



---

# Coherent Accelerator Processor Interface (CAPI) for POWER8 Systems

## White Paper

---

Bruce Wile  
IBM Systems and Technology Group

**Advance**

29 September 2014



© Copyright International Business Machines Corporation 2014

Printed in the United States of America September 2014

IBM, the IBM logo, and [ibm.com](http://ibm.com) are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product, and service names may be trademarks or service marks of others.

All information contained in this document is subject to change without notice. The products described in this document are NOT intended for use in applications such as implantation, life support, or other hazardous uses where malfunction could result in death, bodily injury, or catastrophic property damage. The information contained in this document does not affect or change IBM product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of IBM or third parties. All information contained in this document was obtained in specific environments, and is presented as an illustration. The results obtained in other operating environments may vary.

While the information contained herein is believed to be accurate, such information is preliminary, and should not be relied upon for accuracy or completeness, and no representations or warranties of accuracy or completeness are made.

**Note:** This document contains information on products in the design, sampling and/or initial production phases of development. This information is subject to change without notice. Verify with your IBM field applications engineer that you have the latest version of this document before finalizing a design.

You may use this documentation solely for developing technology products compatible with Power Architecture®. You may not modify or distribute this documentation. No license, express or implied, by estoppel or otherwise to any intellectual property rights is granted by this document.

THE INFORMATION CONTAINED IN THIS DOCUMENT IS PROVIDED ON AN "AS IS" BASIS. In no event will IBM be liable for damages arising directly or indirectly from any use of the information contained in this document.

IBM Systems and Technology Group  
2070 Route 52, Bldg. 330  
Hopewell Junction, NY 12533-6351

The IBM home page can be found at [ibm.com](http://ibm.com)®.

29 September 2014

**Contents**

1	Executive Summary.....	4
2	Business Problem.....	4
3	IBM Solution .....	4
	3.1.1 CAPI Hardware .....	5
	3.1.2 CAPI Software.....	8
4	Target Markets and Segmentation .....	8
5	Strategy for Growth and Adoption .....	10
	5.1 Leveraging the OpenPOWER Foundation.....	11
	5.2 Early Product Offering and Future Directions .....	11
6	Conclusion.....	12

**List of Figures**

Figure 1: CAPI Hardware Ecosystem .....	6
Figure 2: Dedicated Process Model.....	7
Figure 3: Application as a Service Model.....	7
Figure 4: Potential Markets for CAPI .....	9
Figure 5: CAPI Ecosystem Partners and Consumers.....	11

## 1 Executive Summary

IT organizations must provide increased system performance as their workloads grow with demands for big data analysis, social media applications, continuous customer connectivity, and business-specific applications. Increases in processor performance alone cannot satisfy the workload demands, so solutions must also come from system-level advances such as hybrid computing, processing engine customization, and open platform development that enables cross-company innovation.

The Coherent Accelerator Processor Interface (CAPI) on IBM® POWER8™ systems provides solution architects with a new way to gain system-level performance. CAPI connects a custom acceleration engine to the coherent fabric of the POWER8 chip. The hybrid solution has a simple programming paradigm while delivering performance well beyond today's I/O attached acceleration engines.

## 2 Business Problem

While the growth of microprocessor performance via device physics has slowed, organizations still require IT solutions that can process the enormous amounts of data required by today's business initiatives. At the same time, these IT solutions are expected to maintain the accustomed price-performance curve.

IT providers must turn to system-level solutions and partnerships to meet the compute requirements that fuel business innovations, because core performance alone cannot achieve the desired goals. Acceleration engines and hybrid computing are potential strategies for system-level performance gain. Whether the task is searching through large amounts of data for specific patterns in real time, conducting aerodynamic analysis on the next generation of commercial aircraft, or planning for radiotherapy treatments, offloading these computation-heavy tasks to specialized accelerators is an intriguing solution.

However, technical challenges have prevented acceleration engines from achieving mainstream acceptance. A particular algorithm can be accelerated significantly with specialized hardware. Frequently, however, the overhead of controlling and communicating with the accelerator device offsets the raw algorithm speedup resulting in little or no gain in overall application performance. This overhead comes from copying data to pinned pages and communicating with the accelerator device through interrupts and memory-mapped I/O. The programming complexity of dealing with these issues has made hardware acceleration infeasible for many applications.

