THUNDERX Workload Optimized Processor

Addressing the Needs of Next Generation Cloud HPC

Craig Prunty
Cavium Inc
Director , Europe Sales

Data Center R/Evolution

High Performance Computing ➔ Cloud

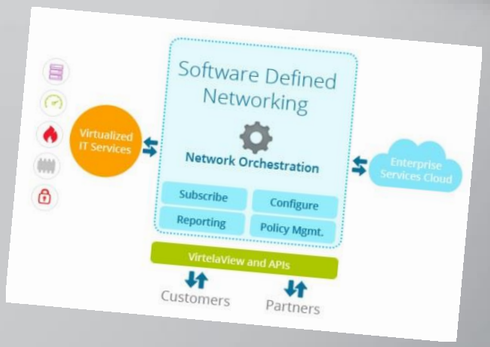
- Virtualization and Scale-out software architecture has changed the computing paradigm
- Compute/$/watt has become the COGS of many business
- Consolidation of work-loads justifies optimization
- IaaS, PaaS, SaaS abstracts HW opening the way for new architectures

Storage ➔ High Performance Big Data

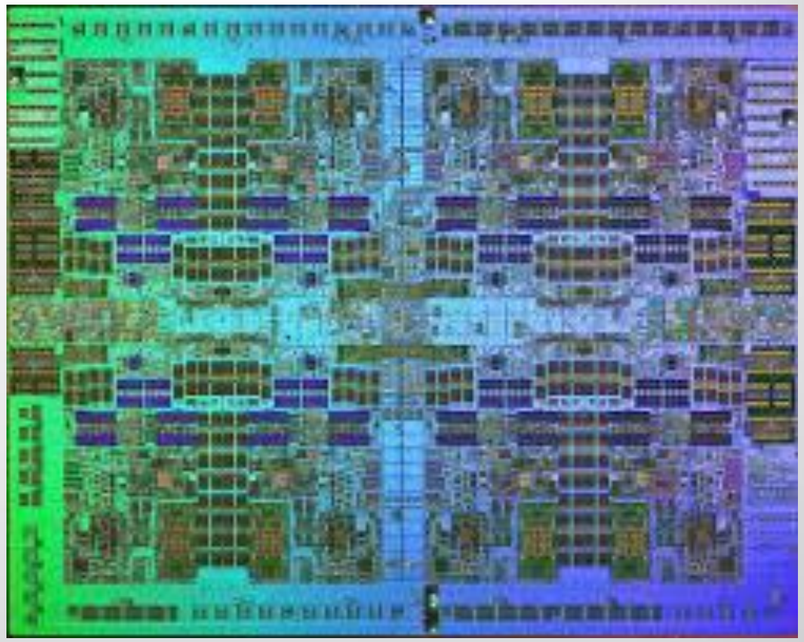
- Analytics on unstructured data is being used to improve business
- Differentiation with optimized Scientific Libraries
- Demand for Real Time – not just Batch

Networking ➔ SDN (Software Defined Networks)

- Driven by multi-tenant datacenter scaling
- Integrated switching enables support of range of topologies



HPC Evolution – Heterogeneous Multicore Architecture



➢ Combines different types of processors

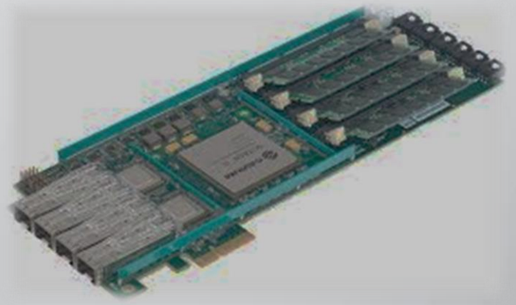
- Each optimized for a different operational modality
  - Performance > Nx better than other N processor types
- Synthesis favors superior performance
  - For complex computation exhibiting distinct modalities

➢ Co-processors

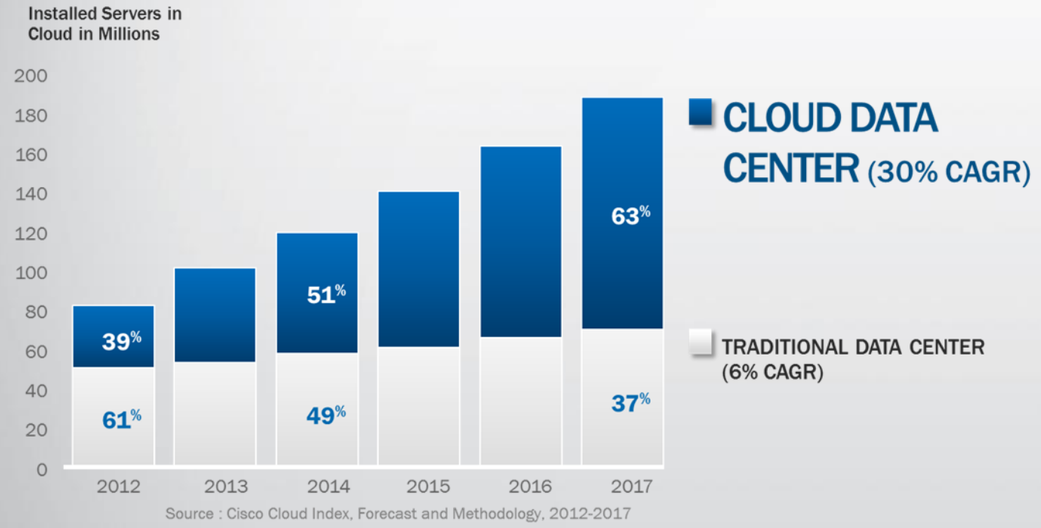
- Graphical processing units (GPU)
- Network controllers (NIC)
- Special purpose components are being applied to general applications

