BEST PRACTICES FOR BIG DATA IN THE CLOUD

Colin Cunningham, Intel
Kumaran Siva, Intel
Sandeep Mahajan, Oracle
03-Oct-2017
4:45 p.m. - 5:30 p.m. | Moscone West - Room 3020

Big Data Talk

Exploring New SSD Usage Models to Accelerate Cloud Performance –

03-Oct-2017, 3:45 - 4:30PM,

Moscone West – Room 3020

1. 10 min – Scott

   Oracle Big Data solution
2. 15 min – Daniel

   NVMe and NVMEoF, SPDK
3. 15 min – Sunil, Case Study

   Apache Spark and TeraSort
4. 5 min - QA

Best Practices for Big Data in the Cloud - 03-Oct-2017, 4:45 - 5:30PM,

Moscone West - Room 3020

1. 10 min - Sandeep

   Oracle Big Data solution
2. 15 min – Siva

   FPGA enables new storage use cases
3. 15 min – Colin, Case Study

   Apache Spark, Big Data Analytics
4. 5 min - QA

Oracle Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Intel Notices & 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. Check with your system manufacturer or retailer or learn more at intel.com.

No computer system can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit http://www.intel.com/performance.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

ORACLE CLOUD PLATFORM

Oracle Cloud Platform

- Develop & Deploy
- Integrate & Extend
- Publish & Engage
- Analyze & Predict
- Secure & Manage

Innovate with a Comprehensive, Open, Integrated and Hybrid Cloud Platform that is Highly Scalable, Secure and Globally Available

ORACLE

Oracle Cloud Platform

Comprehensive | Open | Integrated | Hybrid

- Data Management
- Application Development
- Enterprise Integration
- Data Integration

- Analytics and Big Data
- Content & Experience
- Identity & Security
- Systems Management

Oracle Public Cloud

Oracle Data Center

Built on High Performant Oracle Cloud Infrastructure

Oracle Cloud at Customer

Your Data Center

ORACLE

Oracle Cloud Platform Momentum

14,000+
Oracle Cloud Platform Customers

3,000+
Apps in the Oracle Cloud Marketplace

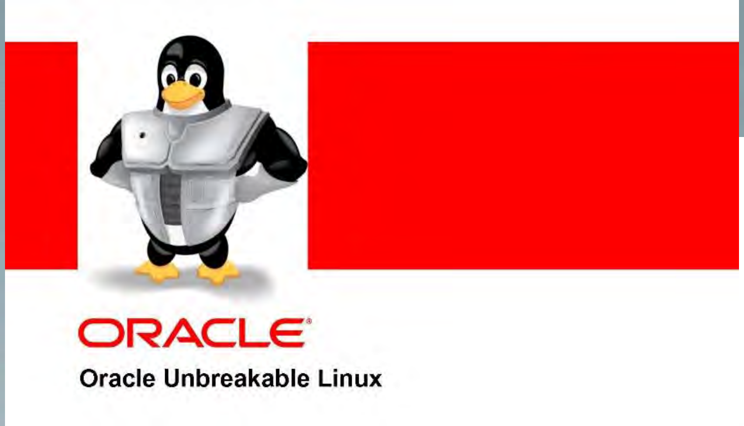
$1.4 Billion
FY17 Oracle Cloud Platform Revenue
(60% YoY Growth)

10 PaaS
Categories where Oracle is a Leader According to Industry Analysts

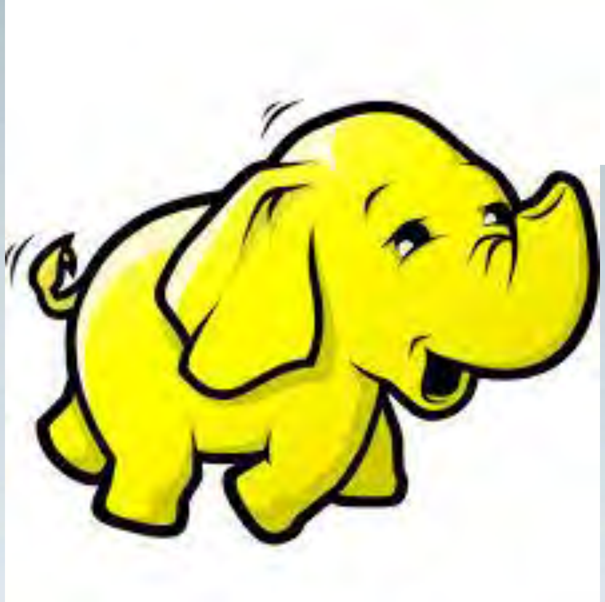
ORACLE

Oracle Big Data as a Service

- Work with a stack you are familiar with
- Maximum portability
- Maximum performance
  - I/O tuned for Oracle Cloud
  - OS tuned for Oracle Cloud
  - Network tuned for Oracle Cloud



Spark



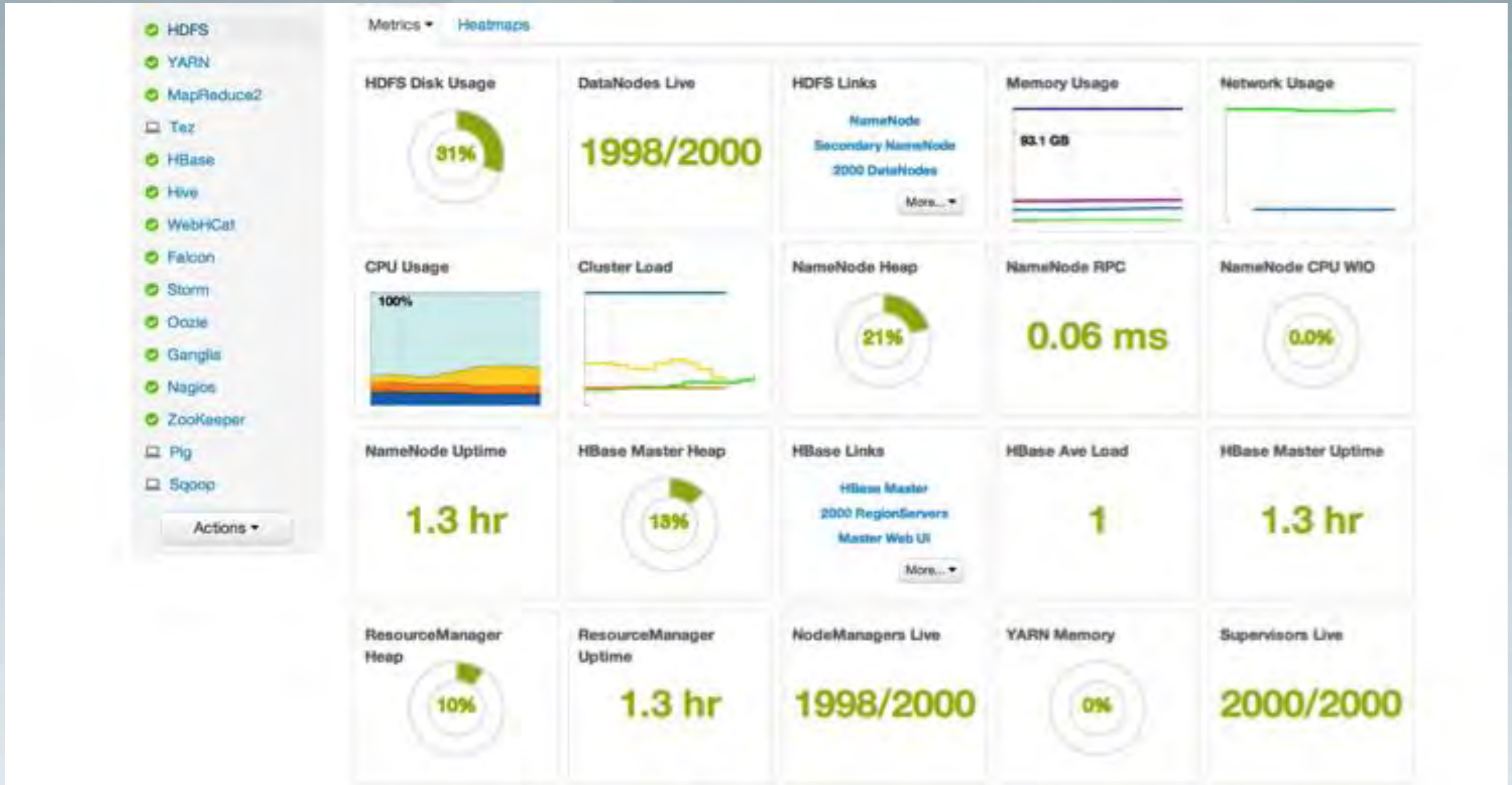
ORACLE

Oracle Cloud Platform does the grunt work

- Elastic and Scalable
  - Big Data clusters are elastic on demand
  - Storage is scaled independently
  - Choose appropriate compute shapes for your workload
- Managed
  - Automated lifecycle management
  - Service monitoring via dashboards or REST APIs

