QxStack VMware® Edition for Apache Spark

Reference Architecture

9th January 2018, Version 1.0

Table of Contents

Legal Disclaimer

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH QUANTA CLOUD TECHNOLOGY (QCT) PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN QCT'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, QCT ASSUMES NO LIABILITY WHATSOEVER AND QCT DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF QCT PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

UNLESS OTHERWISE AGREED IN WRITING BY QCT, THE QCT PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE QCT PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

Quanta Cloud Technology (QCT) may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." QCT reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

All products, computer systems, dates, and figures specified are preliminary based on current expectations, and are subject to change without notice. Contact QCT local sales office or QCT distributor to obtain the latest specifications and before placing your product order.

1. Executive Summary

Customers around the world are striving to get insights from huge amount of data. To get valuable information from data in the reasonable speed, Apache Spark, a middleware framework, is a valid choice for customers' target application. By combining well-designed high-performance hardware with hyper-converged software stack from VMware, QCT provides ***QxStack VMware Edition for Apache Spark*** – a ready-to-use solution based on Cloudera Hadoop and Apache Spark in-memory computing engine with the focus on machine learning related workloads. In this reference architecture, QCT provides guidance for production and experimental variants of virtualized Apache Spark cluster with two storage options, DAS and vSAN, for different use cases.

QxStack VMware Edition for Apache Spark is best suited for various types of customers such as a startup company, health care institution, investment or retail bank, government agency, university, and retail business. With ***QxStack VMware Edition for Apache Spark,*** customers can get a balanced package with these noteworthy aspects:

- **Reliability** - QCT has applied VMware's guidance for Apache Spark/Cloudera Hadoop deployments to the QuantaGrid series server specification, adjusted its parameters, and tested the compatibility and performance of this solution to ensure optimal setup regarding machine learning workloads.
- **Performance** - low latency NVMe-based storage is suitable for response-time sensitive applications. High-core CPU is able to effectively process multithreaded Apache Spark applications written in Java/Scala.
- **Compute cluster elasticity and manageability** - VMware vSAN, vMotion, DRS and HA with monitoring from VMware guarantee the maximum resource utilization and maintain high availability at the same time.
- **Security** - VMware vSphere out-of-the box security features and Cloudera security features form strong defense shield against security threats.

QCT always stays innovative and values what customers care. ***QxStack VMware Edition for Apache Spark*** is designed to fulfill customers' needs and help customers stay ahead.

2. Scope

This reference architecture

- Describes four common use cases based on Apache Spark and a QCT-selected workload to demonstrate the benefits of Apache Spark on virtualized environment.
- Provides a pre-integrated and pre-validated solution- ***QxStack VMware Edition for Apache Spark*** based on Apache Spark and Cloudera Hadoop as a platform for data processing deployed on virtualized environment.
- Covers from planning and design to two best practices for different storage scenarios when customers deploy Apache Spark on Cloudera Hadoop.

3. Target Audience

This document is guidance for the readers who have basic knowledge and understanding of Apache Spark, VMware vSphere, and Cloudera Hadoop. This reference architecture can help the targeted audiences, including CIOs, solution architects, IT administrators, software developers, and data scientists to better understand the benefits and deployment process of the solution.

- CIOs could understand the benefits and estimated the cost of the solution.
- Solution architects could obtain a valuable insight regarding underlying infrastructure stack for the intended application.
- IT administrators could understand the benefits of the manageability and flexibility of the solution.
- Software developers and data scientists could get the information of ready-to-go infrastructure for development and experiments.

4. Glossary of Terms

Term	Description
QxStack	Storage related infrastructure product family from QCT, focusing on solution validation and performance optimization.
Cloudera Hadoop	Popular Apache Hadoop distribution from Cloudera.
Apache Spark	An application programming interface and runtime computation engine centered on a data structure is called the resilient distributed dataset (RDD). RDD is a read-only multiset of data items distributed in the memory over a cluster of machines.
VMware vSphere	An enhanced suite of tool for cloud computing utilizing ESXi hypervisor from VMware.
VMware vSAN	Software-defined storage solution from VMware.
VM	Virtual machine.
DAS	Direct-attached storage where operating system accesses storage hardware through the minimal number of additional layers. It is opposite to network accessed storage.
ML	Machine learning – discipline in computer science which gives computers the ability to learn from data.
MLlib	Apache Spark MLlib is a distributed machine learning framework on top of Apache Spark.
Worker Node	A node or virtual machine where computation on the data occurs. One or more Apache Spark executors run on the worker node.
Spark Executor	A process which performs computation over data in the form of tasks.
SparkContext	A connection to Apache Spark cluster used by Spark driver application.
Spark Driver	A client application submits tasks to the cluster through the cluster manager.
YARN	Yet Another Resource Negotiator – a popular cluster manager used to distribute computation tasks over cluster. It is default cluster manager for Cloudera Hadoop.
HDFS	Hadoop Distributed File System - a distributed reliable storage across cluster nodes.
Sensitive Data Redaction	Data redaction is the suppression of sensitive data such as any personally identifiable information (PII). PII can be used on its own or with other information to identify or locate a single person, or to identify an individual in context.
NUMA	Non-uniform memory access. It is a characteristic of modern 2 or 4 socket servers. Memory access time depends on the memory location relative to the processor. Under NUMA, a processor can access its own local memory faster than non-local memory (memory local to another processor or memory shared between processors).

5. Apache Spark for Data Analysis

Cloudera Hadoop with Apache Spark 2.1 provides ideal combination of rich Hadoop ecosystem and modern high-performance data analysis framework. Apache Spark comes with Resilient Distributed Datasets (RDD) based computation with the following attributes:

- In-Memory - Most of the data inside RDD can be stored in memory for most of the computation time which can enhance the processing performance.
- Immutable or Read-Only – RDD will not be changed and only be transformed to new or different RDD.
- Parallel – allows access and process data in parallel.
- Partitioned — Data records are split into logical partitions and distributed across compute nodes in a cluster.
- Location-Stickiness — RDD's placement can be set in preferences to have compute partitions as close to the records as possible.

