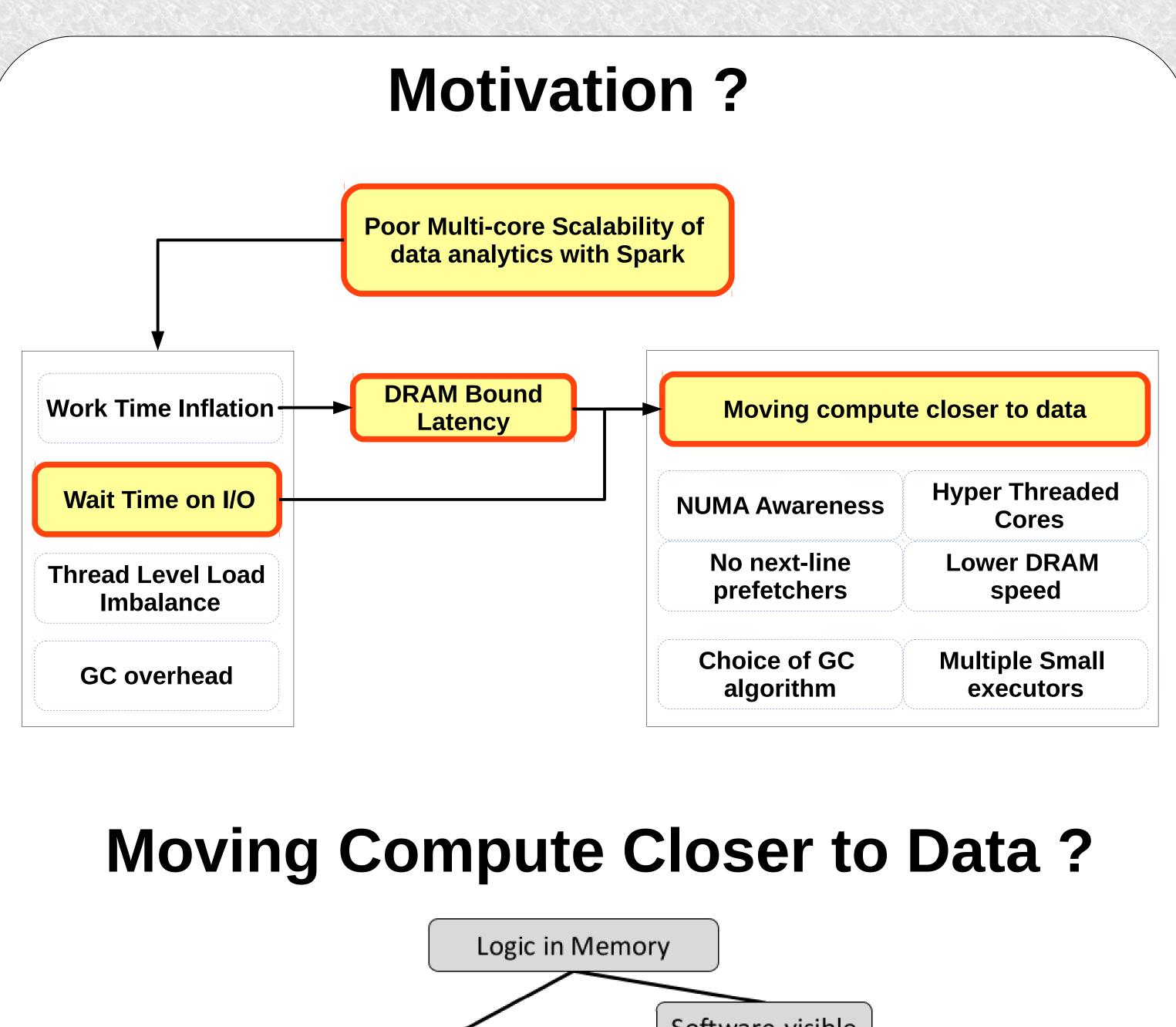# Project Night-King: Improving the performance