SDC 19
SNIA INDIA
May 23-24, 2019
Bangalore, India
STORAGE DEVELOPER CONFERENCE

Optimizing Data Centers for IoT Revolution

Geet Chauhan, Kishor Jadhav
Microchip

Agenda

- Introduction
- IoT Revolution
- Data Explosion
- Data Centers
- Future

Introduction

Introduction

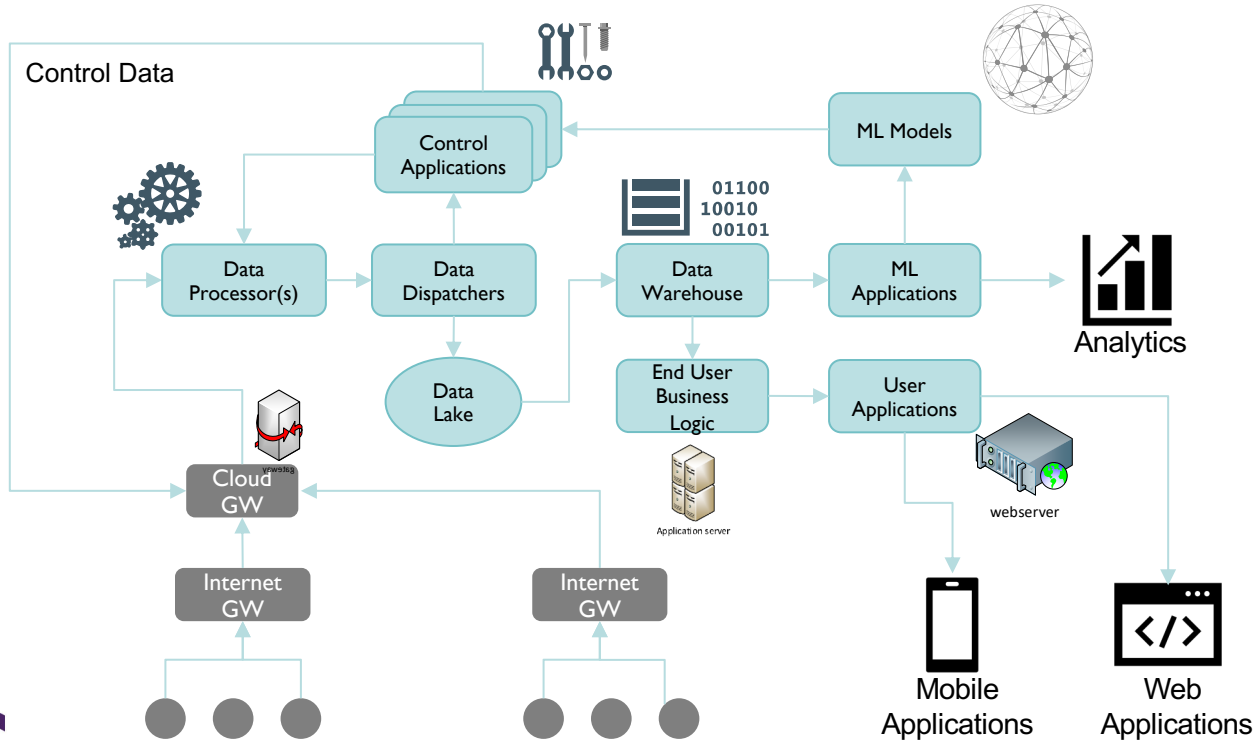
- A brief history of Data Centers
- Big picture of Internet of Things

Evolution of Data Centers

	Legacy	Consolidated	Virtualized
Compute	x86	Distributed Farm nodes	Fully virtualized
Storage	DAS	NAS	SAN, vSAN, Object, SaaS
Network	Ethernet	Ethernet	vLAN, NaaS
Applications	Stand alone	Centralized	Optimized, On Demand
Management	Siloed	Cluster	VM Farm
	Long time ago	Some time ago	Now

SDC 19
SNIA INDIA

Big Picture of Internet of Things



IoT Revolution

Core Tenets of IoT

- Data capture
- Data filtering and pre-processing
- Data storage
- Data analytics and predictions

IoT and Data

- What does it mean in terms of ‘data’?
- What does it mean in terms of ‘data at scale’?
- What do we know about ‘managing big data’?
- What do we do about ‘security’, ‘QoS’, ‘privacy’, ‘vulnerabilities’, ‘standardization’?

