

The background of the slide features a blue-tinted image of industrial robotic arms. Overlaid on this image is a network of glowing blue dots connected by thin lines, suggesting a digital or data network. The text is positioned on the left side of the image.

# **VIRTUALIZATION** AND REAL-TIME SYSTEMS FOR WORK LOAD CONSOLIDATION

Core and Visual Computing Group, Intel®

# LEGAL NOTICES AND DISCLAIMERS

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at [www.intel.com](http://www.intel.com).

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

Any forecasts of goods and services needed for Intel's operations are provided for discussion purposes only. Intel will have no liability to make any purchase in connection with forecasts published in this document.

ARDUINO 101 and the ARDUINO infinity logo are trademarks or registered trademarks of Arduino, LLC.

Intel, the Intel logo, Intel Inside, the Intel Inside logo, OpenVINO, Intel Atom, Celeron, Intel Core, and Intel Movidius Myriad 2 are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries.

\*Other names and brands may be claimed as the property of others.

Copyright 2018 Intel Corporation.

# WORKLOAD CONSOLIDATION IS A KEY STRATEGIC FOCUS



## MANY CUSTOM DEVICES



## EXPANDABLE, HIGH PERFORMANCE COMPUTE



## LOWER COSTS, INCREASE YIELD AND FLEXIBILITY



**INVEST IN WORKLOADS THAT CONSOLIDATE COMPUTE (ATOM → CORE → XEON)**



# WORKLOAD CONSOLIDATION FOR EDGE COMPUTING

6 Enabling Technology Pillars Powered by IA + Workload Accelerators



Manufacturing  
Machine



ALTERA  
Movidius  
an Intel company



Edge Server  
Multi-Function Controller  
Edge Gateway

# DIGITAL TRANSFORMATION AND WORKLOAD CONSOLIDATION

*Technology drives value in businesses in four ways: enhanced connectivity, automation of manual tasks, improved decision making, and product or service innovation. Tools such as big-data analytics, apps, workflow systems, and cloud platforms – all of which enable this value – are too often applied selectively by businesses in narrow pockets of their organization. When used well, digital expands the improvements delivered in one part of an organization across the whole value chain.*

--McKinsey & Company's article, *Finding Your Digital Sweet Spot*,<sup>8</sup> authors 'Tunde Olanrewaju and Paul Willmott

# WHY CONSOLIDATE COMPUTING TASKS?

## BUSINESS DRIVERS



Lower OPEX/CAPEX



Faster Time to Revenue



Aging Workforce



Increased Security Risk

## CUSTOMER DRIVERS



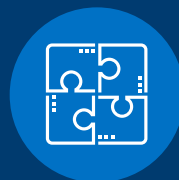
Lower Downtime



Increased Output/Yield



Interoperability for Best of Breed



Increased Flexibility/Portability

# INDUSTRIAL REQUIREMENTS INFLUENCING WORKLOAD CONSOLIDATION

- Worst Case Execution Time
- Bandwidth Capability
- Logical Isolation
- Determinism
- Platform Requirements





# VIRTUALIZATION IS A KEY ENABLER FOR WORKLOAD CONSOLIDATION



Static Locked Configuration



Dynamic Configuration



Elastic Configuration and Orchestration

ISOLATION

VIRTUALIZATION SPECTRUM

SCALABILITY

Separation  
kernels  
(e.g. VxWorks 653)

Real-Time  
Hypervisors  
(e.g. VxWorks HV,  
RTS, ACRN)

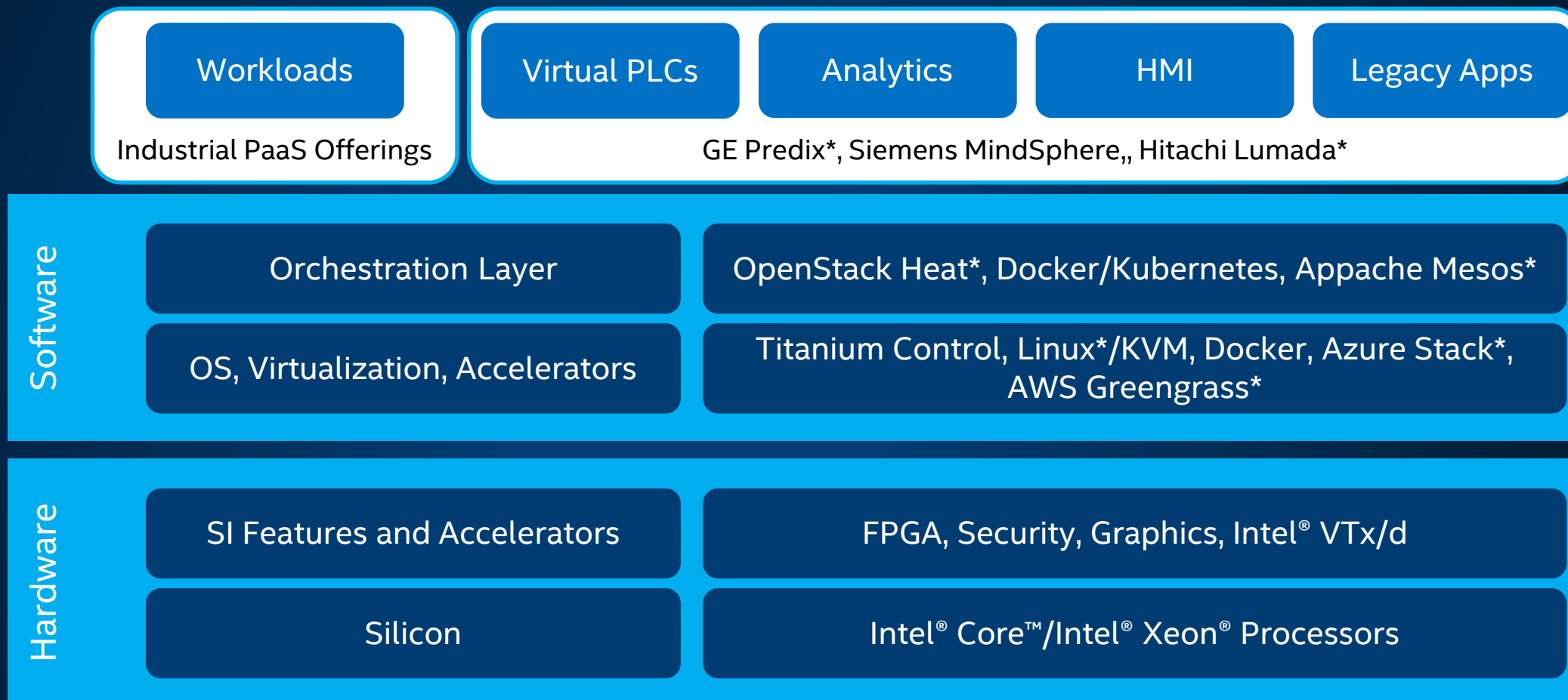
Linux Open  
Virtualization  
(e.g. KVM, bhyve)