Alternatively, software programs running specialized algorithms on general-purpose computing cores avoid the I/O driver and data setup penalties. However, this model falls short in the cost and performance metrics as compared to an application-specific hardware accelerator or FPGA implementation.

## 3 IBM Solution

The Coherent Accelerator Processor Interface (CAPI) on POWER8 systems provides a high-performance solution for the implementation of client-specific, computation-heavy algorithms on an FPGA. This innovation can replace either application programs running on a core or custom acceleration implementations attached via I/O. CAPI removes the overhead and complexity of the I/O subsystem, allowing an accelerator to operate as part of an application. IBM's solution enables higher system performance with a much smaller programming investment, allowing hybrid computing to be successful across a much broader range of applications.

## Advance **Coherent Accelerator Processor Interface (CAPI) for POWER8 Systems**

---

In the CAPI paradigm, the specific algorithm for acceleration is contained in a unit on the FPGA called the accelerator function unit (AFU or accelerator). The purpose of an AFU is to provide applications with a higher computational unit density for customized functions to improve the performance of the application and offload the host processor. Using an AFU for application acceleration allows for cost-effective processing over a wide range of applications. A key innovation in CAPI is that the POWER8 system contains custom silicon that provides the infrastructure to treat the client's AFU as a coherent peer to the POWER8 processors.

Because of CAPI's peer-to-peer coherent relationship with the POWER8 processors, data intensive programs are easily offloaded to the FPGA, freeing the POWER8 processor cores to run standard software programs. Some applications that are well suited for CAPI include Monte Carlo algorithms, key-value stores, and financial and medical algorithms.

CAPI can also be used as a base for flash memory expansion, as is the case for the IBM Data Engine for NoSQL – Power Systems Edition. This innovative product gives the system access to 40 TB of data through the CAPI-connected solution.

The overall value proposition of CAPI is that it significantly reduces development time for new algorithm implementations and improves performance of applications by connecting the processor to hardware accelerators and allowing them to communicate in the same language (eliminating intermediaries such as I/O drivers).

### 3.1.1 CAPI Hardware

IBM's CAPI innovation on POWER8 systems is possible because of a dedicated silicon area on the chip that enables a client-defined hybrid-computing engine to act as a peer to the multiple POWER8 cores on the chip. The unit on the POWER8 chip, called the coherent accelerator processor proxy (CAPP) unit, participates directly in the POWER8 core coherency protocols on behalf of the acceleration engines (AFU). The CAPP unit maintains a directory of all cache lines held by the off-chip accelerator, allowing it to act as the proxy that maintains architectural coherence for the accelerator across its virtual memory space.

IBM also supplies a power service layer (PSL) that resides on the FPGA alongside of the acceleration engine. The PSL works in concert with the CAPP unit across a PCIe connection. This connection is completely managed by the POWER8 chip and the PSL, allowing the client to focus on their accelerator algorithm. The PSL provides a straightforward command → data buffer → command response interface to the client-written accelerator, which grants access to coherent memory. Figure 1: CAPI Hardware Ecosystem shows all of these hardware components.