of Big Data Analytics using Near Data Processing Architectures
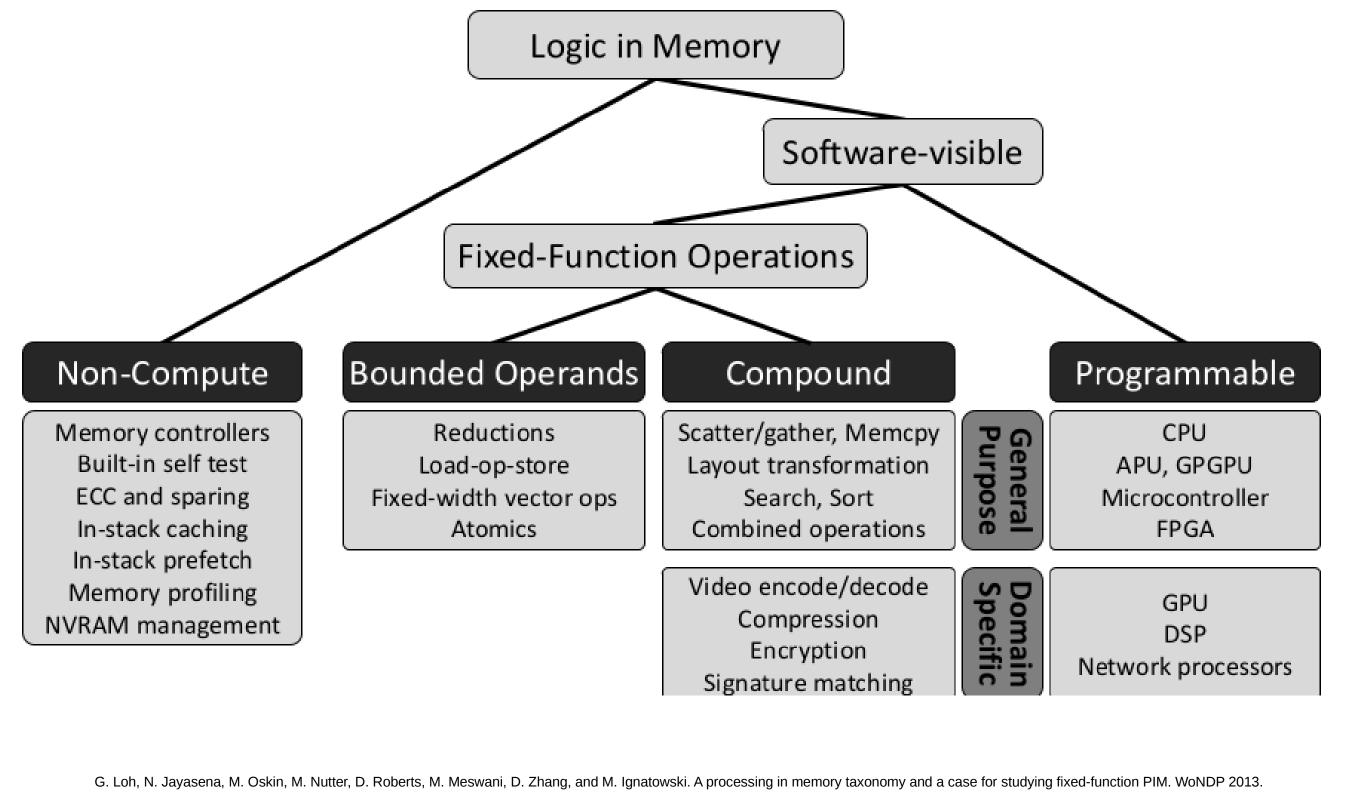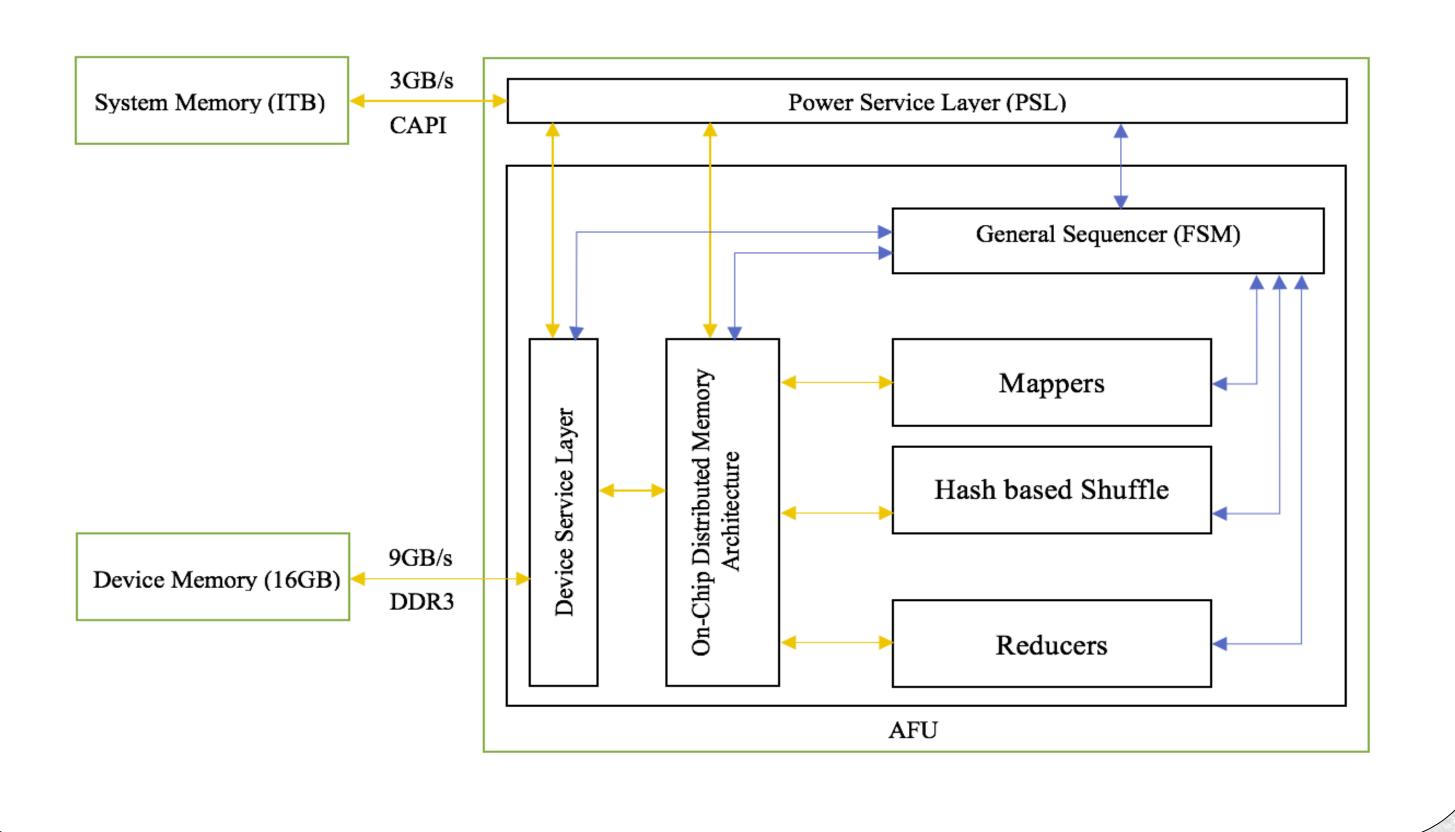