5.1 QCT Selected Workload - Machine Learning

Obtaining valuable information from data becomes one of the critical success factors in term of business development and operation strategies for customers. Thanks to the new findings in the neuroscience, advanced machine learning algorithms, and improved hardware performance, invisible data insights are now available to support complex decision making for customers. Machine learning algorithms running on Apache Spark play a key role in data analysis. The data set is visited multiple times in a loop using database system like queries of data.

Total run time of these machine learning applications could be reduced by several orders of magnitude compared to a MapReduce based computation. Besides the static data analysis, Apache Spark also significantly outperforms traditional MapReduce-based computing for live data processing or live data enrichment use cases thanks to the low machine learning response times.

5.2 Apache Spark Common Use Cases

5.2.1 Streaming Data Processing

Apache Spark in memory computing abilities allows real-time or close to real-time processing of both inward and outward data streams. Examples of such an application could be:

- **Streaming ETL** – Traditional extract-transform-load (ETL) batch processing in data warehouse environments. Data are pre-processed before they are stored in relational database.
- **Data enrichment** – Live data are enriched by previously-stored static data. Online advertising combining historical customer data and live data from customer behavior is a use case of data enrichment and can help the customer to deliver more personalized ads in real time.
- **Trigger event detection** – When rare or unusual patterns of system behavior which deviates from normal operation states occurs, the system with Apache Spark can detect the event and trigger necessary processes. The example could be a single event which cause serious loss of finance, damage of private or public property, or threat of person's life.
- **Complex session analysis** – When the correlation of two or more dynamic events needs to be analyzed, Apache Spark could help to quickly find anomalies. The examples could be the events related to live web sessions.

5.2.2 Interactive Data Visualization

Visualized data help customers to understand events or trends happening in the life system. Operators or administrators could quickly react if the visual outcomes of data analysis are available in the acceptable time frame. Thanks to Apache Spark version 2.0 structured streaming, the interactive queries against live data could be performed.

5.2.3 Trend/Event Prediction

Nowadays, the foundation for modern predictions is machine learning. Apache Spark ecosystem offers MLlib as the first choice in machine learning algorithms framework. Predictive intelligence, customer segmentation, and sentiment analysis are the related use cases. Rapidly evolving is also the use case of network security where Apache Spark helps to discover new threats as soon as attacker enters the system.

5.2.4 Fog Computing

Internet of Things (IoT) related data processing deals with huge amounts of highly-variable data. Complex data analysis executed remotely in the cloud far from IoT data sources doesn't seem to be the right fit for the expected outcomes. Decentralization and analytic speed best matches the concept of Fog computing. Because of the processing speed of Apache Spark, experts predict that Apache Spark has the potential to be the first choice of the platform for IoT computing.

5.3 Apache Spark Running on QCT QxStack VMware Edition

QCT combined well-designed high-performance hardware with hyper-converged software stack in which VMware vSphere and VMware vSAN are mainly adopted as a solid foundation for the Apache Spark infrastructure. QCT proposed the first generation of QxStack VMware Edition with Apache Spark to fulfill the following needs:

5.3.1 On-demand service for tenants - Data isolation

The standalone Apache Spark was not originally designed as a multi-user framework. To configure data isolation, resources allocation and high availability for different tenants are complex tasks. Cloudera provides resource management controls when the same data are shared. VM-based virtualization brings the possibility of data isolation to the higher level, widely accepted in multi-tenant environment. Virtualization with shared storage can mix various workloads to achieve the maximum hardware utilization.

5.3.2 Flexible virtualization infrastructure

The virtualization infrastructure can provide development, test, staging, and production environment on the same hardware. The speed of development, test, and deployment cycle becomes as critical as the service availability itself. Only flexible-virtualized infrastructure would allow customers to add new cluster nodes within few minutes.

5.3.3 High-level security grade and manageability at the same time

Only cloud with security controls in place could provide the highest level of security to protect data and workloads. VMware vSphere and vSAN provide a number of features to support security from hardware to virtual machine level.

6. QxStack VMware Edition for Apache Spark

QCT provides a pre-integrated and pre-validated solution deployed on virtualized hardware based on the two platforms, Apache Spark and Cloudera Hadoop, for data processing. In this reference architecture, QCT hardware is combined with virtualization stack from VMware. Cloudera Hadoop comes with Cloudera Manager – comprehensive tool for Cloudera Hadoop/Apache Spark cluster management. Cloudera Hadoop also provides Hadoop file system (HDFS) to ensure data replication over cluster members. Apache Spark task planning is controlled by YARN resource manager. MLlib is Apache Spark's scalable machine learning library which directly interacts with workload applications and is optimized to run on CPU. Major components of the solution and the selected machine learning workloads are demonstrated in Fig. 1 and Chapter 7.2 while the best practices for the solution deployment are demonstrated in Chapter 8.

QCT provides two storage deployment scenarios listed below to respectively fulfill the requirements of performance and flexibility. Storage deployment scenarios are explained in detail in Chapter 8.1. Each scenario has a variant for production or experimental version of the Apache Spark cluster:

- DAS: directly attached storage provides best storage performance and the highest capacity within limited 1U server.
- vSAN: provides best storage flexibility leveraging software defined storage.