➢ Workload-specific accelerators

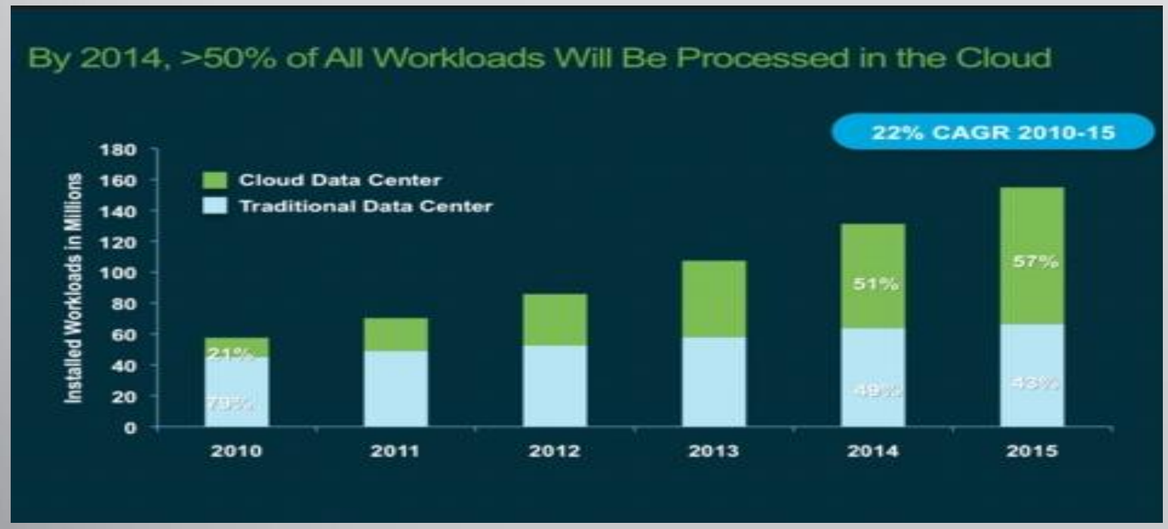
- Speedup specific classes of computational function
- GPGPU



