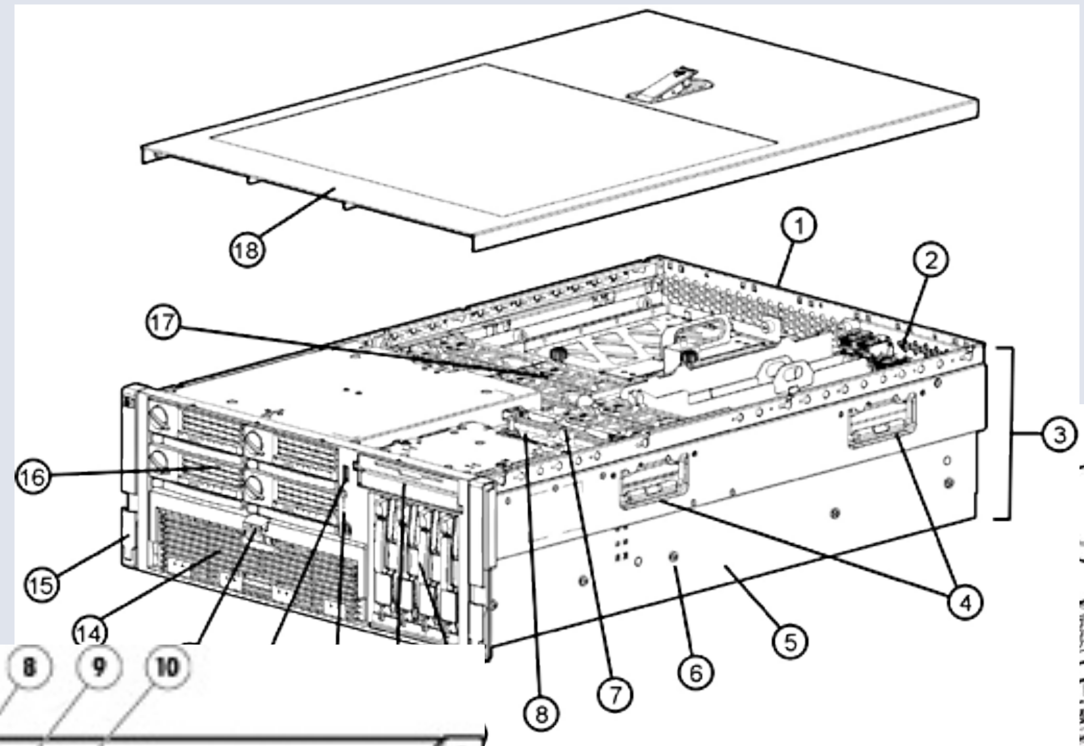

Thrift API Layer
Monolithic Switch
Software
Hardware
Configuration Management
Statistics Packages
Environmental Handling
"Group Hug" Microserver
Modular Enclosure
Network Protocol Handling & Control Logic
API to Switch ASIC
Switching Module
EXISTING TECHNOLOGY
OCP STAGE I
OCP STAGE II

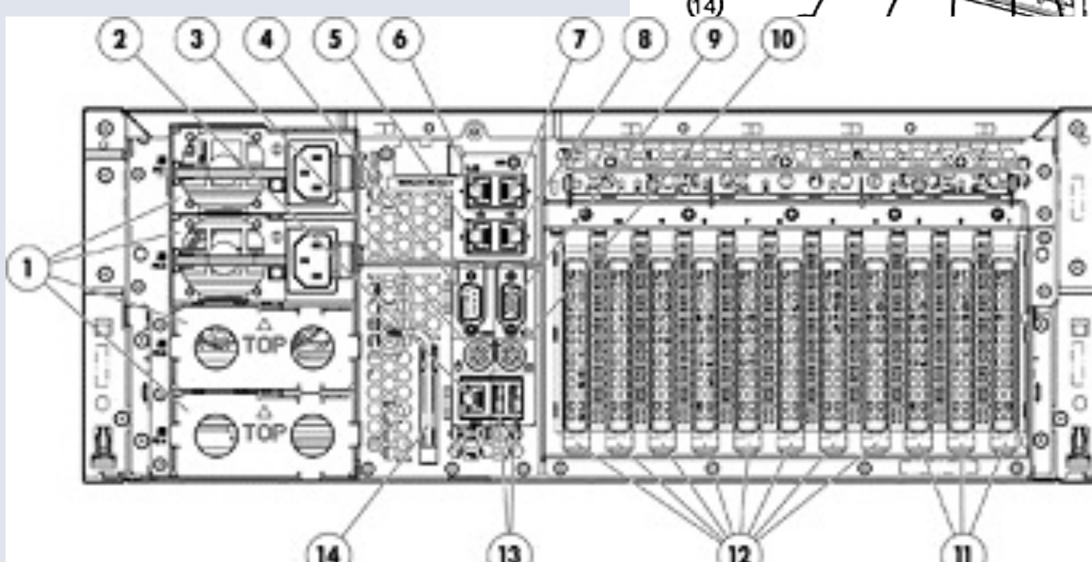
6-Pack

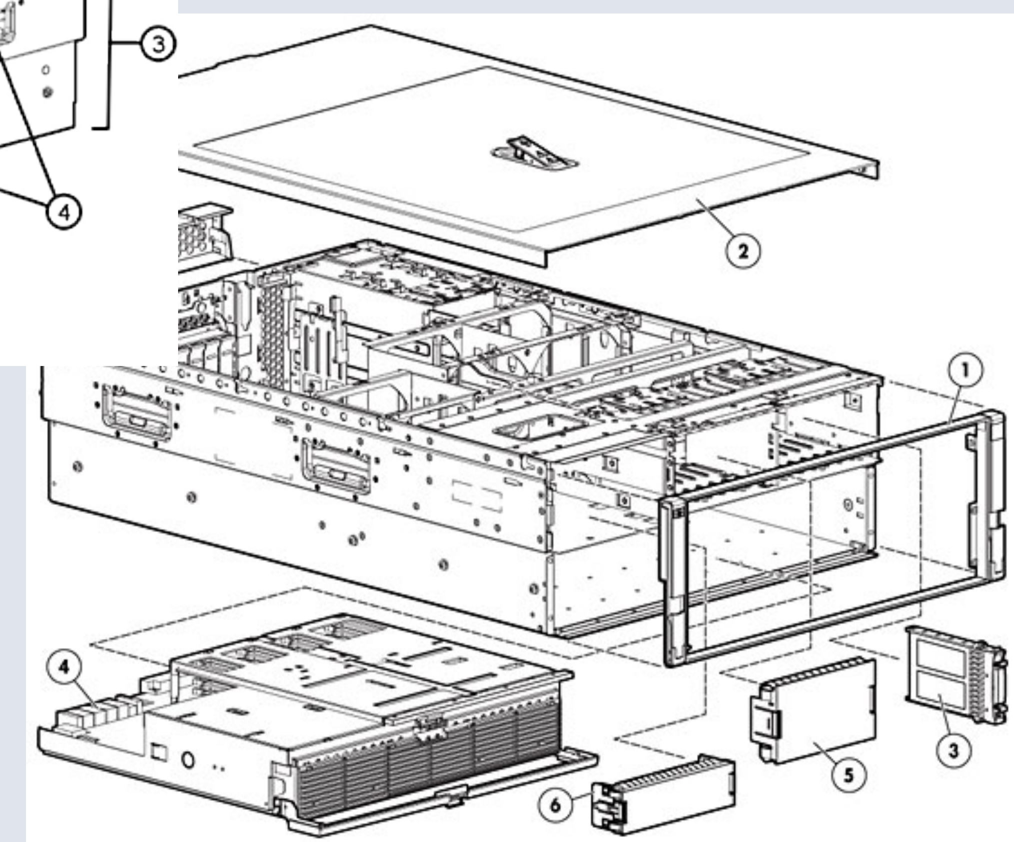
SERVICEABILITY

Complex designs

Typical large datacenter:
1000 Servers per Technician







~~Complex~~ Simple designs

Typical large datacenter:
1000 Servers per Technician

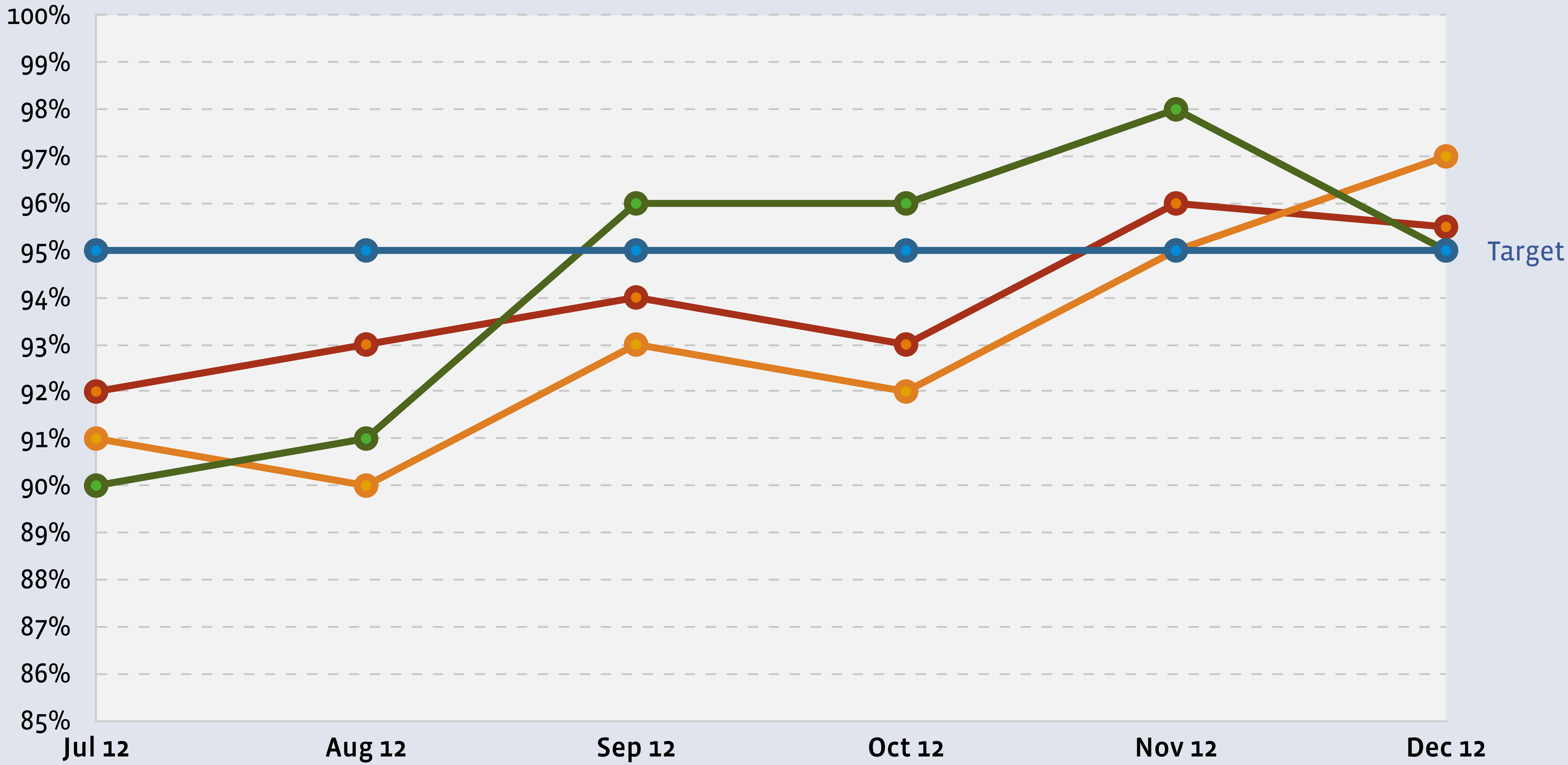
Facebook datacenter:
25,000 Servers per Technician

