Business Intelligence | Data Center | Cloud | Mobility | Security

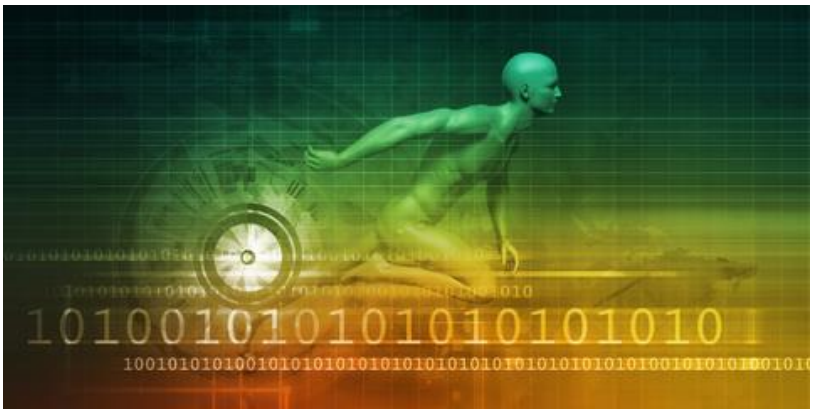
Enterprise Computing Solutions – North America

Transitioning to PCIe/NVMe SSDs
What Sellers Need to Know

Five Years Out

arrow.com

Table of Contents



NVMe and PCIe in the Data Center

The enterprise market for server storage has seen a revolution. Traditional mechanical hard disk drives (HDDs) are giving way to faster solid-state storage technology in the data center. Data center managers now have the option to select the storage performance that optimizes both their workloads and budget.

Solid-state storage technology has also evolved since its introduction. The first SSDs were designed to use the same technology used by existing HDDs to provide servers with immediate storage compatibility with existing servers. This older technology used SATA and SAS bus interfaces and SCSI and AHCI logical device interfaces. Each of these were designed to address the needs of traditional mechanical disk drives – at times purposely adding latency to 'wait' for the mechanical operations of the drive to complete – making them inefficient for the newest SSD storage solutions.

In order to address the needs of solid-state technology, the industry developed the Non-Volatile Memory Express (NVMe) standard. This standard was designed from the ground up to capitalize on the low latency and internal parallelism of solid-state storage devices. The adoption PCIe/NVMe host bus, logical device interface combination to attach SSDs, allow data center server's entire storage subsystem to reach its highest performance.

Arrow's What Sellers need to know about Intel Solid State Drives provides the seller with the basics of the SSD market in the data center, along with the performance advantages of SSDs over HDDs. It also provides an introduction to Intel's SSD product line along with the advantages of selecting an Intel SSD over the completion.

This overview provides the advantages of PCIe/NVMe SSD storage and provides the Seller with the benefits and limitations of adopting them for the servers in the datacenter.

PCIe and NVMe

It is very easy to get lost in the terminology surrounding the new SSD storage. There are two components of the storage subsystem – the host bus (hardware) and the logical device interface (software).

The easiest way to think of the host bus adapter (HBA) is it is the hardware used to interface with the computer. The primary purpose of the HBA is to connect a computer to storage and network devices. Other terms that are used to refer to an HBA are host controllers or host adapters. HBA's include the PCIe bus along with the SAS, SATA, and Fiber Channel bus.

The software component used with the HBA is the logical device interface. NVMe is the software designed specifically to optimize interfacing with solid-state storage technology and optimize the performance of SSD disk drives. NVMe logical device interface compares to the SCSI and AHCI interfaces which are the software components to the older SAS, SATA, and Fibre Channel buses.

Until recently, the only bus designed to run the NVMe protocol was the PCIe bus. Because of this, the term NVMe is occasionally used to describe storage media attached via the PCIe bus. This changed when the industry specification for NVMe expanded to run over other fabrics, such as Ethernet and Fibre Channel.

Storage Host Buses Comparison

A server uses a combination of host buses to communicate with both internal and external storage. Historically, most storage and SSDs used buses such as SATA, SAS, or Fibre Channel for interfacing with the rest of a computer system. In the data center, the SAS interface is a widespread, relatively fast, and reliable. However, SAS and the included SCSI protocol logical interface layer are designed for spinning disk, not flash making them relatively inefficient when compared to the PCIe/NVMe interface.

Find below a comparison of the different host bus found in the data center.