Cloud is Changing Server Industry Dynamics

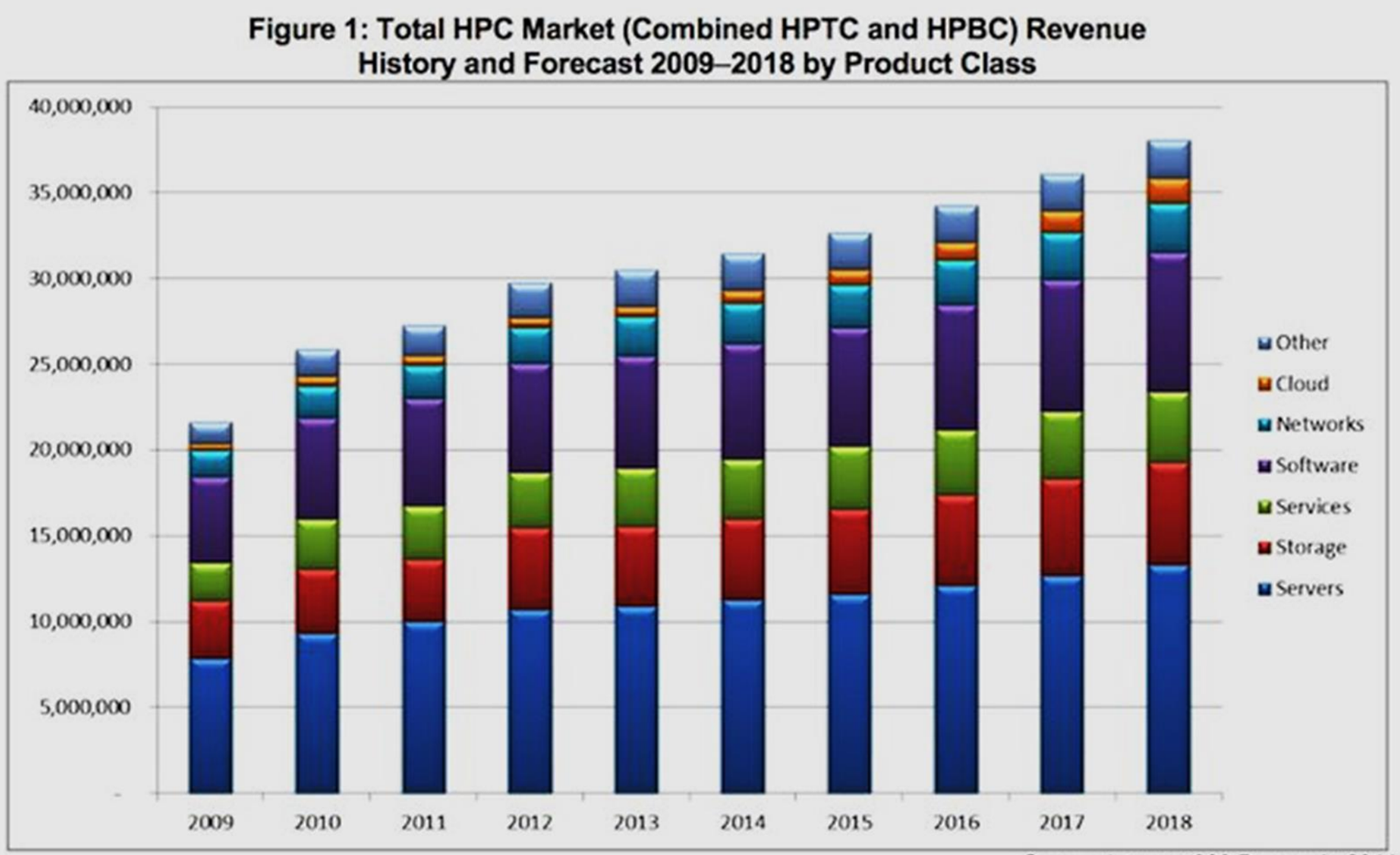


Source : Cisco Cloud Index, Forecast and Methodology, 2012-2017



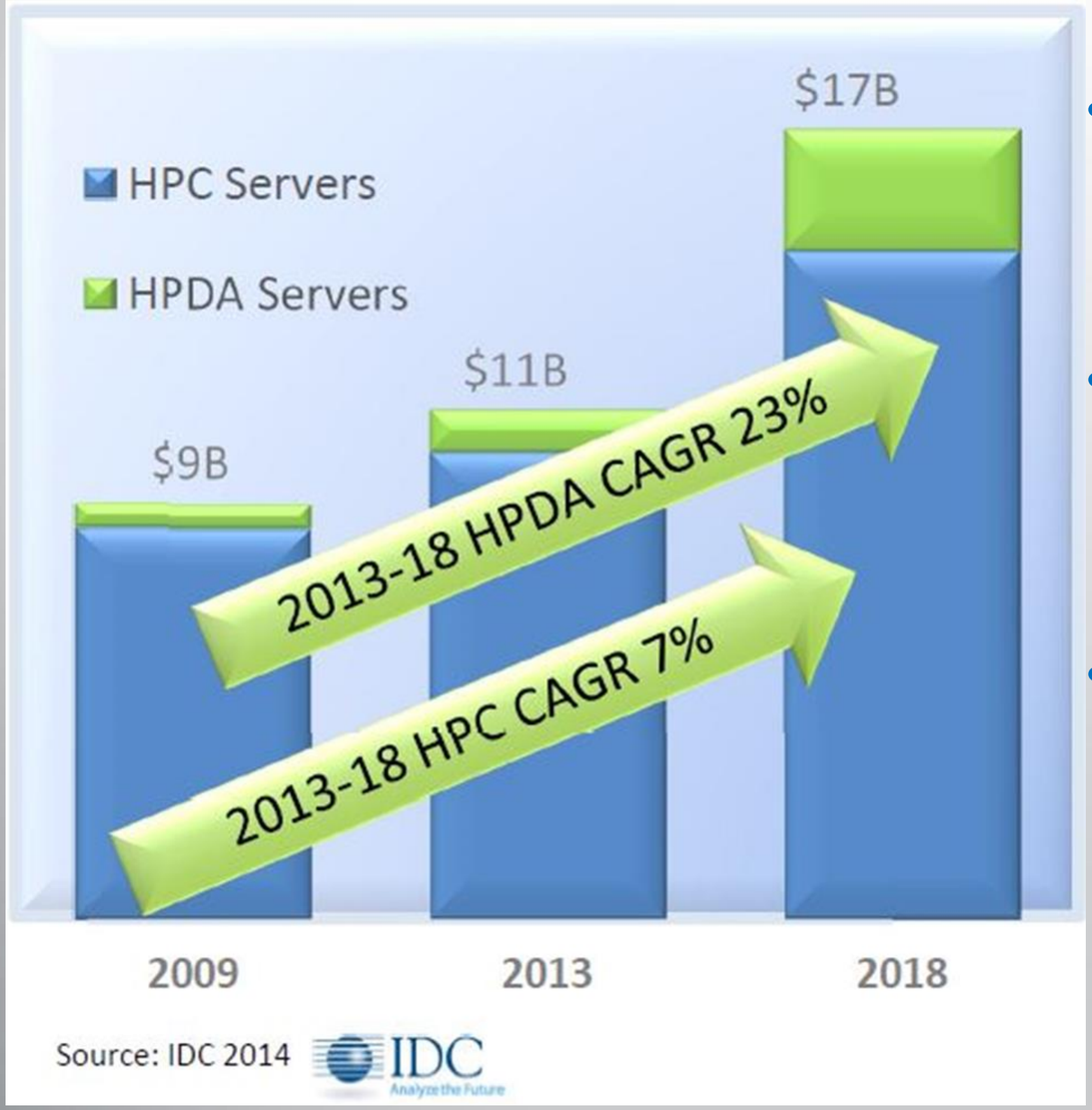
- Cloud deployments driving most of the server growth
- More and More workloads are getting deployed in the cloud