ORACLE

Big Data Cloud Service Management



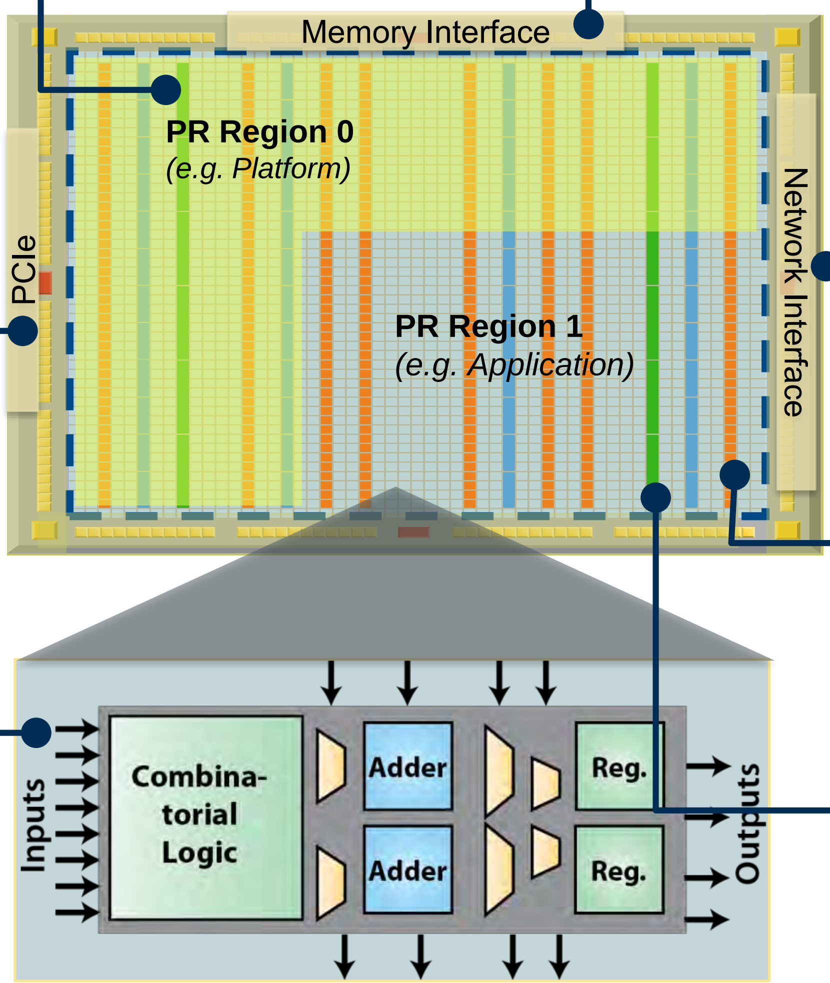
ACCELERATING CLOUD STORAGE WITH FPGA

FPGA Technology Introduction

PARTIAL RECONFIGURATION
- Allows separate regions

PCIE HOST INTERFACE
- Hardened + Soft host interface
- Hardened PCIe controller
- Soft interface allows different use models and drivers

LOGIC ELEMENTS
- Main programmable component
- Millions of logic elements
- Simple logic, adders, and registers
- Interconnect with configurable fabric



MEMORY INTERFACES
- Configurable high performance memory interfaces
- Hardened controllers

NETWORK INTERFACE
- Configurable network interfaces
- Hard/soft interfaces

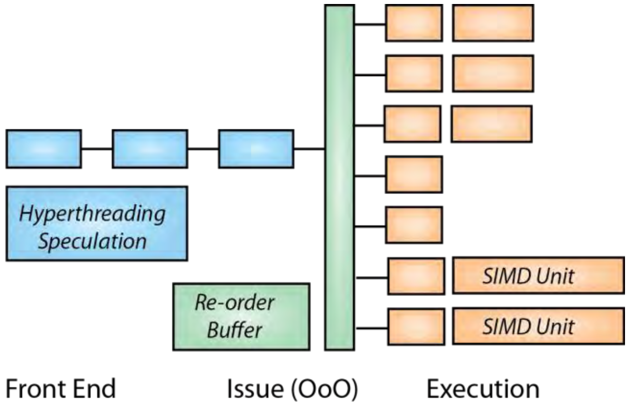
MEMORY BLOCKS
- Thousands of 20Kb memory blocks
- Allows processing to stay on-chip

VARIABLE PRECISION DSP BLOCKS
- Allows FPGA to perform compute intensive functions

Where FPGAs Fit In?

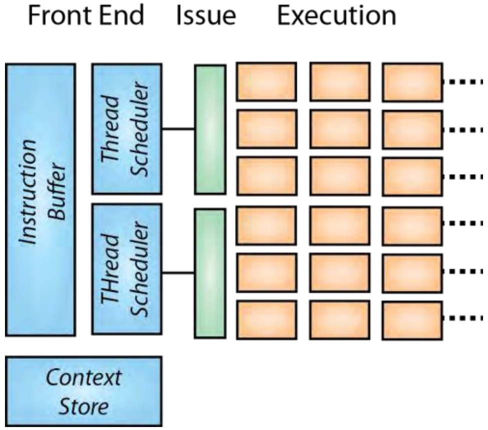
FLEXIBILITY ↔ EFFICIENCY

CPU



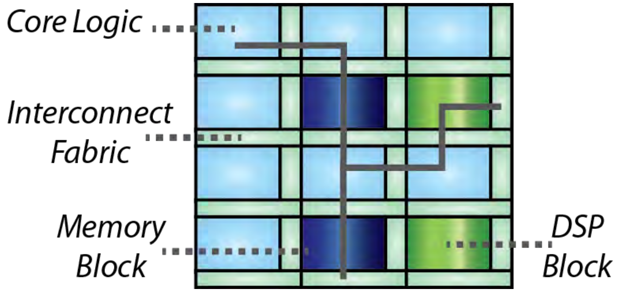
- Balanced architecture: *Good enough most workloads*
- Good single thread & throughput perf.
- Fastest cadence

GPU



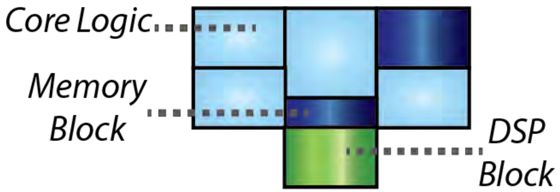
- Focused on compute throughput
- Many low performance threads
- High memory throughput
- Purpose made programming tools

FPGA



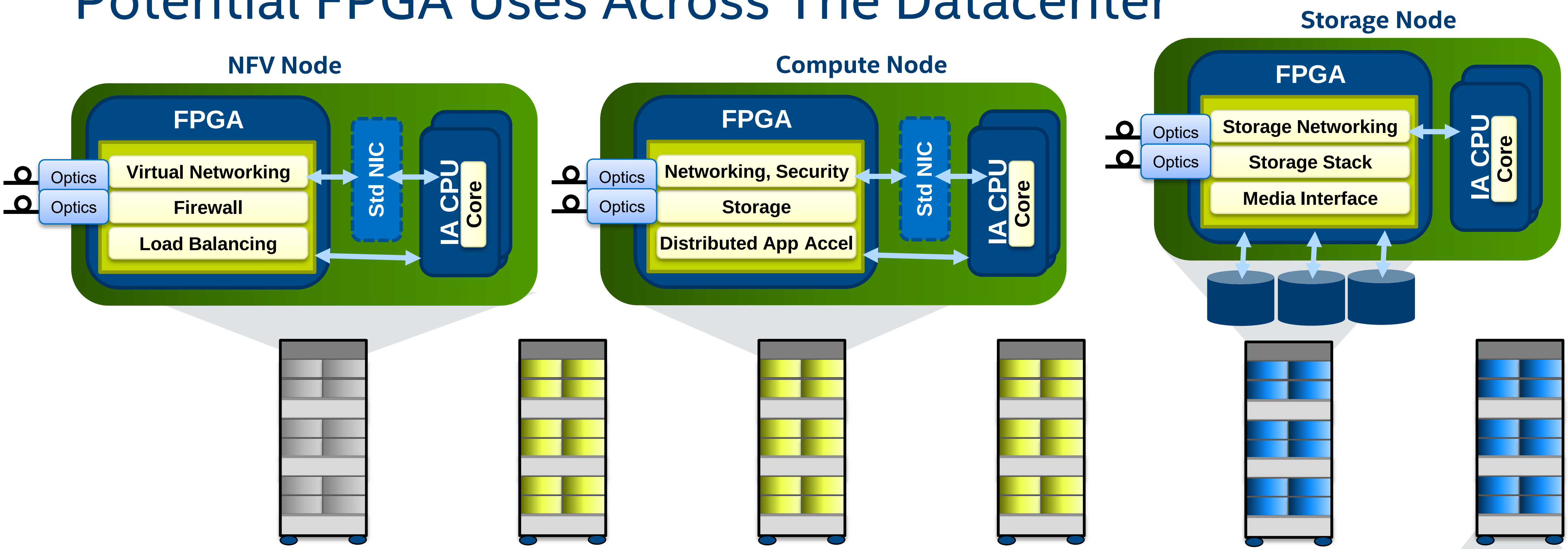
- Full custom pipeline
- Capable of networking and compute
- High memory throughput
- Change cadence in months → rapidly changing needs
- Requires sophistication