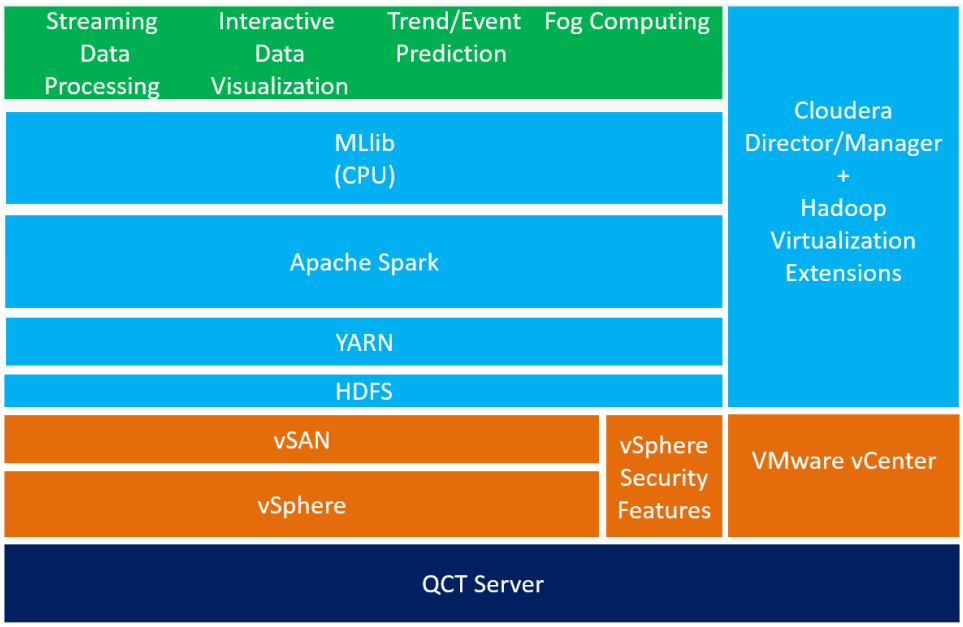

Figure 1. QxStack VMware Edition for Apache Spark - architecture.

The cluster configuration of ***QxStack VMware Edition for Apache Spark*** is composed of three high-end vSAN Ready Node certified servers. Four worker-node virtual machines are hosted on each virtualized host.

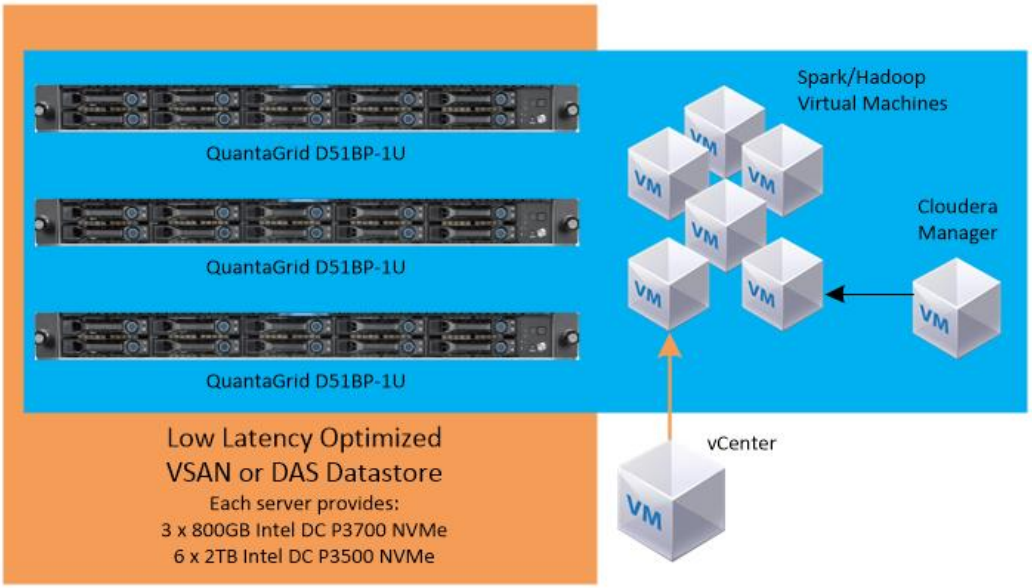

Figure 2. QxStack VMware Edition for Apache Spark - topology.

7. Hardware and Software Components Introduction

7.1 Hardware Components

To design a best-suited solution for the selected workload, QCT tested QuantaGrid D51BP-1U with discreetly selected components, as shown in Table 1.

Table 1. QxStack VMware Edition for Apache Spark hardware components.

Component	Description	Quantity DAS	Quantity vSAN
System	QuantaGrid D51BP-1U	1	1
CPU	Intel E5-2699 v4 @ 2.20 GHz	2	2
DIMM	Up to 640GB RDIMM/1280GB LRDIMM	20	20
Storage Cache	Intel SSD DC P3700 Series; 2.5in PCIe 3.0 800GB	-	3
Storage Data	Intel SSD DC P3500 Series; 2.5in PCIe 3.0 2TB	10	6
ESXi Boot Drive	SATADOM 32GB	1	1
NIC1	Intel XL710-QDA2 40Gb/s XL710QDA2G2P5	1	1
NIC2	S2BP ON 10G LAN/B 82599ES W/BKT(1 IN 1)	1	1
Network Switch	Quanta T5032-LY6	1	1

Hardware Selection Highlights

QCT's servers are designed and optimized for a data center - energy efficiency, solid-built quality, and aesthetic design. Hardware design follows the latest innovations in the industry.

- **Less than 5-millisecond response time of vSAN for Apache Spark I/O requests**
  By adopting all-NVMe design, storage response time stays low even for long IO operations, typically for HDFS.
- **2x Intel E5-2699 v4 CPU with 88 logical cores in total**
  A right fit for Java and Scala multithreaded applications, and enough power to support both storage and compute in the same box.
- **Up to 640GB RDIMM or 1280GB LRDIMM RAM**
  Support for memory-intensive Apache Spark workloads.
- **QuantaMesh T5032-LY6**
  A powerful Spine/Leaf switch to support vSAN traffic.
- **3U minimum cluster size**
  Three hosts as a building block with plenty of power for SMB and the potential to grow up to 64 hosts within one vSAN cluster - suitable for large customers.

7.2 Software Components

Cloudera Hadoop 5.11 with Apache Spark 2.1

Cloudera is one of the leading Apache Spark and Hadoop vendors on the market. Rich Hadoop ecosystem is extended by the newest Apache Spark version 2.1 which is a modern in-memory computing framework to bring a stunning performance in terms of data analysis. With the public cloud support of Cloudera Hadoop, customers can choose to run Apache Spark completely on premise or to expand Apache Spark cluster from private cloud to public cloud.