Efficiency through serviceability

Standing at Machine	OEM REPAIRS				
	Pre-Repair Activities Min	Part Swap Duration Min	Additional Steps Min	Post-Repair Activities Min	Total Repair Time Min
Hard Drive (Non-raid)	2	3	0	2	7
DIMM (Offline)	2	3	0	2	7
Motherboard	2	20	20	2	44
PSU (Hot Swap)	2	5	0	2	9

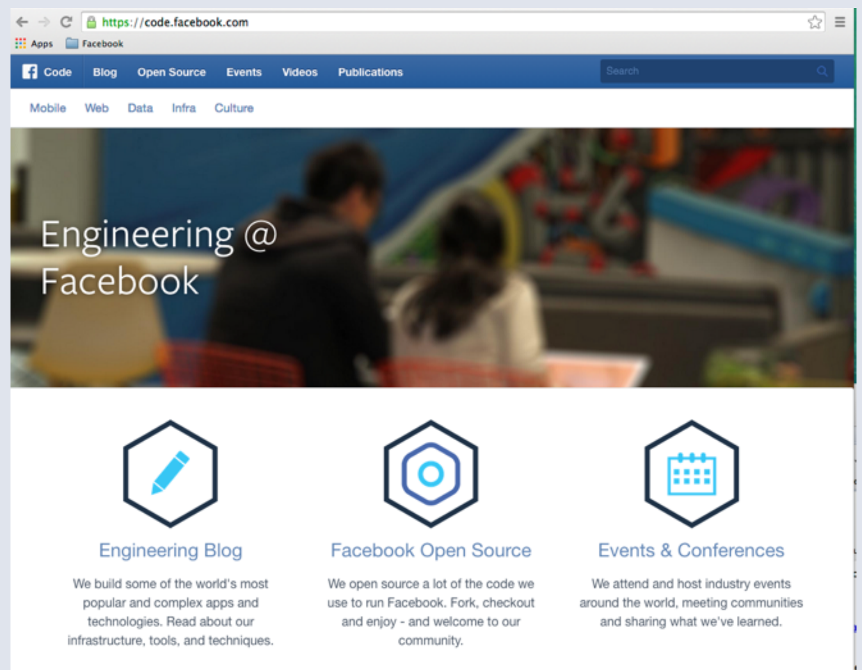
	OCP#1 REPAIRS				
	Pre-Repair Activities Min	Part Swap Duration Min	Additional Steps Min	Post-Repair Activities Min	Total Repair Time Min
Hard Drive (Non-raid)	0	0.98	0	0	0.98
DIMM (Offline)	0	0.82	0	0	0.82
Motherboard	2.5	10.41	2.5	0	15.41
PSU (Hot Swap)	0	0.65	0	0	0.65

First-time-fix repair rates



Let’s engage

OPEN Compute Project

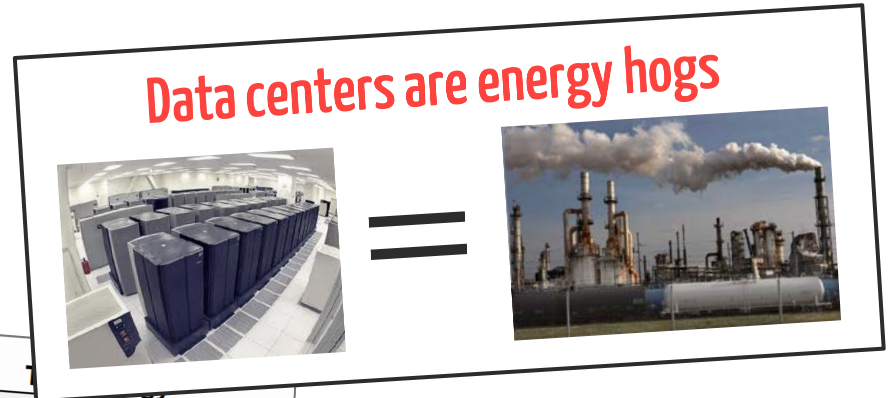


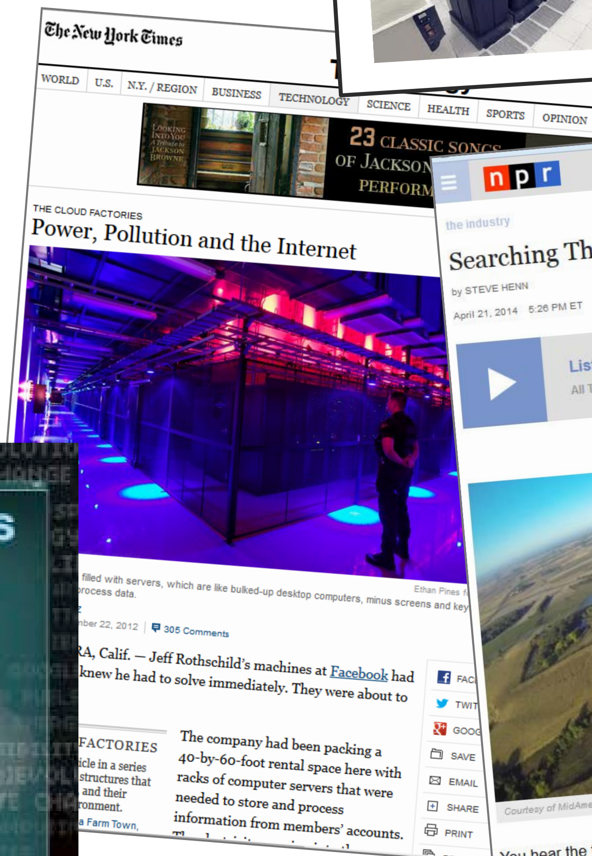
@scale

Data centers & energy: *Did we get it backwards?*

Adam Wierman, Caltech

The typical story about energy & data centers:




The New York Times

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

THE CLOUD FACTORIES

Power, Pollution and the Internet

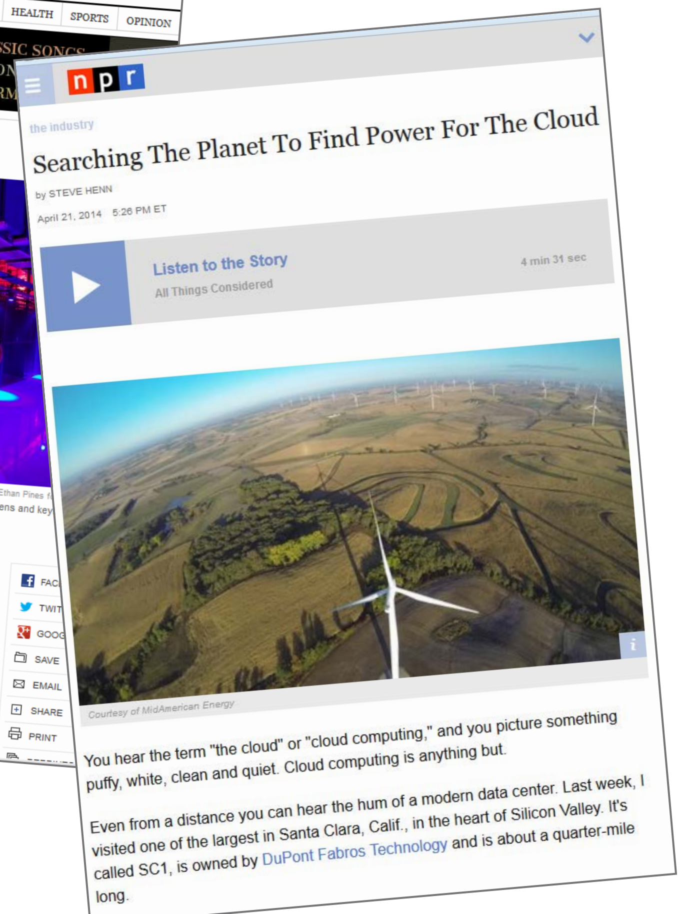
filled with servers, which are like bulked-up desktop computers, minus screens and key

Ethan Pines

305 Comments

RA, Calif. — Jeff Rothschild's machines at Facebook had knew he had to solve immediately. They were about to

The company had been packing a 40-by-60-foot rental space here with racks of computer servers that were needed to store and process information from members' accounts.


npr

the industry

Searching The Planet To Find Power For The Cloud

by STEVE HENN

April 21, 2014 5:26 PM ET

Listen to the Story

All Things Considered

4 min 31 sec

Courtesy of MidAmerican Energy

You hear the term "the cloud" or "cloud computing," and you picture something puffy, white, clean and quiet. Cloud computing is anything but.

Even from a distance you can hear the hum of a modern data center. Last week, I visited one of the largest in Santa Clara, Calif., in the heart of Silicon Valley. It's called SC1, is owned by DuPont Fabros Technology and is about a quarter-mile long.