Container  
Engines  
(e.g. Docker)

Cloud-Like  
Infrastructure  
at the Edge  
(e.g. Openstack,  
Titanium Control,  
Azure Stack)

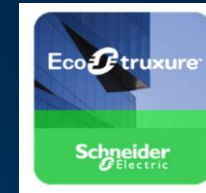


# VIRTUALIZATION INTERCEPT POINTS



# WHAT TO LOOK FOR: NEW PRODUCT CATEGORIES

Public or On-Prem  
Cloud Deployment



Edge  
Appliances



Traditional  
Gateway



Traditional IPC



Edge/Fog Server



Focus Areas

Controllers



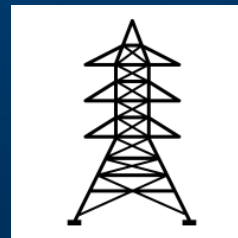
Traditional  
PLCs, MCs,  
CNCs



Multi-function Controller



Devices



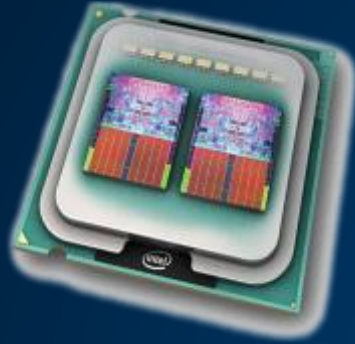
# WHAT IS INTEL® VIRTUALIZATION TECHNOLOGY



Intel® Virtualization Technology (Intel® VT) is a multigenerational roadmap of increasingly powerful enhancements to Intel Processors, Chipsets and I/O devices. It is a complementary technology to virtualization software products that enhances today's virtualization solutions and lays foundation for future platform virtualization. Intel® VT provides hardware assist to the virtualization software reducing its size and complexity enabling lower cost, more efficient, more powerful virtualization solutions.



# QUICK GLANCE OF INTEL® VIRTUALIZATION TECHNOLOGY



Intel® VT-x  
*Processor*

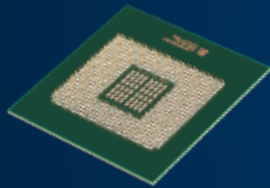
## Intel® VT-x

Hardware assists for robust virtualization

Intel® VT FlexMigration – Flexible live migration

Intel® VT FlexPriority – Interrupt acceleration

Intel® EPT – Memory Virtualization



Intel® VT-d  
*Chipset*

## Intel® VT for Directed I/O

Reliability and Security through device Isolation

I/O performance with direct assignment



Intel® VT-c  
*Network*

## Intel® VT for Connectivity

NIC Enhancement with VMDq

Single Root IOV support

Network Performance and reduced CPU utilization

Intel® I/OAT for virtualization

Lower CPU Overhead and Data Acceleration

# INTEL VIRTUALIZATION PRODUCTS AND TECHNOLOGIES

Intel® VT is a cohesive portfolio of several hardware assist technologies that increase performance and the overall functionality of virtualization software and Intel solutions.

## Technologies

### Intel® VT – x

- CPU Virtualization
- Memory Virtualization

### Intel® VT – d

- I/O Virtualization

### Intel® VT – c

- Virtualization of Network Devices

### Other technologies

- Intel Graphics Virtualization Technology<sup>1</sup>

## Hardware



Wide Range of Processors

## Software



Wide Range of Software\*

## Supported OSes



Wide Range of Operating Systems\*

## Solution Benefits

- Higher Privilege Ring: Reprioritized ring eliminates conflicts simplify hypervisor complexity, and improve capability with unmodified OS
- Hardware Based Transitions: Reduces complexity of software transitions
- Hardware Based Memory Protection: Accelerate Transition and ensure reliability of the process

