# Software Define Storage with a Full Hardware NVM Express Solution

A data center infrastructure becomes more and more complex, where each user needs a different configuration regarding the storage, computing and networking requirements. When a server configuration has resources limitations, adding servers in order to increase the computing and/or storage capacities is not efficient in term of performances, cost and power consumption. The data centric architecture is much more efficient, based on a pool of storage elements, and computing elements. This disaggregated architecture allows scaling dynamically according to user requirements. This is feasible only with low latency interfaces. This paper describes how a full hardware NVMe implementation can answer the low latency requirements of a software defines storage (SDS) infrastructure.

**Software Defined Storage**

Allocating data center resources is one of the main issues for IT managers. In a cloud computing environment, the computing requirement are managed through a virtualization mechanism, therefore users deploy their applications on a virtual machine (VM), and not directly on a physical server. A VM requires computing capabilities, memory and storage. Ideally, the VM will use all the resources in the same physical rack unit.  That assumes that servers come with the right balance of computing cores, memory, storage and networking interface. What happens when users need application with a different balance of resources?

In the case of additional storage, one can use the storage part of another server. Is it efficient? No. Most of current architectures are not designed to provide low latency between a processor and a storage drive from another server, therefore reducing dramatically the performances. The second server will allocate its storage resources for the first server, without using its computing capabilities (CPU+RAM), leading to a lost in term of power consumption and cost. The answer? The disaggregated architecture based on a separate compute and storage elements with low latency interfaces. Each compute element can access to each storage element, with minimal latency penalty.
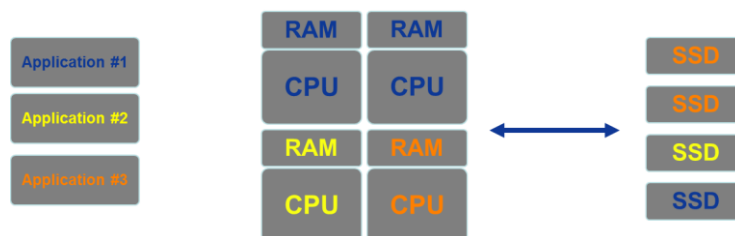


This approach is also called software defined infrastructure (SDI), where the infrastructure refers to compute, storage and network. What are the requirements for software defined storage (SDS)? The main important parameter is a low latency interface. Today's fastest SSDs comes with a latency in the range of hundreds of µs, while storage server come with a latency in the range of ms. An increase of 10x in latency is not acceptable. A new way to interface CPUs to storage elements is definitely required, coming with latencies in the range of dozen of µs.

**NVM Express**

Two technologies provide low latency for storage devices, the nand flash memories and the PCI Express (PCIe) bus interface. The first generation of SSDs combining these 2 technologies was based on the use of a PCIe-to-SATA bridge processor, and a standard SATA SSD architecture. It was efficient in term of time-to-market, but not in term of performances. Then, in 2011, storage industry leaders decided to define a common interface in order to leverage the use of PCIe SSDs in a data center, through a standard interface, and an optimized set of commands and features for NVM data transfers. That led to the Non-Volatile Memory Express specification (NVMe). The full specification is available on www.nvmexpress.org. NVMe is definitely adopted by the market, , and NVMe specification extensions are in progress, such as NVMe over fabrics where the goal is to allow the access between a NVMe SSD and a CPU through a fabrics such as ethernet. Multiple companies developed demos based on 2 servers integrating a PCIe NVMe SSD each. The CPU #1 is accessing the SSD #1 directly, and SSD #2 through the fabrics. The overhead latency for the access to the SSD #1 from CPU #1 is below 10µs. As explained in the 1$^{st}$ part of this paper, it seems that NVMe over fabrics could be an excellent answer to SDS architecture, allowing the access between a CPU and a storage device (local or external), with a very low latency.
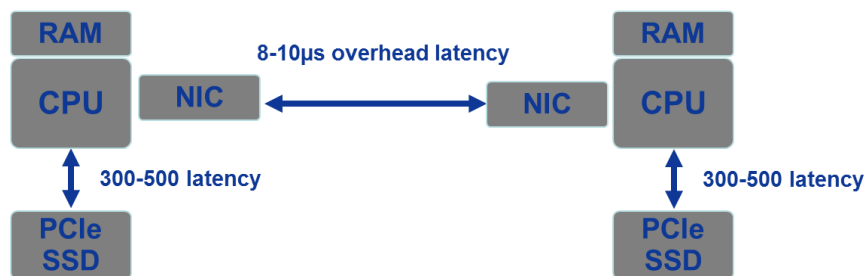