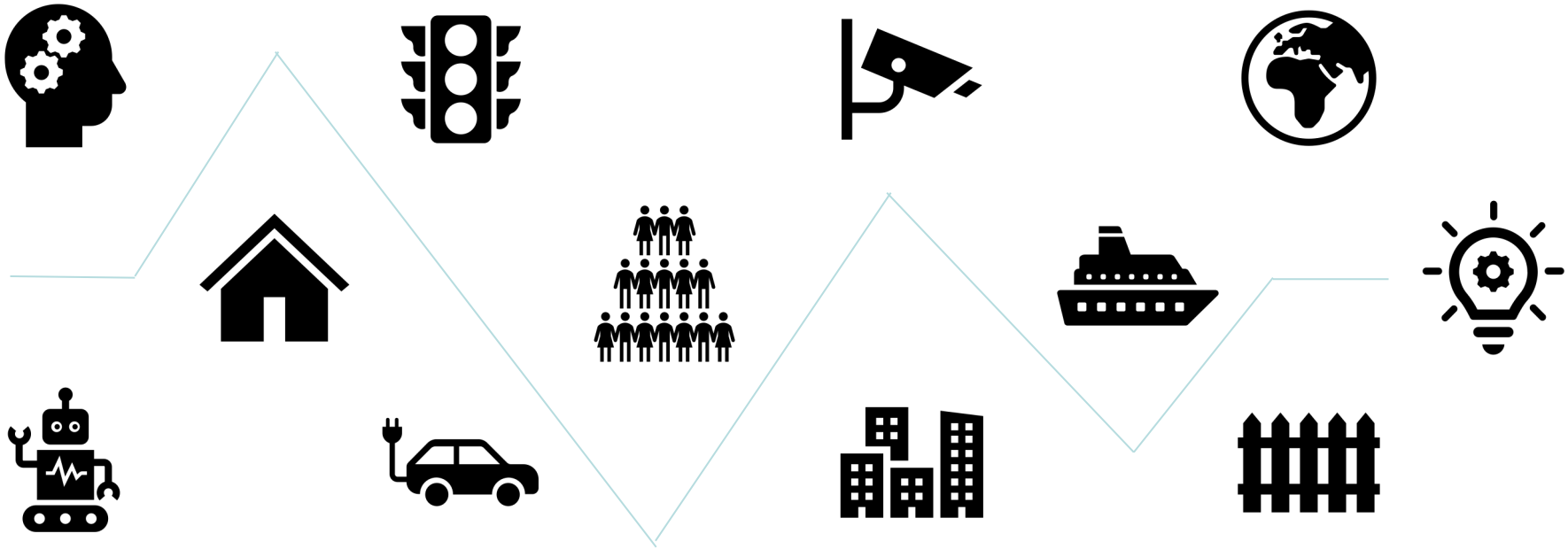
Sensors

- Sensor Networks
  - Multiple sensors and coverage
- Types of Sensors
  - Aerial
  - Terrestrial
  - Underground
  - Underwater
  - Multimedia
  - Mobile

Challenges with Sensors

- Limited storage capacity
- Limited processing power
- Power source
- Power requirements change with range
- Passive sensors generate small amounts of energy
- Distance between data creation spot and data analytics engines
- Signal strength and amplification