1. Intel GVT is an integrated feature that depends on the processor

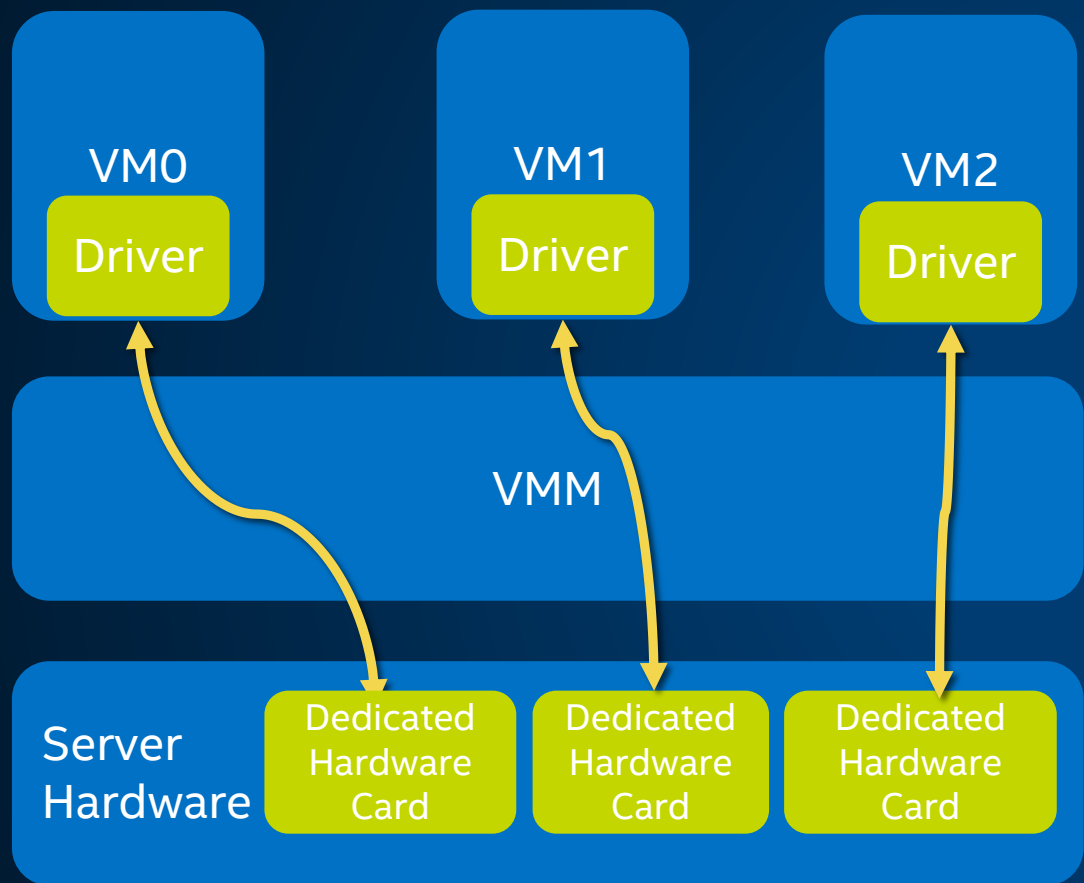
# INTEL® VT-X TECHNOLOGY OVERVIEW

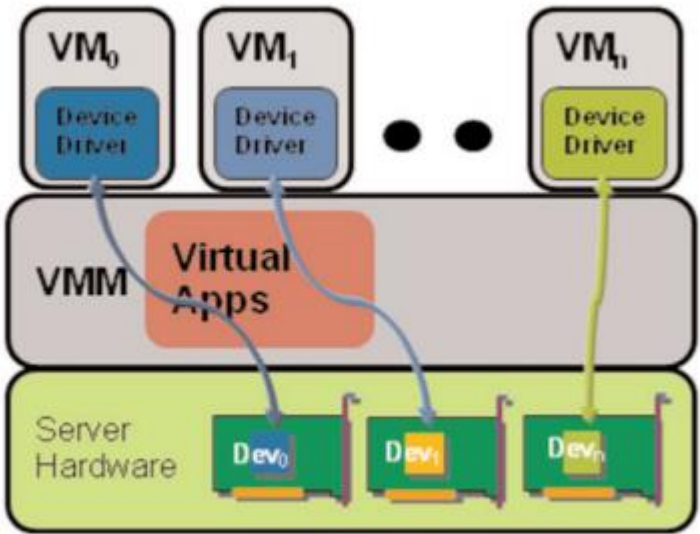
## Intel Virtualization Technology Feature and CPU Mapping

VT-x Base	<b>CPU virtualization</b> features enable faithful abstraction of the full prowess of Intel® CPU to a virtual machine (VM)
Intel® VT FlexPriority	optimizes virtualization software efficiency by improving interrupt handling.
Intel® VT FlexMigration	conduct live virtual machine (VM) migration across all Intel® Core™ microarchitecture-based servers.
Extended Page Table (EPT)	allows a VMM to avoid the VM exits associated with page-table virtualization, which is a major source of virtualization overhead without EPT.
Virtual Processor ID (VPID)	permits the CPU to flush only the cache lines associated with a particular VM
Descriptor-Table Exiting	allows a VMM to protect a guest OS from internal attack by preventing relocation of key system data structures
Pause-Loop Exiting	enable detection of spin locks in guest software and avoid lock-holder preemption
Real Mode Support	This feature allows guests to operate in real mode, removing the performance overhead and complexity of an emulator.