VMware vSAN 6.6

VMware vSAN is a radically-simple, enterprise-class, shared-storage solution for hyper-converged infrastructure optimized for today's both hybrid and all-flash performance. vSAN delivers predictable, elastic, and non-disruptive scaling of storage and compute resources, eliminating costly forklift upgrades. Every Virtual SAN cluster can scale out one node at a time or scale up by adding capacity to the existing hosts. It is capable of achieving over 8 PB of raw storage capacity.

VMware vSphere 6.5

VMware vSphere, an industry-leading virtualization platform, provides a powerful, flexible, and secure foundation for business agility that accelerates digital transformation to cloud computing and the success in the digital economy. VMware vSphere supports both existing and next-generation apps through its

1) Simplified customer experience for automation and management at scale,

2) Comprehensive built-in security for protecting data, infrastructure, and access,

3) Universal app platform for running any app anywhere.

With vSphere, customers can now run, manage, connect, and secure their applications in a common operating environment across clouds and devices.

Software Selection Highlights

Security

- **Secure Boot** – The software can protect both hypervisor and guest operating system by ensuring that the images are not to be tampered and no unauthorized components are loaded at the level of hypervisor and guest operating system.
- **Multiple layer encryption** – vSAN level, VM level, and vMotion encryption

prevent unauthorized access for both data at rest and data in motion. HDFS transparent encryption provided by Cloudera could be further applied inside the virtual machine.

- **Unique encryption keys handling** – VM encryption keys are not stored within virtual machine memory region; thus, the key can be unreachable when malicious attack occurs inside a VM**.**
- **Cloudera Security** – authentication, encryption, authorization, and sensitive data redaction will protect the data inside a virtual machine.

Management

- **Cloudera Manager and VMware vCenter** – industry leading management tools for Apache Spark and Hadoop platforms, and virtual machine infrastructure.
- **Hadoop Virtualization Extensions(HVE)** - HVE consists of hooks and the extensions to data locality components of Hadoop, including network topology, HDFS write/read, balancer, and task scheduling.

8. QxStack VMware Edition for Apache Spark Best Practices

This reference architecture provides the best practices of Apache Spark cluster deployment with two storage scenarios -DAS and vSAN. In QCT solution design, integration, and performance testing, the well-known best practices from Cloudera and VMware are applied on QCT servers. The strategy is to prepare two solution variants focusing on the performance or on the flexibility for data analysis/processing.

QCT is aware that customers would tend to reach the highest performance and minimize the cost related to space and energy to run IT infrastructure. NVMe devices offer unique combination of high performance and low energy consumption in 2.5" space compared to spinning disks, that is, it can provide excellent performance in limited space and further minimize the total cost. Therefore, NVMe SSD is selected rather than spinning hard drives.

If the capacity would be the concern, storage scenario based on VMware vSAN cluster offers data compression and deduplication as the storage scale up or scale out.

Table 2. VM and Cloudera Apache Spark and Hadoop components Mapping.

VM	CDH Components
CDH1 – master node	- HDFS namenode - HDFS HttpFS - HDFS Balancer - Cloudera Management Service: Activity Monitor Alert Publisher Event Server Host Monitor Reports Manager Service Monitor - Spark 2 Gateway - Spark 2 History Server - YARN NodeManager - ZooKeeper Server
CDH2 – master node	- HDFS SecondaryNameNode - HDFS NFS Gateway - Spark 2 Gateway - ZooKeeper Server
CDH3 – master node	- Spark 2 Gateway - YARN Job History Server - YARN Resource Manager - ZooKeeper Server
CDH4-15 – worker nodes	- HDFS DataNode - Spark 2 Gateway - YARN NodeManager

8.1 Storage Deployment Scenarios

8.1.1 DAS Scenario

Direct-Attached Storage (DAS) deployment scenario is designed for customers who expect the best storage performance ratio and the highest capacity within a limited 1U server. Each NVMe device is represented as one independent vSphere datastore. According to the QCT internal testing, NVMe devices can provide more than a double of available throughput for read IO operations and have enough controller IOPS capacity to serve more than one VMDK per device at the same time in comparison to spinning hard drives. However, as few VMDK as possible per single NVMe is recommended due to relatively longer IO operations for HDFS filesystem. VMDK files for all virtual machines are evenly distributed among available datastores in the host. The data-storage high availability of worker nodes is ensured by HDFS replication mechanism. Three virtual machines – master nodes of the production cluster (CDH1-CDH3) with critical Apache Spark services are hosted on different infrastructure for reliability reasons.

To set up an Apache Spark cluster for the purpose of experiment or research, three bare metal hosts are recommended as a minimum configuration to combine master and worker nodes within one infrastructure. Figure 3 shows the storage layout example for the research or experimental cluster. Virtual machines in grey color are master nodes acting in management roles. The odd number of virtual machines is accommodated within the first three hosts of the experimental cluster, which is not in line with general NUMA locality concept and four VMs per host recommended by VMware.

According to the test results, five virtual machines are suggested, because only four virtual machines in worker node role would actually get benefit from correct NUMA placement due to the significant difference in load between master and worker nodes.