ASIC

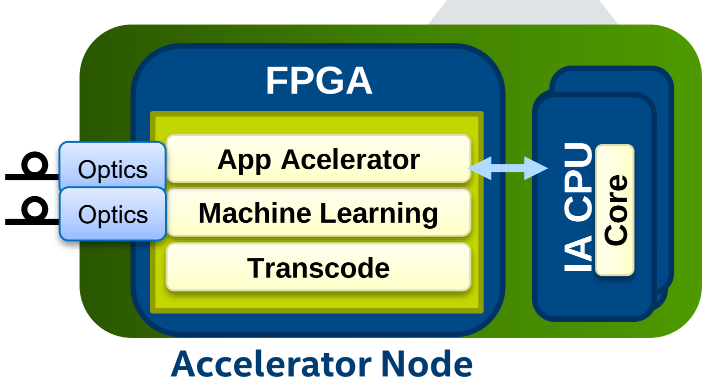


- Fixed function
- High efficiency – only blocks that are needed
- Change cadence in years → needs stable standards
- Expensive: minimum volume for viability
- Requires sophistication

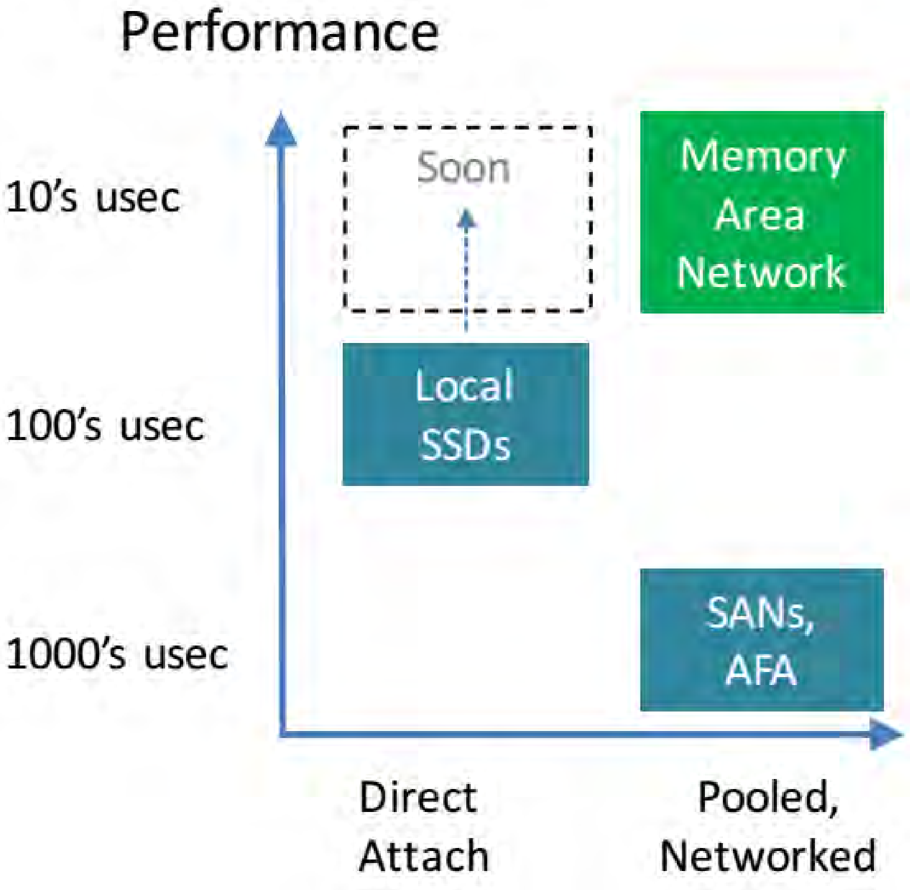
Potential FPGA Uses Across The Datacenter



- **NFV Node**: Provides infrastructure networking functions
- **Compute Node**: Provides Infrastructure including storage networking/stack acceleration
- **Storage Node**: Offloads stack, networking, and interfaces to media HDD, SDD, NVMe
- **Acceleration Node**: Focused on high Large FPGA

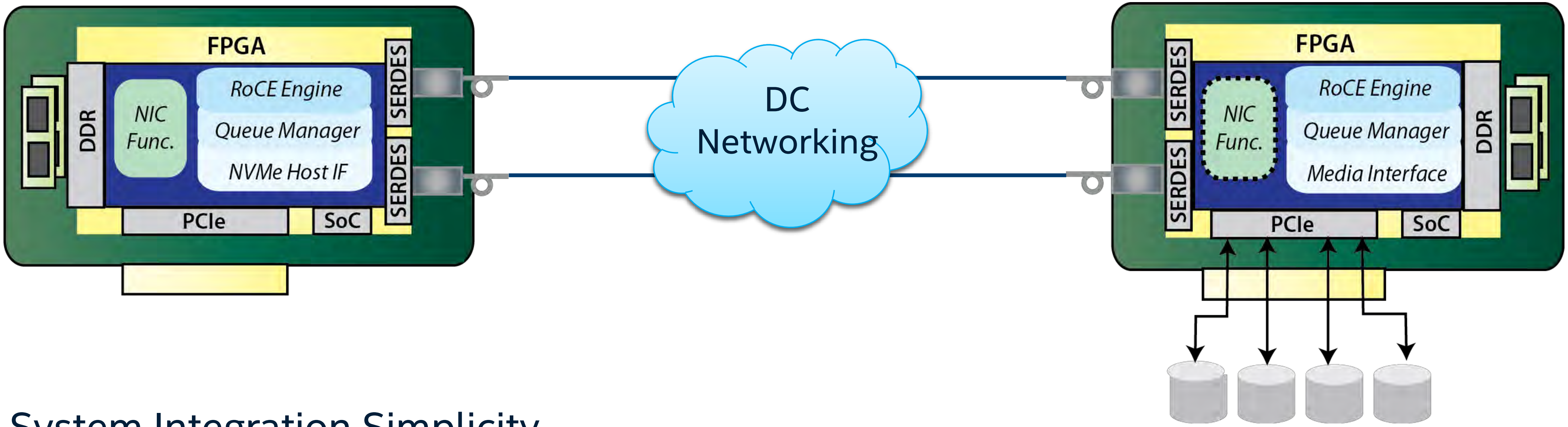


FPGA Value Proposition For Storage



- High Bandwidth & Low Latency
  - Driven by NVMe technology
- Optimized storage stack & networking
  - FPGA can be used to implement storage stack and networking protocol layers
- Rapidly evolving standards
  - FPGA adaptable to different protocols, stacks, and host interfaces

Example FPGA For Initiator & Target

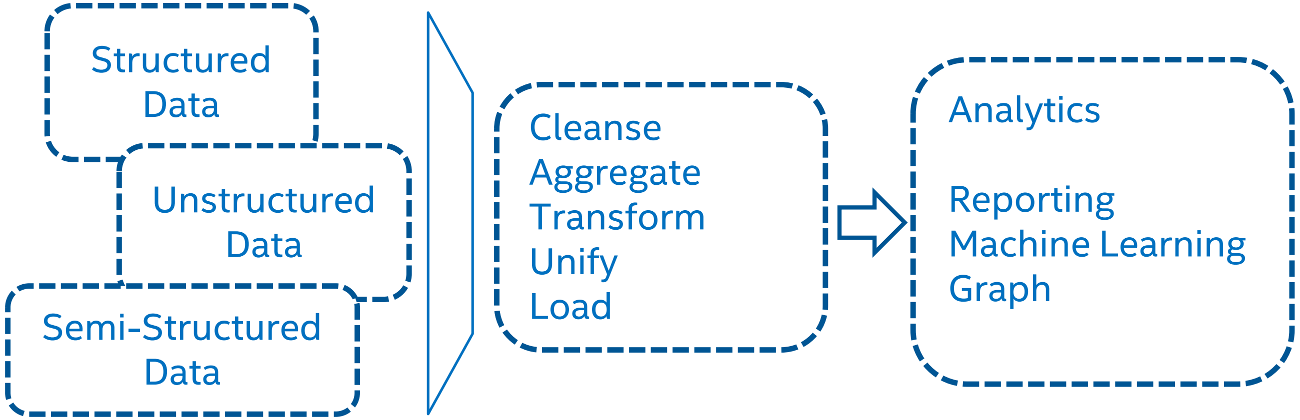


- System Integration Simplicity
  - No Drivers Required, Config SW runs from Console
- FPGA Future Proof Solution
  - FPGA solution can change with Standards, AFU capability (IPSEC, Compression)
- NVMeof compliant
  - Scalable NVMe Solution