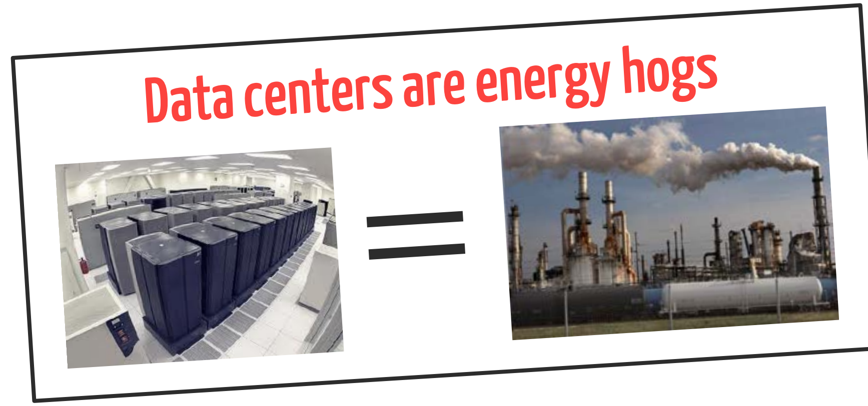
The typical story about energy & data centers:

Data centers are energy hogs

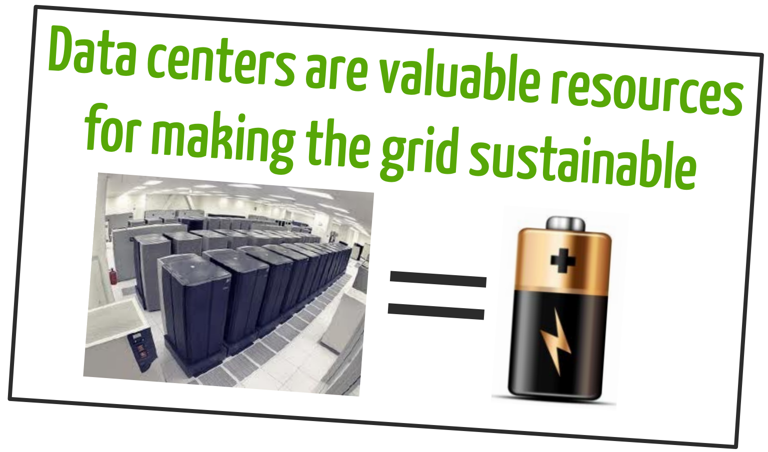
Sustainable data centers

Remember: The cloud is (often) more efficient than the alternative.

The typical story about energy & data centers:



But maybe we got it backwards?



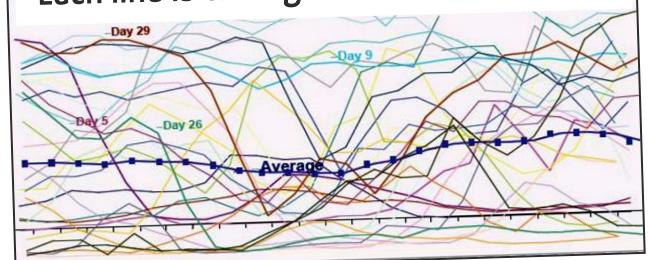
Renewable energy is ~~coming!~~ here!

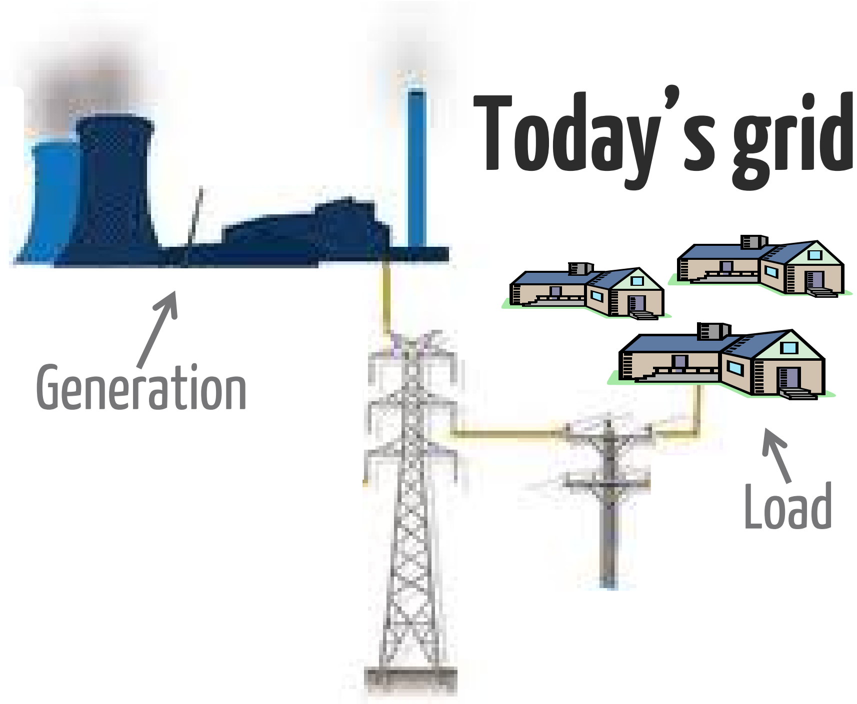
…but incorporation into the grid isn't easy

They are typically

→ Uncontrollable (not available "on demand")
→ Intermittent (large fluctuations)
→ Uncertain (difficult to forecast)

Each line is wind generation over 1 day

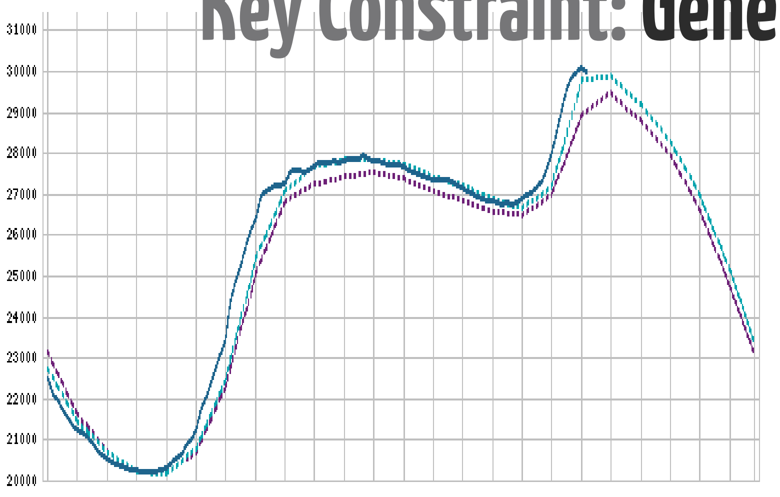




Today's grid

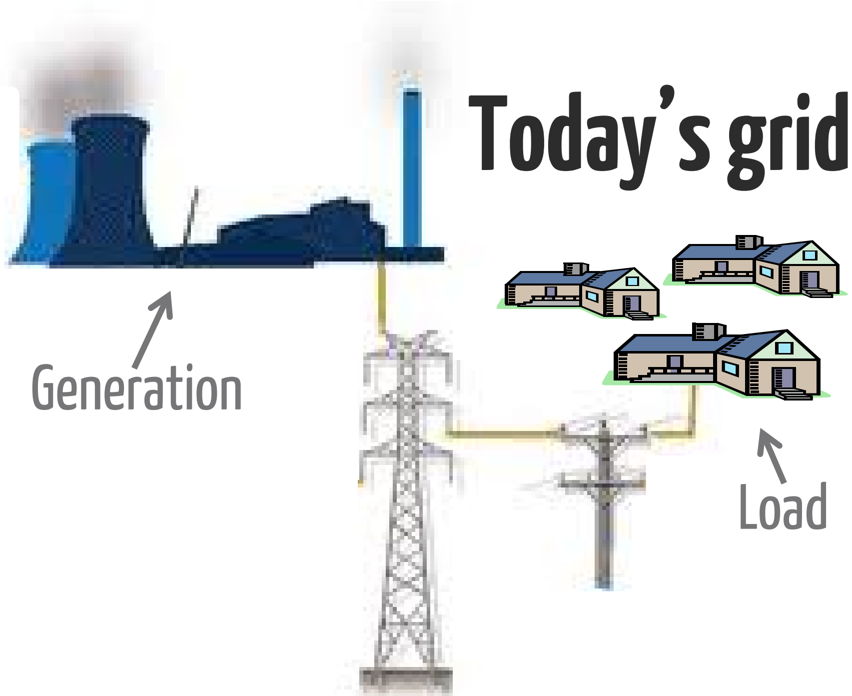
Key Constraint: Generation = Load

(at all times)



low uncertainty

Today's grid



Key Constraint: Generation = Load

(at all times)

⇒ Generation follows Demand

controllable
(via markets)

low uncertainty

Tomorrow's grid

Key Constraint: Generation = Load
(at all times)

less controllable

high uncertainty

low uncertainty

1) **Huge price variability**, *leading to generators opting out of markets!*
2) **More conventional reserves needed**, *countering sustainability gains!*

Key Constraint: Generation = Load
(at all times)

less controllable

high uncertainty

low uncertainty

"*Energiewende* has so far increased, not decreased, emissions of greenhouse gases."

Germany's energy transition

Sunny, windy, costly and dirty

Germany's new "super minister" for energy and the economy has his work cut out

Jan 18th 2014 | BERLIN | From the print edition

SIGMAR GABRIEL has been on a roll. The boss of Germa... entre-left Social ... herded his party into a ... lor Angela Merkel and ... or. He is jovial, ... vith the *Zeitgeist*. ...'s vision of work-life balance, he plans to take Wednesday afternoons off to pick up his two-year-old daughter from her crèche.

Gabriel in search of an energy miracle

But Mr Gabriel, who is mulling a run for chancellor in 2017, will by then be judged on a more daring project. As part of his coalition deal with Mrs Merkel, he is now a "super minister" combining two portfolios, energy and the economy. He is thus in charge of rescuing Germany's most ambitious and risky domestic reform: the simultaneous exits from nuclear and fossil-fuel energy, collectively known as the *Energiewende*, a term that means energy "turn" or "revolution".

More a marketing slogan than a coherent policy, the *Energiewende* is mainly a set of timetables for different goals. Germany's last nuclear plant is to be switched off in 2022. The share of renewable energy from sun, wind and biomass is meant to rise to 80% of electricity production, and 60% of overall energy use, by 2050. And emissions of greenhouse gases are supposed to fall, relative to those in 1990, by 70% in 2040 and 80-95% by 2050.