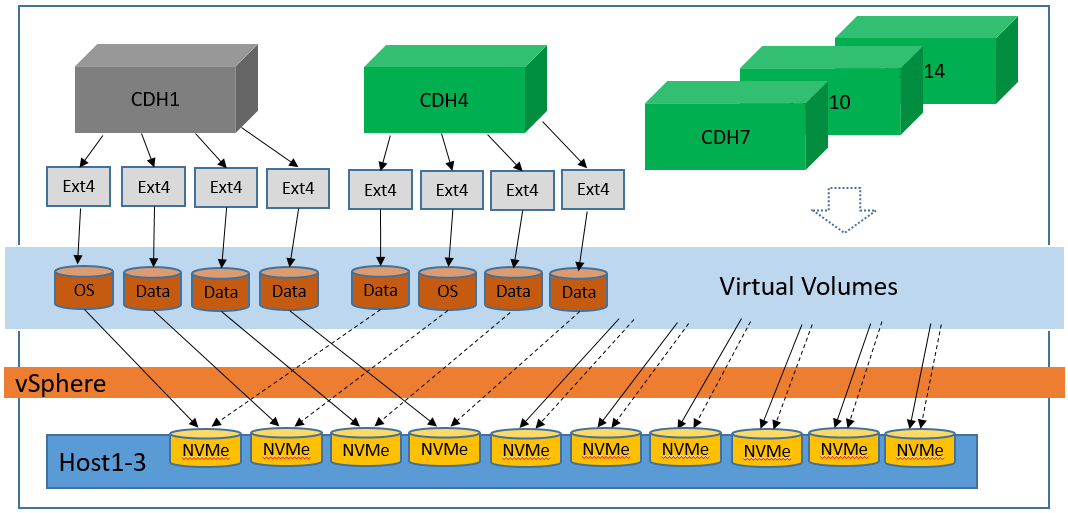

Figure 3. DAS VM storage layout for small research/experimental clusters.

Figure 4 shows the storage layout example for standard production cluster, which emphasize more on high availability than on experimental cluster. The example of the standard production cluster demonstrates 15 VMs, including 3 master nodes hosted on separate infrastructure and 12 worker nodes hosted on three virtualized hosts. The virtual machine planning will be elaborated in detail in Chapter 8.4.

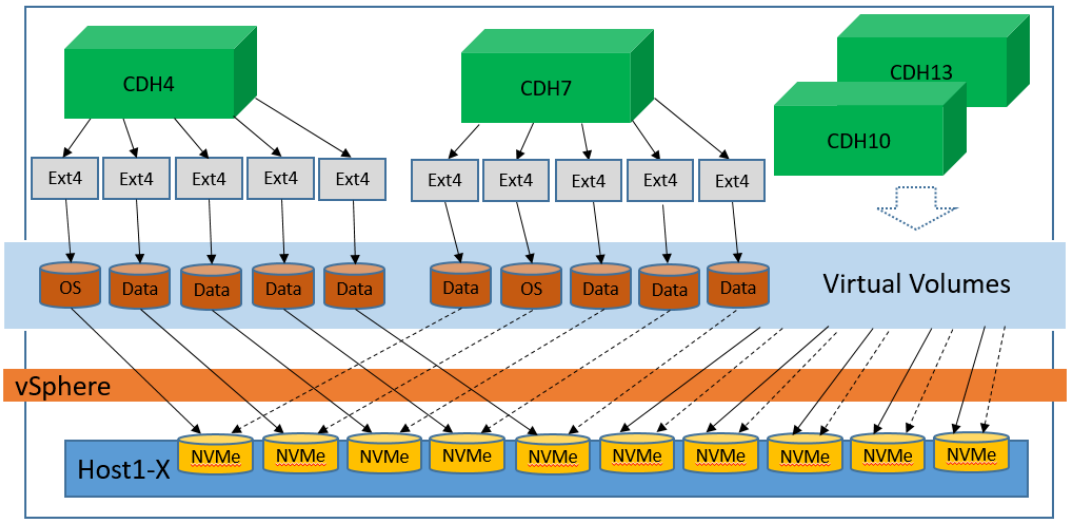

Figure 4. DAS VM storage layout for standard production clusters.

8.1.2 vSAN Scenario

vSAN deployment scenario is designed for customers who expect the best storage flexibility leveraging software defined storage. The possibility to have more than one vSAN storage policy allows customers to create different storage protection levels for virtual machines with different purposes. The worker nodes with the shared data already replicated on the HDFS storage have no need to configure vSAN replicas (failures to tolerate = 0). Management services on the master nodes for Apache Spark

operation are suggested to set a higher number of vSAN replicas (failures to tolerate = 1 and more).

The example of storage layout for research or experimental cluster is shown in Fig. 5. Virtual machines in grey color are the master nodes acting in management roles. Like DAS deployment scenario, QCT accommodated the odd number of virtual machines within the first three hosts, accepting the compromise on general NUMA locality concept. The number of VMDK per VM could stay same for all hosts in the cluster as vSAN itself manages VMDK placement among the hosts. Three disk groups are suggested for QCT's solution to maximize write throughput.

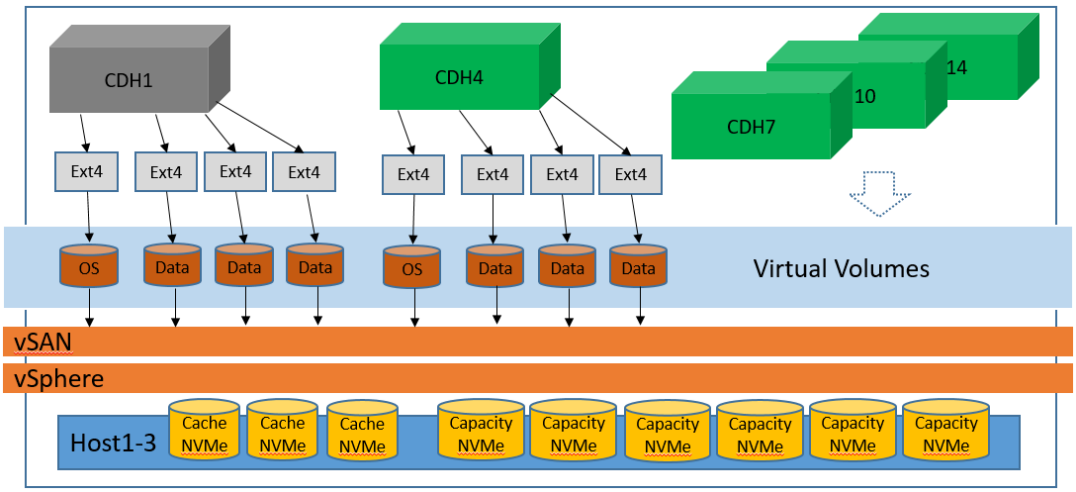

Figure 5. VMware vSAN VM storage layout for small research/experimental clusters.