**IP-Maker NVMe technology**

IP-Maker has developed a NVMe solution, based on a full hardware architecture. It comes as a Verilog IP to be integrated in a FPGA or ASIC. It is compatible with the software drivers (available on www.nvmexpress.org) and it has successfully passed the UNH-IOL compliance tests. The NVMe IP is based on an automatic command processing unit, accelerating the processing time of the NVMe commands, and a multi-channel DMA, delivering a high IOPS range. This optimized version requires neither software nor CPU, leading to reduced gate count and power consumption. Another version comes with a CPU interface in order to add flexibility in the NVMe management, such as the support of vendor specific commands.
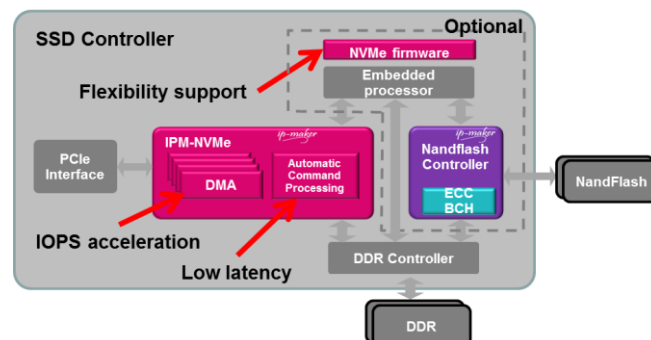
The IP-Maker NVMe IP has been successfully integrated in a high performance reference design, based on a Xilinx FPGA (Zynq and Virtex 7). It uses the PCIe controller built in the FPGA, and a software IP regarding the DDR3 controller. From the host side, it is seen as a NVMe storage drive, where data are stored in the DDR3 memory. The NVMe IP manages data transfer between the host memory and the on-board memory. On a PCIe Gen 2x4 configuration, it reaches about 10µs latency and 385kIOPS, which is the maximum on this PCIe configuration.
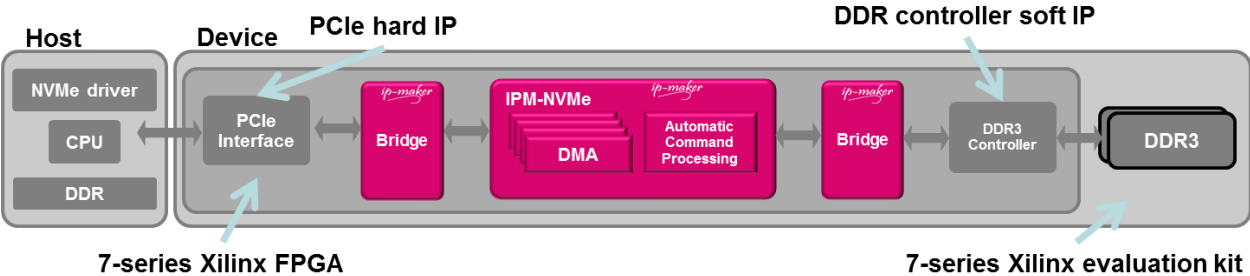


The IP-Maker NVMe IP is also Gen3 capable. The PCIe Gen3 configuration will provide an important performance increase in term of IOPS and latency. Since the latency budget coming from the NVMe IP is only few hundred of ns (thanks to its full hardware architecture), and PCIe Gen3 is the doubling the data rate compared to PCIe Gen2, a latency of about 6-7 µs is expected on a Gen3 x4 configuration.

The IP-Maker NVMe IP provides compatibilities with NVMe over fabrics implementation and it delivers low latency thanks to its optimized design. On the other hand, emerging NVM memories come with latencies in the range of few µs only. The combination of all these technologies will lead in a storage system latency below 20µs, for local SSD as well as external SSD. Software Defined Storage systems can be easily designed without having extra latencies for external SSD. Compared to current external storage systems, this may lead in a 100x latency improvement.
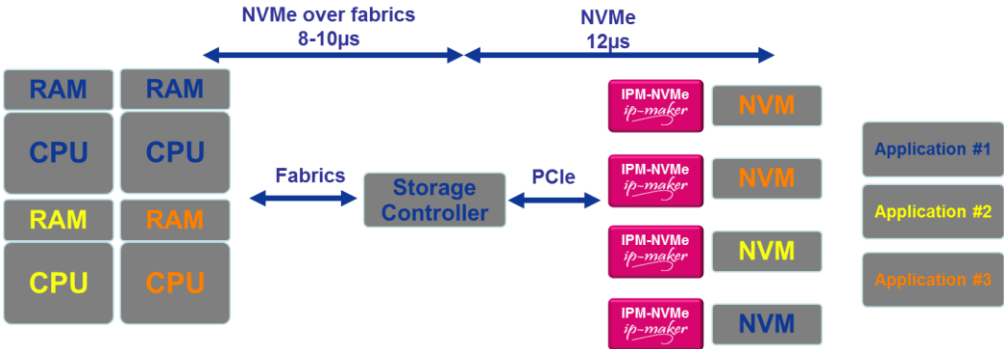


**About IP-Maker**

IP-Maker is a leader in Intellectual Properties (IP) for high performance storage applications. IP-Maker's NVM Express (NVMe) technology provides a unique hardware accelerated solution that leverages the PCIe SSD performances, including ultra-low latency and high throughput. IP-Maker is a contributor to the NVMe specification. The ASIC and FPGA IP portfolio includes NVMe, Universal NandFlash Controller and ECC IP cores. The combination of the IP-Maker technology and its associate services dramatically cuts time-to-market.

Headquarter address: Domaine du Petit Arbois / Avenue Philibert BP50014 / 13545 AIX EN PROVENCE Cedex 4 / France. www.ip-maker.com / contact@ip-maker.com / +33 972 366 513