Cisco Silicon One Web Scale Data Center Study

Introduction

Building a web scale data center is an expensive proposition which takes significant planning and orchestration by large companies. The scale of these data centers varies based on requirements, but they are typically immense construction projects housing 50,000 to 200,000 servers on a single site. Interconnecting all these servers requires a significant amount of fiber, optics, and networking equipment to allow the servers to communicate within the data center, with other data centers, and with the general Internet.

In this paper we will attempt to quantify the power and cost savings customers can enjoy building these massive data centers with Cisco Silicon One™ versus alternate silicon available on the market today.

Contents

Importance of power efficiency

Power efficiency is critical to our web scale customers for several reasons:

- The amount of power available to a data center is established during initial power distribution design and construction. Adding power capacity later is cost prohibitive. Once built, data center processing capacity can only be improved by higher efficiency equipment.
- Data centers are built to remove a certain amount of heat from servers, networking gear, and switches. Today, the heat from this IT equipment is removed by industrial-scale air conditioning. Retrofitting a building to change air handling capacity may not be possible or can become quite costly.
- The air conditioning and chiller equipment used to remove heat requires significant power itself, adding yet more expense.
- Many web scale service providers have committed to powering their facilities exclusively with low-carbon electricity.[1,2,3,4] Higher energy efficiency means less power must be provided from onsite solar, purchase power agreements, or renewable energy certificates.

Smart phones and tablets, increased wireless and wireline speeds, and cloud services have created substantial economic opportunity around the world. The increase in demand for IT services could have adverse environmental impact due to increased energy consumption and Greenhouse Gas (GHG) emissions. Fortunately, cloud services are very efficient, making use of the latest facilities design, equipment, and operating practices, as well as frequently being powered by low-carbon electricity. Meanwhile, companies have discovered business benefits in transferring enterprise IT to this new cloud-compute ecosystem. Because of improved utilization, aggregated cloud services are inherently more efficient than individual, company-scale data centers, exerting a downward pressure on overall energy consumption by IT equipment. Cisco wants to maximize the energy and carbon efficiency of web scale data centers so that both emerging and developed markets benefit from new IT services while minimizing environmental impact from GHG emissions.

Within a web scale data center, minimizing energy consumption from networking equipment is especially important. The more power spent on networking equipment, the less power is available for servers and storage. At first glance, how power is allocated might not seem like a major concern. However, if you consider the fact that these companies generate revenue with services residing on servers and that networking is used to connect them, then we can see that servers are viewed as "revenue generating" while networking is considered a "tax" to deliver those services. Therefore, for every watt spent on networking, that's a watt that can't be applied to servers, thus decreasing the total revenue generated by the data center. From an environmental and business perspective, improving networking energy efficiency is a priority for our customers.