The storage layout example for the standard production cluster emphasizes more on high availability than on experimental cluster. The storage layout demonstrates 15 VMs, including 3 master nodes hosted on separate infrastructure and 12 worker nodes hosted on three virtualized hosts, as shown in Fig. 6.

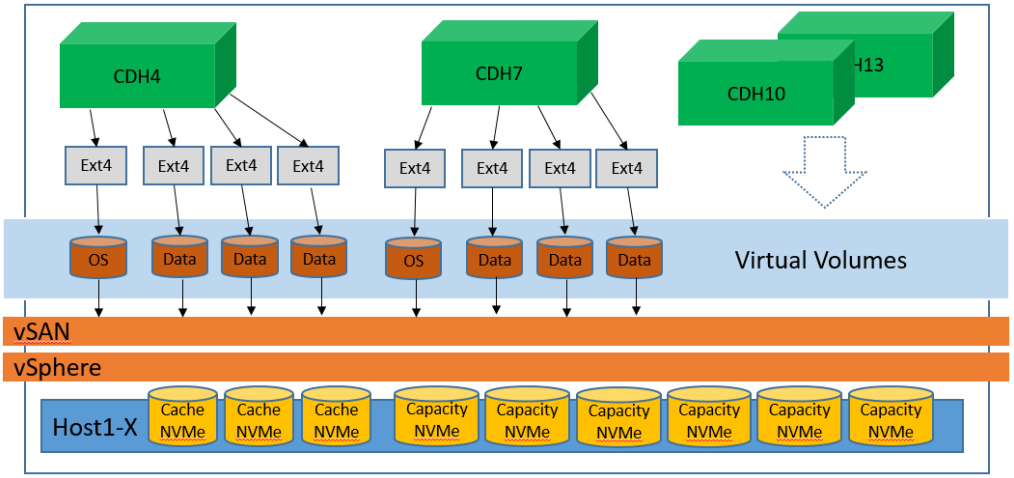

Figure 6. VMware vSAN VM storage layout for standard production clusters.

8.2 Networking Considerations

Thanks to data locality concept followed by both Apache Spark and Hadoop computation engines, data are processed on the virtual machine they belong to. Network bandwidth demand to support Apache Spark cluster nodes communication during computation is therefore not high. Most of the bandwidth for virtual machine networking is used when the computation results are written to HDFS storage or when the constant flow of incoming/outgoing data to/from the Apache Spark cluster is present. VMware vSAN used as the software defined storage however needs to be supported by network path with enough bandwidth, especially when the NVMe devices are used. The following settings are recommended for network infrastructure:

- VMXNET3 network adapter type is used for VM.
- MTU=9000 for jumbo frames is configured for both vSAN and VM data path, including the guest operating system.
- 10Gbps network card/switch/cable supports VM network path.
- 40Gbps network card/switch/cable supports vSAN network path.

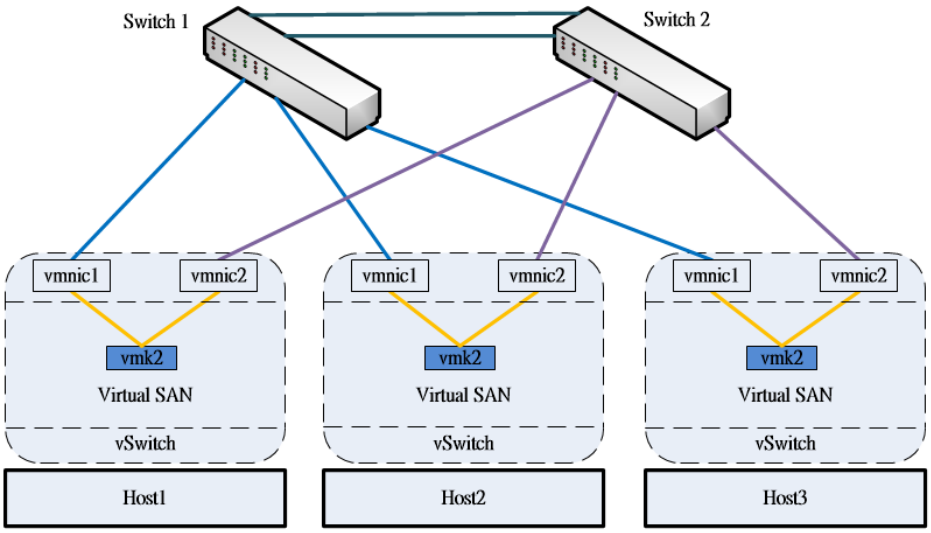

Figure 7. Network topology to support vSAN.

8.3 Virtual Machine Planning

Sizing dilemma is one of the main technical decisions related to Apache Spark on the virtualization platform like VMware vSphere. Historically, both Apache Spark and Hadoop computation engines were designed to run on commodity hardware with low performance per host. When the virtualized hardware is leveraged, Apache Spark computation engine delivers default-setup suboptimal performance with low resources utilization. To find the right size and the right number of virtual machines per host, QCT have undergone an optimization process, including VMware and Cloudera experience application, all-NVMe SKU evaluation in term of the number of CPU cores and available memory, the number of NVMe available in QCT's solution for DAS, and vSAN storage throughput measured by HCIbench. The proposed iterative process is quick sizing guidance for machine learning workloads using Apache Spark, as shown in Fig. 8. After a series of tests, QCT recommends the following initial settings for the optimization process:

- 4 VMs per host for production clusters.
- 100GB vRAM configured for both master and worker nodes.
- 18 vCPUs configured for both master and worker nodes.
- 1 virtual disk for OS boot and 3-4 virtual disks for data. For example: 1x100GB+ 3x300GB (4x250GB).
- Eager-zeroed thick VMDKs along with the ext4 or XFS filesystem inside the guest.
- VMware Paravirtual SCSI (pvscsi) adapter for disk controllers; all 4 virtual SCSI controllers available in vSphere 6.5.
- vMotion used only for evacuation scenarios where a physical server needs to be repaired and the virtual machines are moved away from it for a period of time.