HPC Market Growth



- Notable market effects in the cloud, Ultrascale, and storage segments of the market,.
- HPC Cloud computing predicted to grow at an **18.6%** CAGR from 2009-2018..
- HPTC (High Performance Technical Computation)
- HPBC (High Performance Business Computation)

Data Analytics Growth in HPC



- Well established and growing core HPC market
  - Optimized for processing performance with high bandwidth memory
- High Growth Data Analytics
  - Built on basic batch processing
  - Demand for optimized Real Time
- Unique advantage with Silicon Class Integration (accelerators)

HPC Cloud

Targets **Software** and **Hardware** as a Service

Focus on High-performance utility computing

Target unlimited application resources

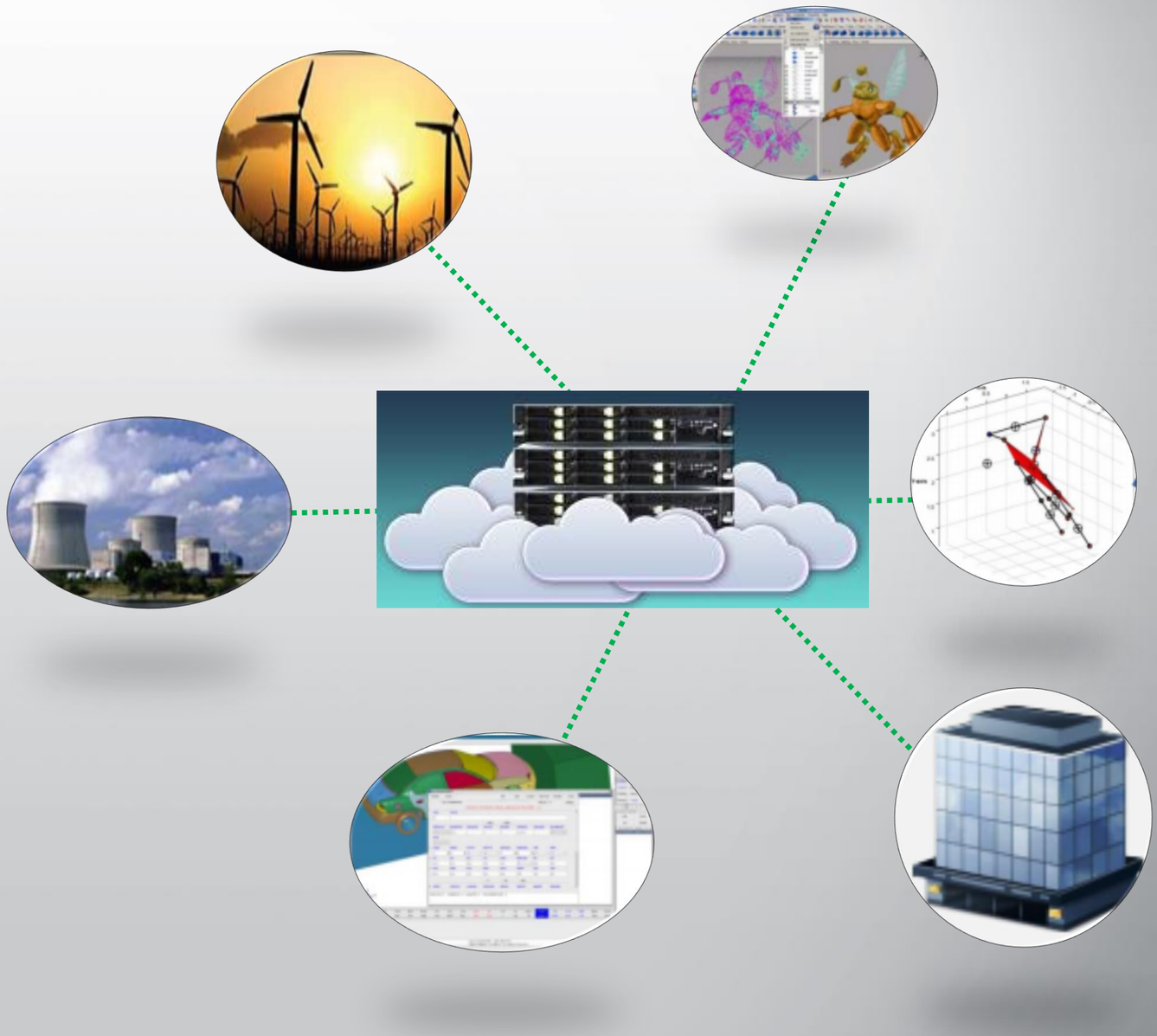
- Instant Resource Availability is critical