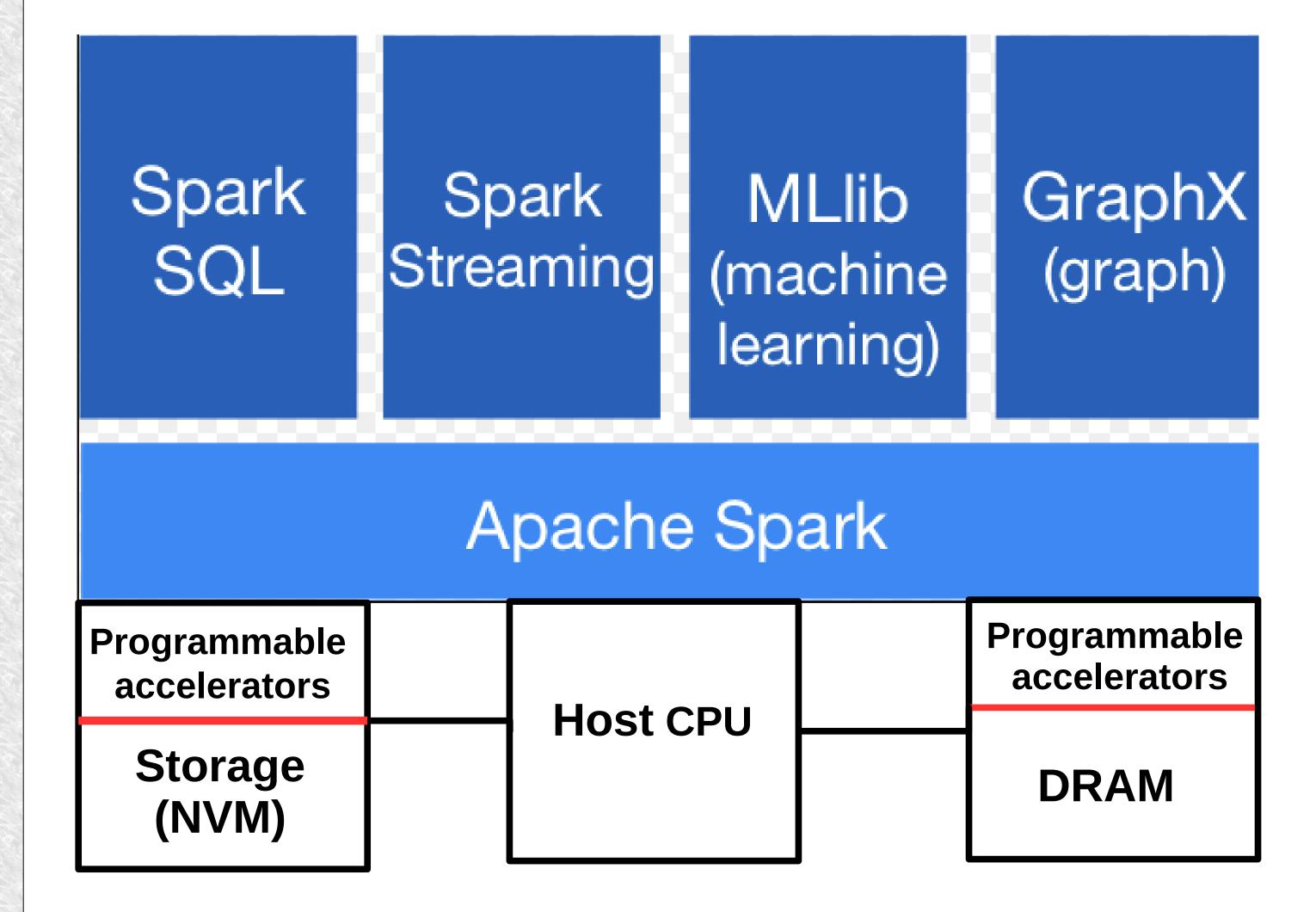
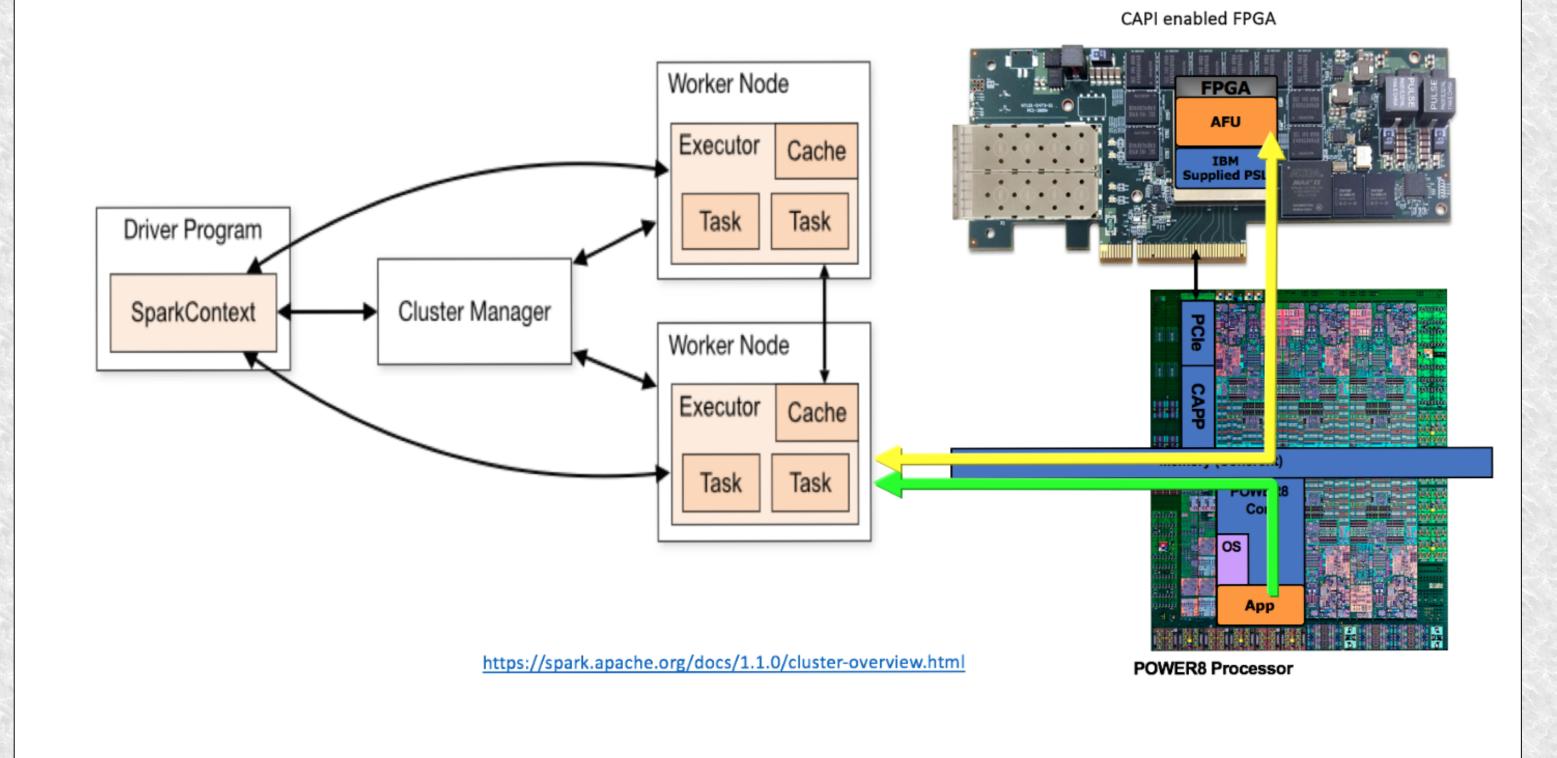## Motivation ?