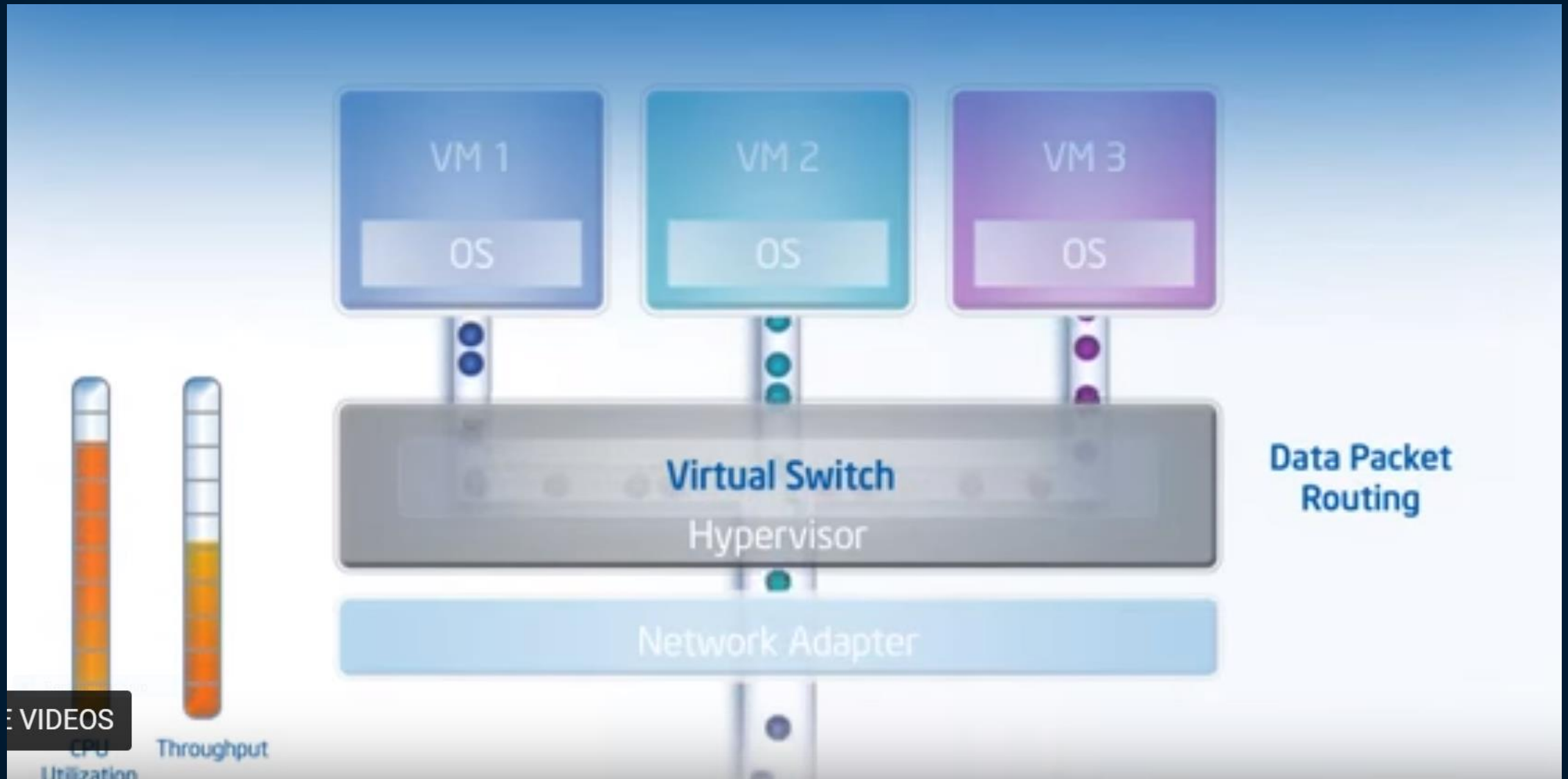


# INTEL® VT FOR DIRECTED I/O (INTEL® VT-D)



Intel® VT-D		allows to
Intel® VT-D		physical
Intel® VT-D		digital
QoS		
Address Translation Services Support	allows PCI-e devices to cache the IOTLB entries (used for DMA remapping) of that device directly in the device itself	
Support for PCI-SIG/IO Virtualization		

# INTEL® VT FOR CONNECTIVITY PROVIDES I/O VIRTUALIZATION



# REAL-TIME SYSTEMS

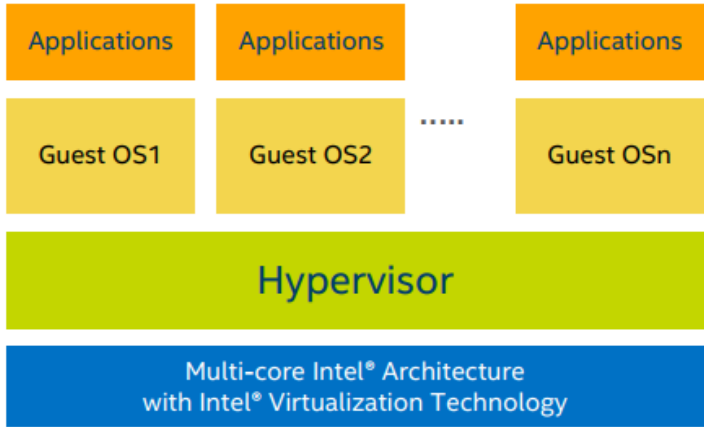
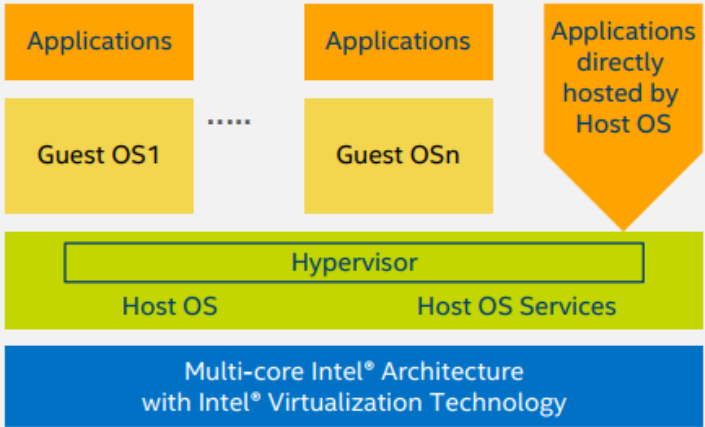


# HARD VS. SOFT REAL-TIME REQUIREMENTS

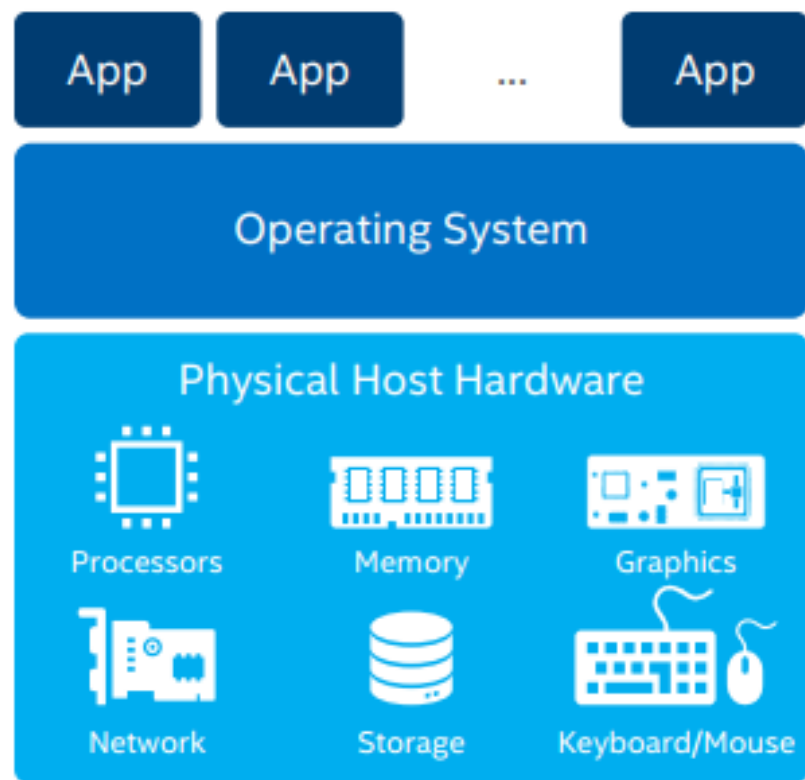
The primary difference between hard and software real-time systems is the consequences of missing a scheduling deadline.

