Importance of Architectural Fidelity

Cisco Silicon One

Simplifying the architecture

Unlike other devices which require disjoint and incompatible semantics to address the service provider and web scale networks, Cisco Silicon One offers a fully unified architecture. At first glance people may assume this means a single device, but what we mean is multiple devices which share the same DNA. This allows Cisco to offer best of breed devices for each role in the network while still offering true convergence across the service provider and web scale networks.

Table of Contents

Networks have always been built with multiple devices, each one having a unique silicon architecture. There are many reasons why this happened, but in the end, it boils down to a few key points:

1. Existing silicon architectures have been around for decades, evolving over time to their current states. Some of them came from small startups which were acquired by larger companies, but the teams continue to work in isolation. Over time the investments in these architectures become so large that too much "technical debt" is accumulated, forcing the continued investment in these aging architectures.
2. Understanding the intricacies of all the networking markets is a daunting task, so most teams only focus on a portion of the market.
3. The common wisdom is that the market segments are totally different, requiring several architectures to address them.

Each of these reasons alone is enough to avoid tackling the problem, but together they create almost an insurmountable hurdle.

Cisco Silicon One has the unique advantage of being built by the leading silicon team in the industry, while being part of the only company with a broad enough understanding of networking to recognize the true requirements across the industry.

After more than a billion dollars and six years of innovations, Cisco Silicon One allows a single architecture to not just span multiple market segments but do so optimally. With one architecture we cover:

1. Service providers and web scale Wide Area Networks (WAN) which require high bandwidth, highly programmable, deep buffered, high scale, and carrier grade routers

2. Web scale data centers which require the highest bandwidth, most power efficiency, and lowest latency switches

An architecture specifies the work partition of a larger task like switching into small blocks and the language these blocks use to communicate with each other and to the outside world

What is an architecture?

The notion of a single architecture can be confusing, and it's often misunderstood. Some naysayers attempt to interpret this as a single chip spanning all market segments and position multiple unique architectures against this interpretation.

With one simple observation, it's easy to realize that one chip can't span the entire space. Web scale requires high bandwidth switches with 12.8Tbps to 51.2Tbps, while service provider access routers require 800G of bandwidth with large buffers. One chip cannot possibly cover this entire space efficiently.

But what if several chips born from the same architecture could?

To better understand this, we must first understand what an architecture really is.

An architecture specifies the work partition of a larger task like switching into small blocks and the language these blocks use to communicate with each other and to the outside world through the Software Development Kit (SDK). Once established, these blocks can be placed into one chip or across multiple chips while still providing a consistent experience to the user.

To be explicit, an architecture is not a single chip, and it's not an implementation of an individual block of the chip, it's how the language the blocks use. Although this definition is simple, it's something that has been missing in the networking silicon industry since its inception.

Silicon One architecture

Looking deeper, Cisco Silicon One defines blocks such as port logic, receive processing, transmit processing, databases, telemetry, traffic management, and control plane. Cisco Silicon One defines how these blocks work together to perform the larger task of routing and switching and how SDK APIs manage these blocks in a consistent fashion.