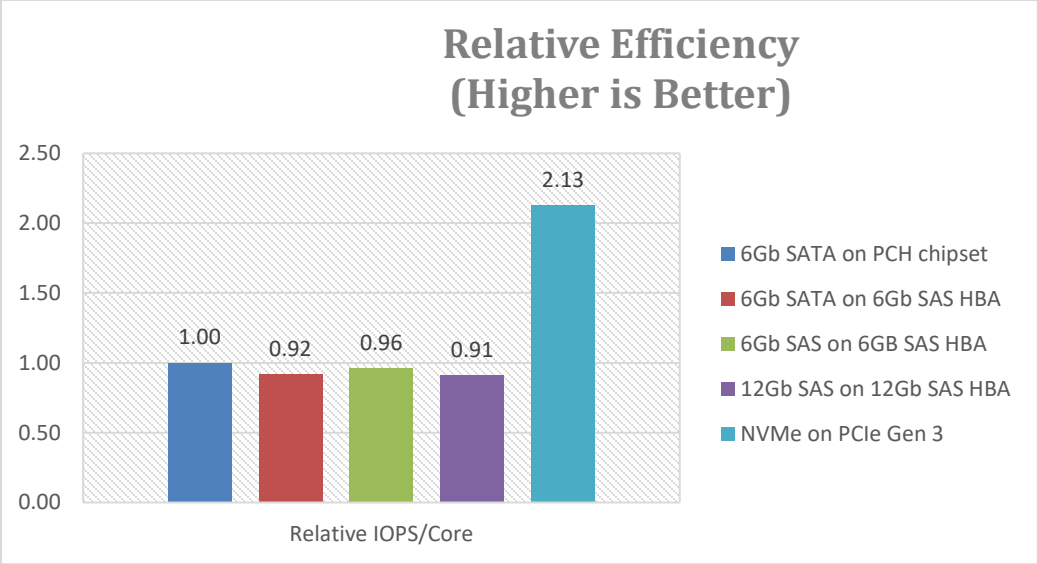
Type	Speed	Disk Drive Attach	Data Paths	Type
PCIe	~3.94GBps	Internal	Four lanes/data paths	Gen 3
SAS	~1.5GBps	Internal & External	Two data paths	12GBps
SATA	~500MBps	Internal	Single lane/data path	6 GBps

Two measures of performance of a host bus and logical interface for storage are input/output operations/second (IOPS) and the amount of CPU resources (measured by the number of CPU clock cycles) it takes to perform and input/output (I/O). The charts below from nvmexpress.org demonstrate that the NVMe on PCIe not only provides more IOPS per core, but also performs each I/O using less CPU clock cycles.

NVMe has streamlined and simplified the command set providing two major benefits.

The first benefit, show in this chart, is that NVMe provides higher IOPS per instruction cycle and lower I/O latency in the host software stack. Relative Efficiency is determined by the interface that provides the most IPOS

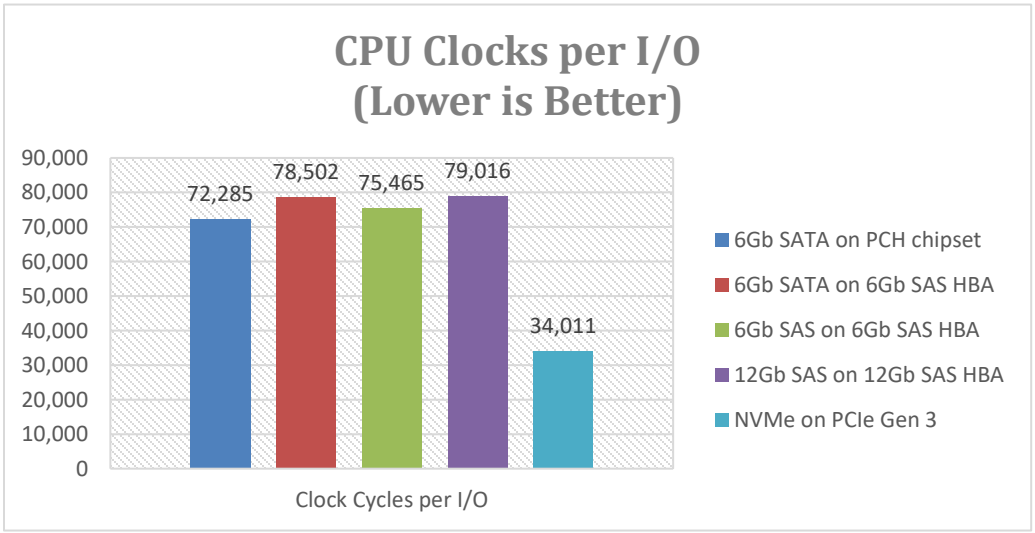
per CPU core. NVMe on PCIe leads the pack by providing 2.13 Relative IOPS/Core – more than 2x of those provided by ether their SATA or SAS counterparts.