German consumers and voters like these targets. But they increasingly dislike their side-effects. First, there is the rising

1) **Huge price variability**, *leading to generators opting out of markets!*
2) **More conventional reserves needed**, *countering sustainability gains!*

Key Constraint: Generation = Load
(at all times)

⇒ Demand must follow Generation
(to some extent)

less controllable

high uncertainty

low uncertainty

Grid needs huge growth in demand response (or storage)

Data centers are a promising option

...they are large loads

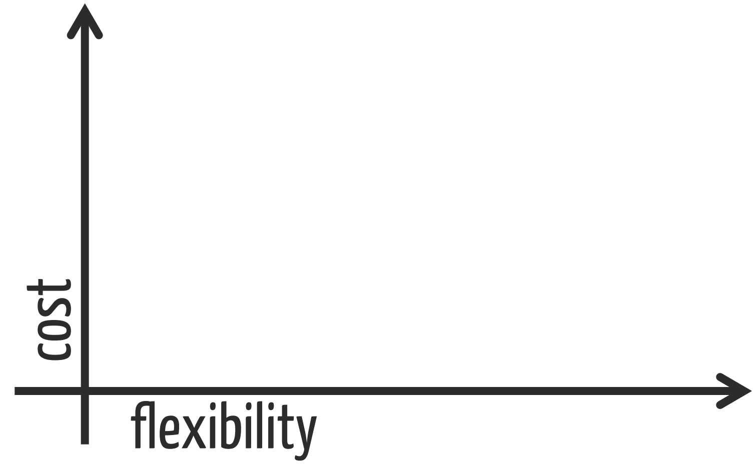
...usage is growing quickly

...highly automated

...they have significant flexibility

500 kW-100 MW each

10-15% growth/year



Data centers are a promising option

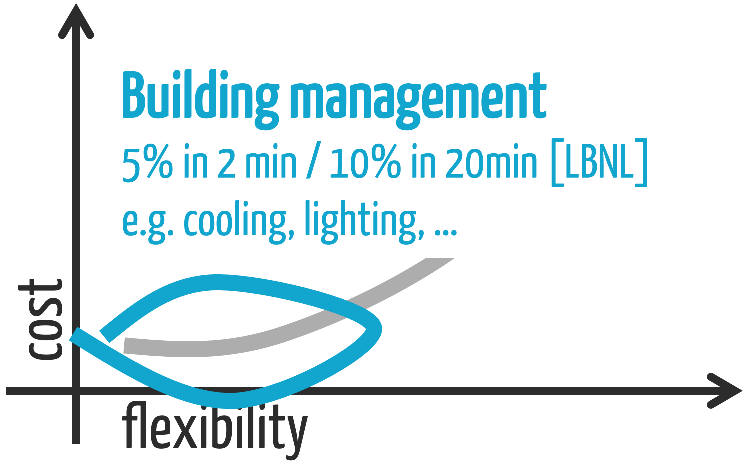
...they are large loads

...usage is growing quickly

...highly automated

...they have significant flexibility

10+ years of research into energy-efficient data centers



Data centers are a promising option

...they are large loads

...usage is growing quickly

...highly automated

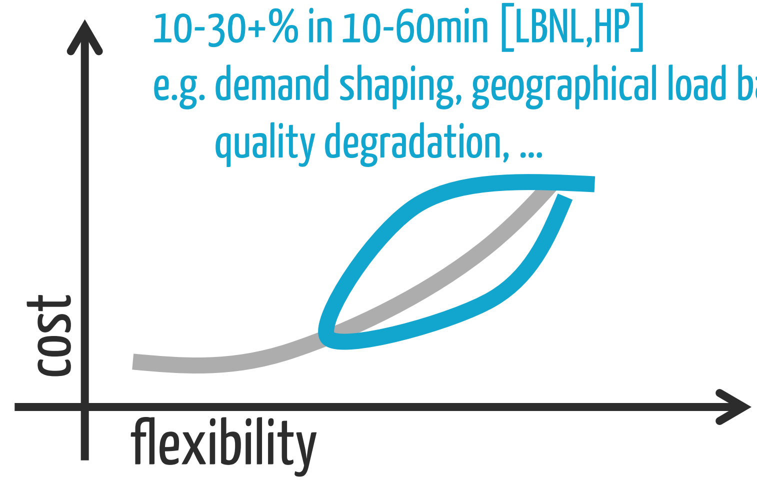
...they have significant flexibility

10+ years of research into energy-efficient data centers

Workload management

10-30+% in 10-60min [LBNL,HP]

e.g. demand shaping, geographical load balancing, quality degradation, ...



Data centers are a promising option

...they are large loads

...usage is growing quickly

...highly automated

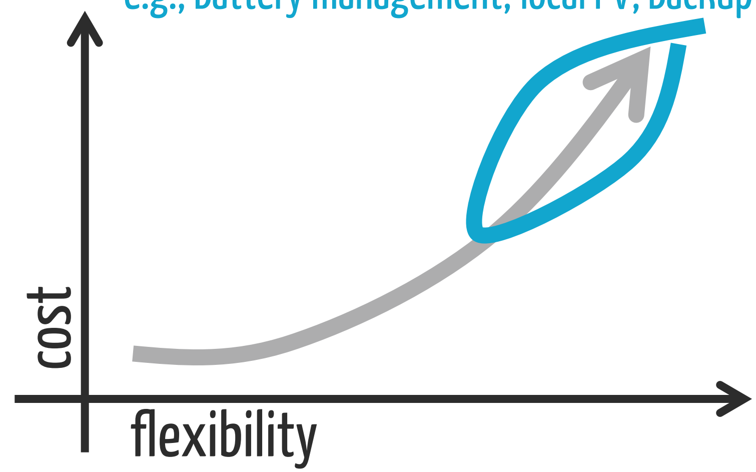
...they have significant flexibility

10+ years of research into energy-efficient data centers

Microgrid management

10-100% in 5-30min

e.g., Battery management, local PV, Backup generation



Data centers are a promising option

...they are large loads

...usage is growing quickly

...highly automated

...they have significant flexibility

10+ years of research into energy-efficient data centers

A new story about energy & data centers:

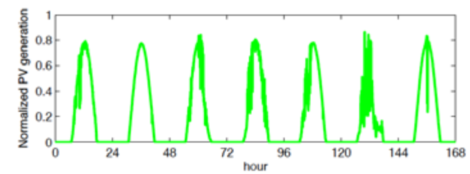
Data centers are valuable resources for making the grid sustainable

=

What is the potential of data center demand response?

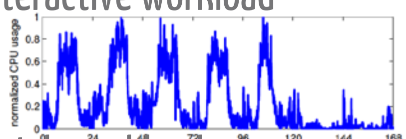
A case study:

PV

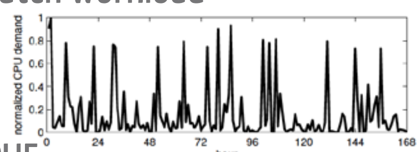


data center

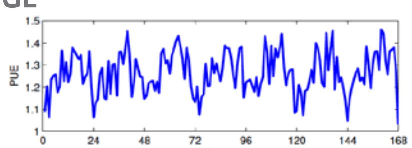
interactive workload



batch workload



PUE

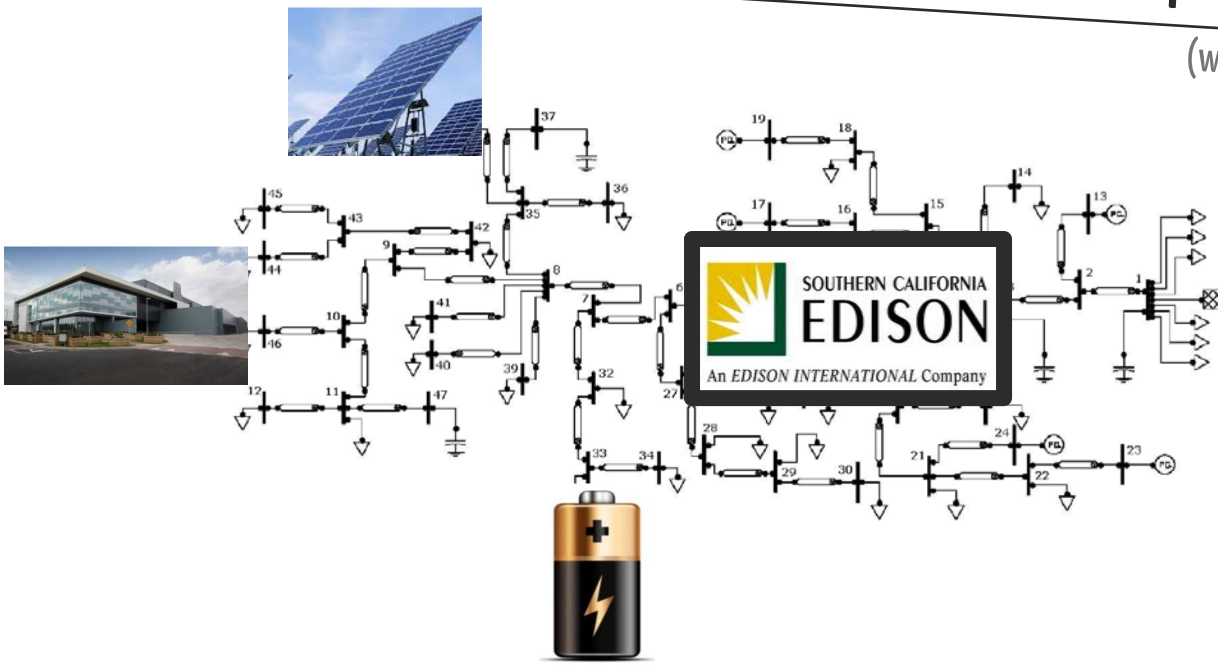


Optimally placed, fast charging rate storage

A case study:

20 MW Data Center with 20% flexibility = ???? kWh fast charging, optimally placed storage

(w.r.t. voltage violations rates)

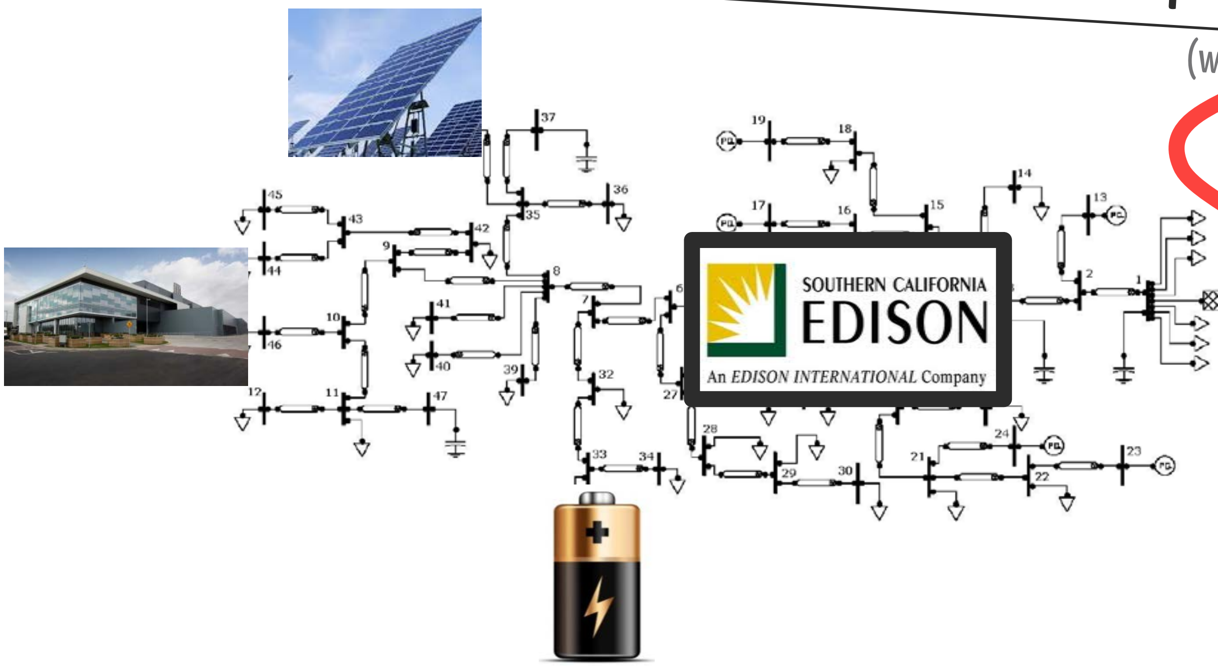


1 MWh if geographical load balancing is used!

A case study:

20 MW Data Center with 20% flexibility = ~~700 kWh~~ fast charging, optimally placed storage

(w.r.t. voltage violations rates)



$1-5 million cost!