- **Hard Real-Time:** scheduling deadlines must be met every single time. Missing the deadline means the system has failed, possibly with catastrophic consequences.
  - Robotic assembly line that require a high degree of timing accuracy
  - Software for dropping control rods in a nuclear power plant
- **Soft Real-Time:** can tolerate missing a deadline occasionally, if an average latency is maintained.
  - A system reporting on the current activity of the assembly line will not have catastrophic results if the information is slightly delayed

# REAL-TIME HYPERVISORS

	Type 1 Hypervisor (Bare Metal or Native)	Type 2 Hypervisor (Hosted)
<b>Installation</b>	Works directly on the hardware of the host and can monitor operating systems that run above the hypervisor	Installed on an operating system and supports other operating systems above it
<b>OS</b>	Completely independent of the OS	Completely dependent on the host OS for its operations
<b>Memory</b>	Small: its main task is sharing and managing hardware resources between different operating systems	Bigger memory footprint: while having a base OS allows better specification of policies, any problems in the base OS affect the entire system even if the hypervisor running above the base OS is secure
<b>Advantage</b>	Any problems in one VM or guest OS do not impact the OS running on the hypervisor  Allows for much better real-time (RT) turnaround	In a scenario where a customer has a legacy OS running a proprietary embedded application but wants to co-host a newer application that the legacy OS cannot support, one can host a newer OS that supports the newer application in the legacy OS environment
<b>Typical Layers</b>	 <p>Figure 5. Different Layers typical for a Type 1 Hypervisor Model</p>	 <p>Figure 6. Different Layers typical for a Type 2 Hypervisor Model</p>

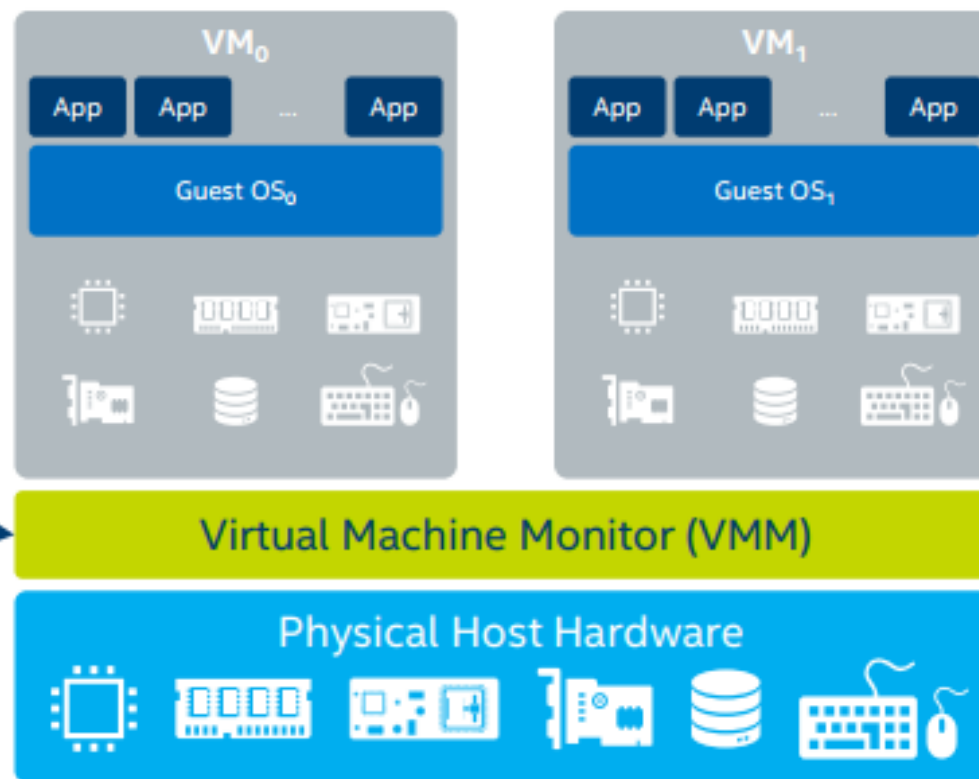
# HYPERVISOR ARCHITECTURE



**Without VMs:**

Single OS owns all hardware resources

A new  
layer of  
software

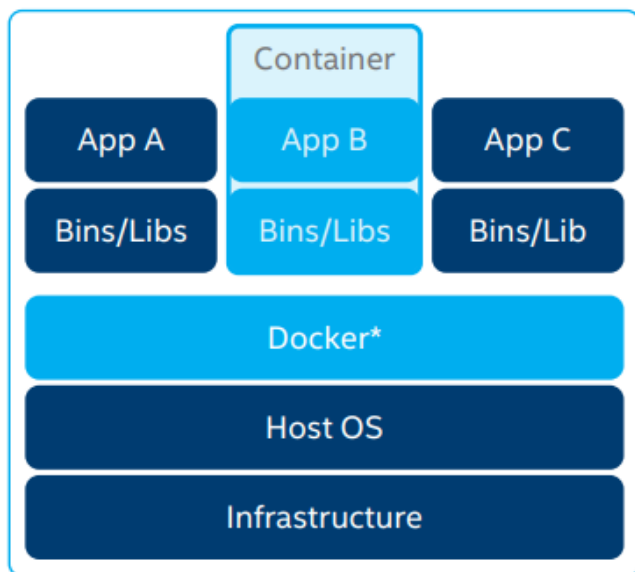


**With VMs:**

Multiple OSES share hardware resources

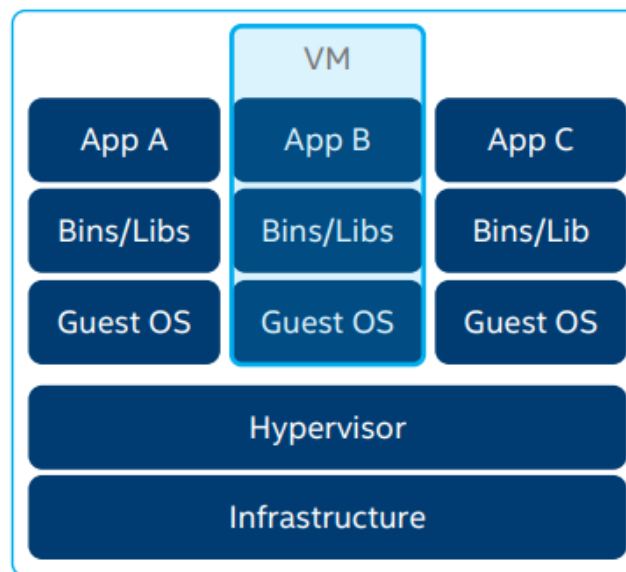


# VIRTUALIZATION VS. CONTAINERIZATION



## Containers

Containers are an abstraction at the app layer that packages code and dependencies together. Multiple containers can run on the same machine and share the OS kernel with other containers, each running as isolated processes in user space. Containers take up less space than VMs (container images are typically tens of MBs in size) and start instantly.



