Overview

The **Drut 2500 Product Series** represents a leading-edge solution for data center operators seeking to enhance dynamic reconfigurability, thereby optimizing scalability, performance, and efficiency. Engineered specifically for GPUaaS cloud providers and private cloud Enterprise customers, this innovative system leverages advanced photonic technology to optimize connectivity between compute, storage, and networking resources. By doing so, it maximizes the return on investment for these critical assets.

The 2500 series is particularly well-suited to address the escalating demands of AI, HPC, and private cloud applications, offering unparalleled agility and performance. Its transformative capabilities empower businesses to adapt swiftly to changing computational requirements, ensuring optimal resource utilization and operational efficiency.

By implementing the Drut 2500 Product Series, organizations can significantly enhance their data center's flexibility and responsiveness, positioning themselves at the forefront of technological advancement in an increasingly data-driven landscape. This solution not only meets current industry demands but also anticipates future needs, providing a robust foundation for sustained growth and innovation in data center operations.

Key Benefits

The Drut 2500 Product Series offers a comprehensive suite of features designed to revolutionize data center operations:

Resource Optimization

- Centralized Asset Management: The system consolidates compute and storage assets into dynamically managed shared pools, significantly reducing idle resources and enhancing overall operational efficiency.

- Advanced Connectivity: Functioning as a sophisticated bridge between assets, the 2500 series enables seamless integration of compute nodes with server resources, ensuring compatibility with existing infrastructures.

Operational Excellence

- Cost-Effective Performance: By minimizing power consumption, cooling requirements, and space utilization, the system substantially reduces operational expenses, leading to improved ROI.

- Scalable Architecture: The modular design facilitates effortless scaling without necessitating extensive infrastructure modifications, allowing for agile expansion as needs evolve.

Future-Ready Technology

- **Adaptable Infrastructure**: The system's architecture is engineered to support the ever-changing landscape of AI/ML workloads, HPC clusters, and private cloud deployments, ensuring long-term viability.

- **Enhanced Security Measures**: Utilizing PCIe remoting technology, the 2500 series guarantees isolated and secure traffic between resources, maintaining data integrity and confidentiality.

This refined feature set underscores the Drut 2500 Product Series' commitment to delivering a comprehensive, efficient, and forward-thinking solution for modern data center challenges.

Use Cases

Delivers superior performance for various business needs:

- **Accelerate AI/ML workflows**: Centralize GPU resources to slash model training and inference times by up to 80%.

- **Supercharge HPC applications**: Achieve sub-millisecond latency and 99.999% uptime for mission-critical workloads.

- **Streamline storage management**: Create flexible resource pools that reduce provisioning time by 60% and improve utilization by 40%.

- **Elevate private cloud infrastructure**: Scale seamlessly to support 10x more concurrent users without compromising performance.

- **Optimize GPU-as-a-Service**: Reduce infrastructure costs by 30% while delivering tailored GPU resources to each user.
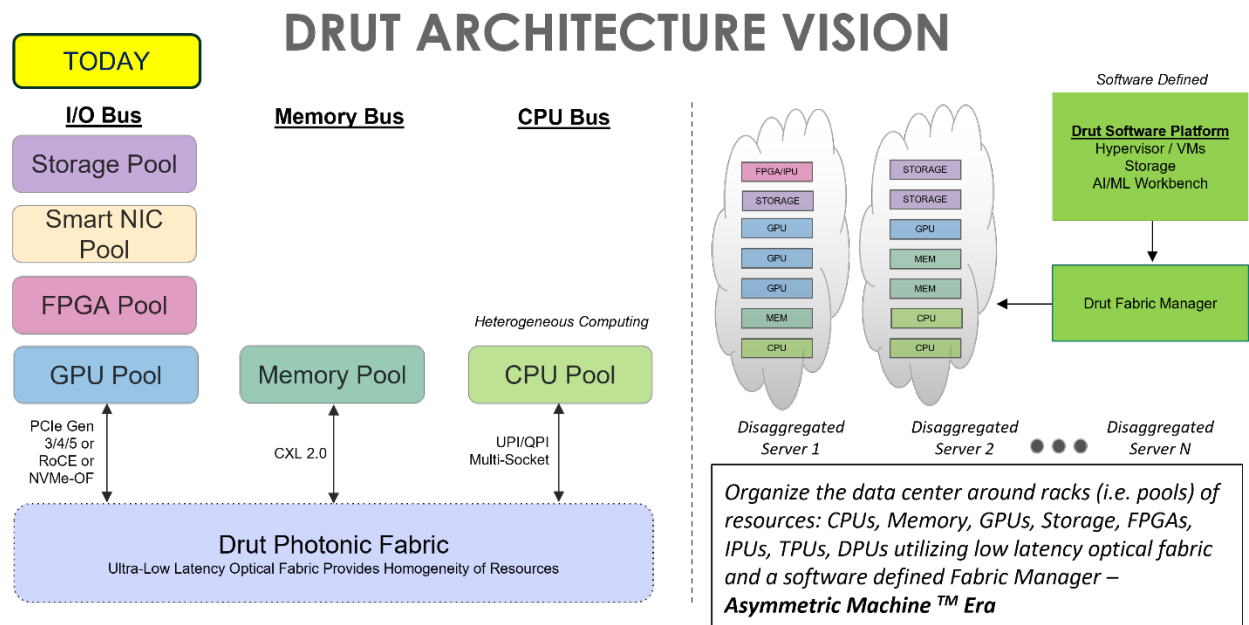
Challenges with Fixed Infrastructure