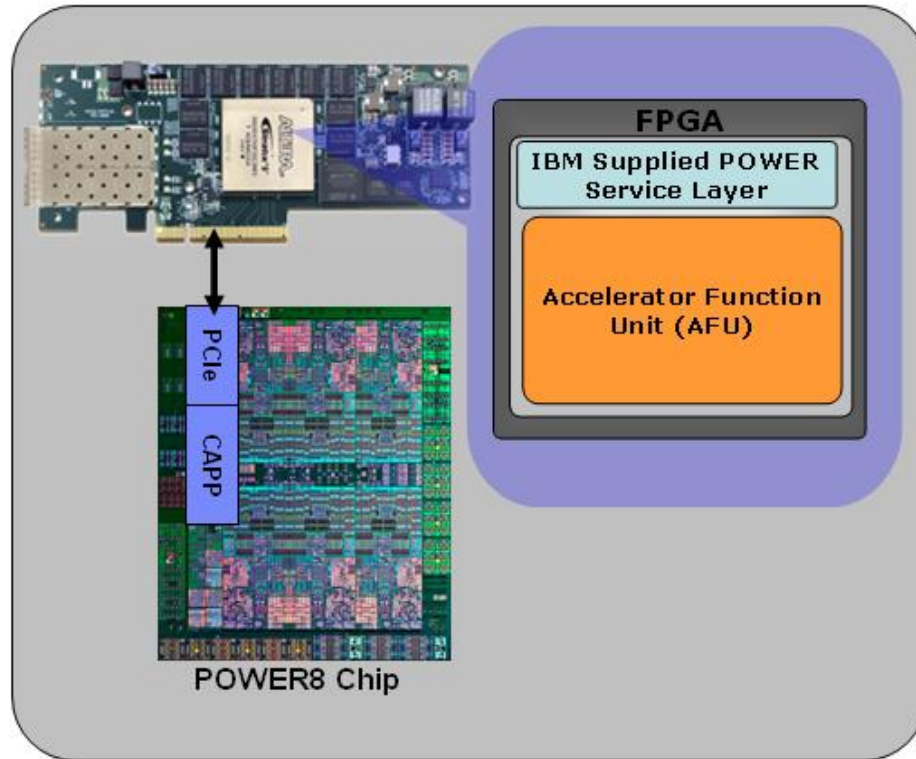


Figure 1: CAPI Hardware Ecosystem

Equally important is the translation function that the CAPP and PSL provide for the accelerator. The accelerator uses the same virtual memory addressing space as the core application that enables it. The CAPP and PSL handle all virtual-to-physical memory translations, simplifying the programming model and freeing the accelerator to do the number crunching directly on the data it receives.

In addition, the PSL contains a 256 KB resident cache on behalf of the accelerator. Based on the needs of the algorithm, the accelerator can direct the use of the cache via the type of memory accesses (reads/write) as cacheable or noncacheable. Fundamentally, the rich set of commands that the accelerator can send to the PSL reflects all of the types of memory accesses available to the POWER8 cores themselves (such as reads, reads with intent-to-modify, reservations, locks, writes, and writes to highest point of coherency).

The accelerator runs hand-in-hand with an application running on the core as shown in Figure 2: Dedicated Process Model.

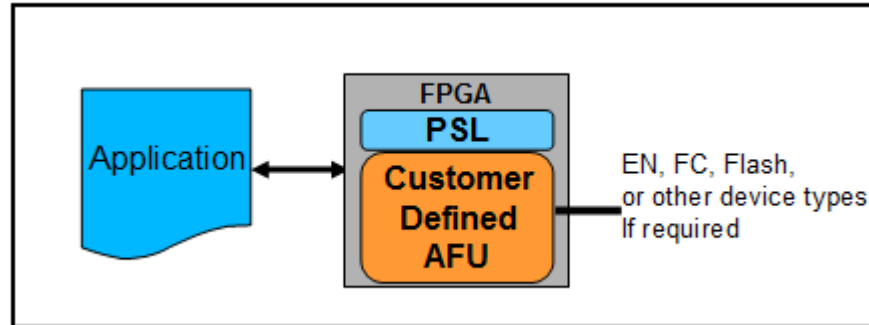


Figure 2: Dedicated Process Model

While the application executes on the host processor, the CAPI model offloads computation-heavy functions to the accelerator. The accelerator is a full peer to the application. The accelerator uses an unmodified virtual address with full access to the application's address space. It uses the processor's page tables directly with page faults handled by system software.

For a CAPI solution, the application might set up the data for the AFU. The application sends the AFU a process element to initiate the data-intensive execution. The process element contains a work element descriptor (WED) provided by the application. The WED can contain the full description of the job to be performed or a pointer to other main memory structures in the application's memory space.

The application can be the master or the slave to the accelerator—whichever mechanism is demanded by the algorithm. For example, as described previously, the application can set up data in memory and then tell the accelerator to start work by sending the address of the data structure. Alternatively, the accelerator can receive incoming packets from the network, perform work on the data, and then inform the application that the processed data is ready for consumption.

The acceleration platform can be integrated into cloud-based services. In this case, the application on the core runs as an Application as a Service (AaaS) for other applications that require the services of the accelerator, as shown in Figure 3: Application as a Service Model. In this case, the AaaS maintains work queues for the accelerator on behalf of each of the requesting threads and performs any maintenance functions needed to inform the accelerator of pending work. This is yet another example of the power of the CAPI innovation because the accelerator can easily follow link lists (pointer chasing) of command-work queues throughout the shared virtual address space.