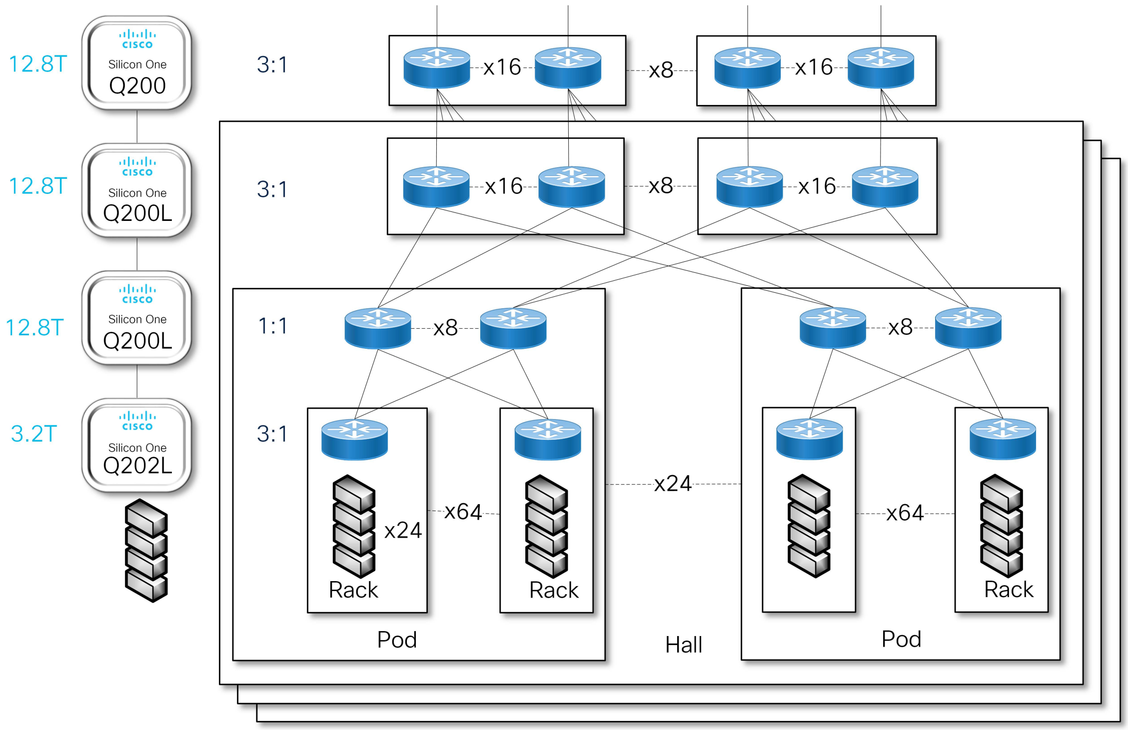

Figure 1. Hyperscale data center topology

Network topology

When trying to quantify power savings, we must first define the network topology. Each web scale vendor builds data centers in a unique way, and this knowledge is considered highly confidential. For this reason, Figure 1 only shows a representative topology housing 110,000 servers, with a 3.2 Tbps Top-of-Tack (TOR) switch, 12.8 Tbps leaf and spine boxes, and a 12.8 Tbps Data Center Interconnect (DCI) box with 3:1 oversubscription in the TOR, spine, and DCI layers.

As shown in Figure 1, we will use Cisco Silicon One Q202L for the TOR switch, the Q200L for the leaf and spine switches, and the Q200 for the DCI router where buffering and large scale are required. In Figure 2, the data center topology is redrawn to better depict the interconnect and oversubscription layers.

Based on this topology we can see that there are 110,592 servers and 6,144 TORs. These connect to 768 leaf nodes with 100GE. These connect to 512 spines with 400GE and the spines connect to 128 DCI boxes with 400GE. The DCI boxes use 400G-ZR to connect to the Wide Area Network (WAN) located within the metro region.

Additionally, we show how to build the same network topology using other silicon available on the market, which clearly highlights some of the key advantages of our solution:

- Cisco Silicon One offers a single unified architecture with a common semantics, so the Software Development Kit (SDK) and P4 forwarding code can be deployed across the entire data center. With other architectures three or four unique silicon architectures are needed, thereby increasing operational complexity.
- A single Cisco Silicon One Q200 can enable a 32x400G router while competitive silicon requires at least two pieces of silicon, driving significant cost and power implications at the DCI layer.

In this paper, however, we will focus more on the implications of the highly power efficient architecture of Cisco Silicon One and what that means to the end customer.