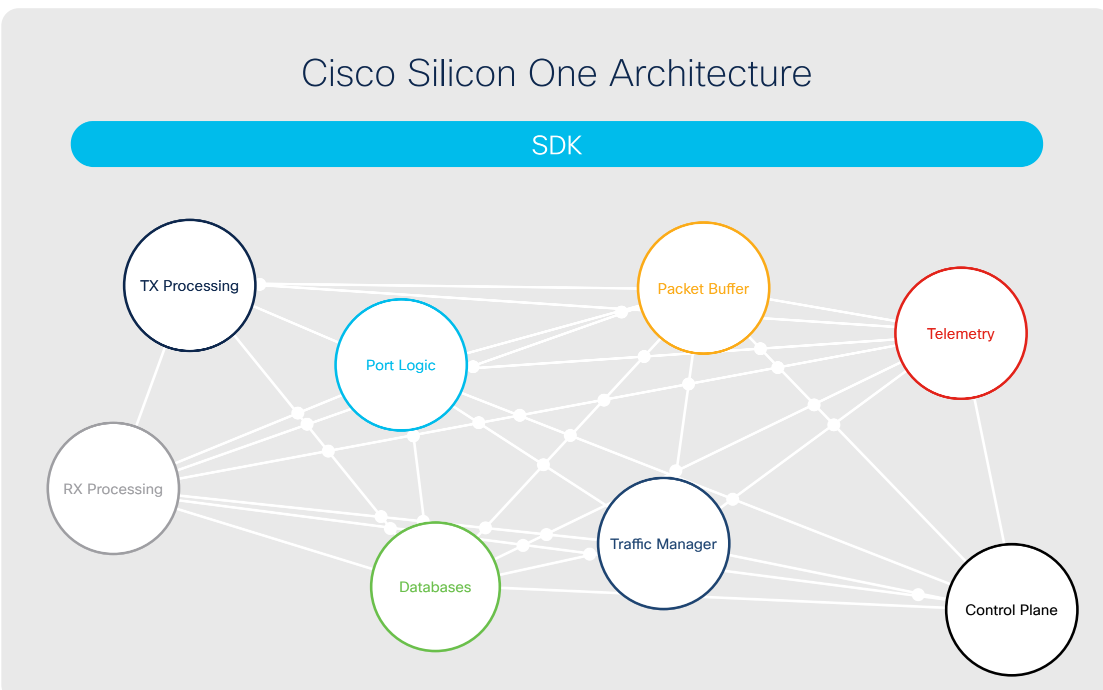

Figure 1. Silicone One architecture

Once defined, creating fully optimized and interchangeable blocks becomes possible. For example, the traffic manager block can be optimized for a simple DiffServ traffic manager supporting just eight queues per port, a traffic manager supporting eight queues per sub-port, or a traffic manager supporting a complex hierarchical scheduling. This general notion of fully optimized blocks can be applied to all the blocks, allowing for a multitude of possible combinations.

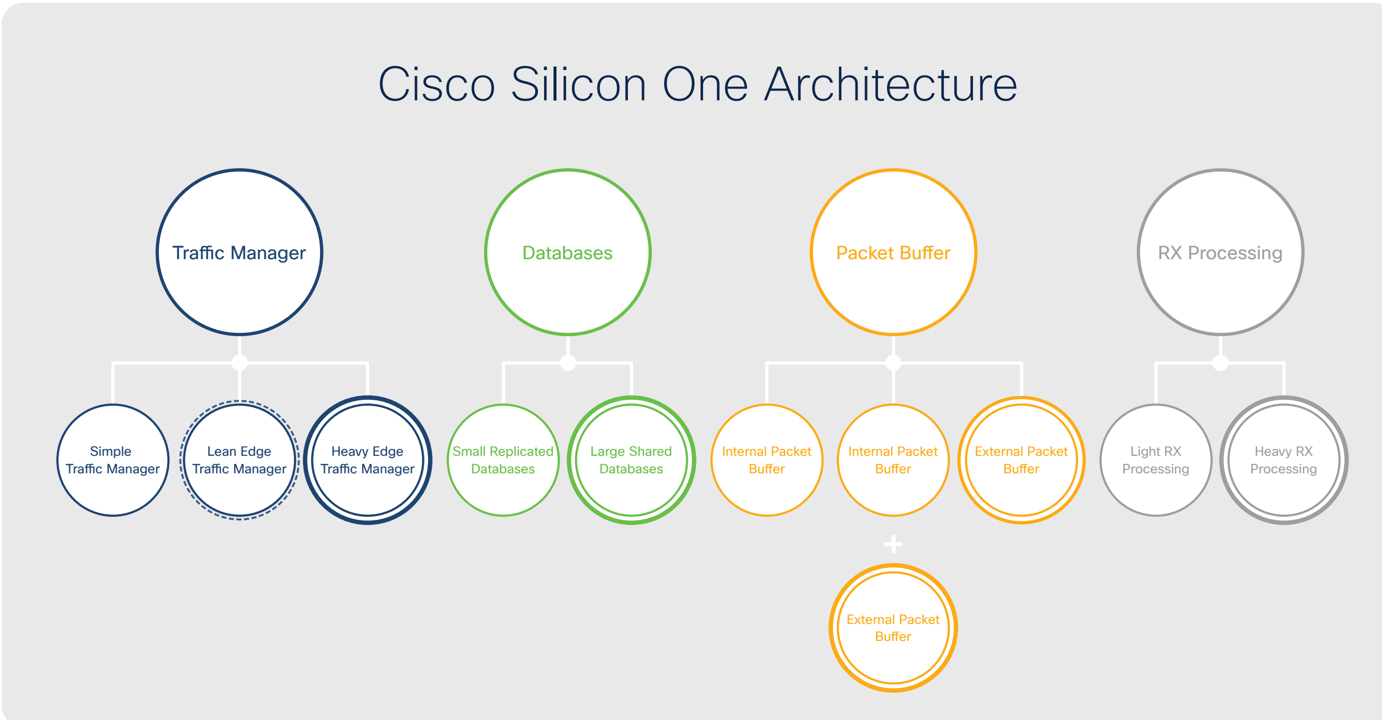

Figure 2. Interchangable blocks

Consider two devices following the same architecture. The first device is a highly programmable, high scale, large buffered routing device targeting the service provider market. The second device is a lean and mean switching device targeting the web scale switching markets. Each device can be optimized for their market but still be within the same architecture.

