

[Newsroom Home](#)[| Newsroom Search](#)[| International Sites](#)[| Email Opt-in](#)

Nailing the IPU: Designing a Processor by
Walking a Mile in Customers' Shoes

'Behind the Builders:' Intel Fellow Brad Burres explains how the infrastructure processing unit moves lower-value chores off the CPU, freeing it for more important workloads.



News

- July 11, 2022
- Contact Intel PR

More Data Center News

Ask Brad Burres what he likes to do in his spare time, and you might be surprised at his answer's simplicity: Along with riding his motorcycle (a Triumph), hiking and leading Boy Scout wilderness excursions, he likes DIY. When you're using a nail gun, he explains, "you don't get distracted, and there are no hard problems to solve."

The irony is that Burres, an Intel Fellow with the Ethernet Products Group, is one of the key people behind Intel's infrastructure processing unit (IPU)

, a domain-specific processor designed to solve a pretty big problem for server customers. While CPU core counts increase to keep pace with increasing demand for compute, so too does the amount of overhead or infrastructure processing tasks. The last thing you want is for your powerful CPU cores to bog down in the sludge of overhead processing instead of running your application.

Think of it like increasing the number of people on a team to get the work done faster. There's a cost associated with scaling your team – you need to account for the time it takes to assign tasks, make sure things are being done properly, manage the team's shift schedule and more. In the data center, infrastructure functions such as networking, storage and security all require processing power, but it doesn't necessarily have to be handled by the CPU. That's where the IPU comes in.

Burres says as much as 30% to 40% of cycles are spent on infrastructure. By moving some of those infrastructure tasks to the IPU, you can free up CPU cycles for more important workloads.

"The goal from any of the big infrastructure providers is to have zero infrastructure cycles running on Xeon®," Burres explains. "If you're Microsoft Azure or AWS or Google Cloud, you do not want to spend any Xeon compute cycles on your infrastructure tasks or across the network."

A Problem Shared is a Problem Solved

Intel's first ASIC IPU was developed as part of a collaboration with Google. The web giant had a specific list of needs that required a deeply intertwined co-development partnership. Together, the teams developed programmable packet processing capabilities to enable a range of use cases as defined by Google. The result is the Intel® IPU code-named Mount Evans, an ASIC that handles hardware offloads and leverages ARM N1 cores for specific workloads that don't require the might of an Intel® Xeon® processor but have to happen somewhere. It's more than just the art of the hand-off, though. Significant work went into making sure Mount Evans could handle infrastructure tasks as efficiently as possible. For example, Mount Evans supports transfer of up to 200 million packets per second (that's packets of information based on real-work workloads, not a benchmark). At the same time, crypto capabilities are employed to ensure the security of the information flying across the data center.

This chip may have been designed in collaboration with Google, but the problems it addresses are by no means exclusive. "Mount Evans is not super specific to Google," Burres says. "The optimization was tailored for the problems Google identified within its own operations, but it's flexible enough to have applications in lots of other places."

Several other customers are eager to leverage the first-generation ASIC IPU for similar needs in their own data center operations.

“The best ideas come from hearing similar problems from multiple customers and having the time to go back with your technical friends and debate how you solve them and look for the commonalities,” Burres says, adding that other companies with similar infrastructure challenges have come up with their own solutions.

The Theory of Evolution: SmartNICs, IPU and Beyond

Though it's only existed in the public domain for a year, the IPU has been taking up space in Burres' brain for years. Work on what became Intel's first IPU started in 2016 with a different product that eventually morphed into what debuted a year ago, and the history of the little chip goes back even further.

In his previous role as SmartNIC lead architect, Burres focused on networking offload, storage offload and overall packet processing, specifically for top-tier cloud service providers. It's all familiar territory for a domain-specific processor that would eventually be designed to do just that. The move from SmartNIC to IPU is a logical evolution, but it's not just a change in terminology. In fact, the difference between the two spells out the future of the IPU more broadly.

“SmartNIC is a transient term. A general NIC is something that doesn't have any intelligence. There's no FPGA or compute, it's just a fixed pipeline. The SmartNIC has intelligence – cores and flexibility – but it's just networking,” Burres explains. “With added functions like storage and telemetry, there's more opportunity for optimization, and the term ‘SmartNIC’ no longer accurately describes what these chips can do.”

Customers are clamoring for Mount Evans IPU inventory. Freeing up general compute power is a big deal for companies whose entire businesses revolve around renting out server capacity, especially in cases where demand fluctuates. Overprovisioning to account for increased need in certain functions is a waste, and throwing in more beefy Xeons is an expensive solution.