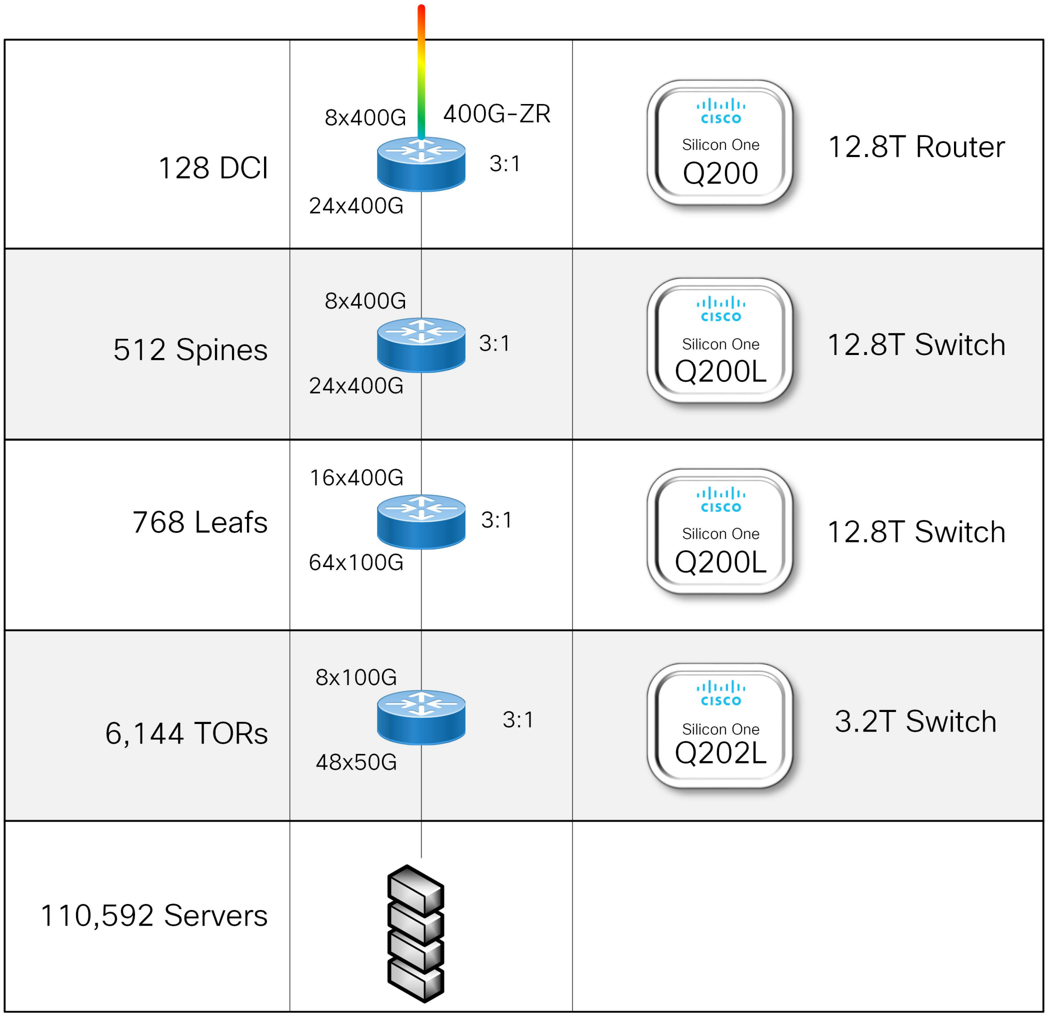

Figure 2. Silicon used for network model

Life of a watt

It's interesting to trace how a watt is consumed from an individual component within the system to see how it's delivered and how it impacts the surrounding area.