The challenge of fixed infrastructure costs has long plagued customers across industries. For decades, businesses have grappled with the financial burden of purchasing equipment that rapidly becomes obsolete, leaving them with outdated technology and sunk investments. This issue has been exacerbated by the rise of GPU services, where the pace of innovation is particularly relentless.

In the realm of GPUs, the standard lifespan of a unit can be as brief as two to three years, with manufacturers introducing new models annually. This rapid turnover creates a constant pressure to upgrade, leading to a cycle of expensive replacements and potential technological lag.

To address these challenges, a paradigm shift towards flexible infrastructure is essential. Such an approach would enable businesses to adapt and redeploy component parts without the need for costly and disruptive forklift upgrades. This flexibility not only helps in managing costs more effectively but also ensures that organizations can stay current with technological advancements without being anchored by outdated hardware.

By embracing a more agile infrastructure model, companies can break free from the constraints of fixed assets, allowing for smoother transitions between technology generations and more efficient resource allocation. This strategy aligns with the dynamic nature of modern computing needs, particularly in GPU-intensive applications where performance demands are ever-increasing.

DRUT ARCHITECTURE VISION

TODAY

I/O Bus	Memory Bus	CPU Bus
Storage Pool		
Smart NIC Pool		
FPGA Pool		*Heterogeneous Computing*
GPU Pool	Memory Pool	CPU Pool

PCIe Gen 3/4/5 or RoCE or NVMe-OF

CXL 2.0

UPI/QPI Multi-Socket

Drut Photonic Fabric
Ultra-Low Latency Optical Fabric Provides Homogeneity of Resources

Software Defined

Drut Software Platform
Hypervisor / VMs
Storage
AI/ML Workbench

Drut Fabric Manager

Disaggregated Server 1	Disaggregated Server 2	• • •	Disaggregated Server N
FPGA/IPU	STORAGE		
STORAGE	STORAGE		
GPU	GPU		
GPU	MEM		
GPU	MEM		
MEM	CPU		
CPU	CPU		

Organize the data center around racks (i.e. pools) of resources: CPUs, Memory, GPUs, Storage, FPGAs, IPUs, TPUs, DPUs utilizing low latency optical fabric and a software defined Fabric Manager –
***Asymmetric Machine ™ Era***

The Drut solution revolutionizes traditional server architecture by adopting a modular, component-based approach. This innovative system deconstructs conventional servers and resources into discrete, disaggregated elements. These individual components can then be dynamically assembled and integrated using Drut's leading-edge software and hardware technologies. This approach offers unprecedented flexibility, allowing for customized configurations that can be easily adapted to evolving computational needs. By enabling the seamless combination of various hardware elements, Drut's solution empowers users to create tailored, high-performance computing environments while optimizing resource utilization and reducing unnecessary hardware redundancies.

Drut 2500 Product Series Solution

The Drut 2500 Product Series represents a significant advancement in PCIe Gen 5 technology, catering to both server and resource applications. This innovative line not only complements existing Drut offerings but also seamlessly integrates with current off-the-shelf technologies. By introducing Gen 5 speeds and increasing density and slot capacity, the 2500 series empowers customers to construct next-generation data center infrastructure optimized for AI/ML workloads.

A typical solution architecture showcases the versatility and power of the 2500 series. In this configuration, four off-the-shelf servers are connected to a resource pool via the Drut iFIC 2500. The resource pool comprises two Drut Photonic Resource Unit 2500s, each equipped with eight GPUs. This setup offers flexibility in connectivity options, accommodating both direct and optical switched cabling to suit various use cases and scaling requirements.