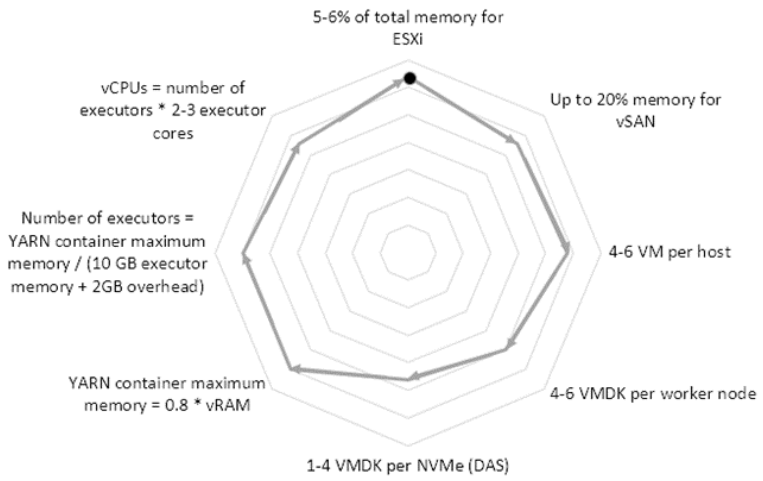

Figure 8. First iteration of VM sizing process with weighted parameters.

8.4 Guest OS Considerations

CentOS is selected as the first choice for "RPM" Linux distribution. It leaves customer a choice to stay with community operating system or to move to a regular paid subscription from RedHat. Other Linux distributions are not the subjects of QCT solution design; however, QCT is open, upon customer's request, to run PoC with alternative Linux distributions. Due to the specific behavior of Cloudera Hadoop on the virtualized infrastructure, several Linux OS parameters were adjusted from their default values, as shown in Table 3.

Table 3. CentOS 7.3 parameters adjusted for VM profile.

Section	Parameter Name	Value
/etc/sysctl.conf	vm.swappiness	1
	net.core.rmem_max	16777216
	net.core.rmem_max	16777216
	net.ipv4.tcp_rmem	4096 87380 16777216
	net.ipv4.tcp_wmem	4096 65536 16777216
	net.core.netdev_max_backlog	250000
/etc/security/limits.conf	* soft nofile	65536
	* soft nofile	1048576
	* soft nproc	65536
	* hard nproc	unlimited
	* hard memlock	unlimited
Huge Pages handling	/sys/kernel/mm/transparent_hugepage/defrag	never
	/sys/kernel/mm/transparent_hugepage/enabled	never
I/O queue scheduler tuning	/sys/block/sda/queue/scheduler	noop
	/sys/block/sdb/queue/scheduler	noop
	/sys/block/sdc/queue/scheduler	noop
	/sys/block/sdd/queue/scheduler	noop
Networking	MTU size for network interface	9000
Filesystem	OS, data	Ext4

8.5 Cloudera Hadoop and Apache Spark Settings for Machine Learning

Machine learning represents a noticeable portion in Apache Spark applied scenarios. SparkTC/Spark-bench software and two popular machine learning algorithms, and SVM and Logistic Regression are used to find optimal settings for Apache Spark cluster. Scala and Java programing languages are used to write applications for Java Virtual Machine (JVM). This reference architecture focuses on JVM-based Apache Spark workloads.

Cloudera Hadoop infrastructure uses default Yet Another Resource Negotiator (YARN) to schedule the tasks initiated by the driver program among Apache Spark cluster worker nodes. Each worker node represents one or more executor processes. The allocated number of executor processes and the allocated memory greatly impact overall cluster performance and hardware utilization efficiency. VMware and QCT's test results reveal that the fewer number of executors with more memory performs better than the more executors with less allocated memory. Figure 9 describes task scheduling in the cluster. Cloudera Hadoop parameters and HDFS parameters are adjusted to provide optimal environment for Apache Spark executors, as shown in Table 4.

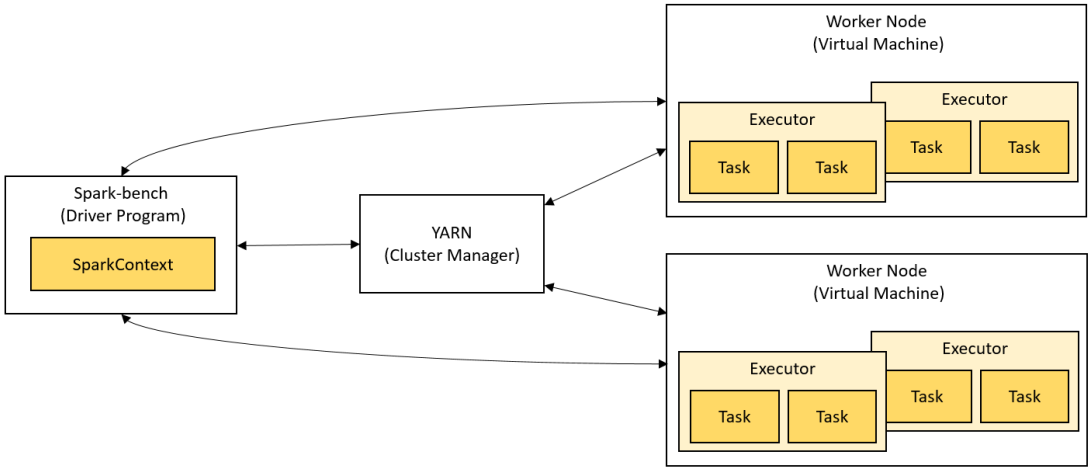

Figure 9. Apache Spark task scheduling.