IoT Application Areas



Data Explosion

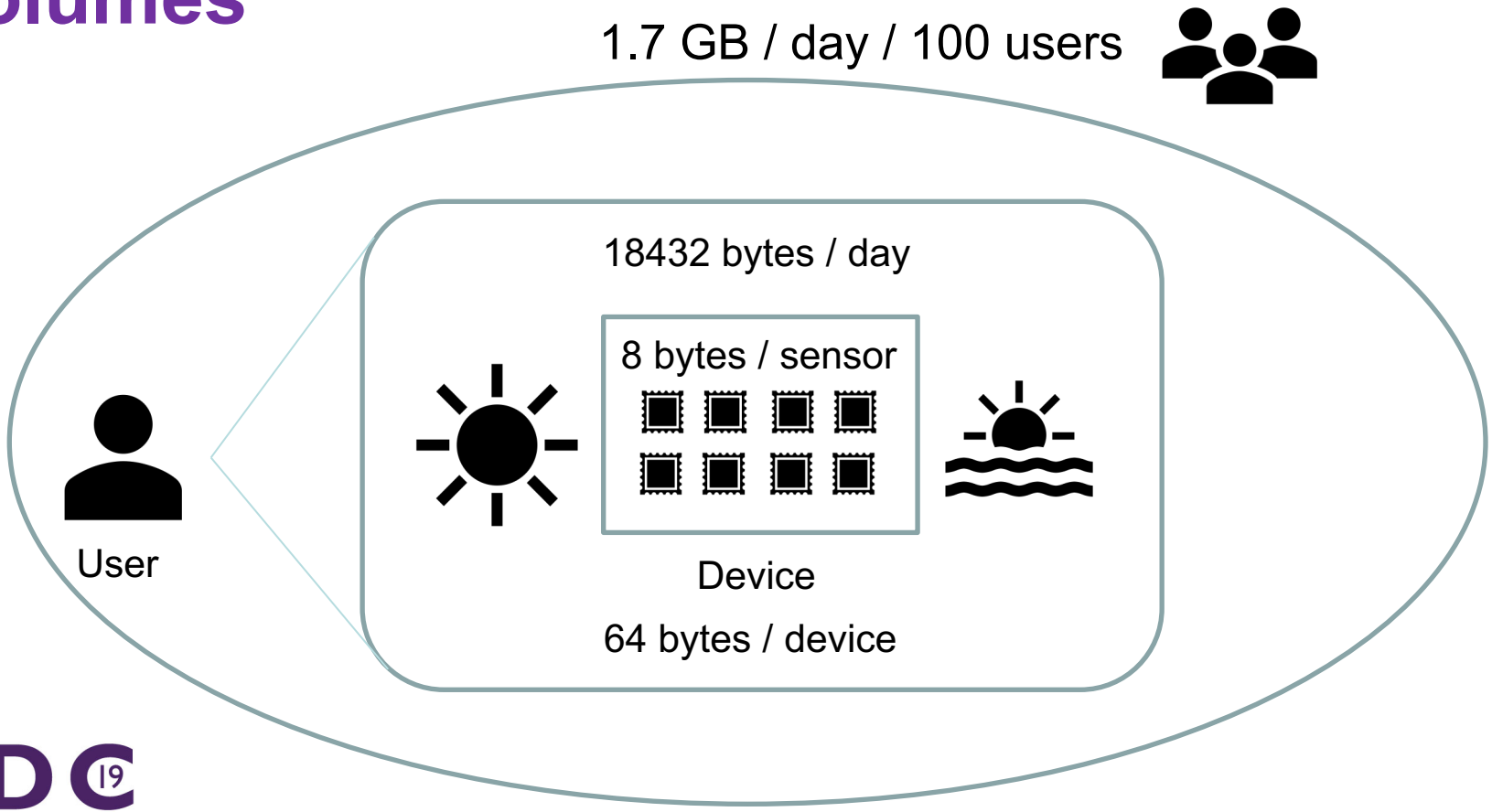
Data Explosion

- IoT hits all the three V's of Big Data premise
  - Volume
  - Velocity
  - Variety
- Example industries
  - Healthcare
  - Agriculture
  - Defense
  - Automotive
  - Retail

Volume

- Increasing use of IoT devices
- Forecast of multi billion devices with multi billion sensors
- Ever increasing number of application areas
- Huge amount of sensor generated data
- Zettabytes of data will be stored per day

Volumes



Velocity

- Speed of data generation / processing
- Types
  - Batch
  - Periodic
  - Near real time
  - Real time

Variety

- Unstructured data
  - Text files & documents
  - e-mail, voice mail
  - Images
  - AV files
  - Sensor data
  - *Application logs*
- Structured data
  - Database
  - Data structures
  - Can be queried into
  - Follows a schema

Data Center Optimization

Areas of Optimization

- Periodicity of data collection
- High bandwidth data transport
- High speed data stores for
  - Storage
  - Retrieval
- Fast data ingestion into ML algorithms

Optimization before Reaching Data Centers

- Increase intelligence in edge devices
  - Make processing closer to generation point
  - Have a micro server in between end point and cloud
- Efficient decision algorithms to chose data variety to be stored
- Smart data filters
- Run smart decision engines on edge devices

Optimization after Reaching Data Centers

- High speed storage switching fabric
- High speed storage controllers
- Data deduplication
- Object storage

Optimizing Data Center Storage

- Data stores
  - High speed flash storage
  - Storage tiering
  - Hyper converged infrastructure
- Data availability
  - High speed storage controllers
  - High speed storage switches
  - Storage controllers with high IOPS

QoS and Efficiency

- Data reliability and integrity
- Quality of Service
- Efficient storage
  - Power efficient storage arrays
  - Thermally efficient

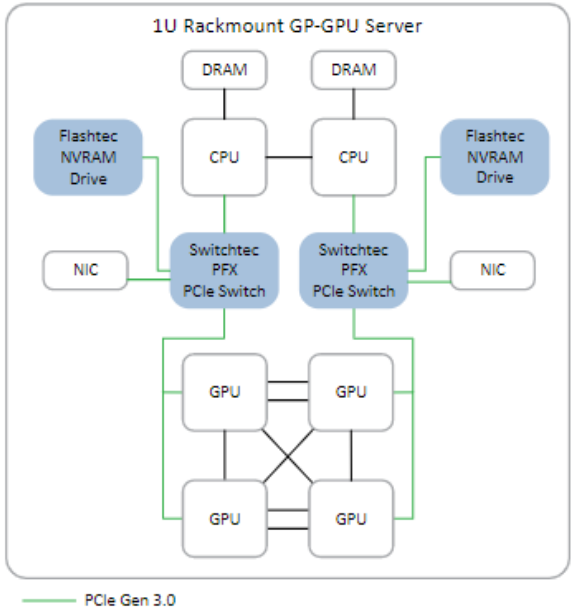
Optimizations for ML Applications

- Low power consumption.
- High throughput and flexibility.
- High use of GPU for different areas like: deep learning, machine learning, image classification, speech recognition, autonomous driving, bioinformatics and video analytics.
- Support the need for growing parallel computing growth.