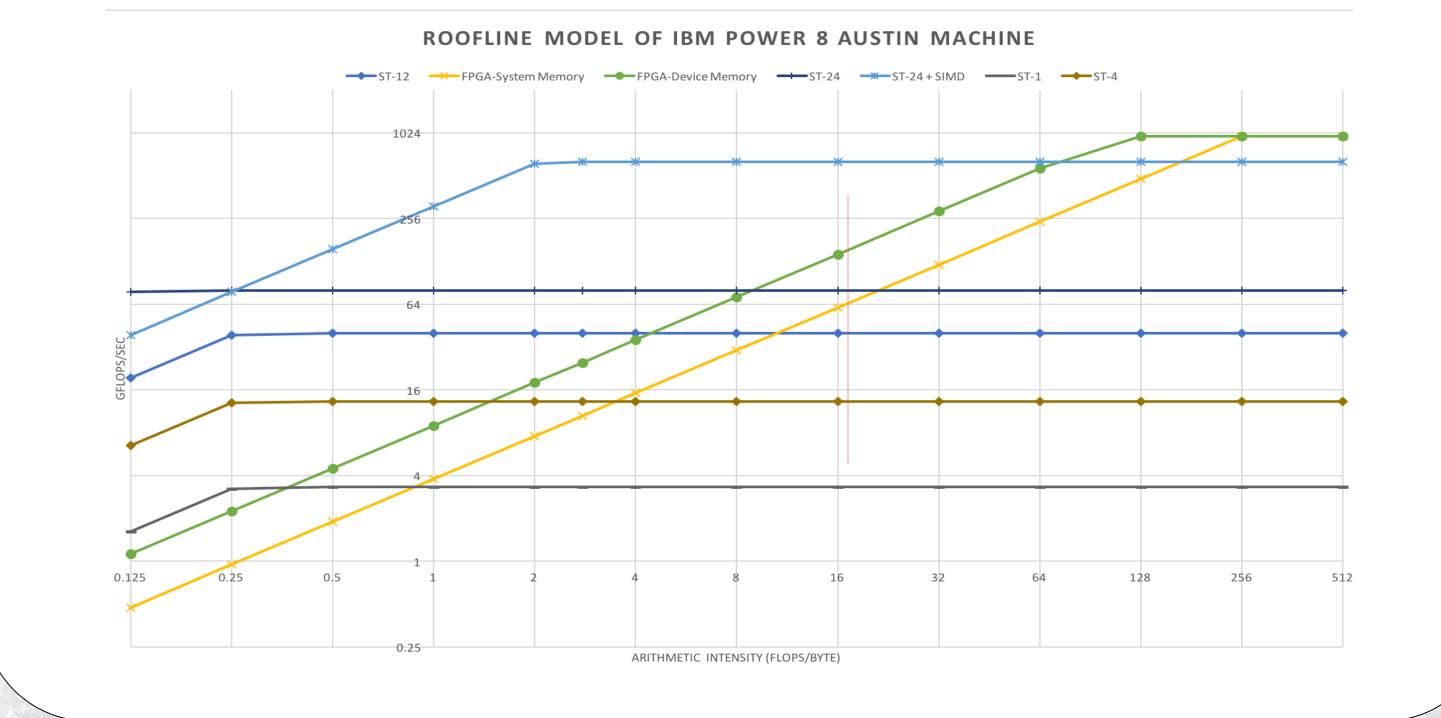Poor Multi-core Scalability of data analytics with Spark