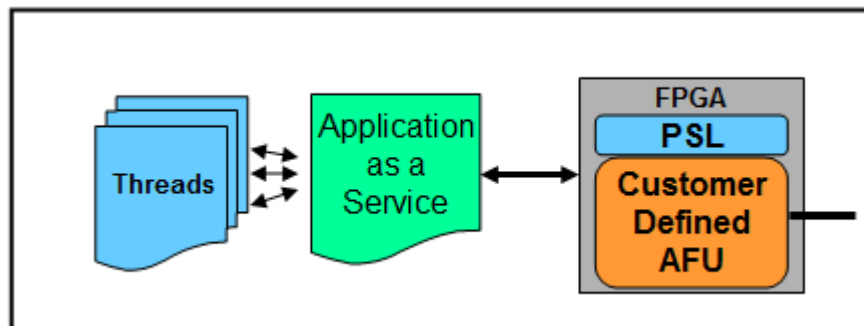


Figure 3: Application as a Service Model

### 3.1.2 CAPI Software

Enablement of a CAPI solution entails the logical connection of an application running on a POWER8 core with an accelerator running in conjunction with the IBM PSL. These connections are made through operating-system kernel extensions and library functions created specifically for CAPI.

The operating-system kernel extensions will be made available to all POWER-Linux little-endian distributions starting with [Canonical's Ubuntu](#). When the system boots with a CAPI device connected on a PCIe port, the kernel extensions recognize the CAPI device. They perform personalized system initializations, including configuration of the connection of the CAPP unit to the PCIe port. Additional kernel extensions perform maintenance, system protection, and communication functions.

The client application uses a CAPI library called "libcxl," which provides multiple functions for the application to connect, call, communicate, and disconnect with an available CAPI device. Libcxl calls include many basic functions, such as those that allow the application to provide the virtual storage address of data structures to the accelerator. It also provides a memory-mapped I/O interface for the application to read/write registers in the accelerator's register space on the FPGA.

## 4 Target Markets and Segmentation

Potential CAPI consumers come from a diverse set of industries and research areas. Workloads with computation-heavy algorithms are a prime target space for CAPI, as are edge-of-network engines and parallel programming challenges.

The market for accelerating complex simulations using hardware accelerators is vast. Figure 4: Potential Markets for CAPI illustrates some of the industries and target markets that use complex algorithms to conduct business. These potential markets are just a subset of the overall market space where CAPI might be applicable.





Figure 4: Potential Markets for CAPI

**Big Data/Database/Compute** – Analytics and algorithms that need to work on large masses of data are candidates for CAPI acceleration. Also included in this space are edge-of-network processing engines, such as video encoding/decoding or Ethernet packet processing. For database applications, CAPI can be used as “slow memory” to give the system a feel of having tens-of-terabytes of main storage (as compared to disk I/O), freeing compute cores of huge device-driver processing overhead and minimizing access latency.

**Social/Media** – To reduce production lags, today’s media data centers invest in high-performance computing (HPC) cluster technology for highly complex tasks. These engines can be implemented as CAPI acceleration. In addition, CAPI can assist in making database connections based on known behaviors and biases, personalizing behavioral patterns, and creating individualized experiences for social media and business.

**Medicine** – Research universities and medical centers can implement CAPI for cancer research by using Monte Carlo algorithms to run radiation therapy simulations. These solutions customize therapy for individuals and bring simulation times down from weeks to hours.

**Finance/Insurance** – High-frequency trading, optimizing order books, and detecting fraud use high-performance computing and specialized engines to explore more potential scenarios and shorten the time required. In these cases, milliseconds matter, and CAPI acceleration engines are a prime space for creating these specialized engines and connecting disjoint data.

---

**Advance Coherent Accelerator Processor Interface (CAPI) for POWER8 Systems**

---

**Visual/Biometric Analysis** – Today’s leading solutions for visual, security, and biometric processing use FPGAs in the I/O model paradigm. These algorithms stand to gain critical performance from CAPI coherency and the removal of device-driver overhead. Solutions include not only security analytics but also retail markets for real-time restocking analysis and assistance for the visually impaired.

**Manufacturing/EDA** – Computer design algorithms and industrial engineering analysis, offloaded to a CAPI device, can provide huge performance benefits. Specific engineering algorithms can be created as end-to-end solutions for manufacturing lines and complex engineering tasks.