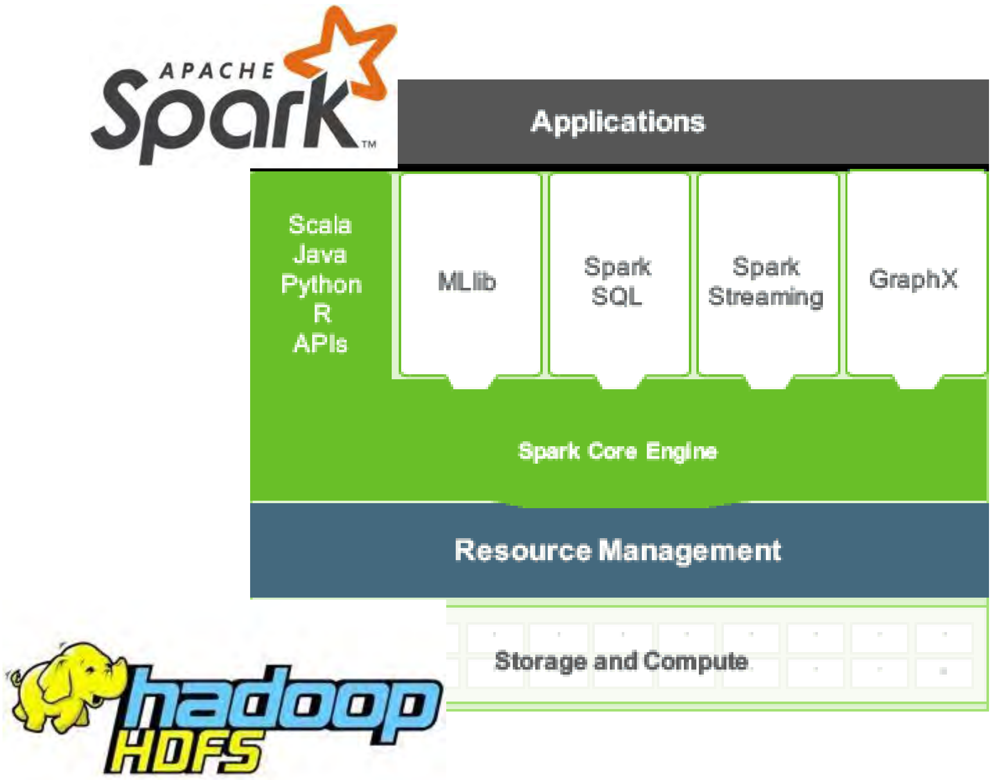
CASE STUDY

IMPROVING PERFORMANCE OF BIG DATA ANALYTICS

The Analytic Workload



We want to execute analytics in the cloud leveraging today's modern Hadoop framework with Spark



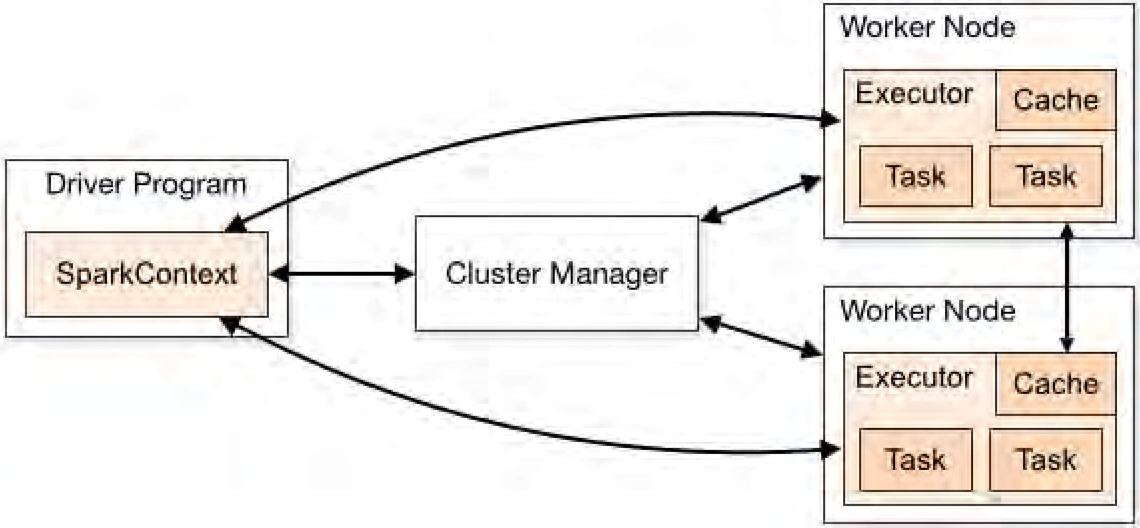
Our study will examine such an analytic pipeline over 3 phases

(1) **Import**
create, write data to HDFS

(2) **Load**
into Hive

(3) **Analyze**
data and report

Spark Resourcing



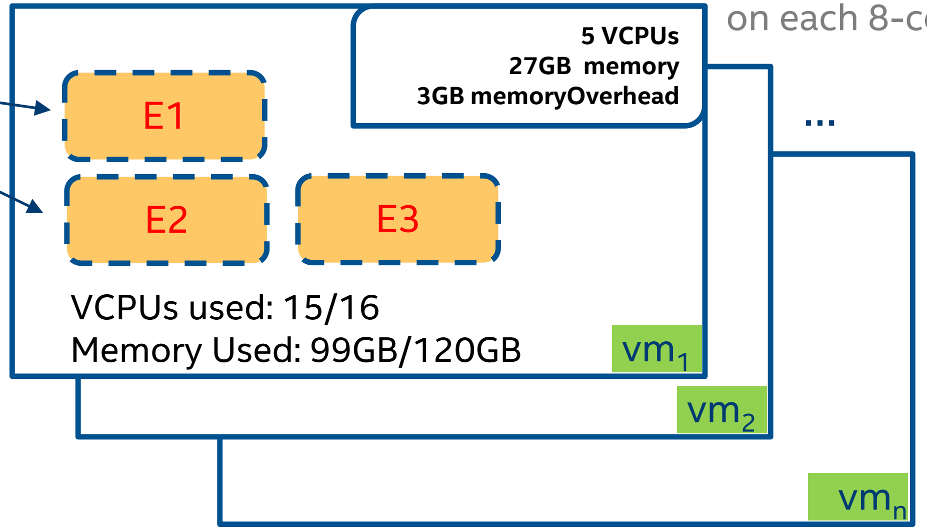
Spark worker nodes live in Hadoop as data nodes

The data node transforms from a physical server into a virtual server in our cloud

Spark Tunings:

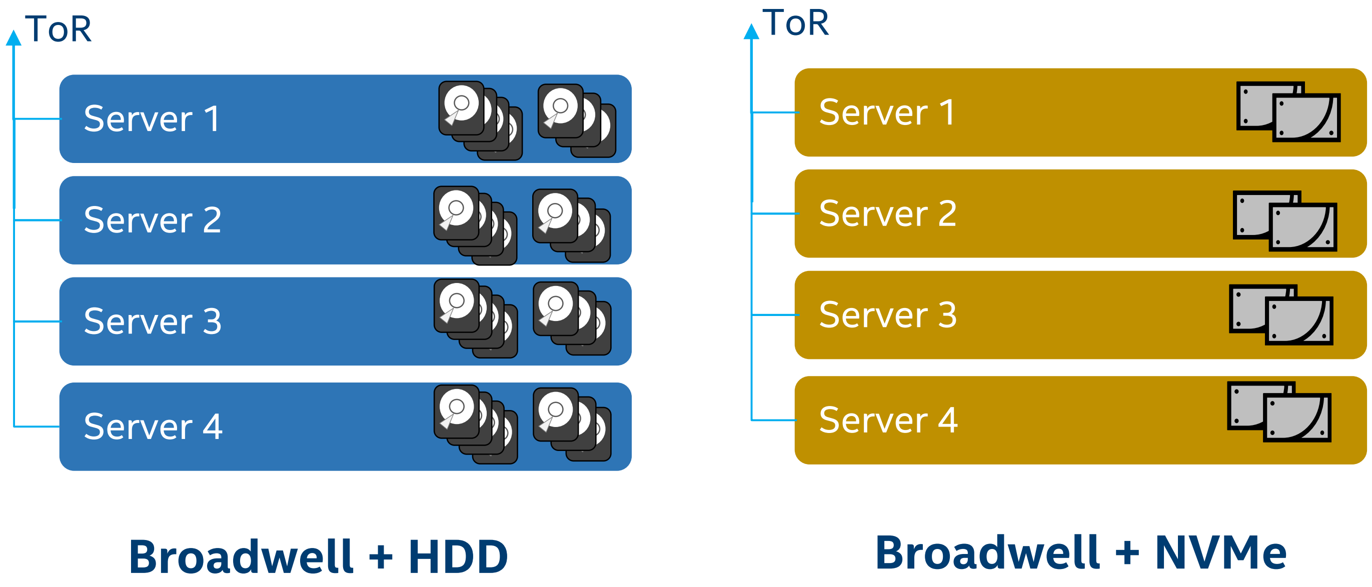
- Number of Spark Executors
- Cores per Executor
- Memory per executor
- Executor Memory Overhead
- Partitions, parallelism, …
- etc., etc.

deployment example having 3 Spark executors on each 8-core/16-VCPU virtual machine