## Virtual Machines

Virtual machines (VMs) are an abstraction of physical hardware turning one server into many servers. The hypervisor allows multiple VMs to run on a single machine. Each VM includes a full copy of an operating system, one or more apps, necessary binaries and libraries – taking up tens of GBs. VMs can also be slow to boot

# WHY LINUX IS NOT A REAL-TIME OS

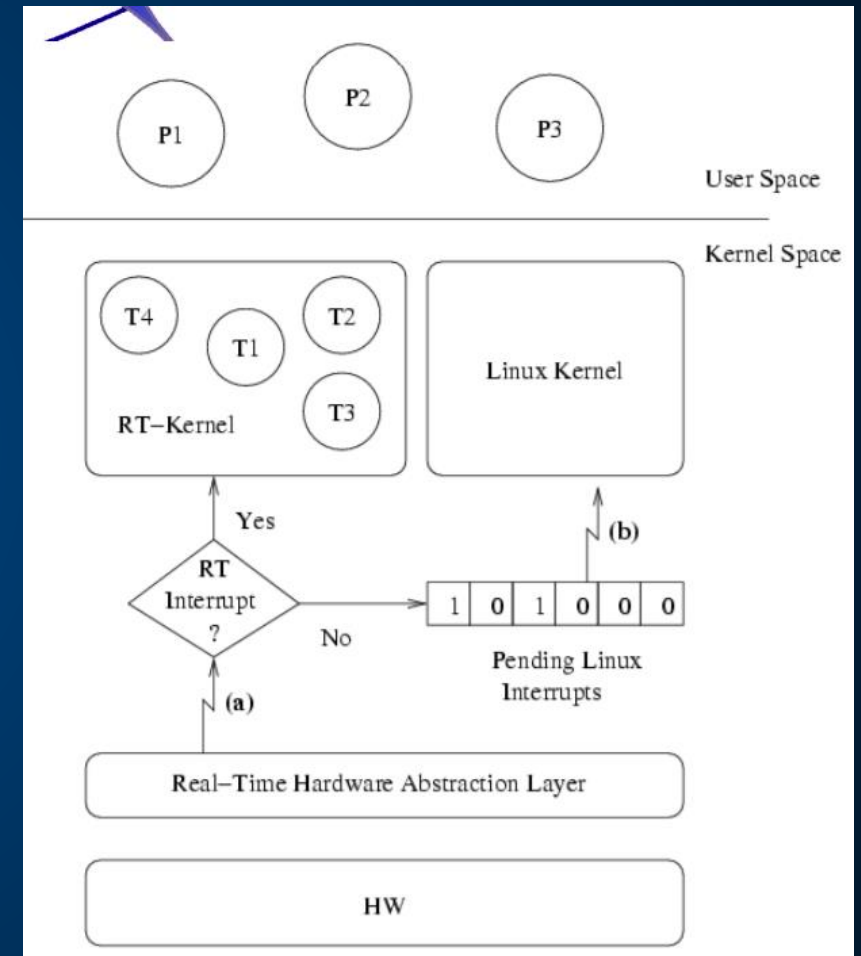
- Coarse-grained Synchronization
  - kernel system calls are not preemptible
- Paging
  - swapping pages in and out of virtual memory is unbounded
  - Latency to external storage is not predictable
- Fairness
  - scheduler may give the processor to a low priority process that has been waiting a long time
- Request Reordering
- Batching

# STRATEGIES FOR SOFT REAL-TIME LINUX

- Scheduling policy
- Preemption Improvement
  - No Forced Preemption (Server):
  - Voluntary Kernel Preemption (Desktop)
  - Pre-emptible Kernel (Low-Latency Desktop)
- Disable support for memory paging
- PREEMPT\_RT Patch – a project whose goal is to implement hard real-time behavior in the Linux kernel. Implements Interrupt

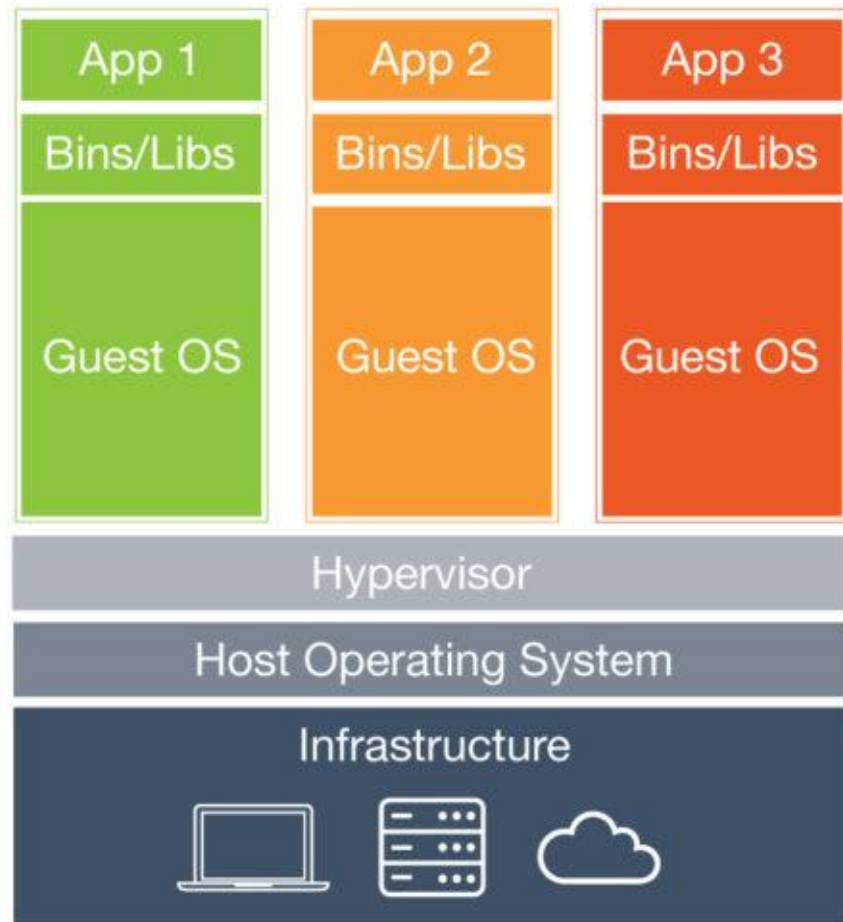
# INTERRUPT ABSTRACTION

- In many applications, only a small part of the process requires real-time performance.
- Run Linux as a low priority process on a small real-time hypervisor