**Oil and Gas Industry Modeling** – Oil and gas companies can use CAPI for HPC technology and for reverse-time migration simulations to minimize the time involved in processing massive amounts of data in order to reduce costs and speed production.

**Weather** – Meteorological forecasting and research require high-speed processing of massive data inputs, so productivity depends on processing speed. Data centers can use specialized engines for weather research, combining the power of offload acceleration with CPUs.

## 5 Strategy for Growth and Adoption

CAPI is applicable for clients who run computation-intensive algorithms. These clients might be OpenPOWER Foundation partners or other partners who create high-demand CAPI acceleration solutions (CAPI-Apps) for downstream POWER8 clients who design their own acceleration engine for a competitive IT advantage. IBM is also creating CAPI solutions for the marketplace, as shown in Figure 5: CAPI Ecosystem Partners and Consumers.

IBM and partner prepackaged solutions are sold to clients and work for that solution only. Prepackaged solutions contain the accelerator algorithm, the application code, and all necessary library extensions. Partners providing CAPI solutions can work with OpenPOWER Foundation card and FPGA providers to personalize the hardware solution to meet their specific algorithm’s need.

Field upgrades for prepackaged solutions can be achieved through predetermined means such as flash updates for the accelerator and application downloads for certified users. However, repurposing of the prepackaged solution can only occur at the solution provider’s discretion.

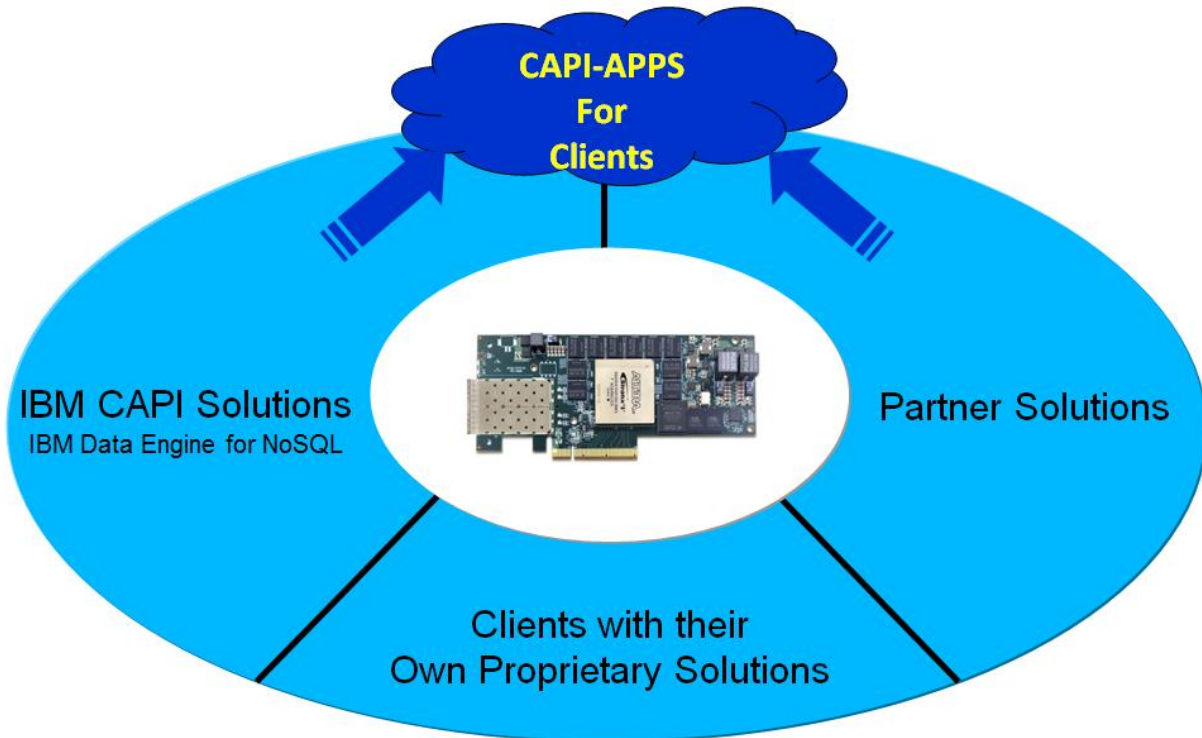


Figure 5: CAPI Ecosystem Partners and Consumers

Alternatively, clients who design CAPI solutions for their own IT advantage procure a reconfigurable CAPI card and the power service layer (PSL). These clients have development teams with application, logic design, and FPGA skills in house. The development teams can create one or more CAPI solutions and can reconfigure the card as necessary to meet their IT requirements.