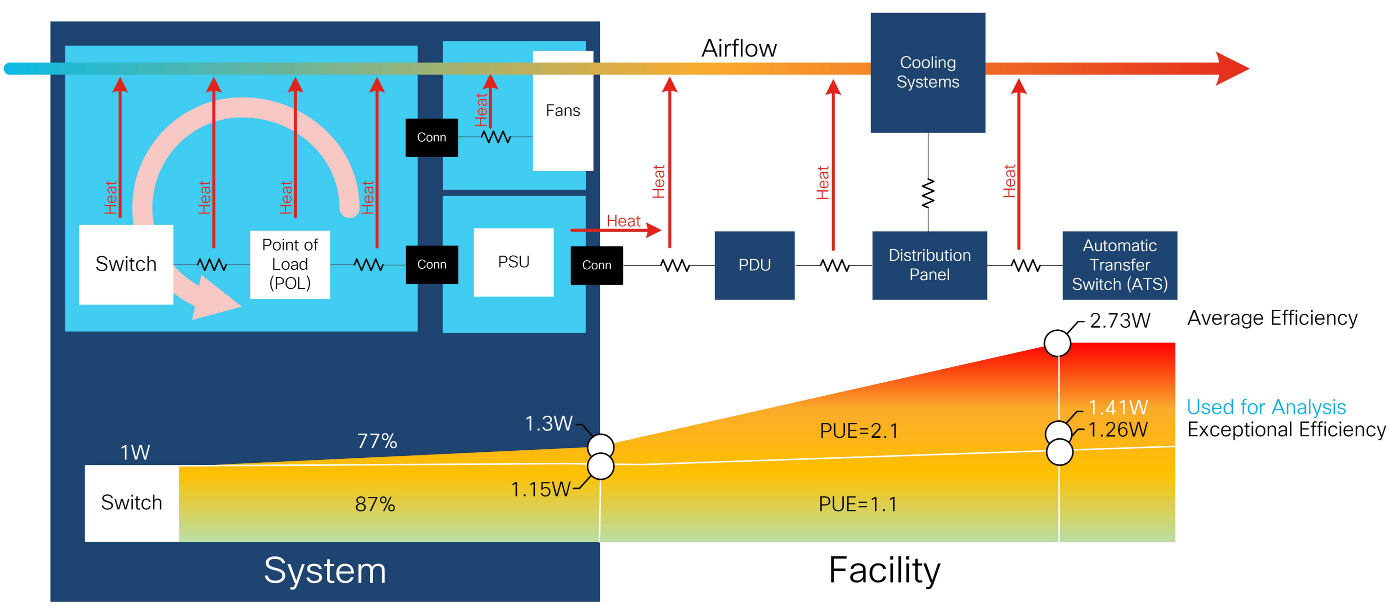

Figure 3. Life of a watt

To the left of Figure 3, we can see a switch that needs to consume power to forward packets. For every watt the switch consumes, the switch needs to pull current through the Printed Circuit Board (PCB) from a Point of Load (POL) power converter. The POL is responsible for converting an internally distributed 12V or 54V to the voltage that the switch needs to operate. As the current flows through the PCB, the resistance of the PCB causes heat to be generated, consuming energy. Additionally, stepping down the voltage in the POL consumes energy causing a loss of efficiency.

The POL then needs to pull current from the Power Supply Unit (PSU) in the system. The PSU is responsible for converting the facility power to the internal voltage used within the system. As seen in Figure 3, the current draw through the PCB and connector and the conversion within the PSU have an additional efficiency loss.

The consumed power generates heat in the system which must be cooled to keep the components within their allowed operating conditions. The systems use fans to pull cold air from the front of the device across the components to exhaust at the rear. The more heat that is generated, the faster the fans need to spin, which increases the fan power in an exponential relationship to airflow.

A secondary critical effect is that as the PCB heats up the resistance of the material increases, causing more heat to be generated. Similarly, transistors in silicon draw more power the hotter the silicon gets. Both effects create a feedback loop. As power is consumed, heat is generated, which increases the amount of power draw dissipating more heat. This means the fans need to run faster to ensure that they don't end up in a positive feedback loop resulting in thermal run-away.

A single watt used by a device requires 1.15 to 1.3W pulled from the facility depending on the efficiency of the system power distribution and thermal solutions and the ambient air temperature and facility altitude.

From a facility perspective, similar sources of power loss exist. Power is delivered from the facility's Automatic Transfer Station (ATS). Power is then distributed through the facility's wiring infrastructure and finally to the Power Distribution Unit (PDU) which delivers the power directly to the system. Each of these steps consumes energy, creating heat in the facility. The facility needs to remove the heat from all the equipment and any heat generated by the facility itself with its cooling systems.