Top diagram from http://spark.apache.org/docs/latest/img/cluster-overview.png

Big Data Run—**Hard Drive vs Local NVMe**

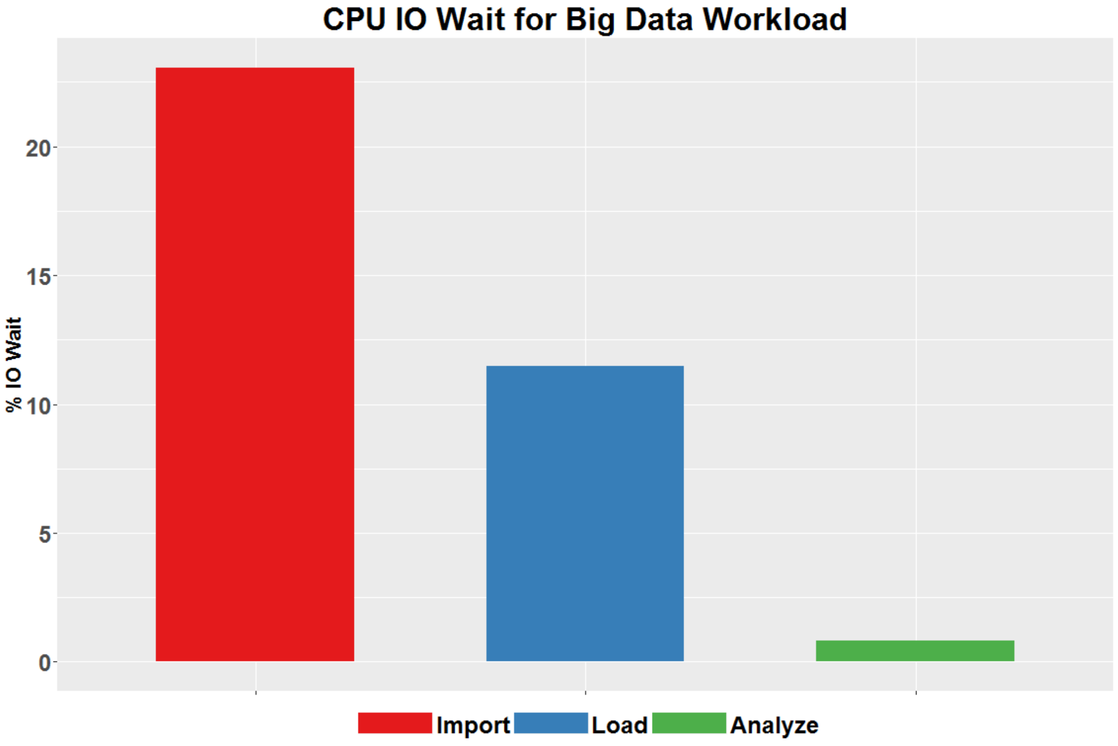


Platform	*2S Intel Xeon E5-2699 v4 768 GB DRAM. 10Gbps Ethernet.*	*2S Intel Xeon E5-2699 v4 768 GB DRAM. 10Gbps Ethernet.*
Storage	*7 x Hard Drives (2TB, 7200 RPM)*	*2 x Intel NAND NVMe*
OS/ Hypervisor	*Centos 7.2 / KVM*	*Centos 7.2 / KVM*
Big Data SW	*Hortonworks Data Platform 2.4*	*Hortonworks Data Platform 2.4*
Big Data Cluster	*18 Datanode VMs (16 VCPUs 120 GB memory) 2 Namenode VMs (4 VCPUs 30 GB memory)*	*18 Datanode VMs (16 VCPUs, 120 GB memory) 2 Namenode VMs (4 VCPUs 30GB memory)*

Big Data Performance Obstacle: **Local Hard Drives**

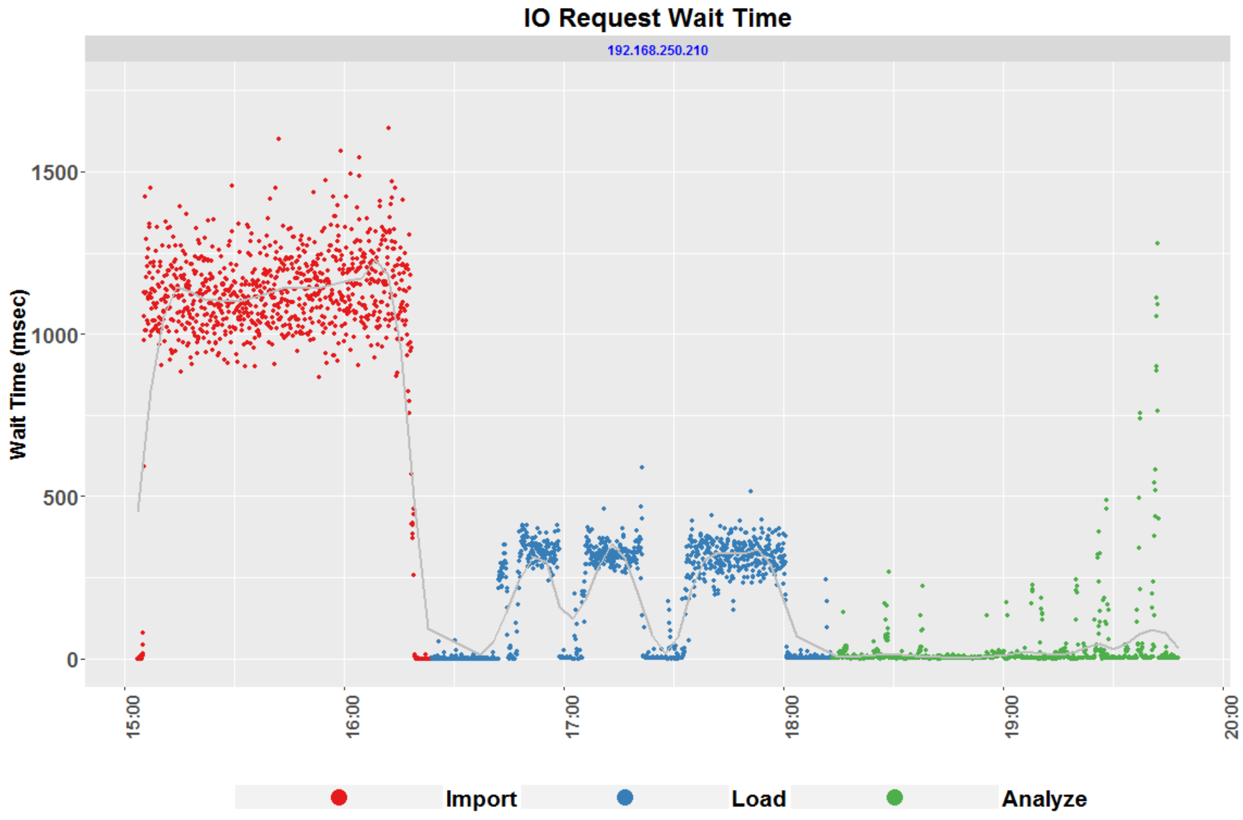
Input/Output

Local HDD leaves CPU waiting for data

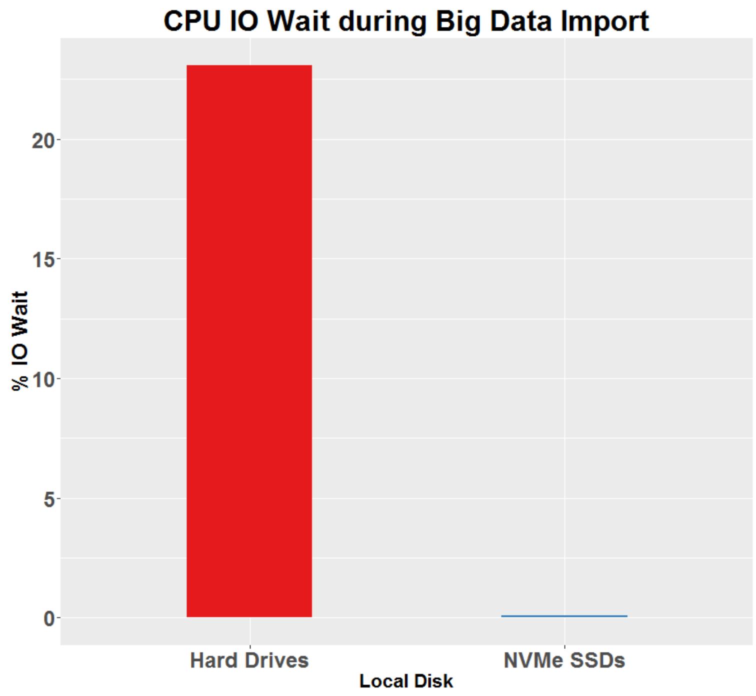


Import and Load Phase most IO-intensive and stand to gain from new storage technologies