Optimizations for ML Applications

- High performance fabric connectivity and composability for multi-host CPU and NVMe SSD systems is critical.
- Dynamic partitioning.
- Multi host SR-IOV sharing.
- Enable real time “composition” or dynamic allocation of GPU resources.

Optimization for ML Applications

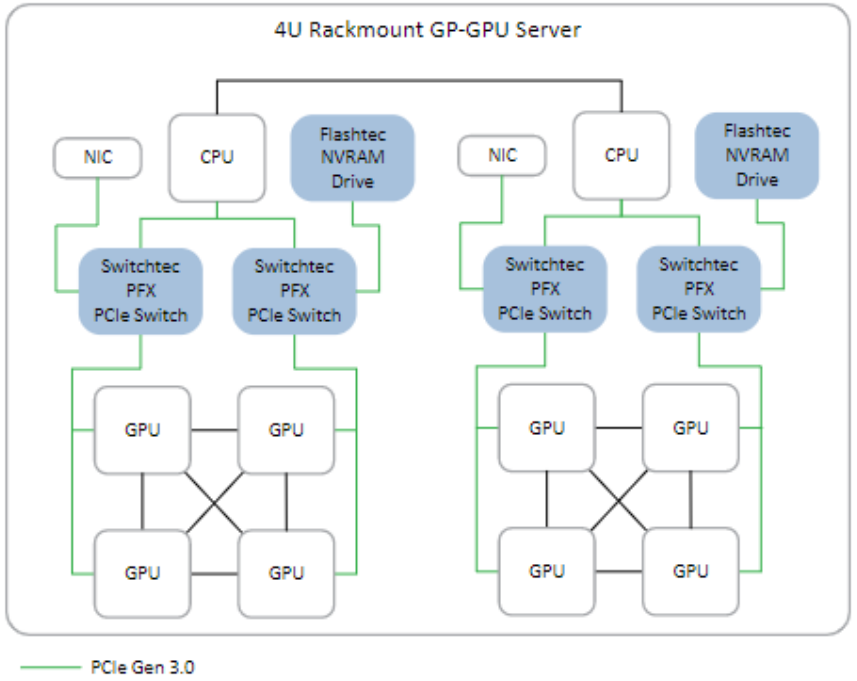


- PCIe advanced switch solutions
- Scalable
- Cost-effective multi host interconnect
- Low latency
- Fabric application programming API