Work Time Inflation

Wait Time on I/O

Thread Level Load Imbalance

GC overhead

DRAM Bound Latency

Moving compute closer to data

NUMA Awareness	Hyper Threaded Cores
No next-line prefetchers	Lower DRAM speed
Choice of GC algorithm	Multiple Small executors

Moving Compute Closer to Data ?

Logic in Memory

Software-visible

Fixed-Function Operations

Non-Compute	Bounded Operands	Compound		Programmable
Memory controllers Built-in self test ECC and sparing In-stack caching In-stack prefetch Memory profiling NVRAM management	Reductions Load-op-store Fixed-width vector ops Atomics	Scatter/gather, Memcpy Layout transformation Search, Sort Combined operations	General Purpose	CPU APU, GPGPU Microcontroller FPGA
		Video encode/decode Compression Encryption Signature matching	Domain Specific	GPU DSP Network processors

G. Loh, N. Jayasena, M. Oskin, M. Nutter, D. Roberts, M. Meswani, D. Zhang, and M. Ignatowski. A processing in memory taxonomy and a case for studying fixed-function PIM. WoNDP 2013.

Programmable Accelerators ?

System Memory (1TB) 3GB/s CAPI

Power Service Layer (PSL)

General Sequencer (FSM)

Device Service Layer

On-Chip Distributed Memory Architecture

Mappers

Hash based Shuffle

Reducers

Device Memory (16GB) 9GB/s DDR3

AFU

Scope of the Project ?

| Spark SQL | Spark Streaming | MLlib (machine learning) | GraphX (graph) |

Apache Spark

| Programmable accelerators | | Host CPU | | Programmable accelerators |
| Storage (NVM) | | | | DRAM |

Experimental Setup ?

Driver Program — SparkContext — Cluster Manager

Worker Node — Executor — Cache — Task — Task

Worker Node — Executor — Cache — Task — Task

https://spark.apache.org/docs/1.1.0/cluster-overview.html

CAPI enabled FPGA

FPGA — AFU — IBM Supplied PSL

PCIe — CAPI — PHM Cm — OS — App

POWER8 Processor

Initial Results ?

ROOFLINE MODEL OF IBM POWER 8 AUSTIN MACHINE

ST-12   FPGA-System Memory   FPGA-Device Memory   ST-24   ST-24 + SIMD   ST-1   ST-4

GFLOP/SEC

1024

64

16

4

1

0.25

0.125 0.25 0.5 1 2 4 8 16 32 64 128 256 512

ARITHMETIC INTENSITY (FLOPS/BYTE)

AHSAN JAVED AWAN
(ajawan@kth.se)

KTH VETENSKAP OCH KONST

BSC

IBM Research