Table 4. Parameters for VM profile.

Section	Parameter	Value
Cloudera Hadoop	Version in Cloudera distribution	5.11
	dfs.blocksize	256MiB
	dfs.replication	3
	yarn.nodemanager.resource.cpu-vcores	18
	yarn.nodemanager.resource.memory-mb	85GiB (raised from default 60 GiB)
	Oracle JDK	1.8
Apache Spark	Version in Cloudera distribution	2.1 release 1
	SPARK_EXECUTOR_MEMORY	12GiB
	SPARK_EXECUTOR_MEMORY_OVERHEAD	2GiB
	SPARK_EXECUTOR_INSTANCES	6
	SPARK_EXECUTOR_CORES	3
	YARN_DEPLOY_MODE	client
	STORAGE_LEVEL	MEMORY_AND_DISK

9. Conclusion

Nowadays, data and its insights are considered one of the key pillars to support successful business. QCT provides innovative and flexible solutions to keep customers in a leading position.

QxStack VMware Edition for Apache Spark is the right fit for the daily operations of machine learning based data processing workloads that provides comprehensive security in different levels, high reliability with thorough certification and parameter tuning, sufficient performance with discreetly selected components. The reference architecture provides the summary of important deployment steps based on the experiences from VMware and Cloudera and is validated by integration and functional testing to ensure time to value and minimize deployment risk.

By adopting this solution, customers can leverage QCT's knowledge and have a simplified path to business insight and innovation.

QCT appreciates any feedback from you. For further inquiry, please visit http://go.qct.io/solutions/software-defined-storage/QxStack-vmware-edition-vsan-readynode/ .

10. Reference

Top Apache Spark Use Cases
https://www.qubole.com/blog/apache-spark-use-cases/
Cloudera Manager
https://www.cloudera.com/downloads/manager/5-11-0.html
Installing or Upgrading Cloudera Distribution of Apache Spark 2
https://www.cloudera.com/documentation/spark2/latest/topics/spark2_installing.html
Virtual SAN Hardware Quick Reference Guide
http://partnerweb.vmware.com/programs/vsan/Virtual_SAN_Hardware_Quick_Start_Guide.pdf
Intel® SSD Data Center Family
https://www-ssl.intel.com/content/www/us/en/solid-state-drives/data-center-family.html
QuantaGrid D51BP-1U
http://www.qct.io/product/index/Server/rackmount-server/1U-Rackmount-Server/QuantaGrid-D51BP-1U
Scaling the Deployment of Multiple Hadoop Workloads on a Virtualized Infrastructure
https://www.intel.com.tr/content/dam/www/public/us/en/documents/articles/intel-dell-vmware-scaling-the-deployment-of-multiple-hadoop-workloads-on-a-virtualized-infrastructure.pdf
Cloudera Reference Architecture for VMware vSphere with Locally Attached Storage
http://www.cloudera.com/documentation/other/reference-architecture/PDF/cloudera_ref_arch_vmware_local_storage.pdf
Big Data Performance on vSphere 6
https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/techpaper/bigdata-perf-vsphere6.pdf

About QCT

QCT (Quanta Cloud Technology) is a global datacenter solution provider extending the power of hyperscale datacenter design in standard and open SKUs to all datacenter customers.

Product lines include servers, storage, network switches, integrated rack systems and cloud solutions, all delivering hyperscale efficiency, scalability, reliability, manageability, serviceability and optimized performance for each workload.

QCT offers a full spectrum of datacenter products and services from engineering, integration and optimization to global supply chain support, all under one roof.

The parent of QCT is Quanta Computer Inc., a Fortune Global 500 technology engineering and manufacturing company.

http://www.QCT.io

United States

QCT LLC., Silicon Valley office
1010 Rincon Circle, San Jose, CA 95131
TOLL-FREE: 1-855-QCT-MUST
TEL: +1-510-270-6111
FAX: +1-510-270-6161
Support: +1-510-270-6216

QCT LLC., Seattle office
13810 SE Eastgate Way, Suite 190, Building 1,
Bellevue, WA 98005
TEL: +1-425-633-1620
FAX: +1-425-633-1621

China

云达科技，北京办公室（Quanta Cloud Technology）
北京市朝阳区东大桥路 12 号润城中心 2 号楼
Tower No.2, Run Cheng Center, No.12, East Bridge Rd.,
Chaoyang District, Beijing, China
TEL: +86-10-5920-7600
FAX: +86-10-5981-7958

云达科技，杭州办公室（Quanta Cloud Technology）
浙江省杭州市西湖区古墩路浙商财富中心 4 号楼 303 室
Room 303，Building No.4，ZheShang Wealth Center
No. 83 GuDun Road, Xihu District, Hangzhou, Zhejiang , China
TEL: +86-571-2819-8650

Japan

Quanta Cloud Technology Japan 株式会社
日本国東京都港区芝大門二丁目五番八号
牧田ビル３階
Makita Building 3F, 2-5-8, Shibadaimon ,
Minato-ku, Tokyo 105-0012, Japan
TEL: +81-3-5777-0818
FAX: +81-3-5777-0819

Taiwan

雲達科技（Quanta Cloud Technology）
桃園市龜山區文化二路 211 號 1 樓
1F, No. 211 Wenhua 2nd Rd., Guishan Dist.,
Taoyuan City 33377, Taiwan
TEL: +886-3-286-0707
FAX: +886-3-327-0001

Germany

Quanta Cloud Technology Germany GmbH
Hamborner Str. 55, 40472 Düsseldorf , Germany
TEL: + 492405-4083-1300

Other regions

Quanta Cloud Technology
No. 211 Wenhua 2nd Rd., Guishan Dist., Taoyuan
City 33377, Taiwan
TEL: +886-3-327-2345
FAX: +886-3-397-4770