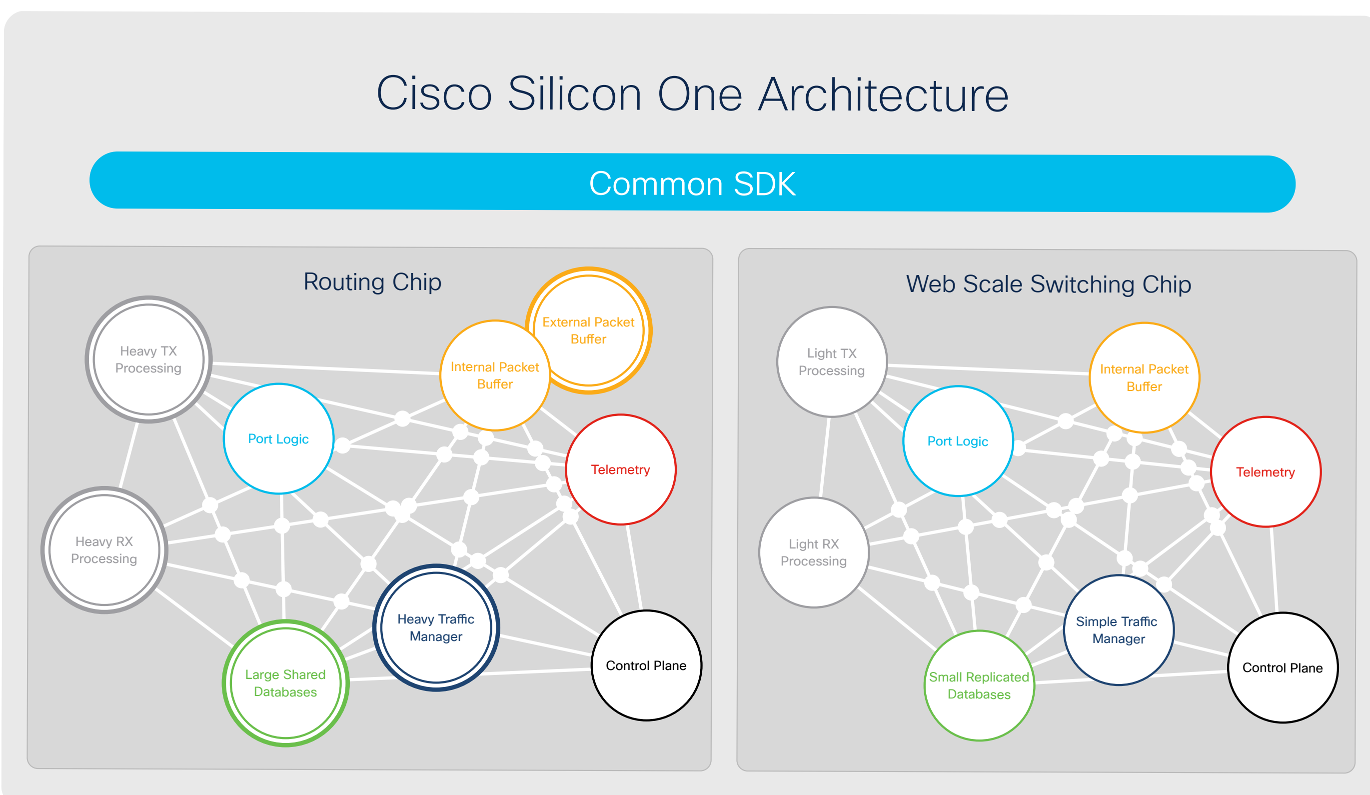

Figure 3. Multiple chips, one architecture

Learn More

Visit Cisco Silicon One

Silicon designers need a wide enough toolbox to build optimized chips, but from a customer perspective the chips look and feel familiar, presenting a common software abstraction layer. Once a customer understands the Silicon One architecture it's easy to migrate from one optimized chip to another.

This paradigm is vastly different from what exists today in the market. Today a customer who understands one architecture has a very steep learning curve to consume a different architecture.

Impact of a converged architecture

It's hard to overstate the implications that a truly converged architecture can have on the networking industry. Customers building and operating networks can enjoy the same experience across their network. There is one implementation of SRv6 stretching from within their web scale data centers and across their entire WAN, enabling them to learn, deploy, and manage one architecture everywhere in their network. Vendors building equipment can build product lines which easily span across markets. Software engineers port a single SDK and write their network operating system once. The summation of these simplifications results in massive Capital Expenditures (CapEx) and Operating Expenditures (OpEx) savings across the entire ecosystem.

On top of these customer advantages, the silicon team within Cisco can focus its vast number of silicon engineers into a single architecture driving true innovation across the web scale and service provider markets instead of replicating development for no true end-customer value.

Conclusion

Although it's tempting to believe that the errors of the past imply fundamental limitations, it's always a good idea to challenge contemporary thinking. Cisco has invested heavily over the last six years to bring a fully unifying architecture to the market while simultaneously delivering best-in-class devices.

09/20