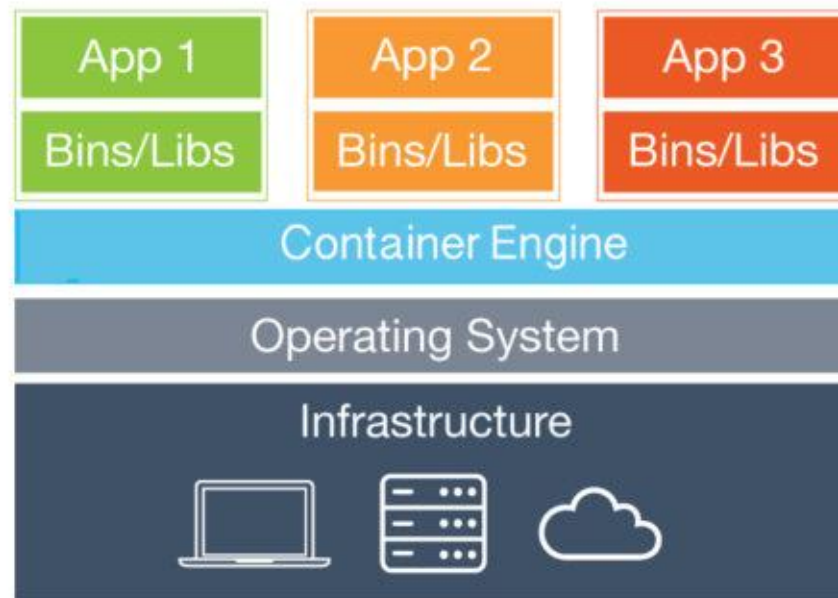




# VIRTUAL MACHINES VS. CONTAINERS



Hypervisor-based Virtualization



Container virtualization

# IA FEATURES LEVERAGED BY TITANIUM CONTROL

Platform Feature	Description	Technology Baseline	Use Cases	Titanium Control Implementation	Benefits
Vt-X	Accelerates virtual machines to near bare metal performance	Xeon / XeonD	Near native virtualized CPU performance	Performance enhancement of VMs, live migration from one Intel CPU generation to another	Performance and scalability
Vt-D	Enables physical NICs and/or GPUs to be mapped directly to virtual machine	Xeon/ XeonD	Native I/O performance	PCI Passthrough and SR-IOV support	Performance and scalability
AVX-512	Enables high performance vector workloads	Xeon Skylake Xeon Scalable Processor	Telecom, AI, high performance storage, encryption and compression	Enhanced KVM performance, guest AVX-512 support	Performance and scalability
Trusted Execution Technology	Used to attest system authenticity and state	TPM 2.0	Secure boot and verified system state	Secure boot, TPM 2.0 storage of communication keys, vTPM 2.0 support in guests.	Security
AES-NI	Accelerates encryption/decryption	Xeon Westmere +	Full disk encryption and faster communications	Linux encryption performance enhancements	Security
UEFI Boot	Used for secure booting	UEFI Spec 2.6+	Secure booting and faster boot	Fast boot, secure boot	Security

# IA FEATURES LEVERAGED BY TITANIUM CONTROL #2

Platform Feature	Description	Technology Baseline	Use Cases	Titanium Control Implementation	Benefits
DPDK	Library to accelerate networking path	Intel DPDK	High speed VM-to-VM networking	Optimized DPDK libraries	Performance and scalability
QuickAssist	Hardware-based compression and encryption	Coletto Creek, VT-d	Exposing QuickAssist engine to VMs and virtualization of QuickAssist across VMs	Support PCI passthrough and SR-IOV access for VMs to QuickAssist accelerators	Security
Enhanced Platform Awareness (EPA)	Set of enabling features that take full advantage of Intel Architecture through OpenStack		Performance / determinism controls such as core pinning, NUMA awareness / controls, hyperthreading awareness / controls, CPU model selection	Suite of features that provide fine grained control for VMs. Can specify core pinning, NUMA affinity to vSwitch and NICs, split NUMA VMs, hyperthreading isolate or require, select CPU model to enable / disable CPU capabilities / instruction sets, huge page sizes.	Performance and scalability