Efficient & Ease of Use

- Flexible, Secure and Dynamically Scalable

Key Applications:

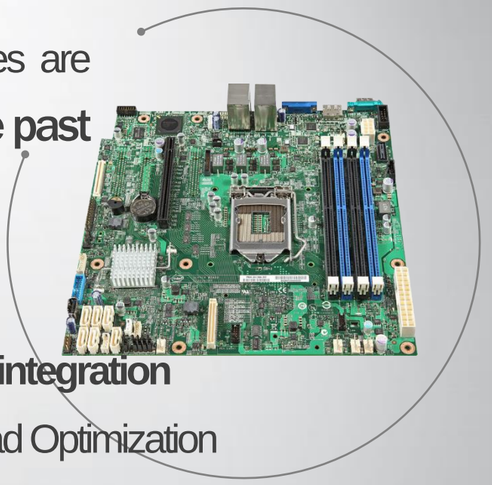
- Workload Management
- Cluster Management
- OS Management



Traditional Approach & Cloud Requirements

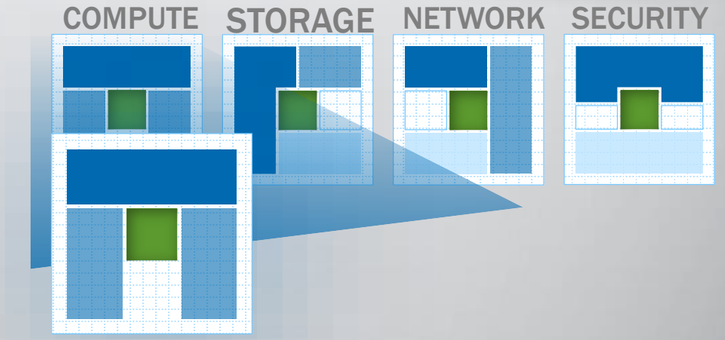
Traditional IT approach:

- Processor + IO Chipset + Memory
- Foundation is modified through addition of specialty cards and options
  - Lack of integration and workload optimization



Traditional Requirements:

- Wide ecosystem/software support
- One size fits all approach
- Reduced personnel requirements through standardization



Cloud Requirements :

- Optimize for specific workloads
- Optimize for Cost, Latency, and Power consumption