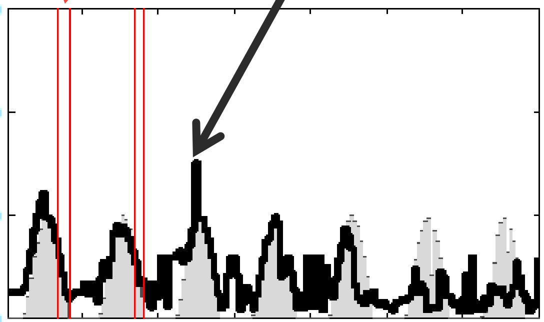
Where are we today?

- Time of use pricing
- Coincident peak pricing
- Wholesale markets
- Ancillary service markets
- Emergency DR

Data centers rarely participate

… and if they do it is highly inefficient coincident peak warnings

peak



Where are we today?

- Time of use pricing
- Coincident peak pricing
- Wholesale markets
- Ancillary service markets
- Emergency DR

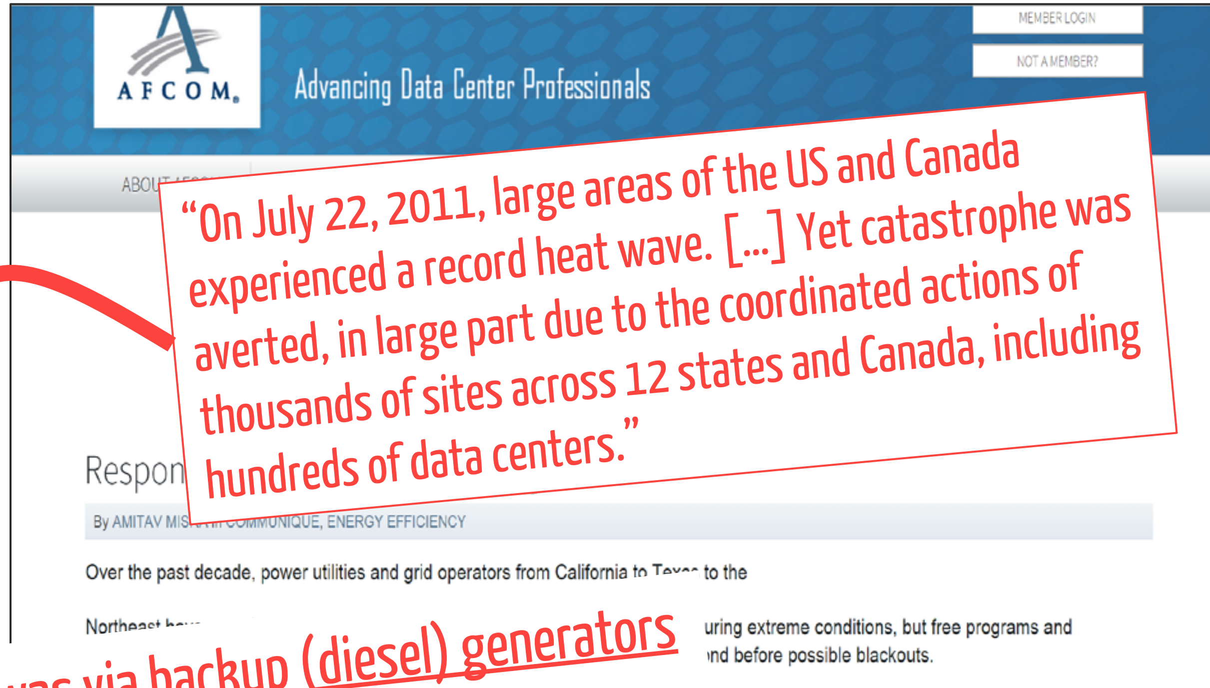
- Risky to participate
- Few opportunities for utility to extract response

For more see [Liu et al 2013]

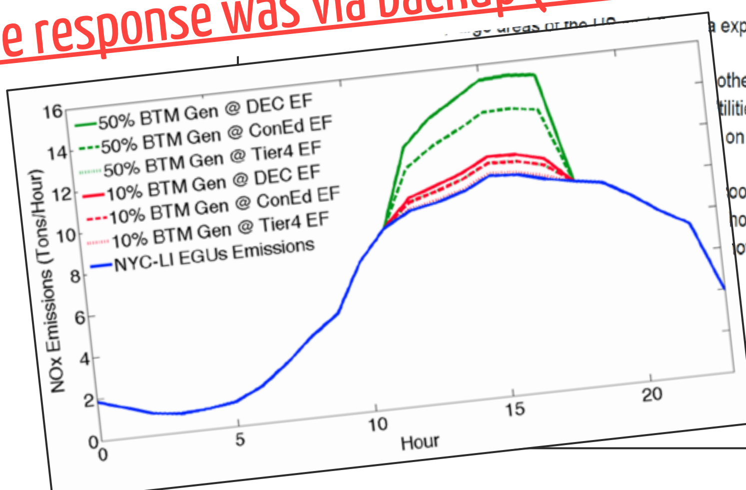
Where are we today?

- Time of use pricing
- Coincident peak pricing
- Wholesale markets
- Ancillary service markets
- **Emergency DR**

“On July 22, 2011, large areas of the US and Canada experienced a record heat wave. […] Yet catastrophe was averted, in large part due to the coordinated actions of thousands of sites across 12 states and Canada, including hundreds of data centers.”



BUT, the response was via backup (diesel) generators



How can we do better?

Engineering: **Algorithm design for data center participation**

[Camacho et al 2014], [Chen et al 2013, 2014], [Ghamkhari et al 2012, 2014], [Aikema et al 2012, 2013], [Irwin et al 2011], [Urgaonkar et al 2013, 2014], [Li et al 2012, 2013], [Liu et al 2013, 2014]

\+

Economics: **New market designs**

[Liu et al 2014], [Chen et al 2014],[Ghamkhari et al 2013], [Li et al 2013], [Wang et al 2014], [Ren et al 2015], [Wierman et al 2014], [Zhang et al 2015]

How can we do better?

Performance is priority #1

Engineering: ~~Algorithm design for data center participation~~

+

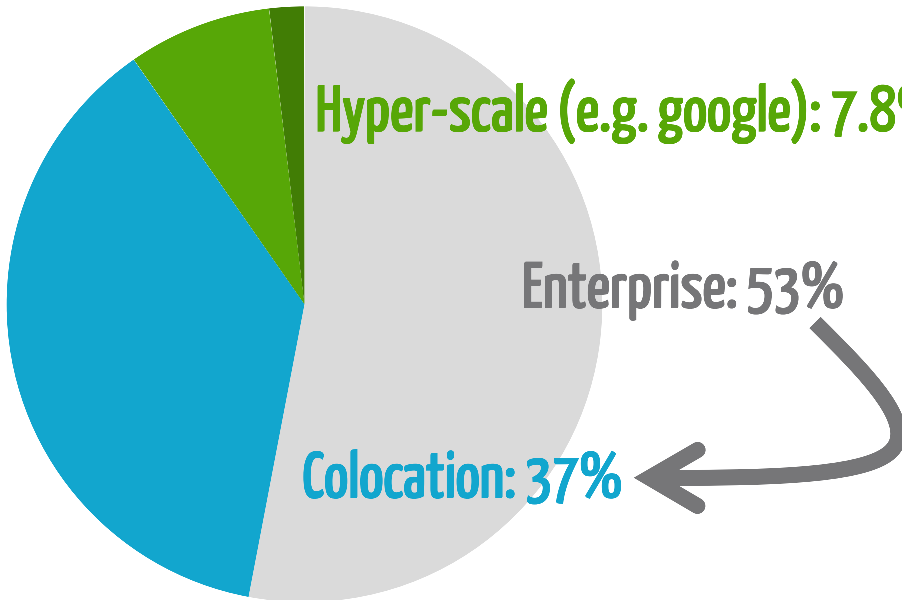
Highly regulated, change is difficult

Economics: ~~New market designs~~

...but, adoption represents a huge challenge

A starting point: Colocated (multi-tenant) data centers

A starting point: Colocated (multi-tenant) data centers



...of total data center industry electricity usage

Why colocated data centers?

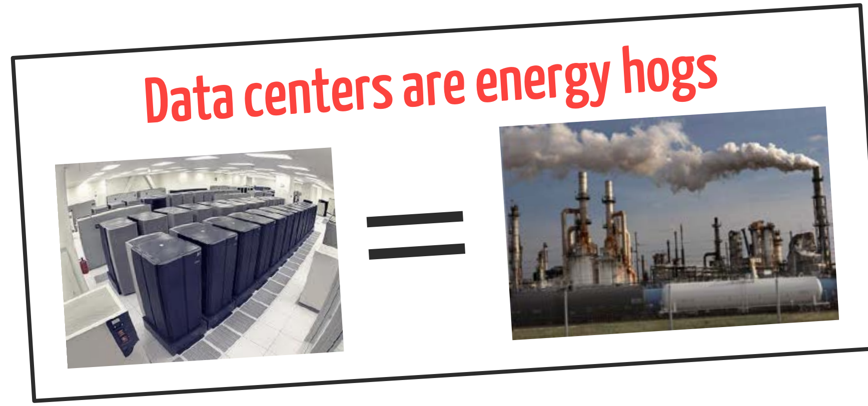
Building operation is separated from computing priorities
...but the data center would like to participate in demand response!

Set up incentives for tenants

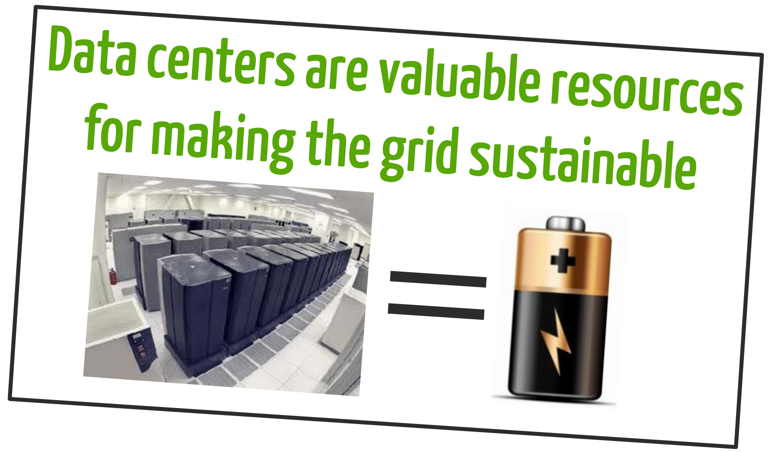
+ On-site generation provides backup!
+ Market power isn't an issue!
+ No regulation – can do whatever they want!
+ Tenants are heterogeneous in workloads!