The second major benefit is NVMe/PCIe uses less than half the number of CPU instructions to process an I/O request when compared to SAS or SATA.

Since the NVMe protocol supports multiple deep processing queues, (an advancement over traditional SAS and SATA protocols), which are designed such that I/O commands and responses operate on the same processor core and take advantage of the parallel processing capabilities of multi-core technologies.

In this chart, CPU Clock Cycles per I/O is shown. NVMe on PCIe uses less CPU resources by about 50% when compared to SAS and SATA.



PCIe/NVME SSD Storage Considerations

It is important to note that the PCIe bus addresses internal storage. Currently, all SSD disks attached using PCIe/NVMe must be internal to the server. This may be a limitation for data centers since their architecture is often based upon networked SAN style storage. This will change in the near future – more on this below in the document.

It is important to analyze customers high availability needs when recommending PCIe/NVMe SSD storage. Even though modern servers with up-to-date operating systems support hot-plug for drives with U.2 form factors, there are some limitation s which will be addressed in future generation servers. At the moment of publishing this document there are no hardware RAID options available for NVMe drives.

The good news is that industry standards for both PCIe & NVMe have been expanded to provide these features. Look for future announcements from server OEMs.

This does not mean that there are no high availability options for PCI\e/NVMe SSD storage. Software RAID is available for via Intel Rapid Storage Technology Enterprise or embedded OS tools like Linux LVM/mdraid or Microsoft® Storage Spaces. Also, it is common to use software to mirror data across systems to provide increased availability. In addition, advances in virtualization of the three major components of a server, compute, storage and network, make failure of an individual part manageable.

Another alternative is to use the approach many hyper-scale data centers use to provide continuous operations. In the event of a failure of any component of the server, workload shifts to another server.

Why NVMe/PCIe over SAS

NVMe/PCIe offers the best storage bandwidth and lowest latency for SSD storage, in part because the PCIe connection offers direct lanes to the server's Intel CPU that allows for advanced data coordination. In addition, since the NVMe protocol was designed from scratch for SSDs, it is optimized for the non-mechanical SSD storage.

What are the advantages of PCIe/NVMe over the SAS interface for SSDs in the data center?

- NVMe provides significantly lower latency and increased efficiency: lower CPU utilization, lower power, lower TCO. (see charts above)
- Allows applications to 'talk' directly to SSD storage via the PCIe bus
- Increased bandwidth: 1 GB/s per lane with up to 1-16 lanes per drive.
- Directly attached to the CPU, eliminates host bus adapter cost and overhead
- PCIe provides scalable performance, flexible form factors and industry stability
- Low power features from both PCIe and NVMe – Typically less that 2W idle and less than 9W maximum
- Designed to scale beyond current standards

- Reinvents storage stack - Moves latency to software
- Replaces legacy interfaces (SATA or SAS) designed for spinning disks

NVMe/PCIe Provides Flexible Form Factors

PCIe/NVMe SSDs come in a variety of form factors unlike enterprise SAS drives that offer only the U.2 2.5in form factor. The following is a breakdown of the form factors, along with notes on each.

M2 Form Factor

- 42, 80 & 110mm Lengths
- Smallest footprint for PCIe
- Used for both boot or to maximize storage density

U2 2.5" or SFF-8639 Form Factor

- 2.5in form factor makes up the majority of SSDs sold today
- Easy to deploy, hot-plug and service

Add-in Card (AIC) Form Factor

- Has maximum system compatibility
- Achieves maximum performance/power per slot (x8 version)
- Provides options for height and length

Server Support for 2.5" NVMe Drives

HPE® ProLiant® servers support NVMe 2.5" SSDs. These drives are front accessible and serviceable, but require a PCIe Express Bay Enablement Kit. The PCIe Express Bay Enablement Kit includes the necessary drive cage, cables, and PCIe NVMe bridge card. HPE systems supported are the HPE ProLiant DL360, DL380, DL580, and ML350 Gen9 Servers. The PCIe hot-add feature planned, but not currently available. Please check HPE's website for updates and additional information.