Cavium: Multi-Core Processor Company

Enterprise and Service Provider

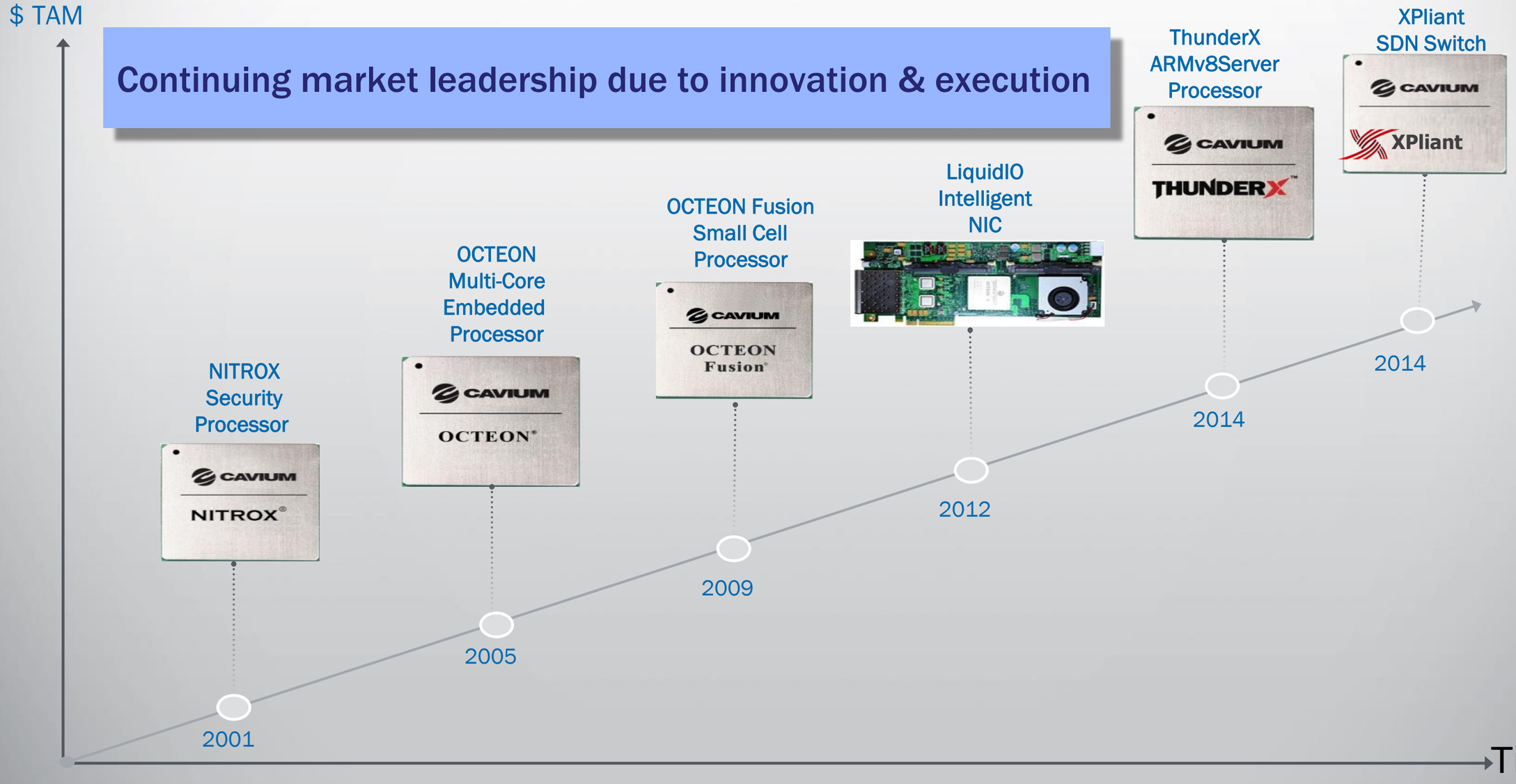
Mobile Infrastructure

Data Center and Cloud

BIG IRON INFRASTRUCTURE

Efficiency + Scalability + Security **LEADERSHIP**

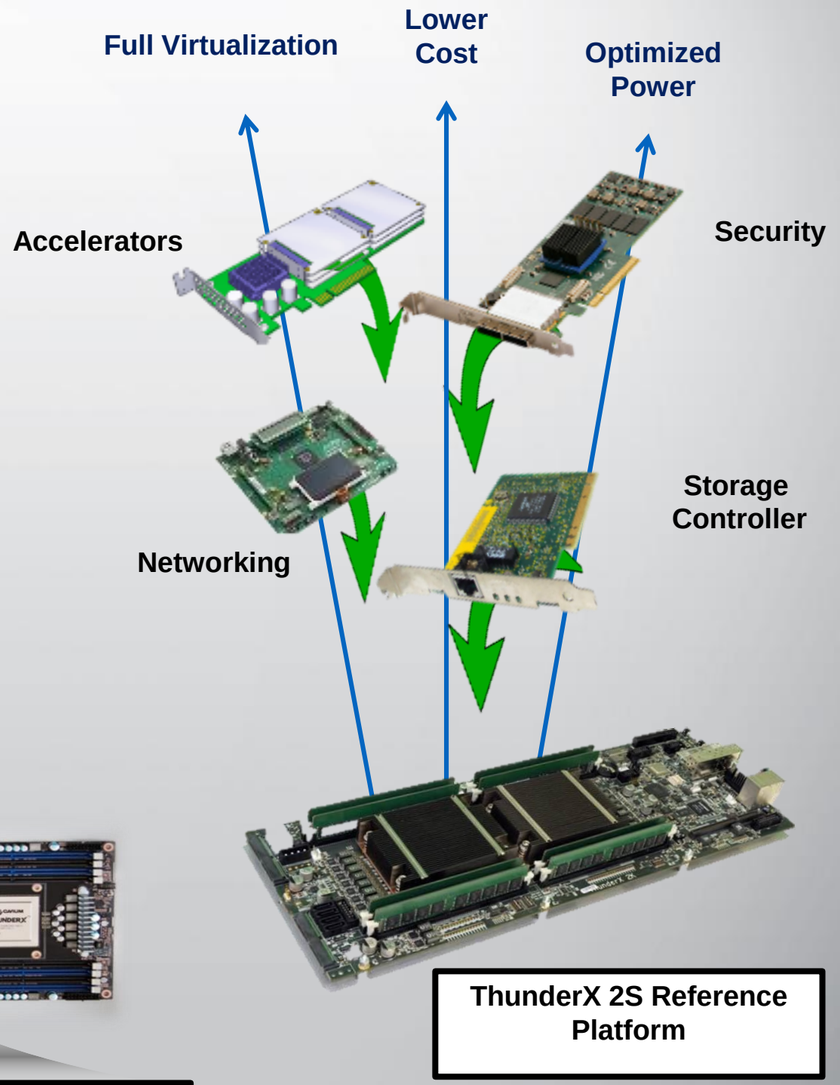
CAVIUM – History of Product Innovation



THUNDERX: Family of Workload Optimized Processors

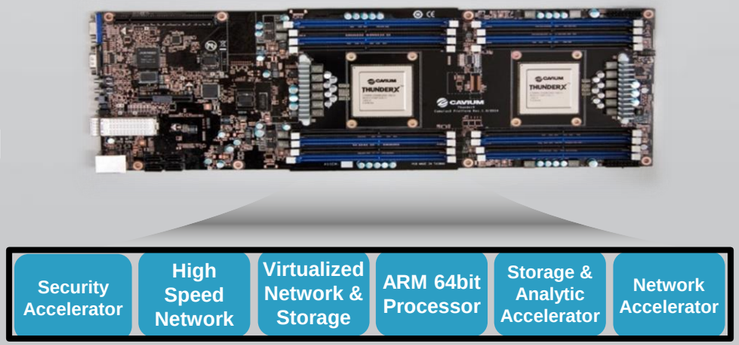
CAVIUM

- Up to 48 full custom ARMv8 cores @ 2.5GHz
- Multi-socket capable with Cavium Cache Coherent Interconnect (CCPI)
- Up to 4x 72-bit DDR3/4 Memory Controllers
  - 1 TB system memory in 2S config
- Family Specific I/O's including 40G/10GE, PCIe Gen3, SATA 6G
- Standards based low latency Ethernet fabric
- virtSOC™: Virtualization from Core to I/O
- Platform : Single & Dual Socket
- Family Specific Accelerators : Storage/Networking / Compute / Security