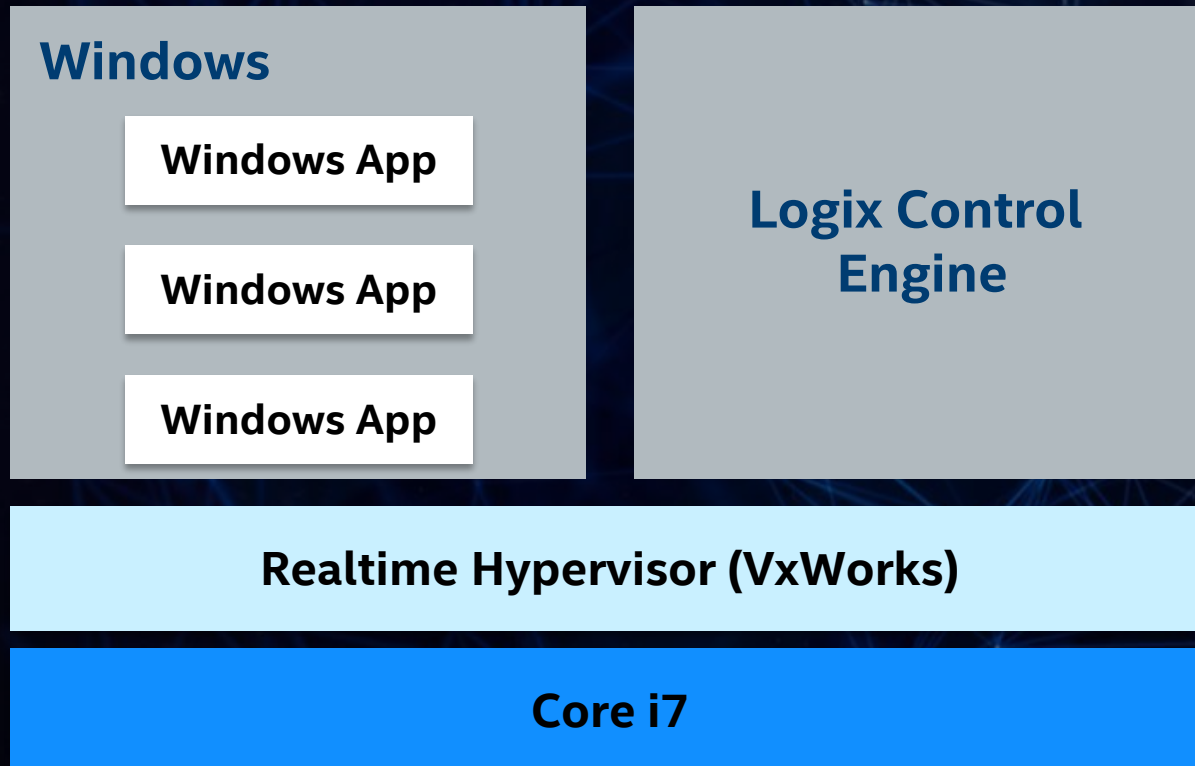
# IA FEATURES LEVERAGED BY TITANIUM CONTROL #3

Platform Feature	Description	Technology Baseline	Use Cases	Titanium Control Implementation	Benefits
BMC / IPMI	Integration with BMCs for hardware monitoring		Discovery of and monitoring of hardware through BMC using IPMI	Integration with BMCs to discover and collect hardware sensor data with capability to alarm on the values.	Open standards
Cache Allocation Technology (CAT)	Ability to isolate portions of cache for improved determinism	All SKUs starting with the Xeon E5 v4 family	Isolating cache of critical VM from noisy neighbor to guarantee performance	Under evaluation / development	Performance and scalability



# MULTI-FUNCTION CONTROLLER WITH VXWORKS

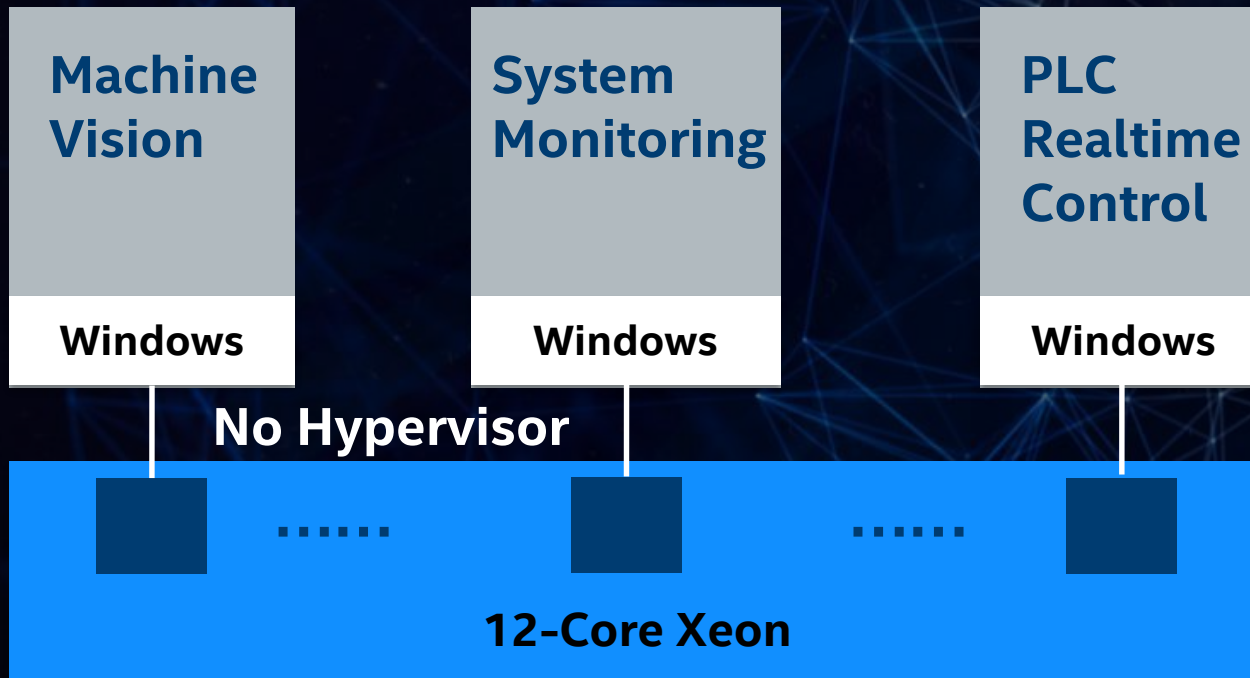
## ROCKWELL COMPACT LOGIX 5480 CONTROLLER



Opportunity	Problems Solved
Core CPU with real-time Hypervisor	<ul style="list-style-type: none"><li>▪ Ability to run Windows Applications alongside real-time Control Engine</li><li>▪ Leveraging existing PLC install base and extending value/revenue stream of assets</li></ul>

# WORKLOAD CONSOLIDATION WITHOUT VIRTUALIZATION

## BECKHOFF CX2072 CONTROLLER



### Opportunity

High-end multi core Xeon CPUs

### Problems Solved

- Maximum performance by direct assignment of workloads to cores (no Hypervisor)
- Differentiation over competitor products
- Workloads assigned statically from Twincat environment

# Open Source Software Stack Supporting High-Availability Cloud Services Optimized For Edge

 UPCOMING EVENT · Akraio Edge Stack Developer Summit · August 23 – 24 · AT&T Labs Research · [REGISTER NOW »](#)

Akraio Edge Stack, a Linux Foundation project in formation, will create an open source software for edge. The AT&T contribution is designed for carrier-scale edge computing applications running in virtual machines and containers to support reliability and performance requirements. Akraio Edge Stack will offer users new levels of flexibility to scale edge cloud services quickly, to maximize the applications or subscribers supported on each server, and to help ensure the reliability of systems that must be up at all times.

[JOIN OUR MAILING LISTS](#)[VIEW THE AKRAINO WIKI](#)[CLICK HERE FOR THE CODE](#)

# INSTALLING RT-TESTS

**pmgtest** – start pairs of threads and measure the latency of the interprocess communication

**cyclictest** – test the latency of switching between processes

**ptsematest** – start pairs of threads and measure the latency of communications through POSIX mutexes.

**svsematest** – start pairs of threads and measure the latency of communications through SYSV mutexes.

**hackbench** – stress test the Linux scheduler using processes communication through sockets or pipes.

**sendme** – sends a signal from a driver to a user and measures the latency

**signaltest** – measures latency of POSIX signals



# MEASURING LATENCY: CYCLICTEST

# Resources

<https://software.intel.com/en-us/articles/intel-virtualization-technology-for-directed-io-vt-d-enhancing-intel-platforms-for-efficient-virtualization-of-io-devices>