Lenovo® provides NVMe/PCIe SSD support in the Lenovo System x®, ThinkServer®, and Flex System™ servers. Implementing NVMe Drives on Lenovo Servers provides the seller with guidance and specifics.

Current Intel® NVMe products

Intel provides a powerful line of PCIe/NVMe SSDs for the data center. Each product line is designed to provide the flexible form factors, performance, and capacity needed in the data center. Here is a look at Intel's current PCIe/NVMe product line:

	Form factor	Capacity (GB)	Performance specs (up to IOPS)	Technologies
Intel SSD DC P3520 Series	PCIe 3.0 x 4 U.2 2.5-inch, Low profile AIC* (HHHL**)	450 / 1.2 TB / 2 TB	375,000 Random Read 26,00 Random Write	Intel 3D NAND, Enhanced Power Loss Data Protection, End-to-End Data Path Protection, Enterprise Reliability Features
Intel SSD DC P3608 Series	PCIe 3.0 x 8 HHHL AIC	1.6 TB / 3.2.TB / 4.0 TB	850,000 Random Read 150,000 Random Write	20nm Intel NAND MLC Flash Memory with High Endurance Technology, Enhanced Power Loss Data Protection, End Data Path Protection, Dual NVMe Controller, Enterprise Reliability Features
Intel SSD DC P3700 Series	PCIe 3.0 x 4 U.2 2.5-inch, Low profile AIC* (HHHL**)	400 / 800 / 1.6 TB / 2.0 TB	460,000 Random Read 175,000 Random Write	25nm Intel NAND MLC Flash Memory with High Endurance Technology, Enhanced Power Loss Data Protection, End-to-end Data Path Protection, Enterprise Reliability Features
Intel SSD DC P3600 Series	PCIe 3.0 x 4 U.2 2.5-inch, Low profile AIC* (HHHL**)	400 / 800 / 1.2 TB / 1.6 TB / 2.0 TB	450,000 Random Read 56,000 Random Write	25nm Intel NAND MLC Flash Memory with High Endurance Technology, Enhanced Power Loss Data Protection, End-to-end Data Path Protection, Enterprise Reliability Features
Intel SSD DC P3500 Series	PCIe 3.0 x 4 U.2 2.5-inch, Low profile AIC* (HHHL**)	400 / 1.2 TB / 2.0 TB	430,000 Random Read 28,000 Random Write	25nm Intel NAND MLC Flash Memory with High Endurance Technology, Enhanced Power Loss Data Protection, End-to-end Data Path Protection, Enterprise Reliability Features

*AIC = add-in card

**HHHL = half-height, half length

In addition, Intel also offers the Intel DC D3700 and D3600 series dual port NVMe/PCIe SSDs. These series SSD's are dual port and are still in the process of being implemented by server OEMs.

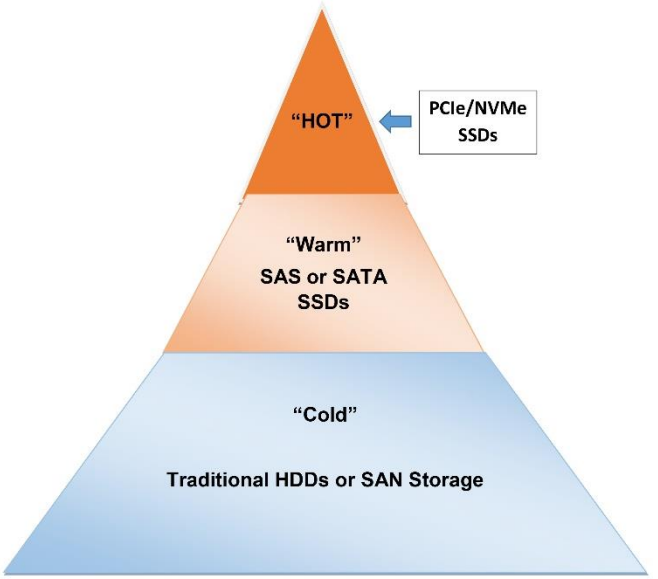
These drives are the first dual port NVMe/PCIe SSDs in the market and their dual ports enable redundant data paths to the application server and storage subsystem. Redundant data paths ensure data access in the event of a storage controller or PCIe switch failure (like SAS). A failure event initiates a failover response from the storage controller of the functioning data path. Upon completion of the failover response, full performance is restored. In order to support new connectivity, systems must have new backplanes, which have a PCIe properly routed to two hosts with or without PCIe switches.