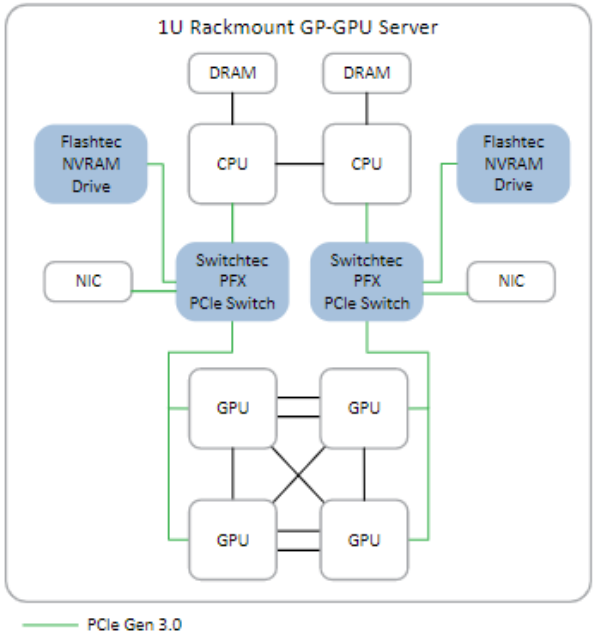
SDC 19
SNIA INDIA

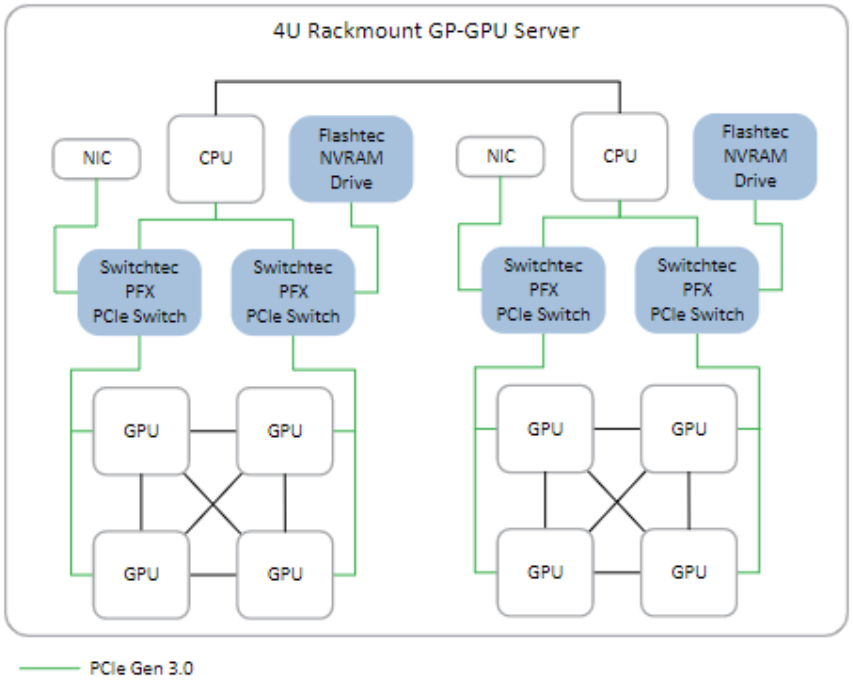
Optimizations for ML Applications

- Scaled view
- Multi node
- High Availability at fabric



Solutions for ML Appliances



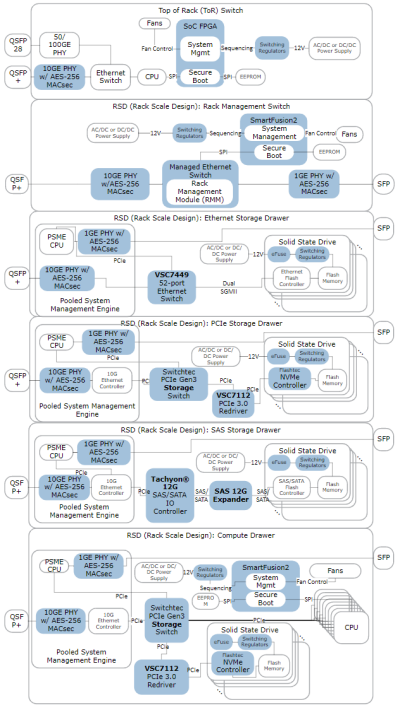


Optimization using Rack Scale Design

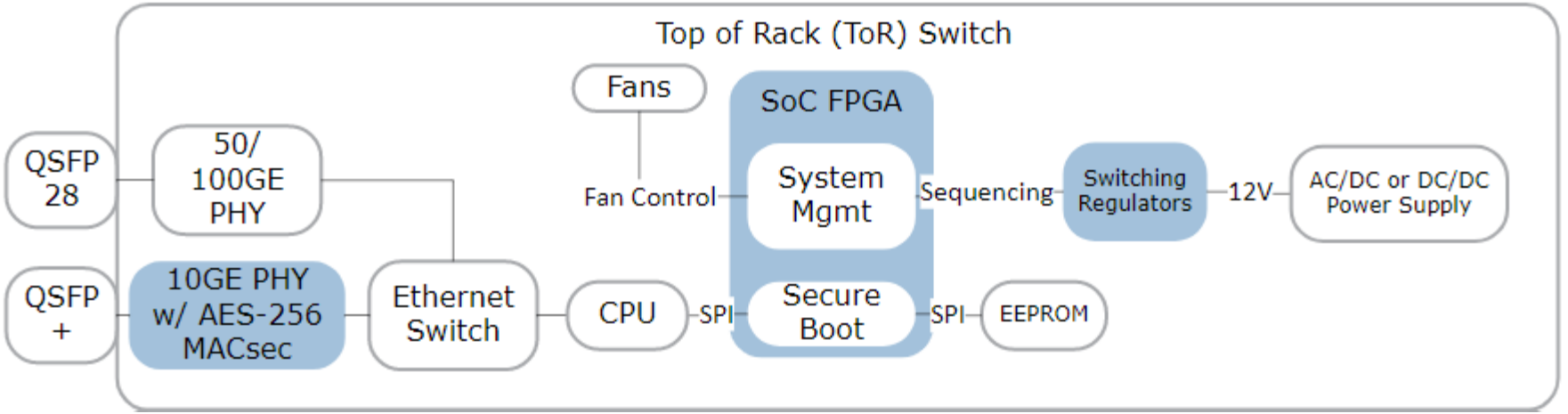
- No single data center fabric is cost, power and performance optimal.
- Media and CPU disaggregation for maximizing platform flexibility, density and utilization.
- Optimize resource utilization and reduce “resource stranding”.
- Using storage switching and pooling solutions help storage disaggregation and dynamic hybrid pool support.
- PCIe switching and dynamic partitioning capabilities enable high speed NVM pools at lowest latency and power.