IO service time likewise suffers

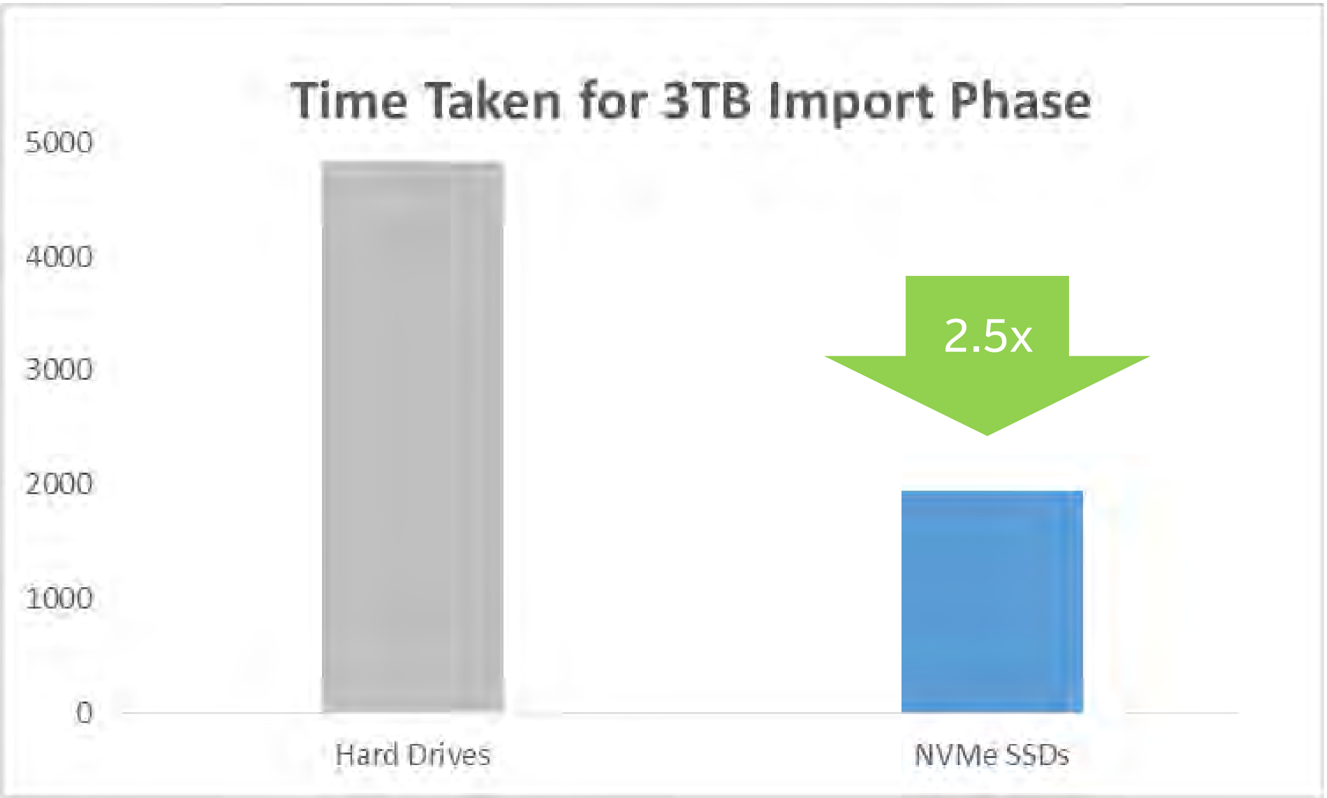


NVMe Impact to BDaaS

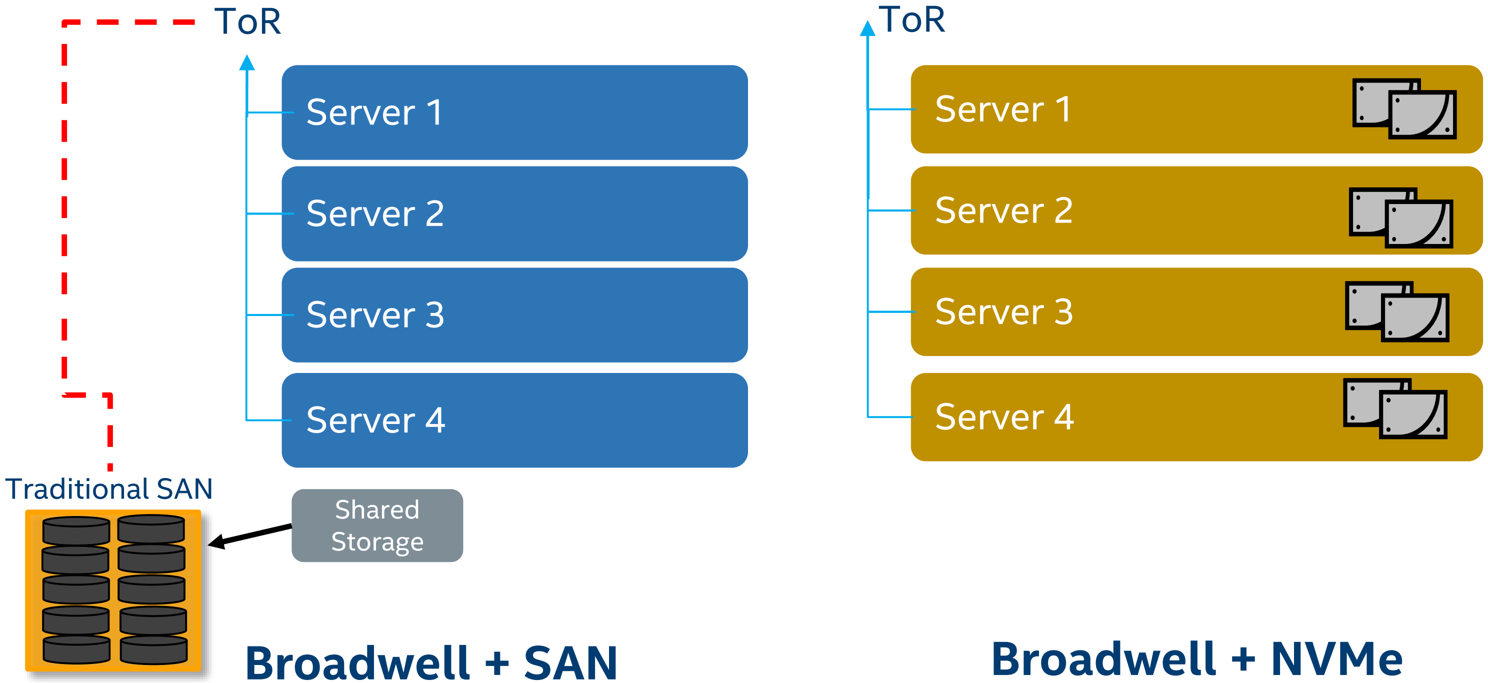


IO wait is virtually extinguished with the adoption of Intel NVMe SSDs

2.5x acceleration of Import phase achieved by replacing hard drives with NVMe SSD on the same cluster



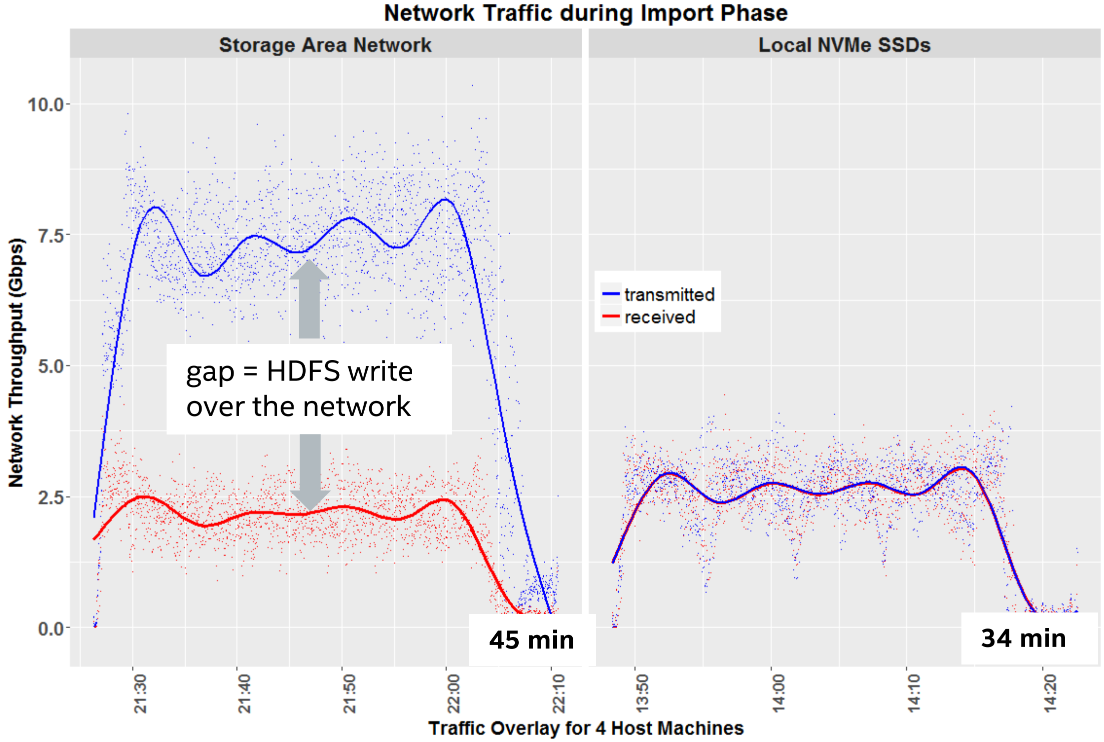
Big Data Run—**Storage Area Network vs Local NVMe**



	Broadwell + SAN	**Broadwell + NVMe**
Platform	*2S Intel Xeon E5-2699 v4* *768 GB DRAM. 10Gbps Ethernet.*	*2S Intel Xeon E5-2699 v4* *768 GB DRAM. 10Gbps Ethernet.*
Storage	*Enterprise Storage Area Network*	*2 x Intel NAND NVMe*
OS/ Hypervisor	*OEL 6.6 /* *Xen*	*OEL 6.6 /* *Xen*
Big Data SW	*Hortonworks Data Platform 2.4*	*Hortonworks Data Platform 2.4*
Big Data Cluster	*18 Datanode VMs (16 VCPUs 120 GB memory)* *2 Namenode VMs (4 VCPUs 30 GB memory)*	*18 Datanode VMs (16 VCPUs, 120 GB memory)* *2 Namenode VMs (4 VCPUs 30GB memory)*