Perfect Fit - PCIe/NVMe SSDs as Cache or Tiered Storage

PCIe/NVMe storage is commonly used to accelerate active I/O either by using it as a cache or as a tier of storage. These are the most popular uses of PCIe/NVMe SSD storage since its introduction due to the higher cost associated with this newer technology. If a customer's budget cannot accommodate an entire PCIe/NVMe SSD solution, a scaled back configuration using PCIe/NVMe SSDs as cache or tiered storage may accelerate performance at a reduced cost.

Caching keeps a copy of active data between the application and other, slower back-end storage. Using PCIe/NVMe storage for cache incrementally increase overall application performance.

Automated storage tiering automatically places data on multiple types of media depending on the activity of the data. The advantage to tiering is "hot" (frequently accessed) data can be stored on faster, more expensive media such as PCIe/NVMe SSDs. "Warm' (less frequently accessed) data can reside on less expensive SAS or SATA SSDs. "Colder' data, which is even less accessed less than 'warm' data, can be made up of low cost traditional hard disk drives or be located on a SAN or other externally attached storage.



As an example, the following storage configuration would accelerate performance over a solution entirely based upon HDDs.

- Boot: Dual SATA SSDs
- "Hot" storage tier: PCIe/NVMe SSDs
- "Warm" storage tier: SATA SSDs
- "Cold' storage tier: traditional HDDs or SAN storage

Performance increases due to using PCIe/NVMe SSD as cache or a storage tier are application and solution dependent. Using a combination of tiered storage type's increases performance over a solution made up entirely with 'slower' SATA or SAS SDDs or traditional HDDS for most applications. It also reduces the storage investment over purchasing an entire storage solution made up entirely of the 'fastest' and more expensive PCIe/NVMe storage.

Types of Applications to Target

PCIe/NVMe storage is the fastest, lowest latency access to data available to servers. Connecting storage using the PCIe bus and NVMe protocol eliminates the need for latency-adding adapters to attach storage. **This makes it ideal for any data-heavy, low-latency content when the increased responsiveness and performance can justify the premium PCIe/NVMe storage commands**. **PCIe/NVMe SSD prices are rapidly falling in line with other SSD storage so be sure to check current prices.**

Applications that are data-heavy and benefit from low-latency include Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) applications. In addition, customers looking to invest in in-memory solutions such as SAP HANA or IBM's DB2 in memory database should also carefully evaluate PCIe/NVMe storage. A solution based on PCIe/NVMe storage may offer the performance boost needed, at a lower cost.

OLTP applications typically handle many short transitions in quick, random bursts, making these solutions well suited to the boost PCIe/NVMe SSDs can provide. OLTP applications include:

- Analytics
- Content delivery (streaming media, video on demand)
- Critical infrastructure applications
- Database query acceleration
- Financial and Ledgers
- Gaming
- Video Surveillance
- Real-time billing
- Real-time monitoring
- Trading

OLAP applications handle bigger chunks of data, typically received from multiple sources. These types of applications require quick responses to complex queries - making them prime candidates to benefit from the performance PCIe/NVMe storage offers. OLAP application workloads include:

- Business intelligence
- HPC applications
- Batch processing
- Data warehousing and report generation
- ERP systems
- High transaction Processing
- Massive Data feeds and reporting

For general-purpose applications and boot drives, SATA SSDs provides adequate performance. Generally, this makes the added performance provided by PCIe/NVMe SSDs not necessary.

Moving the Bottleneck

A server's storage subsystem has traditionally been the slowest component in a server. Replacing mechanical hard drives, and the logical device drivers that were designed for them, with PCIe/NVMe SSD storage provides the lowest latency and highest performance storage subsystem. Removing the storage subsystem as a server's bottleneck may also move the performance bottleneck to another component. Depending upon the workload, network speed is the next logical component that is likely to provide a bottleneck. Upgrading a data center's fabric and servers to 100Gb Ethernet may be the next step in achieving the best overall server performance.