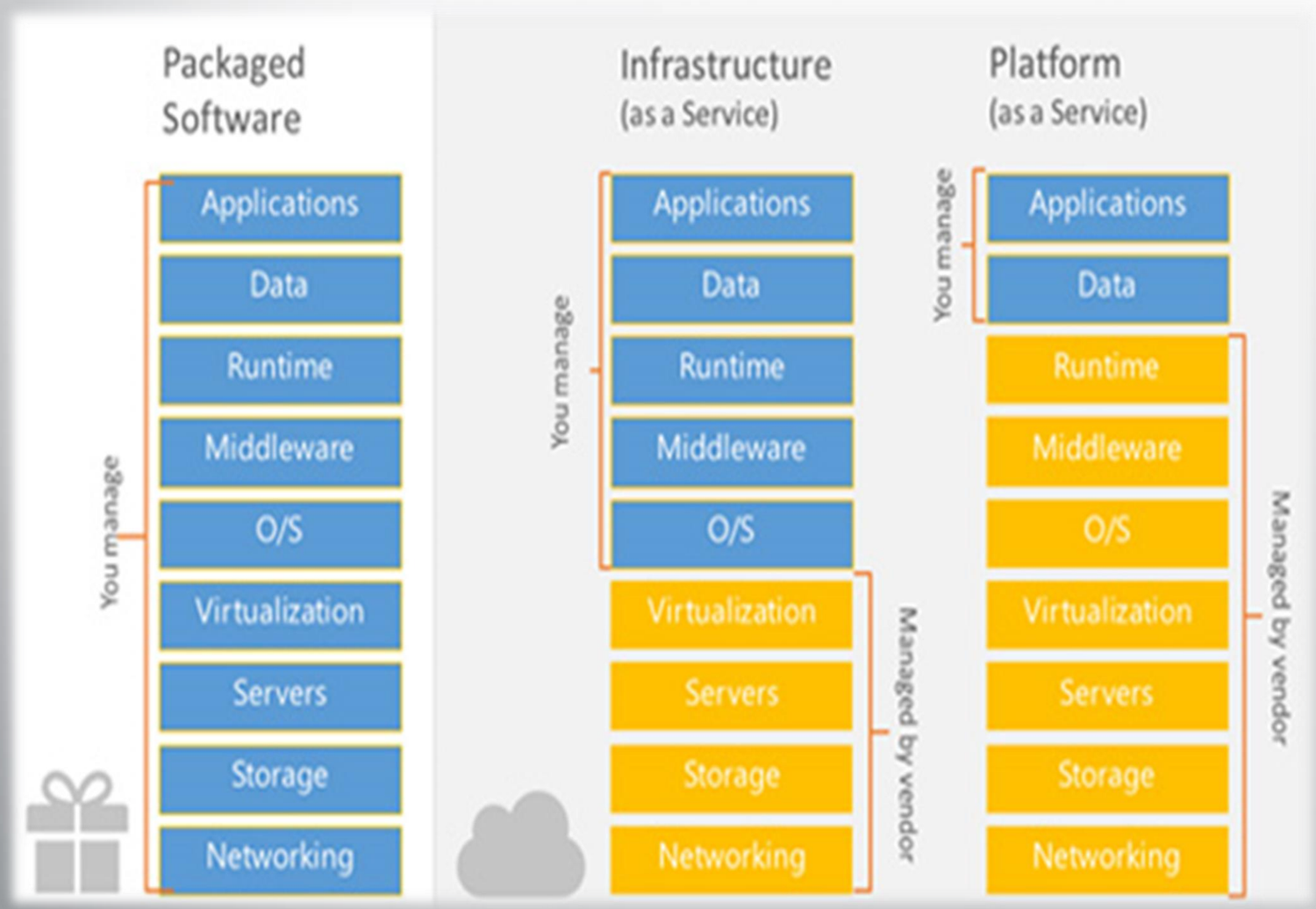


The benefits of this Workload Specific approach

- Efficiency (performance, latency, power, and scalability)
- Best in Class Optimized solution for the specific workload



Software, Application and Platform Trends



Platform:

- Huge clusters with identical configuration
- Resiliency & fault tolerance handled in software
- Highly scale out, commodity servers

Workload Trends:

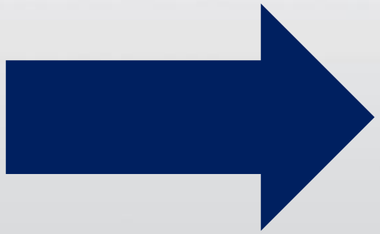
- Most workloads are open source
- Java: Most popular programming environment
- Both static and dynamic content
- Applications are highly parallel and scalable

Software Management:

- More & more SW managed by the cloud vendor

Satisfying HPC Workload Demands

- ✓ Combined Scalar and FP compute (via GPU)
- ✓ Lower System Power + Lower System Cost
  => Improved TCO
- ✓ Low latency & high bandwidth memory subsystem
- ✓ Optimized cluster configuration with integrated networking & switches
- ✓ Workload HW accelerator optimization



Scientific Compute (Rsrch/Govt)

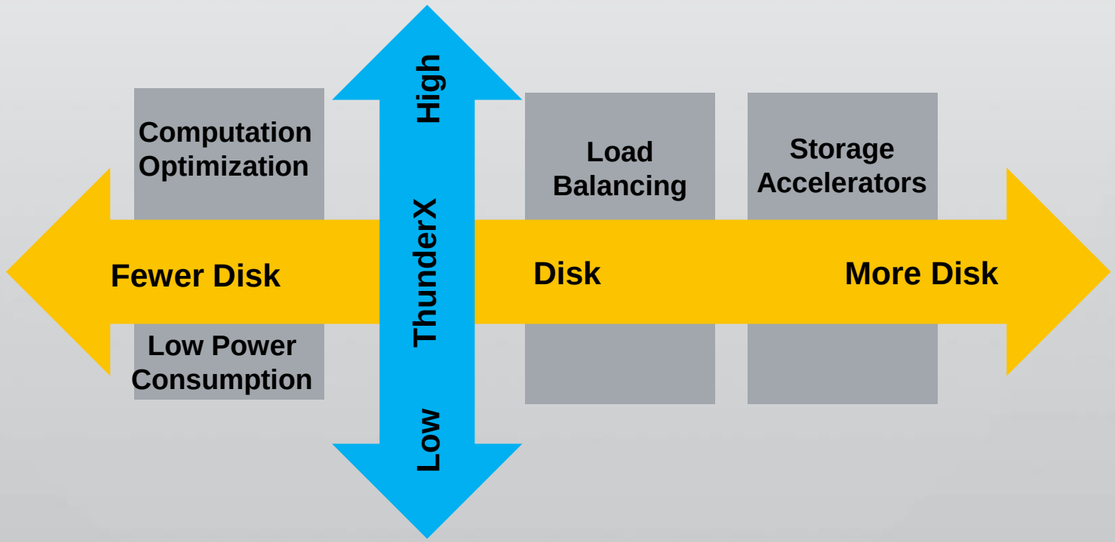
Data Analytics (Fin Services)

Hybrid Clouds (Public & Private)

ThunderX based Hadoop/Big Data Cluster

- Hadoop handles large data volumes
  - Hadoop distributes data across clusters; uses replication to ensure data reliability and fault tolerance
  - Each ThunderX based server in Hadoop cluster stores AND processes data (SOC)
  - Processing and storage cluster are both based of ThunderX platform

- **Hadoop MapReduce Compute Bound Operations and Workloads**
- Clustering/Classification
- Complex text mining
- Natural-language Processing
- Feature extraction



- **Hadoop MapReduce I/O Bound Operations and Workloads**
- Indexing
- Grouping
- Data importing & exporting
- Data movement & transform

ThunderX powered platform meet the requirements of large variety of compute and I/O operation required for cloud HPC based Big Data Solution

Commercial Support for ARMv8 Platforms Growing

CAVIUM

- ✓ Applications
- ✓ Tools & Infrastructure
- ✓ Development Environments
- ✓ Operating Systems

NGINX

allinea

IBM Platform LSF

ORACLE

PathScale

redhat.

CANONICAL

CPU CORE

MEMORY

MEMORY

Industry Leading HPC Systems Partners

CAVIUM

CRAY
THE SUPERCOMPUTER COMPANY

- Public announcement Nov, 2014
- 2 Socket Cluster Configuration & Optimization
- Recognized industry leadership in compilers/tools

lenovo. FOR THOSE WHO DO.

- Public announcement Feb, 2015
- NextScale design team
- Driving HPC EcoSystem, use case and workload focused

E4
COMPUTER ENGINEERING

- EU Partner supporting HPC Labs
- Resell and support Cavium CRB
- Solutions for HPC and Enterprise

Accelerating End User Engagement/Demand

THUNDERX Enabling Next Generation HPC Cloud

CAVIUM

HIGH PERFORMANCE

240GHz compute capacity in 2S
128GB DDR4 memory BW

ARM Eco-System

Commercial Platform Available
Prime for engagement

GREEN HPC DATA CENTERS

Best in class Perf/Watt
Best in class Perf/$

WORKLOAD OPTIMIZED

Workload specific hardware accelerators

SECURE DATA ACCESS

High performance secure communications & application security

CLOUD SCALABILITY

Standards based Ethernet fabric
Low latency + VirtSoc Technology
Massively scalability to thousands of nodes in 2-D/3-D configurations

THANK YOU