“Appetite is huge,” Burres says, before adding that it's not just a case of shipping out silicon. Intel has to make sure its software can handle what customers want to be able to do, like expanding to the edge. “It's not like Xeon, where you sell it to them, and they go off and write the software. We took a big bite, and we want to ensure incremental progress.”

An updated roadmap published in May

laid out the next four years of Intel IPUs, including some details on the second generation, Mount Morgan. It's early days, but the teams working on Mount Morgan and a subsequent third-generation IPU are working with the Xeon team to align and ensure their separate paths are parallel and converge where it makes sense.

“The sky's the limit,” Burres says, when asked what the future holds for IPUs. “The rate of innovation in a new segment is pretty high, which is what makes it so fun. The IPU is flexible enough to support many more use cases. With Mount Evans, we've got networking and infrastructure. Storage is next. After that, you're going to start seeing some of the I/O moving down to the IPU. There's space for innovation all over — we're at the tip of the iceberg.”

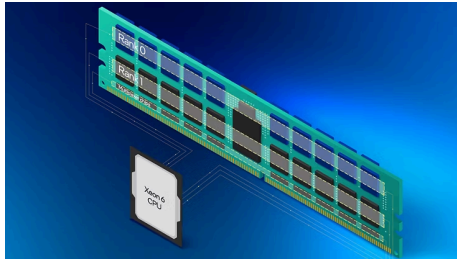
That might scare some people, but for an engineer, this “early and interesting” phase is preferable because it keeps the “overhead tasks” (like a calendar full of meetings) to a minimum. It's not quite as straightforward as a Sunday afternoon with a nail gun, but it's closer than a lot of people get.

Editor's Note: *The first reference to the Intel® IPU was corrected to include its designation as a code name.*

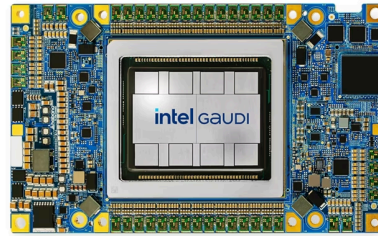
Tags

Data Center & HPC

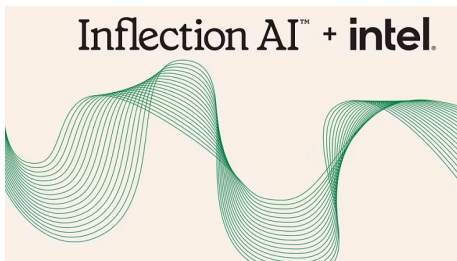
Recent News



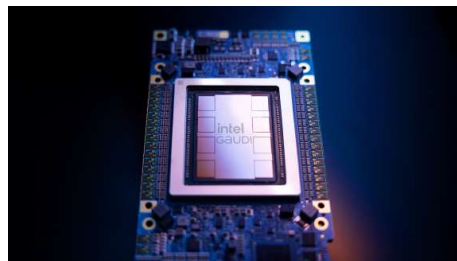
New Ultrafast Memory Boosts
Intel Data Center Chips



Intel Unveils Next-Generation AI
Solutions with the Launch of
Xeon 6...



Inflection AI and Intel Launch
Enterprise AI System



Intel and IBM Deliver Enterprise AI
in the Cloud

About Intel

Intel (Nasdaq: INTC) is an industry leader, creating world-changing technology that enables global progress and enriches lives. Inspired by Moore's Law, we continuously work to advance the design and manufacturing of semiconductors to help address our customers' greatest challenges. By embedding intelligence in the cloud, network, edge and every kind of computing device, we unleash the potential of data to transform business and society for the better. To learn more about Intel's innovations, go to newsroom.intel.com and intel.com

© Intel Corporation. Intel, the Intel logo and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

Company Overview

Contact Intel

Newsroom

Investors

Careers

Corporate Responsibility

Diversity & Inclusion

Public Policy



© Intel Corporation

[Terms of Use](#)

[*Trademarks](#)

[Cookies](#)

[Privacy](#)

[Supply Chain Transparency](#)

[Site Map](#)

[Recycling](#)

Intel technologies may require enabled hardware, software or service activation. // No product or component can be absolutely secure. // Your costs and results may vary. // Performance varies by use, configuration and other factors. // See our complete legal [Notices and Disclaimers](#)

. // Intel is committed to respecting human rights and avoiding causing or contributing to adverse impacts on human rights. See Intel's [Global Human Rights Principles](#). Intel's products and software are intended only to be used in applications that do not cause or contribute to adverse impacts on human rights.