Newly Announced - NVMe over Fabrics

The new standards governing NVMe over fabrics were released on June 9, 2016, allowing NVMe to expand beyond the internal-to-the-server PCIe bus to include fabrics such as Ethernet, Fibre Channel, and InfiniBand.

This is key because the low-latency and performance of NVMe SSDs will soon cater to the multi-host environments that are standard in most data centers.

NVMe over Ethernet is expected to be a major player in the market because Ethernet is the basis for Network Attached Storage (NAS) storage file sharing capabilities and iSCSI-based Storage Area Networks (SANs).

NVMe over Fibre Channel (FC-NVMe) is also expected to be a major player in the market. Fibre Channel was designed for storage from its inceptions. It ensures that all storage data arrives to its network destination without loss at extreme throughput and low latency, while operating at high utilization rates. It also continues to be the most widely used transport for enterprise storage with the Fibre Channel Industry Association estimating 46 million ports in current operation in data centers today.

NVMe over Intel Omni-Path Architecture (OPA-NVMe) will provide HPC workloads with cost-cost effective support of storage in HPC clusters and features such as adaptive routing, traffic flow optimization and packet integrity protection.

InfiniBand provides an interconnect with high bandwidth and low latency. SSDs using NVMe over InfiniBand (IB-NVMe) may be a natural addition to those data centers that currently use InfiniBand.

Broadcom® announced the first NVMe over fabric product on July 19, 2016 with the announcement of the first NVMe over Fibre Channel (16/32 Gb) HBA solution. Look for OEM's including HPE and Lenovo to adopt certified NVMe over fabric adapters in the near future.

While NVMe over Fabrics will allow NVMe storage to seamlessly integrate into the architecture of data centers, but it will also add some amount to storage latency. This means that the currently available PCIe/NVMe storage will likely remain the lowest latency storage SSD solution.

Next Generation - Intel Technology Advances

Intel Optane™ Technology is a premium non-volatile memory media that has the potential to change the face of computing with an unparalleled combination of speed, endurance, and density at a competitive price.

The new soon-to-be-released Intel® Optane™ memory and Intel® Optane™ SSDs based on 3D XPoint media, eliminates the need for transistors used in current NAND based products. Optane technology enables very dense memory arrays that scale with Moore's law. Optane technology marks a monumental jump forward in the same way adopting NAND eclipsed hard disk storage. Latency times are much lower than NAND and remain low as throughput increases.

Look for PCIe/NVMe products based on this technology in 2017.

NVMe/PCIE is the Fastest, Lowest Latency SSD Server Storage

Enterprise storage has undergone a dramatic performance boost with SSD storage. SAS and SATA SSDs use old technology designed for spinning hard drives to connect to the server. The NVMe industry standard, developed specifically for SSD storage, eliminates the inefficiencies of older technology and the PCIe bus interfaces directly with the CPU. This makes PCIe/NVMe SSD storage the fastest SSD storage possible for a server.

NVMe over fabrics and the emergence of Intel Optane 3D Xpoint memory will ensure the continued evolution of SSD storage in the data center, however PCIe/NVMe attached SSD storage will remain the best performing, lowest latency connection for SSD server storage. This makes adoption of PCIe/NVMe storage in the data center a great investment.

Are You Five Years Out?

Most people live in the present. The world of now. But a handful of us work in a unique world that doesn't quite exist yet—the world of Five Years Out.

Five Years Out is the tangible future. And the people who live and work there know that new technologies, new materials, new ideas and new electronics will make life not only different, but better. Not just cheaper, but smarter. Not just easier, but more inspired.

Five Years Out is an exciting place to be. So exciting that, once you've been there, it's hard to get excited about the present. Because we know what's coming is going to be so much better.

Five Years Out is a community of builders, designers, engineers and imaginers who navigate the path between possibility and practicality. Creating the future of everything from cars to coffeemakers.

Are you one of them? Then you're probably working with us.

In Person

800 544 7674
Call to talk or set up a face-to-face meeting with one of our knowledgeable representatives.

Via Email

ecsmarketing@arrow.com
Email us for answers to questions or to start a conversation.

Online

ecs.arrow.com
Visit our website for everything from the latest news to line card information.

Five Years Out

Arrow Electronics, Inc.
Enterprise Computing Solutions
9201 East Dry Creek Road
Centennial, CO 80112, USA