Lot's of work to be done...

The typical story about energy & data centers:



But maybe we got it backwards?



Key points:

1) We need to move beyond a "myopic" focus on a data centers to consider a "system-wide" view of sustainability.
2) It is important to consider more than google-like "hyper-scale" data centers.

Data centers & energy: *Did we get it backwards?*

Adam Wierman, Caltech

cloudera

Five Challenges for Energy Efficient Computing Research

Yanpei Chen, Software Engineer, Performance Team

NSF Workshop on Sustainable Data Centers 2015

Our mission:

Cloudera helps organizations profit from all their data

Cloudera company snapshot

Founded	2008, by former employees of ORACLE YAHOO! facebook Google
Funding	$670M cumulative investment
Employees Today	800+ worldwide
Mission Critical	Production deployments in run-the-business applications worldwide – Financial Services, Retail, Telecom, Media, Health Care, Energy, Government
Diverse Customers	Majority of Fortune 100 companies are Cloudera customers
Cloudera University	Over 40,000 big data professionals trained
Open Source Leaders	Cloudera employees are leading developers & contributors to the complete Apache Hadoop ecosystem of projects

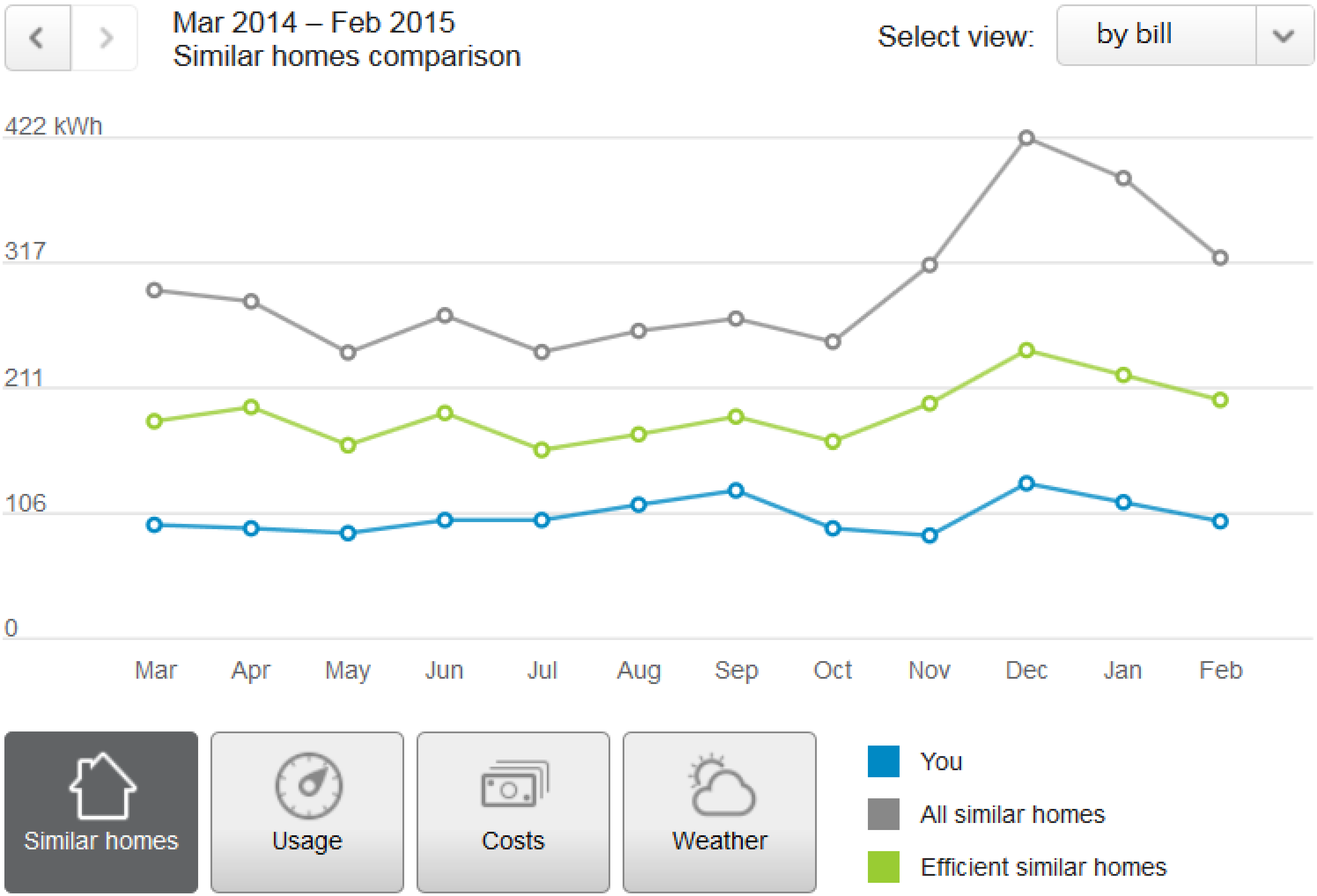
Protecting consumers from fraud

Credit card companies use Cloudera Data Hub to analyze timing, location, $ amount of transactions to distinguish normal and fraudulent behavior *for each customer*. Caught largest fraud case in a provider's history.

Improving neonatal care

The Children’s Healthcare of Atlanta uses Cloudera Data Hub to monitor 24/7 the light, noise, patient vital signs in their neonatal wing. Improved care by adjusting environmental factors. Found ways to improve pain management.

Reducing electricity use



Pubic utilities use Cloudera Data Hub to make visible residential electricity use at a per-hour granularity.

Led to behavior change that saved 2 terawatt-hrs globally in 2013, average 1-3% reduction.

Cloudera company sna

Three of Fortune 100 are consumer Internet companies: Google, Amazon, Apple. Cloudera and big data trace our technical heritage there, and consumer Internet is an important use case. There is also a much bigger world of “big data” beyond these companies!

Founded	2008, by forme
Funding	$670M cumula
Employees Today	800+ worldwid
Mission Critical	Production dep worldwide – Fina Care, Energy, Governm
Diverse Customers	Majority of Fortune 100 companies are Cloudera customers
Cloudera University	Over 40,000 big data professionals trained
Open Source Leaders	Cloudera employees are leading developers & contributors to the complete Apache Hadoop ecosystem of projects

cloudera

What Cloudera sees re energy efficiency

- Energy efficiency important to our largest (~5%) of customers
- All customers rapidly expanding clusters, sometimes 2 expansions per yr
- Hence expect growing interest in energy efficiency

How Cloudera can contribute

- Provide insights/workloads on customer user cases
- Influence how energy efficiency is measured in industry benchmarks
- Channel energy efficiency improvements into open source

My past work

- MapReduce energy efficiency (2009)
- Realized we can't run/measure stuff for real, nor at scale
- MapReduce workload capture & replay (2011) – 5x Cloudera customer workloads
  - Validated Hadoop fair scheduler (2011)
  - MapReduce energy efficiency (2012)
  - TCP incast fix validated on MapReduce (2012)

- Performance engineering at Cloudera - "make things go fast"
  - Worked on MapReduce, SQL-on-Hadoop, Search, Resource Mgmt, HBase

Challenge 2:

How can energy efficiency measurement methods capture realistic behavior?

Many hard design challenges arise from the dynamics of the serviced workload over time. Efficiency and energy efficiency measurement methods can distort technical merit when they measure an unrealistic good case or corner case.

cloudera

Challenge 3:

How do we design for common workloads given that we do not know what is common?

Companies such as Google, Facebook, Twitter have visibility into their own workload as a single case study. Vendors such as Cloudera see many customer workloads but do not and should not have access to their proprietary information. How do we progress and avoid a multitude of point solutions that do not generalize? This is a concern beyond energy efficiency.

Challenge 4:

How do we design for energy efficiency without distorting cluster operations in a way that vastly decreases business value?

Many energy efficiency techniques place constraints on cluster operations that vastly decreases business value. Such constraints relate to how data is placed, when do the jobs/queries get executed, and the performance of the jobs/queries. Energy efficiency must be achieved while still serving the customers' businesses needs.

Challenge 5:

How can we design for energy efficiency at large scale?

Many design challenges are challenging only at scale. Customers ask for proof-points on clusters of 100s of nodes. The demand for scale and the cost of proof-of-concept at scale increases continuously. How do we proceed? This is a concern beyond energy efficiency.

Challenge 1:

How is energy efficiency different from regular computational efficiency?

Even without considering energy, the more efficient system in the traditional sense will lead to the same hardware being able to serve a larger workload. Hence for the same workload increase, slower capacity increase, hence less energy spent. PUE approaches 1 means computational efficiency and data center efficiency converge. So do we need to be concerned with energy efficiency at all, or is it already fully incorporated within traditional measures of computational efficiency?

NSF SDC
Lightning Round

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Tarek Abdelzaher

Professor, UIUC

Cyber-physical Systems

Relevant past work:
Joint optimization of computing and cooling power

Key Challenge 1: *Energy storage*

- Maximize processing per unit of renewable energy
- Exploit energy storage to bridge supply and demand

Key Challenge 2: *Hardware-software co-design for ultra-low-power operation*

- Low-power embedded hardware has better performance per Watt.
- Data centers on embedded processors?

Key Challenge 3: *Data centricity*

- Predict and manage data workflows at minimum cost

Cullen Bash

Director, Hewlett-Packard Laboratories
Software Defined Systems

Relevant past work: Dynamic Smart Cooling,
Sustainable Data Centers, The Machine

Research Challenge: Comprehensive Sustainability Metrics

- Lots of work on thermal and energy metrics…
  … but metrics that include computational work are lacking
- Comprehensive metrics required for optimization.

Operational Challenge: Cost Effective Energy Reuse

- Numerous waste heat reuse examples, but most are building/campus scale.
- Use cases remain limited due to cost and complexity of installation.

Christina Delimitrou

Ph.D. Candidate
Stanford University

- **Relevant past work: Quasar, Paragon**

Challenge: Lack of Predictability at High Utilization

- Can get one at a time, but not both
- Current approaches work around unpredictability to improve utilization

Proposal: Datacenter System Stack for Predictability & Efficiency → Predictability by design, HW/SW co-design

- Resource isolation in hardware, partitioning techniques and/or hardware offloads
- Strip down OS to minimum necessary functions → Protection, not resource management
- Provide feedback to app designers on resource usage
- Do current APIs work? New interfaces?