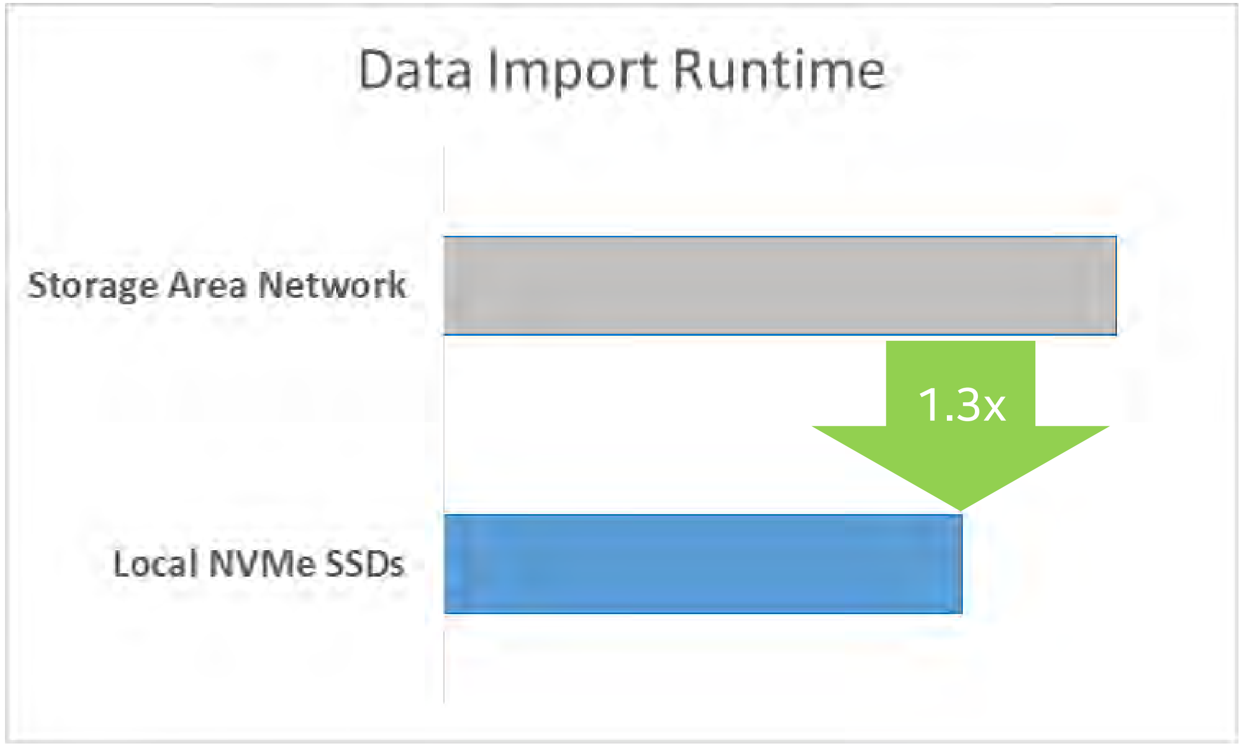
Big Data Performance Obstacle: Network

IO performance suffers if storage over network tests capacity. Local NVMe reduces congestion



At left our 10GbE NIC limited IO when using a SAN appliance. Local NVMe gives Hadoop data nodes a faster path to data

NVMe here yields a 1.3x acceleration in time to deliver our data into HDFS

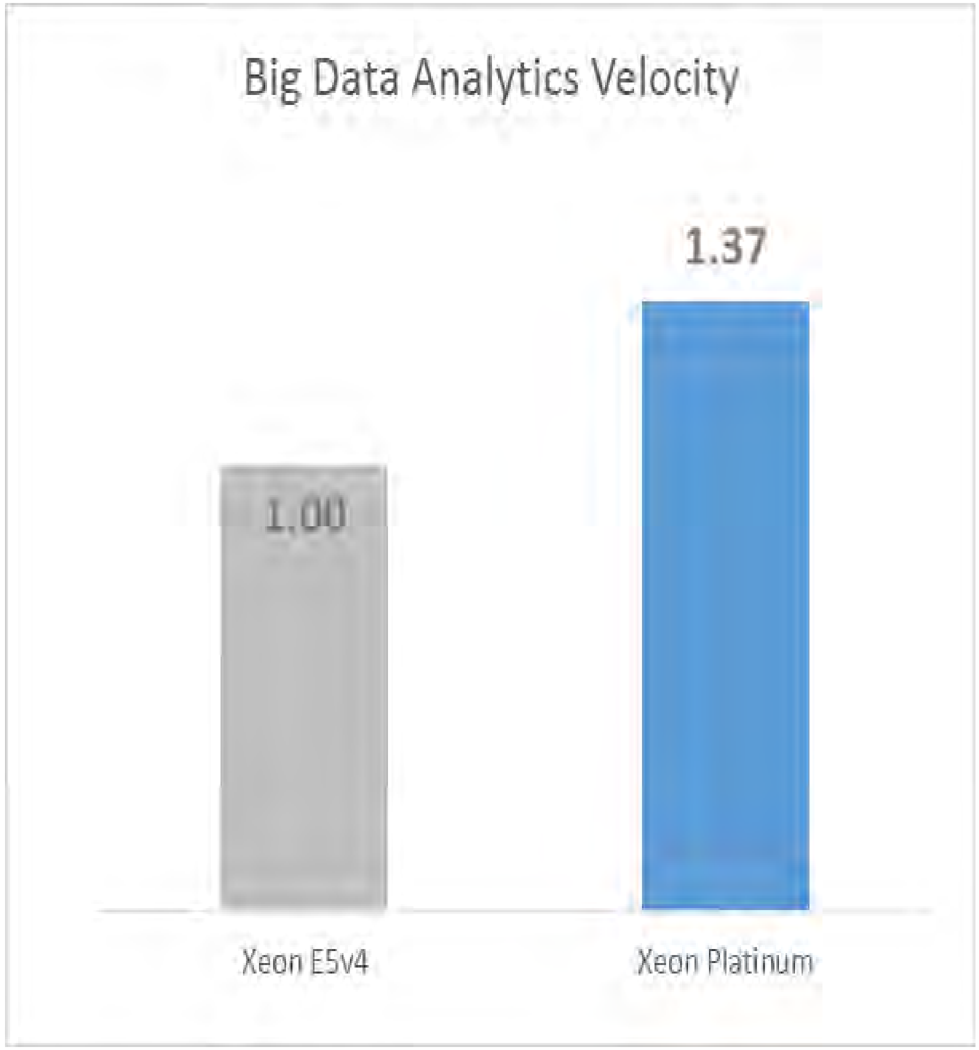


Analytics Scale with Intel Xeon Scalable Processors

Analytics need compute. Machine Learning and SQL ran at 1.37x the velocity over the prior generation using Xeon Platinum

- Skylake has up to 28 cores per socket
- Broadwell has up to 22 cores per socket
- Skylake microarchitecture improvements:
    - improved cache hierarchy for cloud workloads
    - 1.5X memory bandwidth

Intel Xeon Scalable Family lets you do more analytics in every server



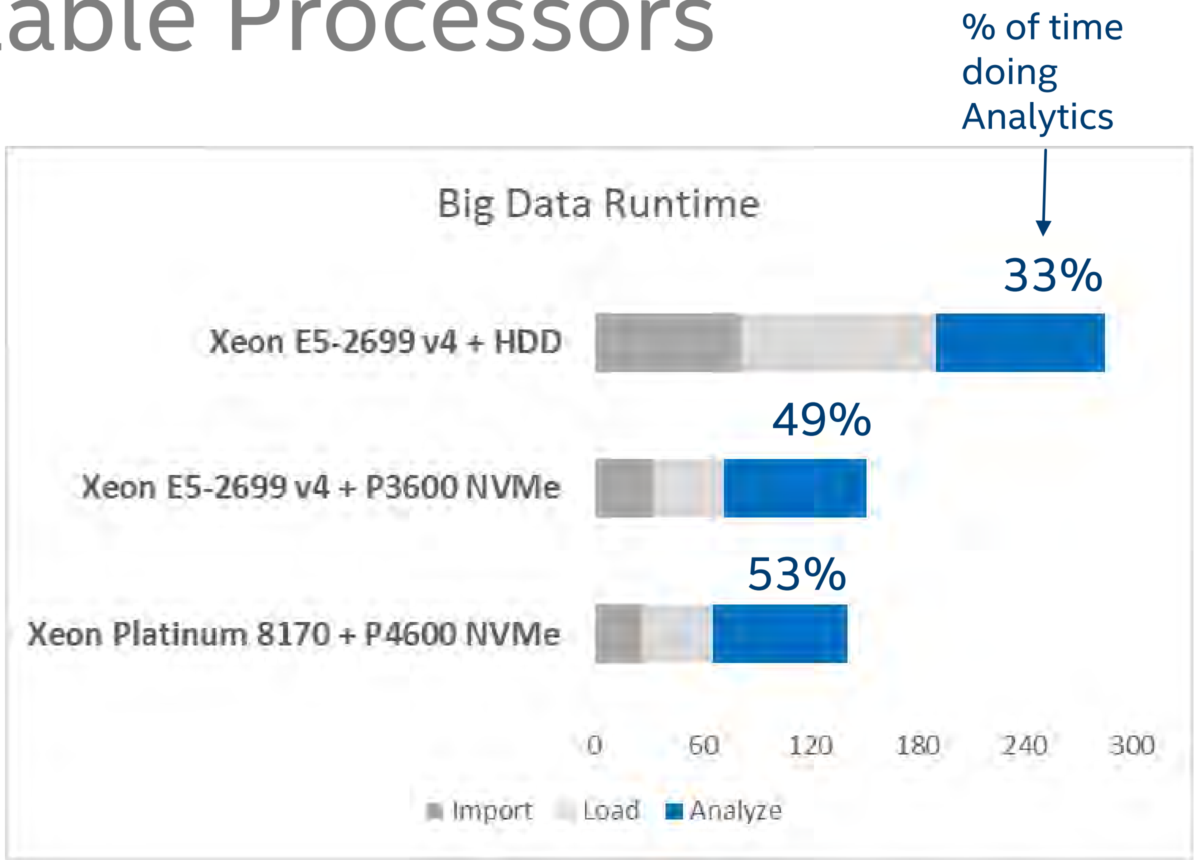
Raise Analytic Productivity with Intel NVMe and Xeon Scalable Processors