Each facility's design is different, but an industry wide term used to describe the facility's efficiency is its Power Usage Effectiveness (PUE). The PUE is the ratio of total facility energy used compared to the energy used by just the compute, storage, and networking IT equipment. The most advanced data center designs have achieved a PUE of approximately 1.1, which means that for every watt delivered to a system, 1.1W is provided to the facility. Typically, data center designs have PUE ranging from 1.1 to 2.1, while some older facilities have a PUE as high as 3.0.

Taking all these effects into account we can show that for every watt consumed in the Application-Specific Integrated Circuit (ASIC) the facility will consume 1.26W to 2.73W. The power consumption of the switching silicon can have a dramatic effect in terms of the total power required in the facility. If we can save 100W on a piece of switch silicon, then we could save 126W to 273W from the facility for every switch deployed.

In this analysis we will assume highly efficient modern system and facility designs. We will assume a total system efficiency of 85 percent and a facility PUE of 1.2. These are close to industry leading numbers and together these mean that for every watt consumed by the switch chip 1.41W will be consumed by the facility.

With Cisco Silicon One we are hyper focused on power efficiency to deal with these important impacts for our customers.

Power savings with Cisco Silicon One

Cisco Silicon One Q200, Q200L, Q201L, and Q202L are highly optimized devices which deliver high bandwidth and low power. To highlight the benefits of using Cisco Silicon One, we developed approximate models for the systems at each level of the network which includes silicon power draw, power plane loss, conversion efficiencies, and fan power draw.

What we found is that across the TOR, leaf, and spine layers Cisco Silicon One may reduce power draw by up to 27 percent, while at the DCI layer power draw maybe reduced by up to 66 percent.

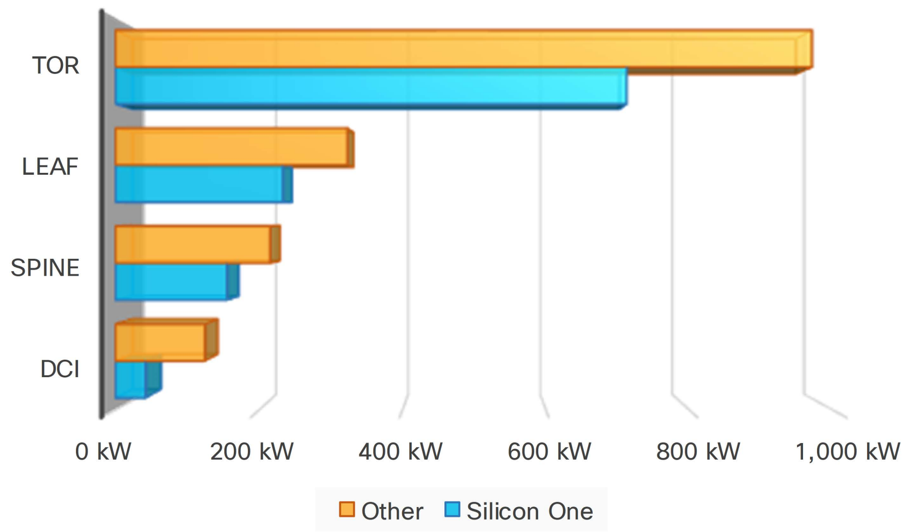

Figure 4. Power network layer

Combining all the layers, we can see that up to a 30 percent reduction in the power may be possible with Cisco Silicon One, saving up to 580kW of power