Qingyuan Deng

Research Scientist & SWE, Facebook
Data Center and Server
Power Management

- Relevant past work: MemScale, CoScale

Challenge 1: increase server / power utilizations

- constraints: least / zero perf. interference & operation complex

Challenge 2: understand applications / services

- closing the gap with service owners
- what do they care: IPC, RPS, 99th-tail, latency, predictability?

Academia collaboration: faculties to work in industry (1~2 years)

- system scale difference
- student interns: inexperienced, NDAs
- NSF could provide incentives encouraging this

Michael Ferdman

Stony Brook University

Design of Efficient Server Systems

- #1 Challenge – establish ***accepted*** goals
  - Is progress without goals progress ?
  - Lots of work, but ***how relevant is it*** ?
    - uArch, utilization, power/cooling, TCO models, workloads, …
      - “datacenters” sessions at ASPLOS (2x!), HPCA, ISCA (2x!)
    - Is Google/Facebook/Microsoft 10% $ savings the only impact ?
- HPC is a nice role model
  - Petascale, Exascale, …
  - Makes technical challenges clear and ***sets a timeline***
    - Enables reasoning about relevance of work

No Goals, No Glory

ARIZONA STATE UNIVERSITY

Relevant work: BlueTool, BlueCenter, GDCSim, TACOMA

Sandeep Gupta,
Professor & Chair, CSE, ASU
Cyber-Physical Systems, Green Computing

Key Challenge 1: Non-Linear Spatio Temporal Variations

Cause: Non-energy proportional systems; variations in environment, workload, energy source; cyber-physical interactions

Need: a) algorithms to guarantee properties in presence of variance, b) non-linear optimizations, and c) managing operations for overall energy proportional system

Key Challenge 2: Green Energy provisioning in geo-scale systems

Cause: Rigid/Opaque electricity infrastructure; intermittency of renewables; non-linear inefficiency in energy storage units

Need: a) Smart grid with "Green API", b) hierarchical energy storage management, and c) new models of geo-distributed energy usage

Key Challenge 3: Discrepancy in simulation and practice

Cause: Lack of validation infrastructure, realistic workloads, energy traces

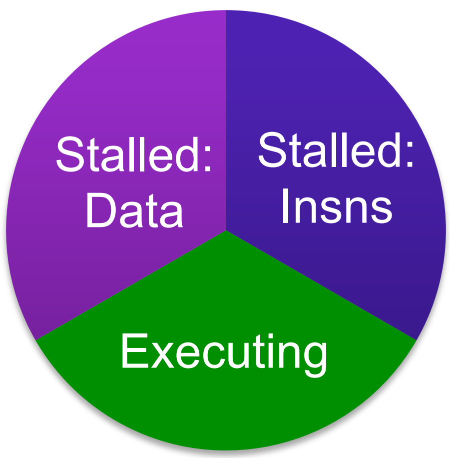
Need: Geo-distributed reconfigurable data center testbeds

Nonlinearities => Non-E.P => Non-Managable@Scale => UnSustainable

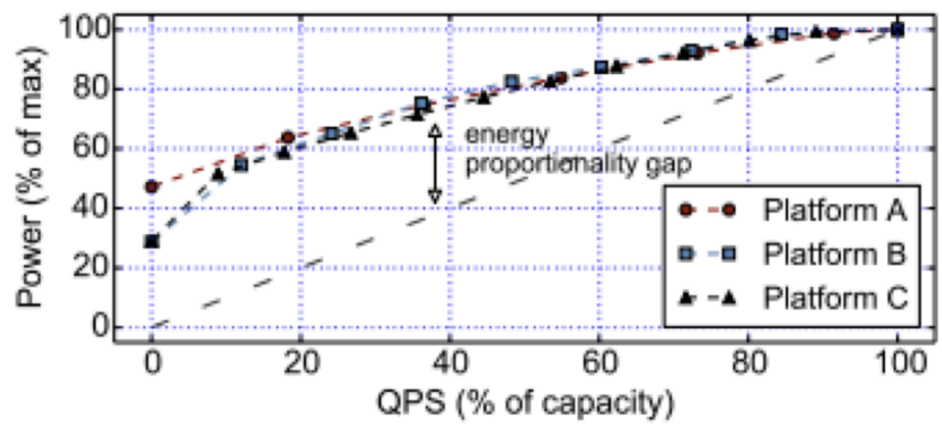
Kim Hazelwood
Director of Systems Research, Yahoo Labs

- Past lives: Google Platforms, Intel Pin Team, Associate Prof @ UVA
- Interests: Datacenter Performance, Power, and Price

My **Performance** Soapbox:
- Datacenters do NOT run SPEC!
- "The Rule of 3s"



My **Power** Soapbox:
- Underutilized machines make me sad



Collaboration between industry/academia is the ~~best~~ only solution

YAHOO!

Magnus Herrlin, Ph.D.

Program Manager
High Tech Systems Group
Berkeley Lab

Relevant past work: Energy and environmental analysis of telecom and data centers

- DOE Center of Expertise for Energy Efficiency in Data Centers
    - National leadership in decreasing energy use in data centers
- DOE Better Buildings Data Center Partners
    - Requires participating data center owners to report and improve their energy performance
- Energy Efficient Data Center Systems
    - Measure and manage
    - High-temperature liquid cooling
    - DC power

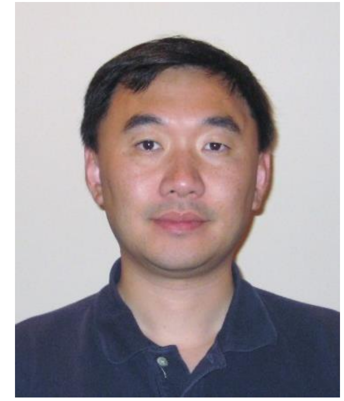
Kate Keahey

Scientist, Argonne National Laboratory
Senior Fellow, Computation Institute, University of Chicago

- Infrastructure Clouds
  - Nimbus: www.nimbusproject.org
  - First open source IaaS implementation
- Infrastructure Platforms
  - Leveraging elasticity to satisfy QoS goals
  - Sensor, social network based applications
- HPC and the Cloud
  - Cloud computing in HPC datacenters
- Experimental Computer Science
  - Leading the Chameleon Project: www.chameleoncloud.org

Jie Liu

Principal Researcher
Microsoft Research
Redmond, WA



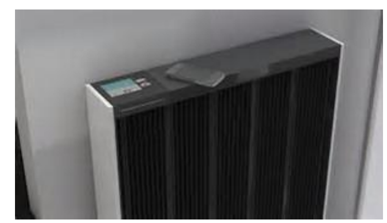
- **Relevant past work:**
  - Data center sensing
  - (VM) power metering and resource alloc.
  - Data furnace
  - Fuel cell powered data centers

Three pawns of sustainable DC

REDUCE REUSE RENEW

Energy Reused Data Centers

- Provision data centers where heat is needed.
  - End up with a low cost, but massively distributed cloudlets connected by slow networks
- It is not suitable for traditional big data workload, but is ideal for cognitive workload on sensor data
  - E.g. Processing 109 hours of video for object recognition generate enough heat to heat a house.
- How to coordinate centralized and distributed clouds
- How to make data and computing secure
- How to create an eco-system



Univ. of Nebraska–Lincoln

Ying Lu

Assoc. Prof., CSE Dept., UNL
Real-Time Systems &
Cluster Computing

- Relevant past work: energy-efficient scheduling in clusters

Key Challenge 1: Energy Efficient Public Clouds Adoption

- Smaller data centers: lack the incentives, resources, and expertise to investigate and adopt energy efficiency measures
- Continue accelerating the cloud computing adoption in both industry and academia

Key Challenge 2: Leverage Renewable Energy in Data Centers

- Intelligently manage workloads and available energy sources in future energy-efficient data centers
- Build data centers that adopt demand response schemes to dynamically manage their electricity loads in response to power supply conditions.

Mike Marty
Senior Staff Engineer
Google Platforms

Google

Focus at Google: high-performance IO

Does end of Dennard scaling spell Doom?

⇒ must get more out of transistors

Key Challenge #1: Increase Server Utilization

- Can double utilization with good control over queues and stragglers

Key Challenge #2: Reduce SWE cost for “bare metal” performance

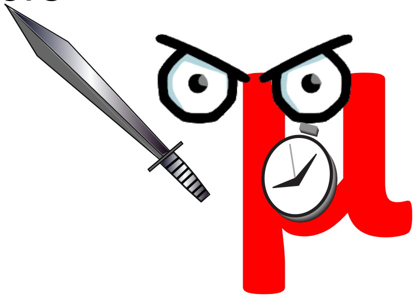
- Performance == Power

Key Challenge #3: Enable energy-efficient cores

- Amdahl’s Law effects may demand more tightly-coupled computing

Key Challenge #4: Figure out fine-grained Hardware Accelerators

- Another potential “killer microsecond” IO device



Hamed Mohsenian-Rad

Assistant Prof., UC Riverside

Energy Management

Coordination with Power System and Smart Grid

Questions: 1) Where do data centers stand in the Smart Grid paradigm for sustainability?

2) What is unique about data center power usage and load flexibility?

Passive Participation:

- Time-of-use pricing, real-time pricing, coincidental peak pricing, etc.
- Local energy resources, solar panels, energy storage, etc.

Active Participation (Interactions):

- Wholesale market bidding, energy, ancillary service, etc.

Key Challenge:

- Gap between Macro (System) level and Micro (Device) level research.
- Limited practicality, missing opportunities, etc.

Chris Page
Sustainability Director, Yahoo

- Key Challenge #1: Innovations in partnership between data centers & utilities: “prosumer”
  - Can DCs/cloud enable greater % intermittent power?
  - Can deployment of DERs improve grid efficiency/reliability?
- Key Challenge #2: Whole-systems impact of rise in mobile on the cloud
  - What’s the impact on: latency requirements, electricity demand (round trip), other?

Raymond Parpart

Senior Consultant, Data Center Strategy
IT Services, University of Chicago

Challenge:

- Can we schedule jobs more intelligently reducing power consumption and/or peak through variation in power consumption?
- How can energy costs be used to influence job scheduling matching system usage to the economics of energy

Top necessity for a sustainable datacenter agenda

Economic models capturing **business costs and benefit** of sustainable/energy-efficient datacenters (DC)

1. **Customizable, living models** for impact on the particular business operating the DC.
2. **Model the cost-benefit to each participant** in the DC operations**.**
3. **Comprehensive model** – flow of materials and energy, projected demands, integration into extra-DC operations such as co-generation, re-cycling, power generation/storage/ distribution.