## 5.1 Leveraging the OpenPOWER Foundation

Multiple OpenPOWER Foundation members are working on new CAPI-App solutions today. These early adopters will provide solutions in many of the target market spaces discussed previously. Look for innovative products from algorithms to cards, boards, and systems that use the collective expertise and core competencies of multiple OpenPOWER Foundation members.

## 5.2 Early Product Offering and Future Directions

CAPI is in the developmental and initial production phases of development. IBM's initial product offering is the CAPI Developer Kit, which will enable clients to begin developing CAPI accelerators and applications. The CAPI Developer Kit will include documentation for FPGA designers and programmers that describes future architecture enhancements beyond the initial CAPI Developer Kit version. Based on client feedback, IBM might choose to implement these features or might change direction. Therefore, the following statements about future plans are subject to change.

## Advance Coherent Accelerator Processor Interface (CAPI) for POWER8 Systems

Table 1: CAPI Future Enablement Direction describes some of the CAPI Developer Kit features and IBM's potential future direction for CAPI.

Feature	CAPI Developer Kit Capability	Future Plans
Virtualization	Dedicated process mode	+ Time sliced (AFU or PSL directed)
Number of AFU engines per PSL	Single AFU	Multiple AFUs per PSL
Translation ordering	Strict/page only	All as defined by Power Architecture
Operating system (target)	Ubuntu 14.10	AIX, Ubuntu, Redhat, and so on

*Table 1: CAPI Future Enablement Direction*

As shown in Figure 2: Dedicated Process Model and Figure 3: Application as a Service Model, the CAPI Developer Kit supports a dedicated process model where the accelerator is connected to one application. In the future, IBM's direction is to enable virtualization capabilities where multiple applications can use a single acceleration engine. In the virtualized mode, an applications' access to the accelerator will be time sliced or context switched via directives from the PSL or from the AFU itself.

The initial product offering restricts the FPGA to containing a single AFU. Thus, there is a 1:1:1 ratio of CAPP to PSL to AFU. As CAPI FPGA sizes grow, a client might want to enable multiple AFUs per FPGA. Depending on the size of the AFUs, a client might have up to a 1:1:4 ratio of CAPP to PSL to AFUs. In this case, AFUs run completely independent of each other and can be paired with different applications and different address spaces. The PSL maintains system and address space integrity, only allowing the individual AFUs to access their own virtual memory locations.

In all future outlooks, there will always be one PSL to every CAPP unit. Future POWER8 chips might have multiple CAPP units per socket, allowing for an even larger number of CAPP-PSL pairs and thereby allowing for more AFUs.

Early implementations allow only for strict page ordering by the PSL when performing translations on behalf of the AFU. Future enhancements might enable all of the translation ordering schemes defined by the Power Architecture®.

As described in section 3.1.2 CAPI Software, CAPI operating-system kernel extensions will initially be available in Canonical's Ubuntu Linux distribution. IBM intends to enable CAPI in AIX as well as other major IBM Power Systems™ operating systems.

## 6 Conclusion

CAPI, the Coherent Accelerator Processor Interface, is an IBM innovation that enables clients to accelerate workloads on FPGAs faster and more simply than ever before. CAPI is available only on POWER8 systems. Through a unique hardware investment and operating-system additions, CAPI on POWER8 processors provides clients with a customizable acceleration engine that runs as a peer to the

**Advance** **Coherent Accelerator Processor Interface (CAPI) for POWER8 Systems**

---

POWER8 cores. It accesses memory using the same programming methods and virtual address space as the application that calls it.

The CAPI solution provides much faster data preparation time when compared to a traditional, I/O attached FPGA by removing the need for a device driver and its code stack. CAPI-attached devices can perform functions such as pointer chasing that an I/O attached device cannot do.

With vast market opportunities for acceleration offload engines and hybrid computing, CAPI will grow into a diverse set of markets. To enable these opportunities, clients can create CAPI solutions for their own IT organizations, while other partners can build prepackaged solutions to resell to clients. IBM will also sell CAPI-based IT solutions, starting with the IBM Data Engine for NoSQL - Power Systems Edition.