# DynamicXcelerator Architecture

Photonic Resource Unit PRU 2500
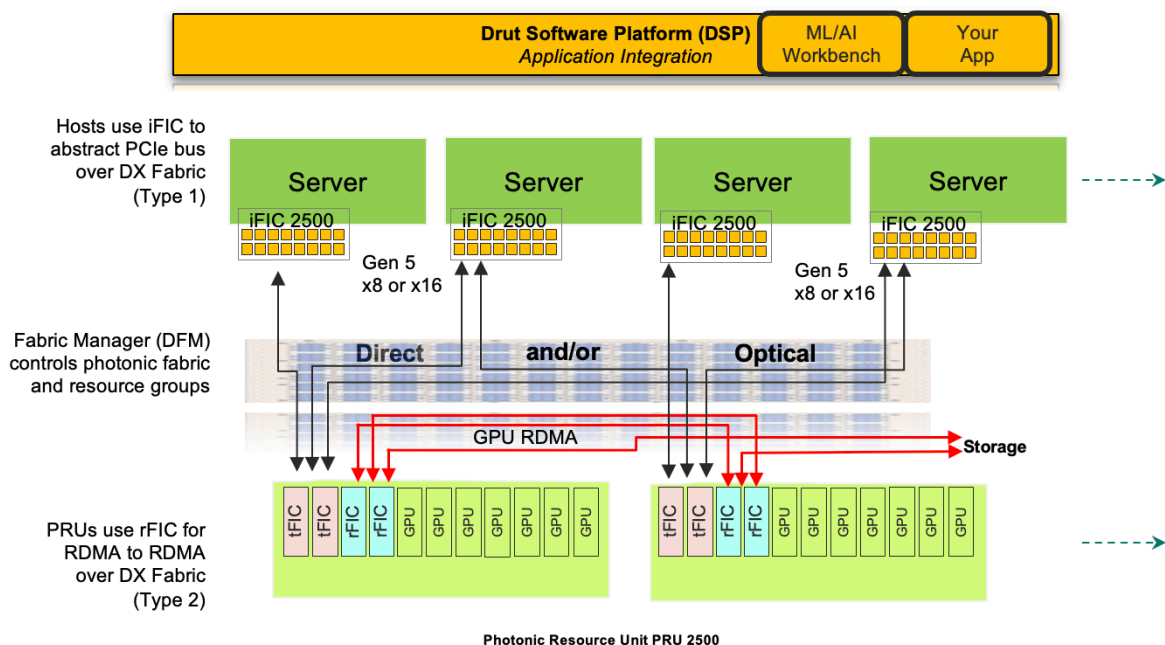
The 2500 series stands out for its ability to enhance existing solutions while paving the way for future advancements in data center technology. Its compatibility with current systems, coupled with its forward-looking design, positions it as a crucial component for organizations seeking to leverage the full potential of AI and machine learning in their infrastructure.

For those interested in delving deeper into the technical specifications and capabilities of the Drut 2500 series, comprehensive datasheets are available for review at drut.io.

The primary Drut products that make up this solution are the Photonic Resource Unit (PRU 2500), and the Fabric Interface Card (FIC 2500), both shown below. The Photonic Switch from the 1000 series product remains in the 2500 solution as a rate-agnostic layer 1 switch.

Photonic Resource Unit 2500
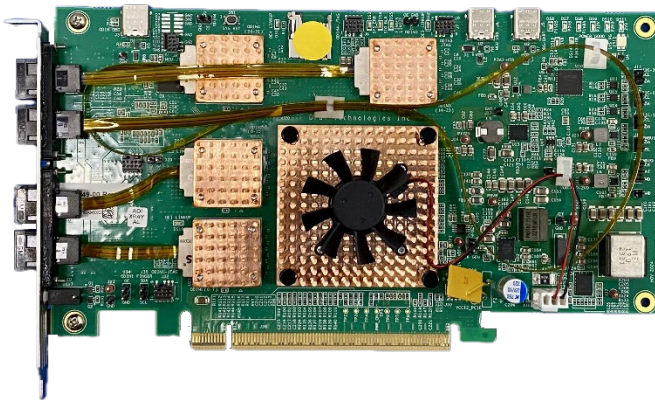
Fabric Interface Card 2500

The Drut 2500 series offers significant advantages in GPU-centric workloads and resource management:

Direct GPU Communication

The architecture enables direct GPU-to-GPU and GPU-to-storage communication, bypassing the server's compute complex. This design offers two key benefits:

- **Enhanced Performance**: By eliminating the need to traverse the server's compute complex, data transfer speeds are significantly increased, reducing latency and improving overall system efficiency.

- **Cost Reduction**: The direct communication pathway allows for a substantial reduction in server expenditure. Organizations can opt for less powerful (and thus less expensive) servers without compromising on GPU performance.

Flexible GPU Management

The Drut system introduces a novel approach to GPU failure handling and replacement:

- **Simplified Troubleshooting**: When a GPU appears to fail, it can be easily removed from the server and replaced. This process often reveals that many perceived hardware failures are actually software-related issues.