Karthick Rajamani
IBM Research
karthick@us.ibm.com

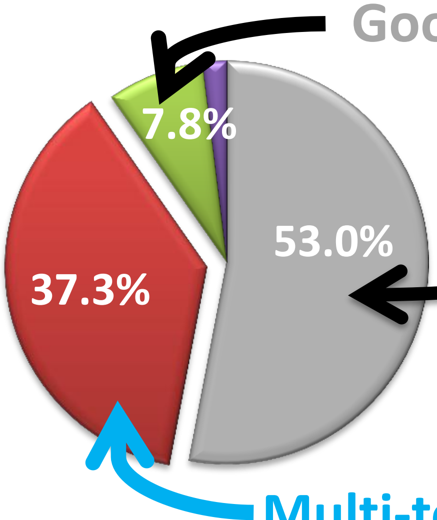
Cloud computing technologies and architectures
Relevant past work: Server and data center energy management.

Partha Ranganathan
Google

Shaolei Ren

Assistant Prof., FL International Univ.

- **Relevant projects:**
  - Power management in multi-tenant data centers
  - Improving water efficiency in data centers
- **Challenge 1:** Multi-tenant data centers are common but have been rarely studied
  - Tenants manage their own servers, while data center operator manages facility
  - How to coordinate tenants' power management for sustainability?



- **Challenge 2:** Drought is here and don't forget water footprint
  - Most data centers use cooling towers and hence are water-consuming
  - How to improve data center water efficiency without compromising other important metrics (e.g., cost, performance)?

Ali Saidi
ARM

Key Challenge 1: Slowing of memory improvements

- CPUs move to smaller and smaller process geometries but DRAM capacity & bandwidth have stalled
- Bandwidth, Capacity, Price: pick 2 (or maybe just 1)

Key Challenge 2: Increasing utilization while meeting SLA

- The opaqueness of the application/kernel boundary cause load imbalance and queuing
- Avoiding this queuing requires having idle time in system

Anand Sivasubramaniam

Professor, Penn State

http://www.cse.psu.edu/~anand

Relevant Recent Work:

Energy Storage for Cap-Ex (ASPLOS '12, ASPLOS'14, Sigmetrics'12) and Op-Ex (ISCA'11, Sigmetrics'11) savings

Key Challenge 1: Energy Storage – The what and where?

- Let's not settle for less and take what is given!

  Datacenters, the new "Tesla"

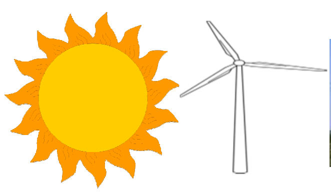
- Trade-offs:

  Power vs. Energy, Backup vs. Demand-Response

Key Challenge 2: Energy Storage – The How?

- Energy Storage needs to become one more resource

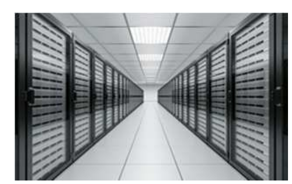
  Empower the software – to bank? which bank? for whom? when and how much to withdraw? …



Supply

Energy Storage

Demand



ReRout Lab
Computer Systems
Research That Reaches Out

Dr. Christopher Charles Stewart
www.cse.ohio-state.edu/~cstewart

Our planet is not stationary. Sustainable software will adapt.

Time to budget

Denard Scaling shrunk transistors without increasing power density per transistor

The overarching research challenge for 40 years has been, "*how can faster transistors speed up software?*"

Denard is done! Faster transistors increase power. The new research challenge, "which software provides most value per joule?"

Budgeting is important & hard

N software teams produce value v_i and power p_i, *which software should run in a data center that has power budget P?*

FCFS fails horribly. Clearly, we need models to predict value and power.

Greedy approaches work well provided good models. Until...

Sustainable budgets will change over time at every granularity

The efficiency of eco-system services varies over time

Research directions for sustainable computing

Cooling off	Cooler than a polar bear's toenails
Software optimized for commodity hardware	Adapt to diverse instruction sets, core counts, and cache topologies.
Optimization frameworks for a fixed budget	Online algorithms and migration
Operating systems that make the common case fast	OS that lets software choose the common case
Speed at all costs	Approximate & incremental computation

Prezi

Integrated Resource Management

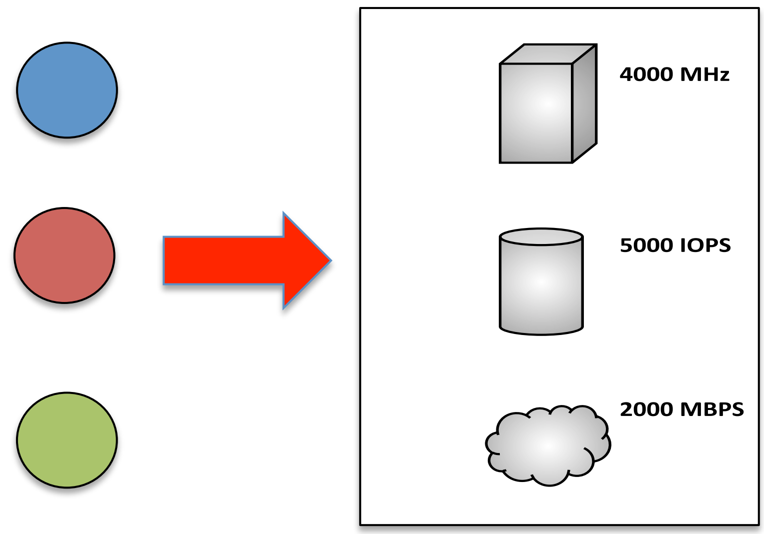
Resource management and scheduling in data centers

Datacenter models: What level of detail?

Workloads: Trace vs model ?

Objective functions

QoS models

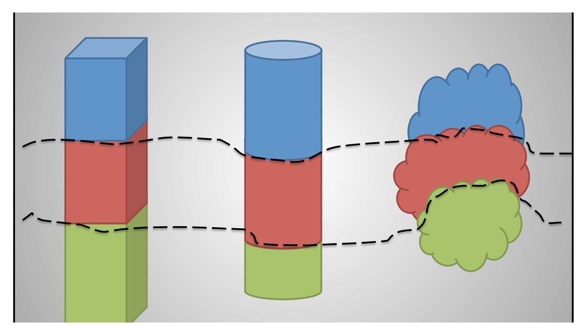


Converged Infrastructure

Computing + Storage + Networking

Heterogeneous Resources

Across stack at multiple scales



Integrated efficiency /fairness models

Peter Varman, Rice University

VMware and Sustainable Datacenters

Dahlia Malkhi, Michael Wei, Ravi Soundararajan

Key Challenges:

Continuing to do more with less (short term)

- More efficient resource usage throughout the datacenter
- Small-footprint VMs, fault tolerance, increased consolidation

Democratizing the Datacenter OS (medium term)

- Networking/storage virtualization enabling use of commodity components with lower power footprint
- Virtualizing more devices: bringing additional functionality to public clouds

Federating the Cloud (long term, speculative)

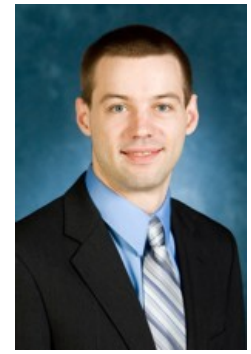
- Cooperative services among clouds
- "AirBnB" for Cloud

Xiaorui (Ray) Wang, Assoc. Prof. @ Ohio State

- Relevant past work
  - Feedback power control for server, rack enclosure (MPC), DC (SHIP), and CMPs
  - Thermal energy storage, data center network (DCN), power attack, GreenWare
- KC1: Maximize DC perf within power/thermal constraints
  - Dark Silicon: Many server cores must remain off.
  - Power Oversubscription: Host more servers within a DC.
  - Temporarily boost DC perf? **Data Center Sprinting**
    - **Phase 1 :** Safely overload circuit breakers (CB) for immediate sprinting
    - **Phase 2:** Additional energy from UPS batteries and, **Phase 3:** thermal tanks
- KC2: Minimize non-IT (cooling, DCN) power
  - How to coordinate emerging liquid cooling and free air cooling?
  - Proactive thermal prediction: Offline CFD analysis + online sensor readings
  - Optimize DCN power by consolidating traffic flows.

Thomas Wenisch

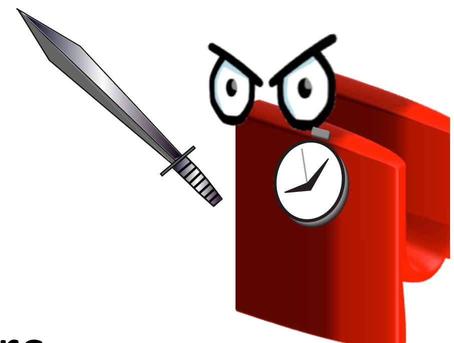
Assoc. Prof., U. Michigan
Server Architecture & Energy Management



- Relevant past work: PowerNap, Power Routing, MemScale

Key Challenge 1: Killer Microseconds

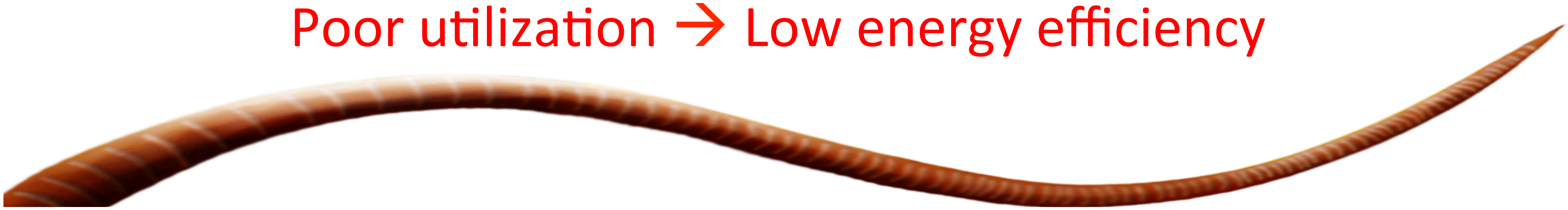
- SW and HW are great at handling ms and ns-scale stalls… … but no great mechanisms for µs-scale stalls
- µs-scale stalls common due to I/O, Flash, etc.



Key Challenge 2: Managing the Tail at Scale

- Rare events cause latency spikes in 99% tail
- Scale magnifies tails – must wait for the stragglers

Frequent stalls → Queuing delays → Poor utilization → Low energy efficiency



Qiang Wu

Facebook Inc., infrastructure software engineer

Key Challenges

1: Intelligent power over-subscription

2: Resource management for heterogeneous platform

3: Ensuring reliability and safety

4: Optimizing w/o increasing operational complexity

Prashant Shenoy
UMass

Renewables and Data Centers

- **How to reduce energy and carbon footprint of DCs?**
- Much emphasis on energy of IT infrastructure
- Renewable cooling in data centers
  - Direct air, evaporative cooling, hardware aspects
- Local renewables and grid interactions
- Challenge: run a data center using intermittent sources