Customers build and use Big Data platforms to enhance and expedite decision making, derive insights and take action

Shorten the time to move and process data and expand opportunity for impactful analysis

Intel NVMe and Xeon Scalable Processors do just that. Leaving you time to innovate—

"Go off and do something wonderful"



Hortonworks Data Platform 2.4 on CentOS 7.2 with KVM data node VMs (16VCPU/120GB RAM) 2 name node VMs (4VCPU/30GB RAM) running executing 3TB data set on analytics workflow as:

(1) 4 x 2S Intel Xeon E5-2699v4 with 7x 1.8TB SATA HDDs
(2) 4 x 2S Intel Xeon E5-2699v4 with 2x 1.5TB P3700 NVMe
(3) 4 x 2S Intel Xeon SP 8170 with 2x 1.5 TB Intel P4600 NVMe (3D NAND)

Big Data Performance Summary

- Intel NVMe SSDs populate HDFS faster
  - 2.5x faster than with hard drives by removing the IO bottleneck
  - 1.3x faster than with a storage area network by relieving network congestion

- Skylake accelerates Machine Learning and SQL by 1.37x

- Do more Analytics with Intel NVMe and Xeon Scalable Processors
  - 53% of time in big data pipeline doing analytics, +20% over the previous generation

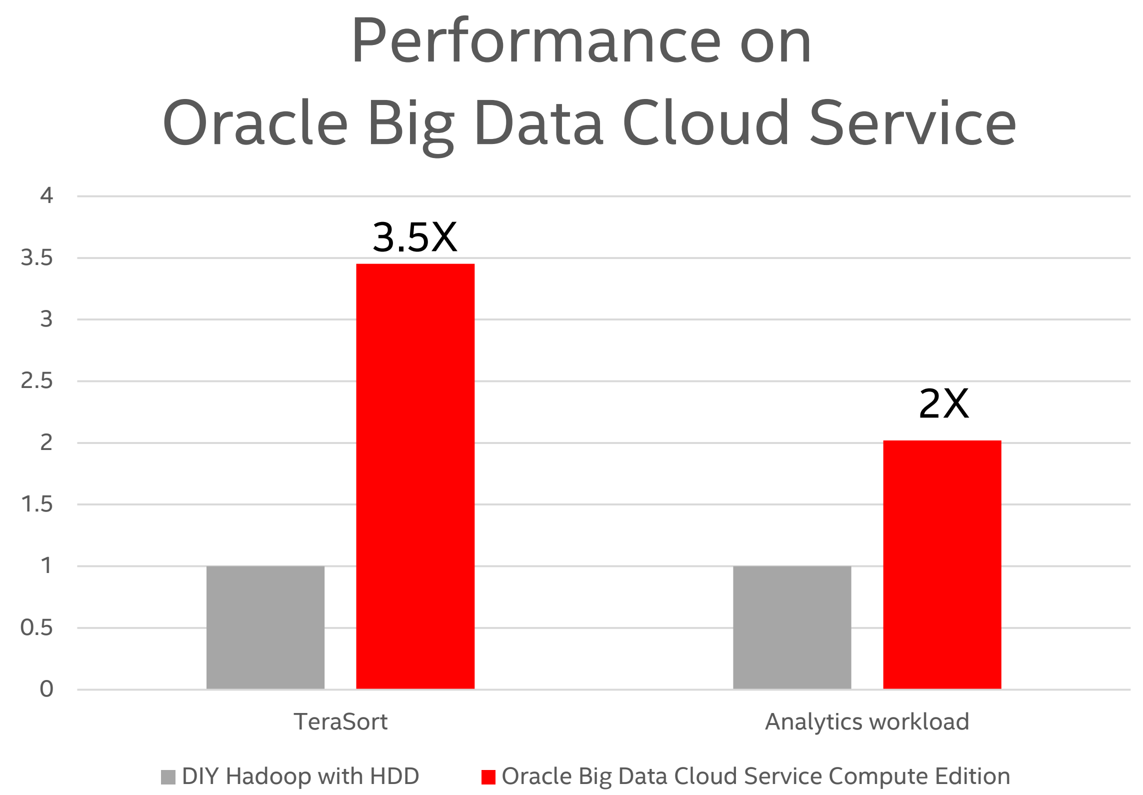
BRING CLOUD PERFORMANCE TO THE NEXT LEVEL

Oracle and Intel

Customers adopting both Hadoop and cloud services at a rapid rate

Oracle and Intel cloud partnership has extended to optimize Big Data

You can do it yourself but Oracle has made Big Data simple

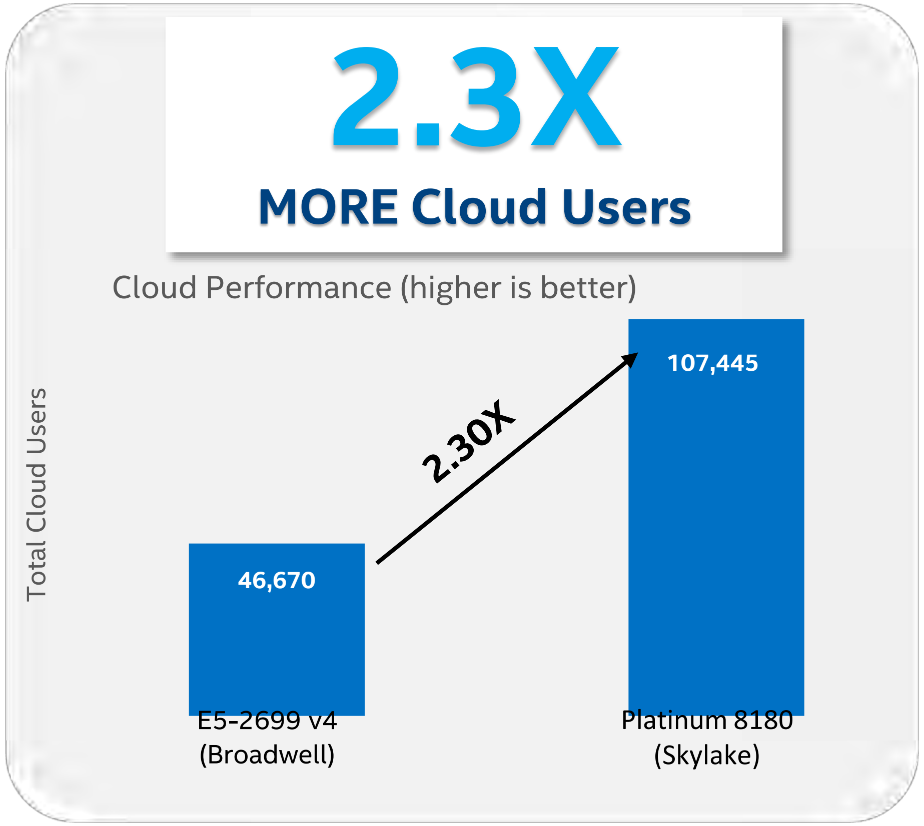


Up to 3.5X faster performance on Oracle BDCSCE over DIY Hadoop

Both clusters utilized 18 data node VMs (16VCPU and 120GB RAM) and 2 name node VMs (4VCPU and 30GB RAM), on 4 x 2S Intel Xeon E5-2699 v4 running either Oracle Big Data Cloud Services Compute Edition or DIY (Hortonworks Data Platform 2.4 on CentOS 7.2 with KVM, with 7x2.0TB 7200RPM drives per system)

Intel® Xeon® Scalable Processor optimized cache architecture

Increase cloud applications capacity, Higher VM density, Better isolation



- Intel® Xeon® Scalable processor features a new/optimized cache hierarchy
- Increased cloud performance for memory bandwidth-intensive applications
  - Larger non-inclusive Mid-Level Cache (MLC/L2) – **4X** increase (vs prior gen), enables applications to have more dedicated resources, and reduces impact of other applications on shared Last-Level Cache (LLC/L3
  - Larger total combined cache (MLC+LLC) available on Xeon® Platinum 8180 vs. Xeon® E5-2699 v4 (prior gen)

Optimized Cache Hierarchy delivers Improved Application Performance in the Cloud

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to http://www.intel.com/benchmarks. Configurations: Xeon® E5-2699 v4 processors setup: 16 Oracle WebLogic applications, 4x Intel® Memory Latency Checker v3.4 instances running on 1 HW thread per instance running on 2-socket Intel® Xeon® E5-2699 v4 processors (22-core per socket). Xeon® Platinum 8180 processors setup: 22 Oracle WebLogic applications, 4x Intel® Memory Latency Checker v3.4 instances running on 1 HW thread per instance running on 2-socket Intel® Xeon® Platinum 8180 processors (28-core per socket). For both Broadwell and Skylake, we used Oracle WebLogic Server 12.2.1.0.0, Oracle Java HotSpot™ 64-bit server VM on Linux version 1.8.0_102, Red Hat Enterprise Linux server release 7.3. Both platform has 768GB of DDR4 2133 MHz memory.*Other names and brands may be claimed as the property of others.