Overcoming Traditional Data Analytics Performance Bottlenecks with Inline Acceleration

Mike Strickland, Director, Data Center Solution Architect
Intel® Corporation
October 2017
Accelerate Unmodified Big Data Analytics with Intel® Programmable Acceleration Card

- **Intel® Big Data Analytics Frameworks and API Libraries**
  Accelerate innovation in big data analytics with frameworks built on software-defined infrastructure with open standard building blocks.

- **Intel® Frameworks and Libraries or D/B standard interfaces Integrated with FPGAs**
  Run unmodified customer applications, for both Intel Xeon® processor and FPGA support, with end-to-end virtualization and security.

- **Accelerate Relational, NoSQL, and Unstructured Data**
Inline Acceleration for Unmodified Customer Applications
Streaming/Networking & Storage $\rightarrow$ FPGA $\rightarrow$ CPU (avoid two data copies)

Single Multi-function Accelerator

Integrate to Intel® Frameworks and APIs
- Run unmodified customer applications
- End-to-end security and virtualization

Lookaside Acceleration
- PCIe lookaside algorithm acceleration
- Networking + streaming + data access
  - Inline acceleration and protocol acceleration
- Compression, filtering, encryption
- Fast lookups/hashing
NoSQL: System & IO Acceleration Opportunity
Source: rENIAC CEO

- Connection Management
- Compression/Encryption
- Book Keeping
- Data Encode/Decode

System & IO: 75%

Business Logic: 25%
rENIAC Distributed Data Engine/Switch (rDS)
4X+ Cassandra acceleration (Source: rENIAC)

Overview
- No customer application change
- Plug-in card with 10GbE
  - Proxy tier or on database server
- Distributed cache, proxy for reads and writes
- Predictable latency for SLAs
- Roadmap for storage compaction

Inline Acceleration
✓ Networking/CQL acceleration
✓ Data access acceleration
✓ Compression
✓ Hashing
Swarm64 Relational Database Acceleration
Two Workloads: Traditional Data Warehousing, Real Time Data Analytics

Database accelerate with a plugin

**Acceleration Overview**
- 10X+ single table inserts/s for real time data analytics
  - With modest tuning, 15M PostgreSQL INSERT/s*
- 2X+ optimized queries for data warehousing
  - Using industry standard TPC-DS benchmark
- 3X+ storage compression
  - Data & tables managed by Swarm64

**Inline Acceleration**
- No customer application change
  - Storage engine plugin: PostgreSQL, MySQL, ...
- Query Engine accelerate INSERT, SELECT, ...
- More io bandwidth, mem depth from compress.
- Optimized indexing

*Note*: Dual Intel® Xeon® E5-2695 v4 processors, (8) 32GB DDR4-2400, (8) 512GB NVMe SSD.