Optimization using Rack Scale Design

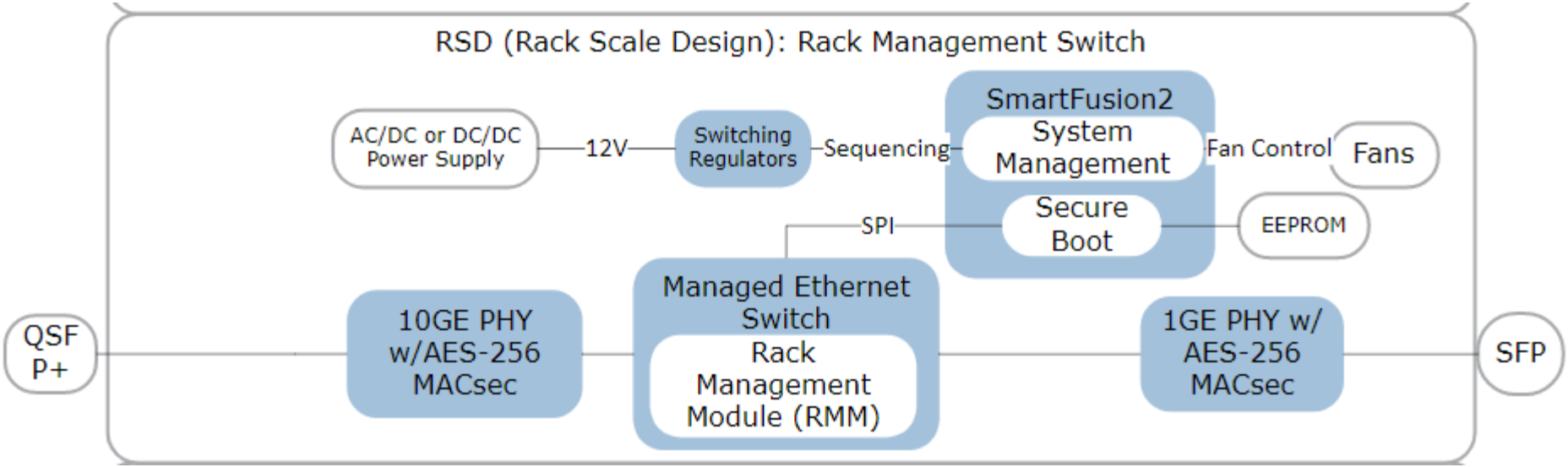
- Top of Rack (ToR) Switch
- RSD – Rack Management Switch
- RSD – Ethernet Storage Drawer
- RSD – PCIe Storage Drawer
- RSD – SAS Storage Drawer
- RSD – Computer Drawer