- **Cost-Effective Maintenance**: The ability to hot-swap GPUs without physically removing them from the entire system saves both time and money. This feature eliminates the need for extensive downtime and reduces the risk of damaging other components during the replacement process.

By implementing these innovative features, the Drut 2500 series not only enhances performance but also offers substantial cost savings and operational flexibility, making it an attractive solution for organizations looking to optimize their GPU-intensive workflows.

Simplified Deployment and Management

The 2500 series solution revolutionizes data center operations by harnessing the power of photonic connectivity, delivering unparalleled low-latency and high-bandwidth resource sharing. This leading-edge technology offers several key advantages in deployment and management:

Dynamic Resource Orchestration

Drut's Fabric Manager (DFM) serves as the central nervous system of the 2500 series, orchestrating the PRU 2500 to allocate resources with unprecedented flexibility and efficiency. This intelligent system enables:

- **Real-time GPU allocation**: Instantly provision and reassign GPU resources to meet fluctuating computational demands.

- **Dynamic FPGA distribution**: Seamlessly allocate FPGA resources for specialized workloads and accelerated processing tasks.

- **Adaptive storage management**: Efficiently allocate and reallocate storage resources to optimize data access and utilization.

Seamless Integration

The iFIC 2500 component integrates effortlessly with existing infrastructure, minimizing disruption and maximizing compatibility. Key features include:

- MPO16 fiber connectivity: High-density, high-performance connections between the iFIC 2500 and PRU 2500.

- Backward compatibility: Seamless integration with legacy systems, protecting existing investments.

- Scalable architecture: Easily expand capacity without overhauling the entire infrastructure.

Comprehensive Management Interface

DFM provides a robust, user-friendly platform for end-to-end orchestration of the 2500 series ecosystem. This centralized management solution offers:

- Real-time provisioning: Instantly allocate and reallocate resources to meet changing demands.

- Intuitive monitoring: Gain deep insights into system performance and resource utilization.

- Automated optimization: Leverage AI-driven algorithms to continuously fine-tune resource allocation for maximum efficiency.

By combining these advanced features, the 2500 series dramatically simplifies data center operations while delivering unprecedented flexibility and performance.

Why Choose the 2500 series solution?

The innovative system architecture of the 2500 series delivers a paradigm shift in data center capabilities, offering a suite of advantages that redefine operational efficiency and scalability:

Adaptive Infrastructure

The modular design of the 2500 series empowers data centers with unprecedented flexibility:

- Dynamic scaling: Seamlessly expand or reconfigure resources to meet evolving workload demands without disrupting ongoing operations.

- Workload-optimized configurations: Tailor the infrastructure to specific application requirements, maximizing performance and efficiency.

- Rapid deployment: Quickly integrate new modules to address sudden spikes in computational or storage needs.

Optimized Total Cost of Ownership

The 2500 series architecture is engineered to deliver substantial long-term economic benefits:

- Reduced operational expenses: Streamlined management and improved energy efficiency significantly lower day-to-day costs.

- Enhanced resource utilization: Intelligent allocation ensures optimal use of all system components, minimizing idle resources.

- Extended hardware lifespan: Modular upgrades allow for targeted improvements without wholesale replacements.

Future-Proofed Operations

Built with an eye toward the future, the 2500 series ensures sustained relevance in the face of rapidly evolving technologies:

- Scalable architecture: Designed to accommodate exponential growth in data processing and storage requirements.

- Technology-agnostic platform: Adaptable to emerging computational paradigms, including quantum and neuromorphic computing.

- Continuous innovation support: Regular firmware and software updates unlock new capabilities and performance enhancements.

This forward-thinking system architecture not only addresses the immediate needs of modern data centers but also positions them for long-term success in an ever-changing technological landscape.

Summary

The Drut 2500 series represents a quantum leap in data center technology, offering a comprehensive solution that redefines scalability, security, and cost-efficiency for next-generation operations. This revolutionary product line seamlessly integrates leading-edge connectivity with intelligent resource management, empowering organizations to:

- Adapt with agility: Dynamically respond to fluctuating workloads and emerging technological trends.

- Optimize performance: Maximize resource utilization through real-time allocation and load balancing.

- Fortify security: Implement robust protection measures at both the hardware and software levels.

- Drive innovation: Leverage a future-ready platform that evolves alongside emerging computational paradigms.

By harnessing the synergistic capabilities of the 2500 series, businesses can confidently navigate the complexities of the digital era, transforming potential challenges into opportunities for growth and technological leadership. This groundbreaking solution not only addresses the immediate demands of modern data centers but also lays the foundation for sustained success in an increasingly data-driven world.