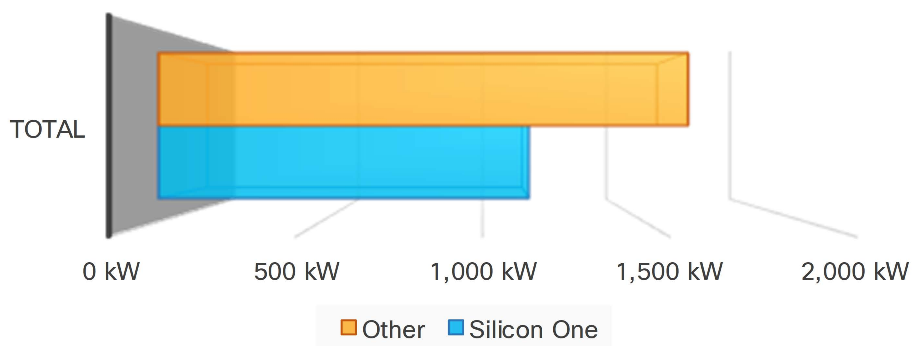

Figure 5. Total power for the network

According to the U.S. Energy Information Administration (EIA) the average commercial price per kilowatt hour (kWh) is 10.67 cents[5]. Based on the 580kW savings over a 10-year period, $5.4 million can be saved in electricity[6]. To help drive down costs, many web scale providers build facilities in low cost electricity regions like Iowa, Nevada, Washington, New Mexico, and Texas. According to a study by the Site Selection Group in 2017, the price per kWh can be as low as 4.68 cents in these areas[7]. Building the facilities in these regions may bring the 10-year electricity cost down to $2.4 million.

Above and beyond the savings in electricity, building a data center with Cisco Silicon One also saves you money during construction. Because the power is significantly lower than other solutions, the amount of electrical wiring and distribution infrastructure can be reduced. According to Tuner & Townsend's data center cost index 2019[8], building a 30MW data center costs between $7.1 and $8 per watt in the United States. This means that an additional $4.1 million to $4.6 million can be saved in construction costs. Assuming the average between these

For more information

Visit [Cisco Silicon One]

points, the construction cost is $4.35 million. Putting these together we show that simply by using Cisco Silicon One customers may save a between $6.8 million and $9.75 million over a 10-year period depending on the location of the web scale data center.

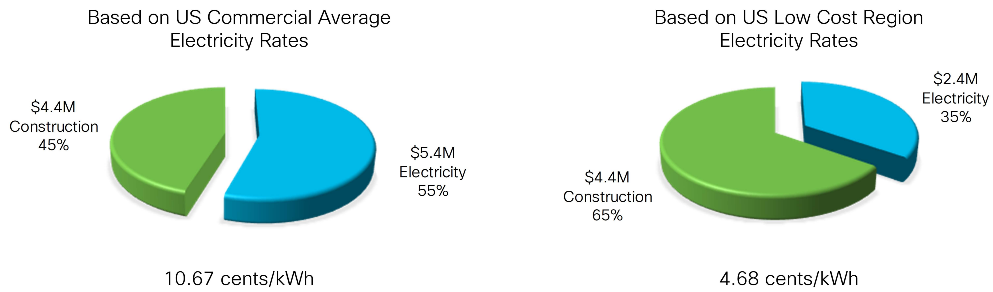

Figure 6. Breakdown of 10-year cost savings

Another way to think about this is to convert the 580kW power savings into additional revenue-generating servers. A typical server consumes between 400W and 1,000W of power. Assuming a 500W server and facility PUE of 1.2, an additional 966[9] servers could be deployed at the facility. With a facility being able to house 110,000 servers this is nearly a one percent increase in server count, and therefore revenue from the facility by one percent.

Conclusion

Because of the power efficiency of Cisco Silicon One, customers can potentially save up to $9.75 million over a 10-year period or increase their revenue potential by one percent. In addition, Cisco Silicon One provides industry leading performance, buffering, programmability, and scale while delivering an end-to-end architecture spanning from the web scale TOR through the WAN.

References:

1. Microsoft believes buying clean energy...
2. Apple Inc. Officially in the green
3. Facebook Sustainability, 02 / 08 Renewable Energy
4. Amazon Renewable Energy
5. https://www.eia.gov/energyexplained/electricity/data-and-statistics.php
6. 580 kW * (24 hours * 365 days * 10 years) * 0.1067 dollars
7. https://info.siteselectiongroup.com/blog/power-in-the-data-center-and-its-costs-across-the-united-states
8. https://www.turnerandtownsend.com/en/perspectives/data-centre-cost-index-2019/
9. 580,000 watts / 1.2 PUE / 500W