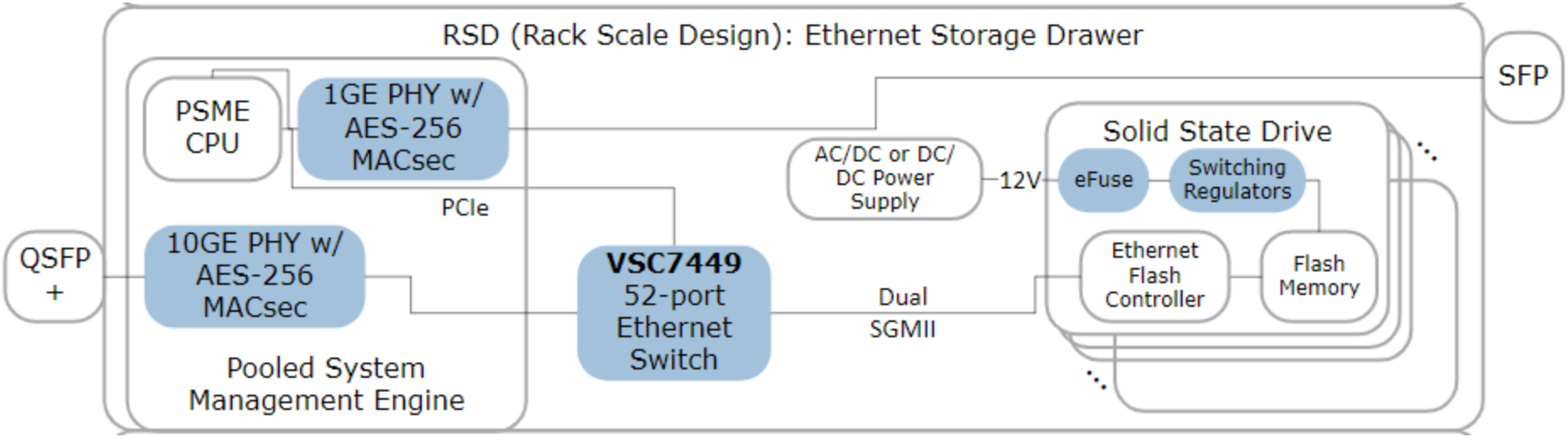
RSD – ToR Switch



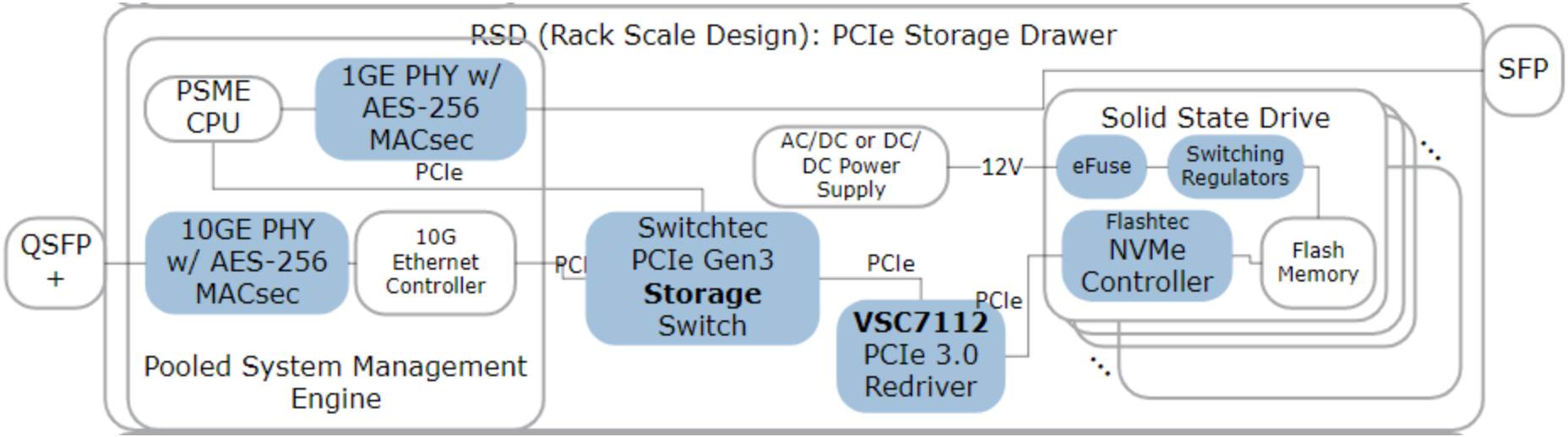
RSD – Rack Management Switch



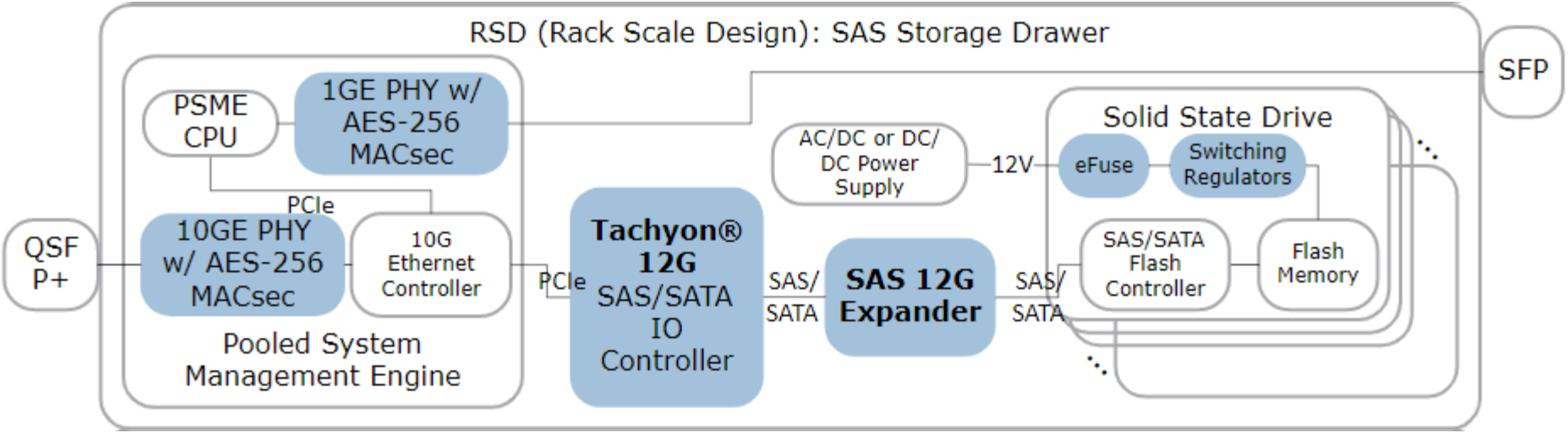
RSD – Ethernet Storage Drawer



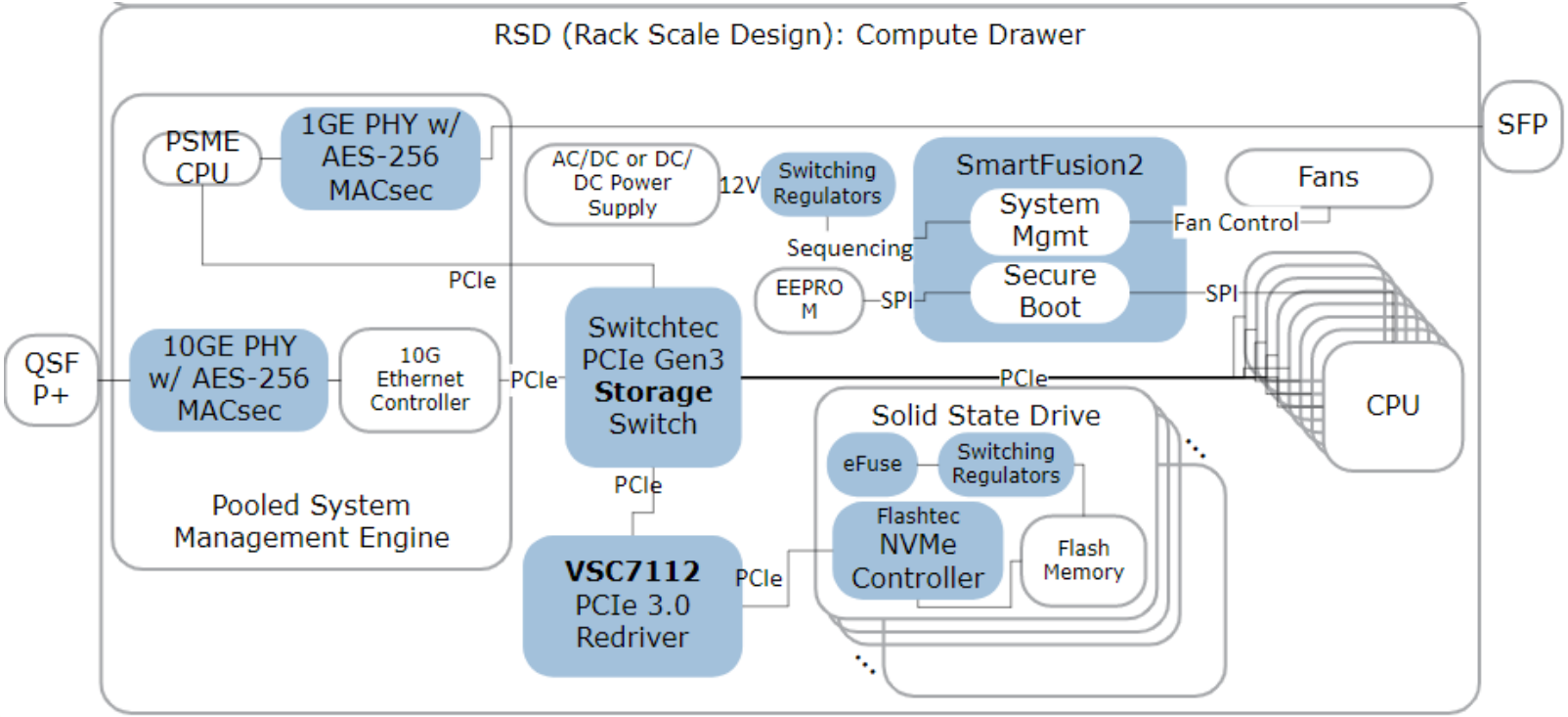
RSD – PCIe Storage Drawer



RSD – SAS Storage Drawer



RSD – Compute Drawer



Future

IoT and Edge Computing

- Need for decision in milliseconds require pushing towards computing and decision making at the edge.
- Technology induced latency is a curse.
- Sensing, learning, decision making and reacting faster than ever requires moving on the spot analytics closer to data creation.
- In future, most of the sensor data generated would never reach a data center.
- The goal is to reduce dependency on traditional data centers for the intelligence.

Edge Computing Application Areas

- Data gathering, filtering and pre–processing
- Analytics
- Security
- Location Based Services
- Asset management (autonomous logistics machinery)
- Mining with Robots

And more …

Data Centers Would Still Evolve

- Use of Persistent Memory in data center storage stacks
  - Typical NVDIMM target application areas
  - Low latency lookups
  - Byte level data processing
  - Fast in-memory workloads and faster IOPs

Data Centers Would Still Evolve

- Use of Computational Storage in data centers
  - Using Computational Storage Device (CSx); Drive, Processor and Arrays
- Gen-Z adoption in the industry

Thank You