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
- DRAM Bound Latency
- Wait Time on I/O
- Thread Level Load Imbalance
- GC overhead

Moving Compute Closer to Data?

- Logic in Memory
- Software-visible
- Non-Compute
  - Memory controllers
    - Built-in self test
    - ECC and spiring
    - In-stack caching
    - In-stack prefetch
    - Memory profiling
    - NUMA management
- Fixed-Function Operations
- Compound
  - Reductions
    - Load-op stores
    - Fixed-width vector ops
  - Memory
    - Scatter/gather
    - Memory layout
    - Transformation search
    - Sort
    - Combined operations
  - Graph processing
- Programmable
  - GPU
  - APU
  - Microcontroller
  - FPGA
  - Network processors

Programmable Accelerators?

- System Memory (SMM)
  - JGIBs
  - CAPI
- Device Memory (DMM)
  - JGIBs
  - DDR3

Scope of the Project?

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

Experimental Setup?

- Storage (NVM)
- Host CPU
- Programmable accelerators

Initial Results?

- AHSAN JAVED AWAN
  (ajawan@kth.se)