

BIGSTREAM Hyperacceleration for Apache Spark for Samsung SmartSSD CSDs powered by Xilinx FPGAs

HYPERACCELERATION WITH THE SMARTSSD CSD

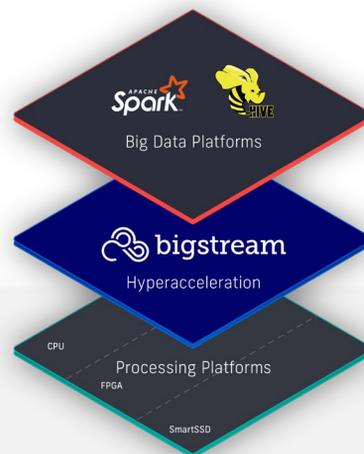
Organizations look to Apache Spark for essential workloads in their data environment. As they attempt to scale, though, they typically face challenges of performance and cost. Hardware acceleration represents an opportunity to solve these challenges but introduces the challenge of intricate programming and altering Spark code.

Bigstream Hyperacceleration solves these problems, introducing software to utilize specialized hardware without requiring any Spark code change.

The Samsung SmartSSD Computational Storage Device (CSD) represents a shift in the compute-storage paradigm, processing data on the storage device, drastically reducing the volume of data that needs to move to the CPU.

SCALING WITH DATA VOLUME

A single server can contain multiple SmartSSD CSDs (up to 24 U.2 SmartSSDs in a typical 2U server). Each SmartSSD CSD can run query acceleration in parallel, producing almost linear speedup, even on a low-end or highly over-subscribed host CPU. While NVMe SSDs are fast, most enterprise SSDs are limited to only four lanes of PCI-Express by the U.2 drive connector, creating a bottleneck. By processing data directly on the storage layer, SmartSSD can perform operations using wider interfaces than the U.2 connector supports, or even access storage via a newer PCI-Express generation than supported by the host CPU. By comparison, unaccelerated SSDs stop delivering additional throughput when the host CPU runs out of PCI-Express lanes.



KEY BENEFITS

- > **Faster Insights** – Jobs run up to **10x faster** without changing your Spark code, increasing data science productivity and the ability to meet tight SLAs
- > **Enriched Analytics** – incorporates fuller data sets and more data sources by eliminating the bottlenecks forcing you to make compromises
- > **Lower TCO** – instead of scale-out and scale-up (adding or upgrading servers), optimize your existing Spark environment with hyperacceleration, lowering hardware and operating expenses.

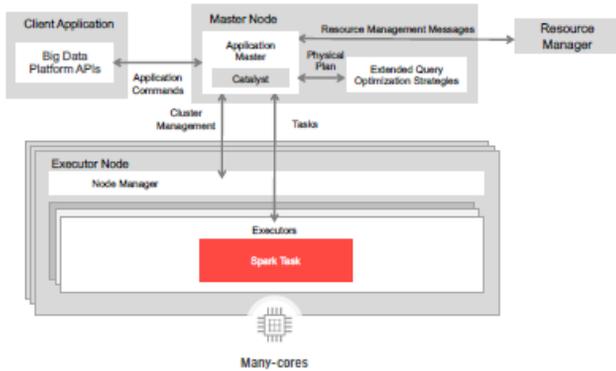
HOW IT WORKS

The solution combines advanced computational storage with a complete software stack for integration into existing analysis workflows. Queries are automatically optimized by Bigstream software to allow efficient processing of data in situ, by the SmartSSD CSD itself, thus avoiding large data movement between storage and CPU and speeding up insight generation.

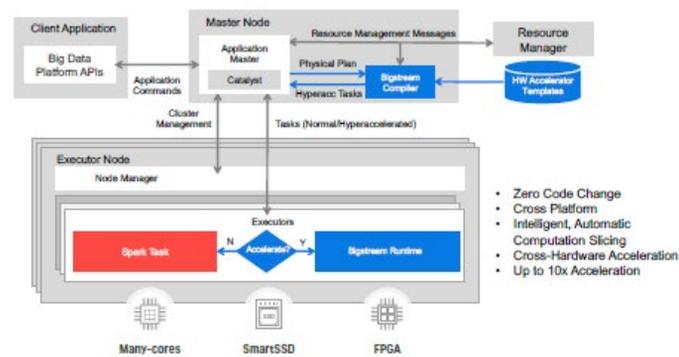
SEAMLESS ACCELERATION WITH SPARK

- > Standard Spark (strategies)
- > Zero code change
- > SW/HW easy to install – unlike CPU-attached accelerators, SmartSSD CSD connects using the existing SSD PCI-Express lanes. This configuration frees up PCIe lanes for other uses, such as networking, or additional storage throughput.

UNACCELERATED



WITH BIGSTREAM



Along with Bigstream Hyperacceleration, the SmartSSD CSD accelerates Spark performance by up to 10x at a far lower TCO than typical scale-up and scale-out approaches.

TAKE THE NEXT STEP

Visit Bigstream > www.bigstream.com
www.xilinx.com/smartssd

	Compatibility	Apache Spark 2.1.1, or Spark 2.2.1
	Host Operating System	Ubuntu 16.04.5 LTS, or Kernel version 4.4.0-135-generic
	Data File formats supported for acceleration	Row: CSV, JSON Coming soon: Column: ORC, Parquet, AVRL
	Form Factor	2.5"
	Capacity	2TB, 4TB, 8TB
	Spec Compliance	NVMe spec rev. 1.2, PCIe base specification rev. 4.0
	NAND flash memory	Samsung V-NAND
	Acceleration Engine	Xilinx Kintex Ultrascale+ "KU15P" FPGA
	Power Consumption	Up to 25W
	Physical Dimensions	69 x 100 x 14 mm
	Weight	400 grams

